Discovery and development of antibiotics by genome mining, protein characterization and biosynthetic pathway manipulation

Cumulative Dissertation

to fulfil the
Requirements for the Degree of
"doctor rerum naturalium" (Dr. rer. nat.)

Submitted to the Council of
the Faculty of Agricultural Sciences, Nutritional Sciences,
and Environmental Management
of the Justus Liebig University of Giessen

by I Dewa Made Kresna, M.Sc

1st Reviewer : Prof. Dr. Till F. Schäberle
Department of Natural Product Research
Institute for Insect Biotechnology
Justus-Liebig-University Giessen

2nd Reviewer : Prof. Dr. Marc F. Schetelig
Department of Insect Biotechnology in Plant Protection
Institute for Insect Biotechnology
Justus-Liebig-University Giessen

Date of defense :

Declaration of Authorship

I hereby declare that:

1. This dissertation submitted is a work of my own, written without any illegitimate help by any third party and only with materials indicated in the dissertation.
2. I have indicated in the text where I have used texts from already published sources, either word for word or in substance, and where I have made statements based on oral information given to me.
3. At any time during the investigations carried out by me and described in the dissertation, I followed the principles of good scientific practice as defined in the "Statutes of the Justus Liebig University Giessen for the Safeguarding of Good Scientific Practice".

Ich erkläre hiermit:

1. Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe.
2. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.
3. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis" niedergelegt sind, eingehalten.

Giessen, June 2022

I Dewa Made Kresna, M.Sc

Acknowledgement

This thesis brings me to the end of my Ph.D and I would like to thank all the people who supported me during this adventurous journey.

First of all, I would like to thank the mighty God *Ida Sang Hyang Widhi Wasa* who creates this wonderful nature.

I would like to thank my parents Dr. I Dewa Ketut Anom and Ni Gusti Ayu Puspita, M.Sc who have supported me mentally and physically and motivated me to keep learning.

I would like to thank my little family, my wife Vasty Angel Mait, S.Psi for your warm heart, support and understanding throughout the period of my doctoral study.

I would like to thank the DAAD (Deutscher Akademischer Austauschdienst) for the Ph.D scholarship and the Federal Republic of Germany for letting me living this beautiful country. Your huge support has been more than what I could ever expect.

I would like to express my heartfelt gratitude to Prof. Dr. Till F. Schäberle. I am very grateful to be part of your group. You have shown me the freedom to think, to feel and to do science experiments. You have been a supervisor, a mentor, a teacher and a friend for me. Thank you for letting me be part of many interesting projects and thank you for believing in me.

I would like to thank the students, postdocs and secretary of Prof. Till F. Schäberle working group: Dr. Nils Böhringer, Dr. Zerlina G. Wuisan, Alexandra Bender, Dr. Riyanti, Dr. Luis Linares-Otoya, Dr. Ute Mettal, Dr. Yang Liu, Lei Wang, M.Sc, Dr. Katja Fisch, Dr. Jean Marie-Pohl, Leo Padva, M.Sc, Jil-Christine Kramer, M.Sc, Dr. Walter Balansa, Pascal-Marcel Geisler, M.Sc and Prof. Dr. Wael Abdel-Mageed. You have been very kind to help and share your expertise, experience, guidance and priceless knowledge. You have provided me the feeling of having a family during my study in Germany.

I would like to thank the other Ph.D students, technical assistance and postdocs from IME Fraunhofer and JLU Giessen: Dr. Martin Gand, Dr. Michael Marner, Dr. Maria Patras, Stephan

Brinkmann, M.Sc, Mona Abdulahi, Dr. John Heep, and Ing. Christoph Hartwig. You have been supporting me with your help, knowledge and expertise on many occasions.

I would like to thank the people and research groups that have let me be part of Darobactin collaboration projects: The working group of Prof. Dr. Kim Lewis, Dr. Kenichi Yokoyama, Dr. Yvain Nicolet, Dr. Silke Reiter and Dr. Steven Schmitt.

I would like to thank you for the opportunity to co-supervise and be helped by several lab course students, internships and master students: Egzona Lami, Giliana Mulaku, Hyrie Ajdini, Friscasari Gurusinga, Flaviana Susanto, Sanjata Joshi, Rouba Sehli, Farhana J. Bithy, Ibrahim Abdelhafiz, Omar Moushli, Nicolas R. Castillo and Kulsum Zaidi.

I would like to thank my little Indonesian community in Giessen: Roberto Utomo, Dr. Zerlina G. Wuisan, Winda, Flaviana C. Susanto, Frisca Gurusinga, Cakra and Willy. Your presence, parties and our game-board activities have always filled my longing for Indonesia everyday culture environment.

Lastly, I would like to thank Justus Liebig University of Giessen and IME Fraunhofer Giessen for providing me with the opportunity to study, do experiments and experience the fruit of science.

Abbreviation List

A (domain)	Adenylation (domain)
AA	Amino Acid
AAA	Agar Activity Assay
ACP	Acyl Carrier Protein
AMP	Adenosine monophosphate
AT	Acyltransferase
ATP	Adenosine triphosphate
BGC	Biosynthetic Gene Cluster
BPC	Basic Peak Chromatogram
C (domain)	Condensation (domain)
CoA	Coenzyme A
DAHP	3-Deoxy-$_D$-arabino-heptulosonate-7-phosphate
DCM	Dichloromethane
DH	Dehydratase
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
E (domain)	Epimerase (domain)
EDTA	Ethylenediaminetetraacetic acid
EIC	Extracted Ion Chromatograms
ER	Enoylreductase
E4P	Erythrose-4-phosphate
FAD	Flavin Adenine Dinucleotide
Fre	Flavin reductase
GA	Gibson Assembly
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	Histidine tag
HPLC	High Performance Liquid Chromatography
HR	Homologous Region
HRMS	High Resolution Mass Spectrometry
IPTG	Isopropyl β-$_D$-1-thiogalactopyranoside
Kan	Kanamycin

KR	Ketoreductase
KS	Ketosynthase
LB	Luria Bertani
LCMS	Liquid Chromatography Mass Spectrometry
MBP	Maltose Binding Protein
MeCN	Acetonitrile
MeOH	Methanol
MS	Mass Spectrometry
MT	Methyl Transferase
m/z	Mass to charge ratio
NADH	Nicotinamide Adenine Dinucleotide
NRPS	Nonribosomal Peptide-synthetase
NMR	Nuclear Magnetic Resonance
OD	Optical Density
RaS	Radical SAM
RBS	Ribosomal Binding Site
RiPP	Ribosomally synthesized and Post-translationally modified Peptide
rpm	Revolutions per minute
rRNA	Ribosomal ribonucleic acid
SAM	*S*-adenosylmethionine
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel
PCP	Peptidyl Carrier Protein
PCR	Polymerase Chain Reaction
PEP	Phosphoenolpyruvate
PKS	Polyketide synthase
POP	Polyoxyperuin
ppm	Parts per Million
PPT	Phosphopantetheinyl
Pre-peptide	Precursor peptide
TE	Thioesterase
TFA	Tri Fluorotic Acid
UPLC	Ultra Performance Liquid Chromatography
UPRS	Unspecific Protease Recognition Site
4-HBA	4-hydroxybenzoic acid

Sigle and three letter code of amino acids

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

Contents

Abstract

After the golden era of antibiotic discovery in the 1940s-1960s, we are currently facing an emerging antibiotic crisis as bacteria develop resistance to antibacterial agents in use. The declining antibiotic drug development pipeline aggravates this. The development of resistance after exposure to an inhibiting compound like an antibiotic is a natural consequence of the stress response of living organisms to survive. Even worse, the resistance gene(s) can be transferred between bacteria resulting in a fast spreading. Natural products produced by bacteria and fungi still significantly contribute to the discovery and development of new antibiotics. However, the antibiotics discovery pipeline is not adequately filled, due to: (1) re-discovery of known compounds, (2) a majority of bacteria is unculturable under laboratory conditions, and (3) the biosynthetic gene clusters (BGCs) encoding active compounds are not expressed.

In this work, the biosynthesis of the ambigols produced by *Fischerella ambigua* was characterized *in vitro*. The ambigols show antibacterial, antifungal and molluscicidal effects, and inhibit cyclooxygenase and HIV reverse transcriptase. Enzymes involved in the biosynthesis were cloned and heterologously expressed in *E. coli* cells. 3-deoxy-7-phosphoheptulonate (DAHP) synthase (Ab7) is involved in chorismate biosynthesis by the shikimate pathway. Chorismate in turn is further converted by a dedicated chorismate lyase (Ab5), yielding 4-hydroxybenzoic acid (4-HBA). The stand-alone adenylation domain Ab6 is necessary to activate 4-HBA, which is subsequently tethered to the acyl carrier protein (ACP) Ab8. The Ab8-bound substrate is chlorinated by Ab10 in meta position yielding 3-Cl-4-HBA, which is then transferred by the condensation (C) domain to the peptidyl carrier protein and released by the thioesterase (TE) domain of Ab9. The released product is then expected to be the dedicated substrate of the halogenase Ab1 producing 2,4-dichlorophenol as the monomeric building block for ambigol biosynthesis, catalyzed by the previously characterized P450 enzymes Ab2 and Ab3.

Furthermore, to discover new compounds possessing antibiotic properties, a genome mining approach was implemented resulting in discovery of polyoxyperuins (POPs) and monocyclic (mc-) darobactin. The *pop* BGC (PKS-NRPS-hybrid) was identified from our in-house genomic database originated from the marine-derived *Streptomyces* sp. s120. To characterize its products, homologous expression and comparative MS analysis of extracts from the native

producer and a knockout mutant led to the identification of the metabolites polyoxyperuin A seco acid (**1**) and polyoxyoeruin A (**2**) corresponding to the *pop* BGC. Furthermore, by overexpression of a regulatory element, i.e. a LmbU-like transcriptional activator, the production yield of **1** and **2** was increased, enabling isolation and structure elucidation using high-resolution mass spectrometry and NMR spectroscopy. Compound **1** exhibited a low antibiotic effect against *Micrococcus luteus*, while **2** showed a strong Gram-positive antibiotic effect in micro-broth-dilution assays.

A BGC showing high similarity to the darobactin BGC was mined from NCBI public database. The BGC was identified from a chromosomal genomic sequence of *Sulfidibacter corallicola* M133^T. Since the natural producer is not in hand, a heterologous expression approach was used to express the BGC and to generate its product. Therefore, the BGC was generated synthetically, cloned in the pRSFduet expression vector and transferred to *E. coli* as a host. Comparative MS analysis of extracts from the heterologous producer and control led to the identification of two ionized masses corresponding to the putative product of the *mcd* BGC. The product is predicted to be the core peptide McdB (W-R-W-S-W-P-F) with one additional oxygen and loss of two hydrogens. The modification is suspected to correlate to the formation of a *C-O-C* (ether bond) like it is the case for darobactin. The product was named mono cyclic (mc-) darobactin, since the latter carries an additional ring formed between two carbon atoms.

Moreover, as darobactin is a lead compound for a new antibiotic, a method to increase its production yield, a comprehensive understanding of the biosynthesis and derivatization to obtain more active derivatives and get first insights into its structure-activity-relationship were performed. Using a heterologous expression system, production yield was increased 10-fold and the production time decreased 5-fold. Furthermore, the minimum BGC was identified to consist of only *darA* and *darE*. By an *in vivo* study, modification of the heterologously expressed tagged DarA by DarE was investigated. The result suggested that DarE catalyzes both ring formations in DarA. Modifications occurred before the core peptide was released from the DarA-leader and both ring modifications could happen independently. For the derivatization of darobactin, its biosynthesis was manipulated by mutating codons encoding the residues in the core peptide of DarA. In that way the ribosomally synthesized and post-translationally modified peptide (RiPP) darobactin A was derivatized yielding 69 derivatives of which 37 showed inhibitory activity against *E. coli*. From the initial test, one of them shows better antimicrobial activity against *P. aeruginosa* and *A. baumannii* strains than Darobactin A and B.

1. Introduction

1.1. Natural product antibiotics and their resistance mechanism

An antibiotic is a substance that can kill or inhibit the life of bacteria. In nature, antibiotics are mostly produced by microbes, bacteria and fungi (a few are from higher animals and plants) to kills other microbes as a result of food competition or self-defense mechanisms. Natural product antibiotics can be divided based on their structures into several classes as the following.

i) Aminoglycoside: Their structures contain aminocyclitol ring to which amino sugars are attached by glycosidic linkages e.g. kanamycin, gentamicin, amikacin, tobramycin, neomycin, and streptomycin.[1] They exhibit bactericidal activity by interrupting protein synthesis, binding irreversibly to the 16S ribosomal RNA receptor on the 30S subunit of the bacterial ribosome.[2]

ii) Tetracyclines: Their structures comprise a linear fused tetracyclic nucleus (four rings designated A, B, C, and D) to which a variety of functional groups are attached e.g. tetracycline, rolitetracycline, lymecycline, clomocycline, methacycline, doxycycline, and minocycline. They inhibit protein synthesis by preventing the attachment of aminoacyl-tRNA to the ribosomal acceptor (A) site (30S subunit).[3]

iii) Amphenicols: Their structures character contain phenylpropanoid structure e.g. chloramphenicol, thiamphenicol, azidamfenicol, and florfenicol. They inhibit protein synthesis by binding to the 50S subunit of the bacterial ribosome.[4]

iv) Polypeptides: Their structures are made of amino acid building blocks with some modifications including rings formation e.g. gramicidin, bacitracin, polymyxin, bacteriocins, nisin, epidermidin, tuberactinomycins, enniatins, and lipopeptides. Polypeptides fall into two subclasses which are non-ribosomally peptide synthetized peptides and ribosomally synthesized peptides. Most of them have been identified to bind bacterial membrane components such as lipopolysaccharide and membrane proteins and disrupts the membrane permeability.[5]

v) β-lactams: Structurally, they possess the β-lactam ring e.g. penicillin, cephalosporins, monobactams, carbapenems, carbacephems and monobactams. They bind to bacterial penicillin-binding proteins and disrupt peptidoglycan synthesis.[6]

vi) Polyketide: including macrolide which are consist of a macrocyclic lactone of different ring sizes, to which one or more deoxy-sugar or aminosugar residues are attached e.g. erythromycin, lipiarmycin and methymycin. Macrolides act as antibiotics by binding to

bacterial 50S ribosomal subunit and interfering with protein synthesis.[7] Other examples of polyketide antibiotics are ansamycin, rifamycin, fusidic acid and mupirocin.

vii) Glycopeptides: are actinomycete-derived antibiotics with unique tricyclic or tetracyclic heptapeptide cores that are usually glycosylated and sometimes have additional lipophilic fatty acid side chains[8] e.g. vancomycin, teicoplanin, telavancin, ramoplanin and decaplanin, corbomycin, complestatin and bleomycin. They inhibit the transglycosylation and/or transpeptidation steps involved in cell wall biosynthesis.[9, 10]

viii) Others: structurally unique and different to other antibiotic structures described above e.g. lincosamides, cycloserine, phosphonates and pleuromutilins. Their mode of action are various such as inhibiting protein synthesis and cell wall biosynthesis.[11, 12, 13, 14, 15]

Antibiotics have been widely used to fight against bacterial infections. Antibiotics are probably one of the most successful forms of chemotherapy in the history of medicine.[16] The period from the 1940s to 1960s was the golden era of antibiotic discovery and from the 1970s onwards, discovering new antibiotics has declined that only a few new antibiotics filled the clinical trials pipeline[17] (Figure 1). The pressure imposed by the use of antibiotics in the long term guides the evolution of bacteria to become resistant as a result of natural selection. The crisis of antibiotic resistance has numerous causes and influences and is the cumulative effect of widespread and extensive use of antibiotics for over 60 years for numerous purposes e.g. in hospitals, chronic care facilities, day-care centers, in a very wide variety of patients including notably the newborn, the immunosuppressed, the elderly and significantly as a supplement in aqua- and other animal-culture.[18] The development of resistance itself is a complex and multi-step process of the emergence of increasingly fit resistance genes through genetic change, often with subsequent movement through populations on a variety of mobile genetic elements. There is essentially no gene in any bacterium that cannot be moved to another bacterium.[18]

The antibiotic resistance mechanism employed by evolved bacteria can be classified as a target or bullet-related. Targets can be: (i) protected by modification e.g. mutations making it insensitive to antibiotic action such as mutations in RNA polymerase conferring resistance to rifampin; (ii) modified by an enzyme e.g. methylation of an adenine residue in 23S rRNA making it insensitive to macrolides); (iii) replaced e.g. ribosomal protection proteins conferring resistance to tetracyclines; (iv) protected at cellular or population levels (formation of a protective barrier by secretion e.g. of large amounts of exopolysaccharides). The bullet can be: (i) modified so the efficiency is lost e.g. acetylation of aminoglycosides, (ii) destroyed e.g. the

β-lactam antibiotics by the action of β-lactamases, and (iii) pumped out from the cell e.g. efflux pump mechanisms of resistance.[16]

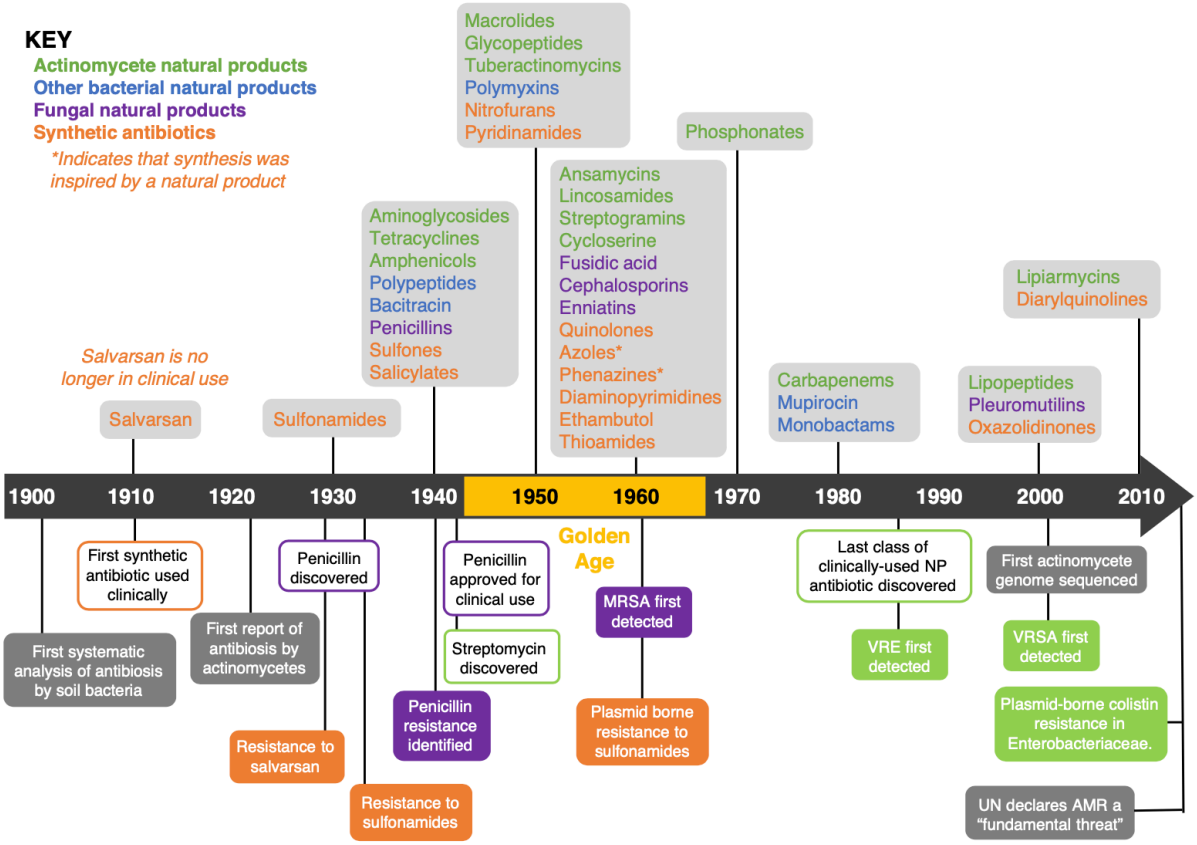


Figure 1. Timeline showing when new antibiotic classes reached the clinic. The antibiotics are colored per their source: green = actinomycetes, blue = other bacteria, purple = fungi and orange = synthetic. At the bottom of the timeline are key dates relevant to antibiotic discovery and antimicrobial resistance, including the first reports of methicillin-resistant *S. aureus* (MRSA) strains, vancomycin-resistant enterococci (VRE), vancomycin-resistant *S. aureus* (VRSA) and plasmid-borne colistin resistance in Enterobacteriaceae. *Picture source*: [19].

1.2. Natural product-based discovery and development of new antibiotics

As bacteria evolve and will become resistant to an antibiotic if they are exposed to it, it is necessary to keep on developing and discovering new antibiotics. Natural products are still an important source and represent privileged diverse structures that originate from natural selection. Traditionally, culture-based approaches have been the most effective for identifying compounds with the empiric ability to prevent cell growth *in vitro*. However, the limitation of this approach is the frequent rediscovery of known compounds (Figure 2).[20] To dereplicate

known compounds already at an early stage of the discovery process, mass spectrometry (MS)-based dereplication strategies that rely on the cosine similarity of the MS fragmentation pattern were developed, e.g. the online tool GNPS (Global Natural Products Social Molecular Networking), to enable dereplication of known compounds in raw extracts.[21] Furthermore, advances in next-generation DNA sequencing technologies showed that under laboratory conditions most of the biosynthetic potential encoded in microbial genomes is not expressed and most members of the microbial communities cannot be easily cultivated.[20] To get a hand on living cultures of the previously uncultivated bacteria and to trigger expression of silent BGCs, approaches such as *in situ* cultivation, by using diffusion chambers like the iChip, or co-culture incubations have been developed. In that way, the bacterial cells are exposed to growth factors and signaling molecules from their environment.[20]

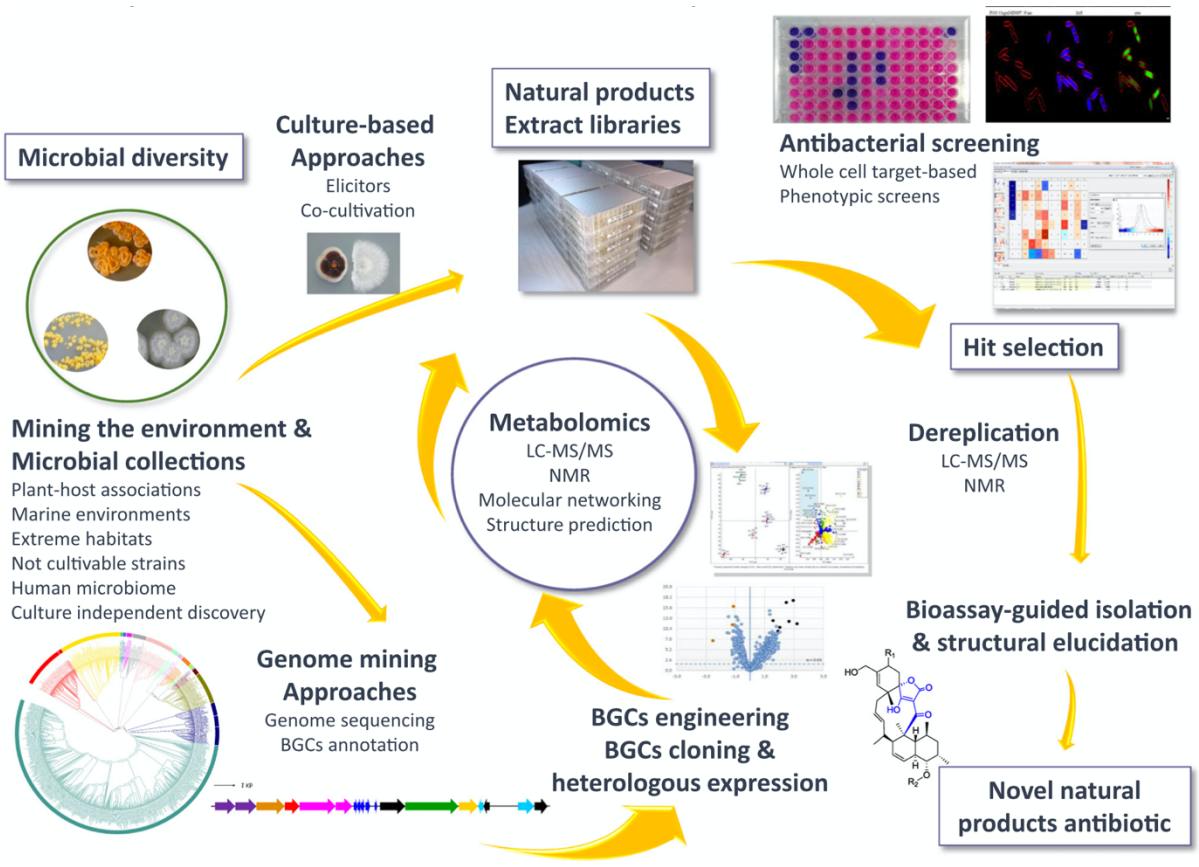


Figure 2. Current multidisciplinary approach used in the discovery of microbial natural products antibiotics: this integrates new genome diversity mining and culture-based technologies with all the analytical omic-based tools, traditional bioactivity screening and the bioengineering of biosynthetic gene clusters. *Picture source*: [20].

Another approach that has been used to discover new compounds is genome mining. Genome mining utilizes biological information to predict the function of proteins and their relationship with other proteins encoded in genes creating a biosynthetic gene cluster (BGC). Genes belonging to the same BGC express proteins involved in the biosynthesis of specific compounds. Several bioinformatic tools e.g. antiSMASH[22], RODEO[23] and eSNAPD[24] helps annotate and group genes from a raw genomic sequence to a proposed BGCs. The genome sequence could be from a local database obtained from genomic sequencing of the desired strain or public database. The most popular website integrating several public databases are National Center for Biotechnology Information (NCBI). Several database resources of NCBI as of 4th September 2021 are (1) genomes (515 million records): Nucleotide, BioSample, SRA, Taxonomy, Assembly, BioProject, Genome and BioCollections; (2) Genes (167 million records): GEO Profiles, Gene, GEO DataSets, PopSet and HomoloGene; and (3) Proteins (1.4 billion records): Protein, Identical Protein Groups, Protein Clusters, Structure, Protein Family Models and Conserved Domains.[25] After all, from a genomic sequence of the desired strain, one can have an overview of what BGCs the organism has and could be a basis decision whether the founded BGCs would be executed further in the lab for characterization of its products.

However, as the information of new compounds, proteins, genes and BGCs are daily reported, it is not always that fast for the BGC annotating tools to catch up this information and integrate them into the software either by updating the old version or making a new one. In this case, screening desired BGCs belonging to a specific desired compound class has to be done manually. Fortunately, a powerful searching tool BLAST[26] has been developed and integrated into the NCBI website. BLAST is a searching tool used as a standard to find regions or similarities between biological sequences locally or against the public databases. The tricky part is rather what sequence would be the query. If we directly use the key protein sequence in the biosynthetic pathway of the desired compound, the BLAST result would more likely show known BGCs or derivatives (similar to the query) from all organisms which possess this sequence. If the goal is to find a new compound (not derivative) but still belong to the same type/class, a more detailed background about the key enzymes in the biosynthesis of the desired compound e.g. amino acid sequence motifs/signatures are required. Motifs from the same enzyme class are usually conserved. The residues within the motifs have essential roles in the functionality of the protein that they are preserved throughout evolution. Motifs can be found by aligning similar proteins obtained from BLAST using the desired protein sequence as a query. BLASTp search using motif as a query will give hit result with a higher chance of new

proteins from the same class/type. Thereafter, the hits can be analyzed, evaluated and curated to determine the completeness and border of the BGC before executing the work further in the lab.

1.3. Homologous and heterologous expression system

Once a BGC of interest has been identified *in silico*, the next question is how to successfully express the genes to obtain the corresponding metabolite? Two approaches that have been mostly adopted are homologous and heterologous expression. Heterologous expression systems use foreign host species to express the BGCs of interest. A suitable heterologous host is mostly phylogenetically closely related to the original organism, and is in the best case well investigated, e.g. its cultivation conditions and DNA manipulation techniques are established[27]. The BGCs could be integrated either into the chromosome of the host strain, or be present on a suitable expression vector. In contrast to chromosomal integration, a plasmid carrying the BGC could provide many copies of it, which could increase the production yield and decrease fermentation time needed to reach a high titer. However, the host is in most cases forced to live in an environment containing antibiotics to maintain the plasmid. Another limitation is that the plasmid size is limited. Therefore, for large BGCs like PKS, NRPS or PKS-NRPS hybrids, chromosomal integration might be more favorable. Heterologous expression is also very useful in the case where the original strains possessing the BGC of interest are uncultivable in the lab or the BGCs were identified from metagenomic sequences. In this case, (if no metagenomics library carrying the information is available) the BGCs have to be synthetically generated. Although many companies offered DNA synthesis product, the size is limited. In essence, while the benefits of heterologous expression are the more controllable growing conditions and established DNA manipulation techniques, the downsides are: (1) silent BGCs (unknown reason – possible unknown regulation or uncomplete BGC), (2) rare codons and codon bias, (3) special proteins e.g. chaperones are necessary for proteins involved in the biosynthesis to be folded in the right form, (4) missing precursors and (5) the end products could be toxic for the host[28] (Figure 3). Many approaches have been employed to tackle these problems, such as introduction of promoters or knock out of repressors, co-expression of rare tRNA, helper proteins, precursors and the genes conferring resistance to the metabolite. Homologous expression instead, relies on the wild type strain to express the BGCs of interest. The wild type strain does not possess the problems that can be encountered in the heterologous expression system, except that the BGCs might be silent (not expressed under the conditions used). The

downsides, however, are that not all wild type strains are easy to handle and can require a long cultivation time. Even though time consuming, optimization of production conditions can be envisaged. If the strain is genetically accessible, further encounter strategies can be used, e.g. introduction of a suitable promoter or knock out of putative repressor genes.

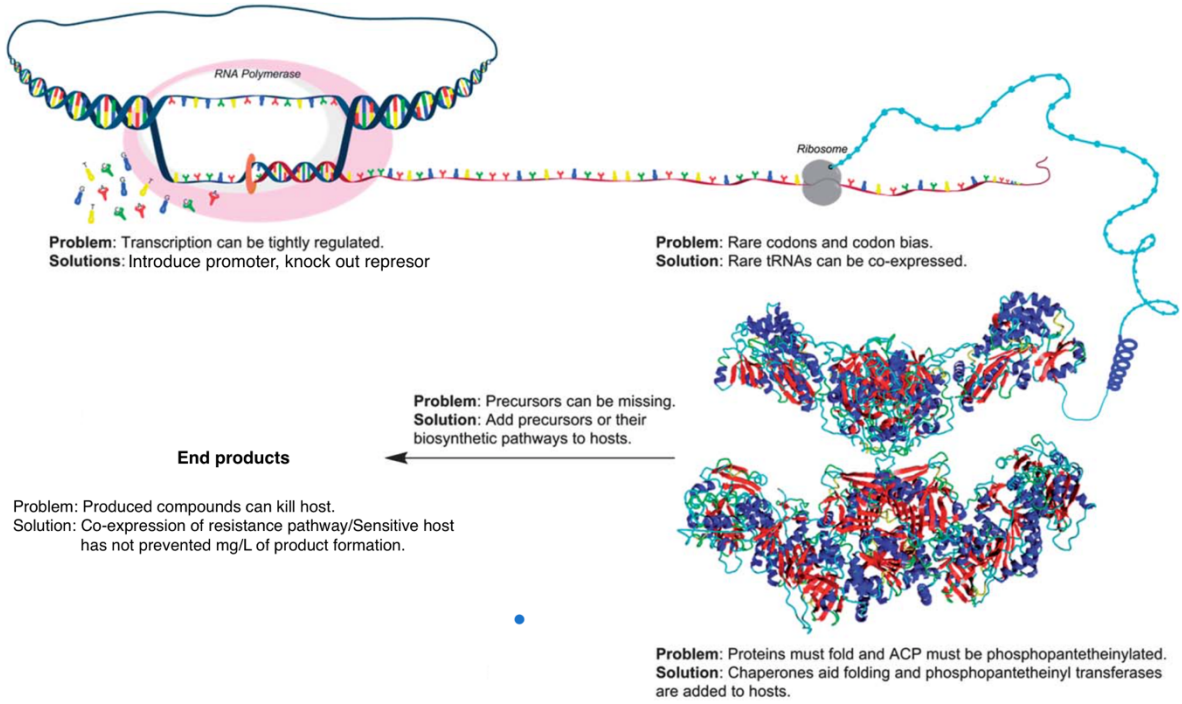


Figure 3. Heterologous expression is a multistep process. Problems can be encountered at each step, like transcription, translation, protein folding, post-translational modification, precursor availability, or product toxicity can prevent successful heterologous expression. This figure highlights some of the common problems and approaches to addressing them. *Picture source*: [28].

1.4. Derivatization of existing antibiotics

Discovering new druggable antibiotics is challenging. After the golden era of antibiotics, the rediscovery of known compounds based on the traditional activity screening approach has been a major issue. Accessing unknown BGCs based on genome mining approach is promising to discover new compounds but many times this ends up either not active compounds or active but not druggable. In order to combat the emerging antibacterial resistance to antibacterial agents, since the mid-1970s industrial approaches to the development of new antibacterial agents have been dominated by the paradigm of chemical modification of existing drug classes to improve their properties e.g. more potent, less toxic, etc.[29] To rationally design a

modification, a study of structure-activity relationship (SAR) of the compound of interest is fundamental. The huge possibility to modify the existing compound of interest lead to the challenge of not only producing the possible derivatives but also to screen which possible derivatives are likely to deliver the expectation that not everything needs to be synthesized/produced. One alternative to reduce these extensive time consuming, cost, and work labor is by *in silico* modelling. This approach requires supercomputer power to model calculate the behavior and properties of the possible derivatives e.g. binding to its target in a computer-simulated environment. Generation of the derivatives could be done by synthesis, semi-synthesis or synthetic biology. Pure synthesis means that the final product is generated without using the living organism as a workhorse, instead, by several chemical and physical reactions. Precursors for chemical reactions are usually purchased. Semi-synthetic approaches employ biological machinery like cells (*in vivo*) or pure enzymes (*in vitro*) in combination with chemical and physical reactions. Synthetic biology uses solely the biosynthetic pathway within the cell and exploits cells as a workhorse. Manipulation of the original biosynthetic pathways could produce derivatives. As an example, the codon responsible for the integration of a specific amino acid residue into the core peptide of a RiPP compound can be altered, or the order of modules in a PKS-NRPS assembly-line machinery can be changed.

1.5. Aim of the thesis

The aim of this thesis is to characterize and manipulate the biosynthetic pathways of bioactive secondary metabolites, and to pursue genome mining to discover new compounds. *In silico*, *in vivo* and *in vitro* experiments will be performed to characterize specific proteins involved in the respective biosynthesis. Proteins encoded by genes from known BGCs of interest would be investigated *in silico* to obtain their predicted properties including their role in the biosynthesis. Verification of the predicted function would be done by heterologous expression, purification and *in vitro* characterization. Missing parts in the biosynthesis would also be investigated to extend our understanding about the specific functions of given proteins in the respective biosynthesis. When *in vitro* characterization is unfavorable due to limited or unknown conditions, an *in vivo* approach would be taken, e.g. using *E. coli* cells as a host. Based on the understanding gained by these experiments, analogue derivatization of the final compound would be performed by manipulation of the biosynthetic pathway. Furthermore, all the knowledge obtained from the results will be used to pursue a genome mining approach to

discover new compounds. The detailed approaches, which will be persuade to achieve these aims, are listed in the following.

No.	Objective	Approaches
1.	Characterization of the missing parts in ambigol biosynthesis	Heterologous expression, purification and *in vitro* characterization of proteins involved in the biosynthesis
2.	Characterization of the radical SAM enzyme DarE and analysis of the obscure protease activity in Darobactin biosynthesis	a. Cloning and heterologous expression of DarE to perform *in vitro* and *in vivo* characterization of its catalytic activities b. Mutation of the precursor peptide DarA to allow *in vivo* characterization of the putative DarE protease activity
3.	Generation of Darobactin derivatives to assess the DarE flexibility and to identify compounds that show improved activity against human pathogens	a. Library generation of the heptapeptide with altered amino acid sequences. This will be used to (i) analyze the flexibility of DarE and to (ii) screen the bioactivity of new derivatives b. Selected derivatives will be produced by heterologous expression and purified to enable structure elucidation and MIC determination
4.	Discovery of BGCs encoding for putative new antibiotic compounds by genome mining and biotechnological generation of the corresponding compounds	a. *In silico* analysis of DNA sequences to discover BGCs of interest b. DNA manipulation to enable homolog or heterologous expression of the respective BGC c. Production, purification and characterization of the product(s)

1.6. Outline of chapters

1.6.1. Chapter I: In vitro characterization of 3-chloro-4-hydroxybenzoic acid building block formation in ambigol biosynthesis

The ambigol biosynthetic gene cluster (BGC) has been previously proven. Two out of the ten enzymes encoded within the BGC have been characterized *in vivo* and *in vitro* – Ab2 and Ab3 as cytochrome P450 enzymes, which catalyze the biaryl and biaryl ether formation to mature the building blocks yielding ambigol. However, little is known about how these building blocks are biosynthesized. In this chapter, the heterologous expression and subsequent *in vitro* characterization of the so far undescribed enzymes encoded in the BGC is described.

1.6.2. Chapter II: Genome-Mining-Guided Discovery and Characterization of the PKS-NRPS-Hybrid Polyoxyperuin Produced by a Marine-Derived Streptomycete

Genome mining is a promising approach and has been employed to discover new compounds as it has the advantage of exploiting hidden biological information treasure. A PKS-NRPS-hybrid BGC originated from a marine-derived Streptomycete was identified. To identify its products, homologous expression was a more favorable approach than heterologous one due to the following reasons: (1) heterologous expression has common problems e.g. host suitability, missing precursors, not complete BGC, etc.; (2) cloning large size BGCs is still very challenging; (3) the strain is genetically accessible. Nevertheless, the regulation regulating the expression of the BGC was investigated in the first place. To ensure the expression of the BGC, a strong promoter upstream a putative regulator (overexpression strain) was introduced and thereby enabling the identification of two highly concentrated ions in the LCMS profiling of the culture extract from the transgenic strain in comparison to the control. Further verification by knocking out an important gene in the BGC abolished the production of two ions indicating the connectivity to its products. The production of the overexpression strain led to the characterization of new compounds polyoxyperuin seco acid and polyoxyperuin A. The former compound has a linear open peptide chain within the structure, while the latter one has a closed ring. Polyoxyperuin seco acid exhibited a low antibiotic effect against *Micrococcus luteus*, while polyoxyperuin A showed a strong Gram-positive antibiotic effect in a micro-broth-dilution assays.

1.6.3. Chapter III: Homologous expression of BGCs in Streptomyces strains

This chapter describes the step-by-step protocol to do genomic DNA recombination in *Streptomyces* strains. The approach relies on a double-crossover recombination event, which occurs during DNA repair, or the DNA replication. A plasmid that cannot propagate in the host strain is used to introduce foreign DNA. An antibiotic resistance marker is co-introduced to enable phenotypic observation of the recombination event. This general approach can be used for many applications, e.g. gene knock out experiments, or insertion of a promoter.

1.6.4. Chapter IV: Optimization of heterologous Darobactin A expression and identification of the minimal biosynthetic gene cluster

Darobactin is a gram-negative antibiotic that was firstly reported in 2019 to be produced by a bacteria belonging to the genus *Photorhabdus*. Because of its low toxicity in the infected mouse model, darobactin has become a highly promising lead structure for the development of antibacterial therapeutic agents in the drugs pipeline. For this reason, increasing the production yield is of interest to improve the compound supply. Since the BGC has been identified, heterologous production has been possible. Optimization of many parameters is described in this chapter such as host strains, source of BGCs, manipulation within the BGC, etc. Furthermore, the minimum BGC consisting solely of two genes is reported.

1.6.5. Chapter V: The radical-SAM enzyme DarE catalyzing the formation of the intramolecular C-C and C-O-C rings in darobactin biosynthesis

The minimum BGC of darobactin has been previously identified. It consists of the precursor peptide encoding *darA* and *darE* that encodes the post translationally modifying enzyme DarE. The heptapeptidic darobactin has two fused intramolecular rings within its structure, whereby it was hypothesized that the necessary modifications are catalyzed by DarE. To obtain insights into the functionality of this unusual enzyme, it was investigated by an *in vivo* approach. Additionally, heterologous expression approaches were performed to enable purification of the tagged enzyme for subsequent *in vitro* studies.

1.6.6. Chapter VI: Systematic investigation of the DarE flexibility – generation of specific and random darobactin derivative libraries

As darobactin is a promising structure in the antibiotic drugs discovery and development pipeline, derivatization of the natural product to obtain more active variants and to get insights into the structure-activity-relationship (SAR) is of high interest. The flexibility of DarE (the only modification enzyme in the biosynthesis) towards its substrate was systematically investigated. Therefore, each amino acid of the heptapeptide was exchanged against all other proteinogenic amino acids and analyzed if DarE was able to introduce the intramolecular rings. The resulting derivatives library was qualitatively assayed against gram-negative *E. coli* cells. Second, a random library was generated, which needs to be analyzed by high throughput methods. Additionally, based on literature research, a specific variant was created possessing better activity against *Pseudomonas aeruginosa* and *Acinetobacter baumannii* than Darobactin A and B.

1.6.7. Chapter VII: Genome-mining-guided discovery and characterization of a RiPP BGC originated from Sulfidibacter corallicola M133^T

The aim is to develop and discover new RiPP antibiotics using a genome mining approach. Darobactin is a potent RiPP antibiotic that selectively kills gram-negative pathogens. The BGC of darobactin consists of *darA-E*. DarE catalyzed the formation of two rings in the core peptide of DarA (precursor peptide). Thereafter, unidentified peptidases cut before and after the modified core peptide releasing mature darobactin, which is transported out of the cell by involvement of the transporter complex DarB-D. To discover other darobactin-like compounds, DarE as the key enzyme in the biosynthesis was used as the basis for genome mining. Initially, conserved motifs of DarE were investigated by aligning DarE-like proteins obtained from public databases. A selected conserved motif was used as a query for BLASTp search against public databases integrated in the NCBI website. The hit results were further investigated to locate the presence of precursor peptides. This culminated in the discovery of a new, considered as complete, BGC originated from *Sulfidibacter corallicola* sp. M133^T. By heterologous expression of the BGC and LCMS profiling of the extracts, new ions corresponding to the product of the BGC were identified and named monocyclic-darobactin (mc-darobactin).

1.7. References

1. Germovsek, E. *et al*. What do I need to know about aminoglycoside antibiotics? *Arch. Dis. Child. Educ. Pract.* **0**, 1–5 (2016).
2. Ramirez, M. S. and Tolmasky, M. E. Aminoglycoside modifying enzymes. *Drug Resistance Updates* **13**, 151-171 (2010).
3. Chopra, I. and Roberts, M. Tetracycline Antibiotics: Mode of Action, Applications, Molecular Biology, and Epidemiology of Bacterial Resistance. *Microbiol. Mol. Biol. Rev.* **65**, 232-260 (2001).
4. Fernández-Martínez, L. T. New Insights into Chloramphenicol Biosynthesis in *Streptomyces venezuelae* ATCC 10712. *Antimicrobial Agents and Chemotherapy* **58**, 7441-7450 (2014).
5. Hancock, R. E. W. and Chapple, D. S. Peptide Antibiotics. *Antimicrob. Agents Chemother.* **43**, 1317-1323 (1999).
6. Mora-Ochomogo, M. and Lohans, C. T. β-Lactam antibiotic targets and resistance mechanisms: from covalent inhibitors to substrates. *RCS Med. Chem.* **12**, 1623-1639 (2021).
7. Dinos, G. P. The macrolide antibiotic renaissance. *British Journal of Pharmacology* **174**, 2967-2983 (2017).
8. Butler, M. S. *et al*. Glycopeptide antibiotics: Back to the future. *The Journal of Antibiotics* **67**, 631-644 (2014).
9. Kahne, D. *et al*. Glycopeptide and lipoglycopeptide antibiotics. *Chem. Rev.* **105**, 425–448 (2005).
10. Arhin, F. F. *et al*. in Antibiotic Discovery and Development (eds Dougherty, T. J. & Pucci, M. J.) 301–346 (Springer US, New York, 2012).
11. Morar, M. *et al*. Structure and Mechanism of the Lincosamide Antibiotic Adenylyltransferase LinB. *Structure* **17**, 1649–1659 (2009).
12. Lambert, M. P. and Neuhaus, F. C. Mechanism of D-Cycloserine Action: Alanine Racemase from Escherichia coli W[1]. *J. Bacteriol.* **110**, 978–987 (1972).
13. Prosser, G. A. and Carvalho, L. P. S. de. Kinetic mechanism and inhibition of Mycobacterium tuberculosis D-alanine: D-alanine ligase by the antibiotic D-cycloserine. *FEBS Journal* **280**, 1150–1166 (2013).
14. Chekan, J. R. *et al*. Molecular basis for resistance against phosphonate antibiotics and herbicides. *Med. Chem. Commun.* **7**, 28-36 (2016).
15. Paukner, S. and Riedl, R. Pleuromutilins: Potent Drugs for Resistant Bugs—Mode of Action and Resistance. *Cold Spring Harb Perspect Med* **7,** a027110 (2017).
16. Aminov, R. I. A brief history of the antibiotic era: lessons learned and challenges for the future. *Front. Microbiol.* **1**, 1-7 (2010).
17. Katz, L. and Baltz, R. H. Natural product discovery: past, present, and future. *Journal of Industrial Microbiology and Biotechnology* **43**, 155-176 (2016).
18. Prescott, J. F. The resistance *tsunami*, antimicrobial stewardship, and the golden age of microbiology. *Vet. Microbiol.* **171**, 273-278 (2014).
19. Hutchings, M. I. *et al*. Antibiotics: past, present and future. *Current Opinion in Microbiology* **51**, 72-80 (2019).

20. Genilloud, O. Natural products discovery and potential for new antibiotics. *Current Opinion in Microbiology* **51**, 81-87 (2019).
21. Wang, M. *et al*. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology* **34**, 828-837 (2016).
22. Blin, K. *et al*. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research* **49**, 29-35 (2021).
23. Tietz, J. I. *et al*. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nature Chemical Biology* **13**, 470-478 (2017).
24. Reddy, B. V. B. *et al*. eSNaPD: A versatile, Web-Based Bioinformatics Platform for Surveying and Mining Natural Product Biosynthetic Diversity from Metagenomes. *Chemistry & Biology* **21**, 1023-1033 (2014).
25. Sayers, E. W. *et al*. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **50**, D20–D26 (2022).
26. Altschul, S. F. *et al*. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410 (1990).
27. Li, L. *et al*. Refactoring biosynthetic gene clusters for heterologous production of microbial natural products. *Current Opinion in Biotechnology* **69**, 145-152 (2021).
28. Stevens, D. C. *et al*. The role of transcription in heterologous expression of polyketides in bacterial hosts. *Nat. Prod. Rep.* **30**, 1391-1411 (2013).
29. Chopra, I. *et al*. Discovery and development of new anti-bacterial drugs. *Pharmacochemistry Library* **32**, 213-225 (2002).

2. Chapter I:

In vitro characterization of 3-chloro-4-hydroxybenzoic acid building block formation in ambigol biosynthesis

Kresna, I D. M., Linares-Otoya, L., Milzarek, T., Duell, E. R., Mohseni, M. M., Mettal, U., König, G. M., Gulder, T. A. M. and Schäberle, T. F.

Published manuscript: *Organic & Biomolecular Chemistry* **19**, 2302-2311 (2021). doi: 10.1039/D0OB02372H

Summary: This manuscript presents the extended understanding of ambigol biosynthesis, which previously had been partially characterized. The genes encoding the enzymes involved in the biosynthesis were heterologously expressed and purified by affinity chromatography. Subsequently, the pure enzymes were characterized *in vitro*. The results allowed to comprehend the sequence how the enzymes work at each step in the ambigol biosynthetic pathway. Furthermore, insights were obtained how formation of side products is prevented.

Contribution: I Dewa Made Kresna co-designed and planed the strategies and experimental set up. He planned and performed the experiments for the expression and *in vitro* analysis of all the enzymes (Ab6, Ab8, Ab9 and Ab10) except Ab7. He analyzed the results and put them into a discussion. He wrote the initial draft of the manuscript as well as the following revision and contributed to the preparation of the figures together with the co-authors.

Organic & Biomolecular Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d0ob02372h

In vitro characterization of 3-chloro-4-hydroxybenzoic acid building block formation in ambigol biosynthesis†

I Dewa Made Kresna, [a,b] Luis Linares-Otoya,[a,b] Tobias Milzarek, [c] Elke R. Duell, [c,d] Mahsa Mir Mohseni,[e,f] Ute Mettal,[a,b] Gabriele M. König, [e] Tobias A. M. Gulder [c,g] and Till F. Schäberle *[a,b]

Received 27th November 2020,
Accepted 22nd February 2021

DOI: 10.1039/d0ob02372h

rsc.li/obc

The cyanobacterium *Fischerella ambigua* is a natural producer of polychlorinated aromatic compounds, the ambigols A–E. The biosynthetic gene cluster (BGC) of these highly halogenated triphenyls has been recently identified by heterologous expression. It consists of 10 genes named *ab1–10*. Two of the encoded enzymes, *i.e.* Ab2 and Ab3, were identified by *in vitro* and *in vivo* assays as cytochrome P450 enzymes responsible for biaryl and biaryl ether formation. The key substrate for these P450 enzymes is 2,4-dichlorophenol, which in turn is derived from the precursor 3-chloro-4-hydroxybenzoic acid. Here, the biosynthetic steps leading towards 3-chloro-4-hydroxybenzoic acid were investigated by *in vitro* assays. Ab7, an isoenzyme of a 3-deoxy-7-phosphoheptulonate (DAHP) synthase, is involved in chorismate biosynthesis by the shikimate pathway. Chorismate in turn is further converted by a dedicated chorismate lyase (Ab5) yielding 4-hydroxybenzoic acid (4-HBA). The stand alone adenylation domain Ab6 is necessary to activate 4-HBA, which is subsequently tethered to the acyl carrier protein (ACP) Ab8. The Ab8 bound substrate is chlorinated by Ab10 in *meta* position yielding 3-Cl-4-HBA, which is then transfered by the condensation (C) domain to the peptidyl carrier protein and released by the thioesterase (TE) domain of Ab9. The released product is then expected to be the dedicated substrate of the halogenase Ab1 producing the monomeric ambigol building block 2,4-dichlorophenol.

Introduction

The ambigols A, B, C, D and E (**1–5**, Fig. 1) are polychlorinated natural products produced by the terrestrial cyanobacterium *Fischerella ambigua*.[1–3] The observed activities for **1** encompassed antibacterial, antifungal and molluscicidal effects, as well as inhibition of cyclooxygenase and HIV reverse transcriptase.[1] The distinctive structural feature of the polyhalogenated compounds **1–5** is the presence of dichlorophenol units in which the chlorine atoms are in relative *meta* position.[4]

[a]*Institute for Insect Biotechnology, Justus Liebig University of Giessen, Heinrich-Buff-Ring 26–32, 35392 Giessen, Germany. E-mail: till.f.schaeberle@agrar.uni-giessen.de*
[b]*Department of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Ohlebergsweg 12, 35392 Giessen, Germany*
[c]*Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technical University of Munich, Lichtenbergstraße 4, 85748 Garching, Germany*
[d]*Institute for Infectious Diseases and Zoonoses, Ludwig-Maximilians-University Munich, Veterinärstraße 13, 80539 Munich, Germany*
[e]*Institute for Pharmaceutical Biology, University of Bonn, Nußallee 6, 53115 Bonn, Germany*
[f]*Kinderklinik, Hämatologisch-Onkologisches, Universitätsklinikum Bonn (AöR), Venusberg-Campus 1, Geb. 31, 53127 Bonn, Germany*
[g]*Chair of Technical Biochemistry, Technische Universität Dresden, Bergstraße 66, 01062 Dresden, Germany*
† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0ob02372h

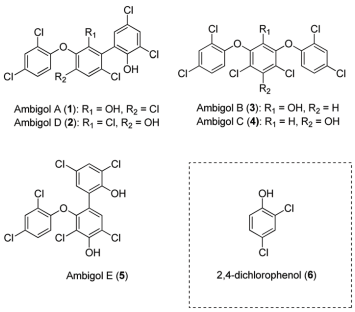


Fig. 1 The structures of ambigol A (**1**), D (**2**), B (**3**), C (**4**), E (**5**) and the proposed building block 2,4-dichlorophenol (**6**).

Therefore, 2,4-dichlorophenol (**6**) has been proposed as a biosynthetic precursor of the ambigols.[4]

The ambigol biosynthetic gene cluster (BGC) consists of ten genes (Fig. 2a) and was proven to be functional for ambigol biosynthesis by heterologous expression.[4] The aromatic phenol building block is assumed to be derived from the shikimic acid pathway. This hypothesis is supported by the presence of the genes *ab7* and *ab5*. In this respect, bioinformatic analysis suggested that the gene *ab7* encodes a 3-deoxy-7-phosphoheptulonate (DAHP) synthase, which catalyses the first dedicated step in chorismic acid biosynthesis.[4] The putative chorismate lyase Ab5 then converts chorismic acid to 4-hydroxybenzoic acid (4-HBA) (**7**, Fig. 2b), which is expected to be chlorinated by the putative halogenases Ab1 and Ab10, yielding compound **6**.[4] Biaryl coupling of this precursor was revealed to be catalysed by the cytochrome P450 enzymes Ab2 and Ab3, encoded in the BGC.[4]

While Ab2 and Ab3 have been recently investigated in detail, a comprehensive understanding of the halogenation mechanism (putatively catalysed by Ab1 and Ab10) is still lacking. Further biosynthetic gaps include the role of the acyl carrier protein (Ab8), the Non-Ribosomal Peptide Synthetase (NRPS) module (Ab9) and the predicted alone-standing adenylation domain (Ab6). In this work, we aim to complement the existing biosynthetic proposal by providing further insight into the formation of the precursor **11** as dichlorophenolic building block. In this sense, we used a chemoenzymatic *in vitro* approach to study the role of the enzymes Ab1, Ab6, Ab7, Ab8, Ab9, and Ab10.

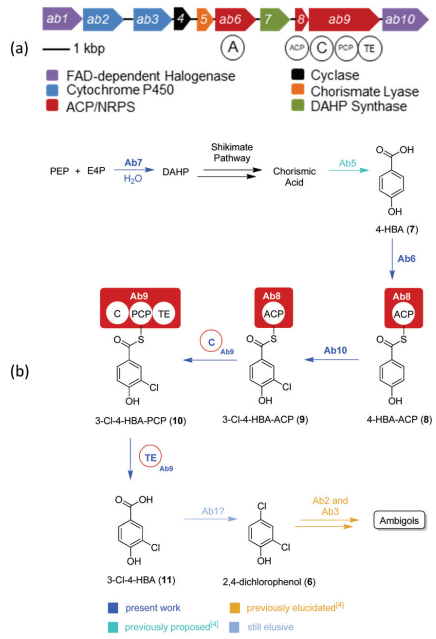


Fig. 2 (a) BGC encoding the ambigols. The domains of carrier protein/NRPS modules are given below the genes. A: adenylation; ACP: acyl carrier protein; C: condensation; PCP peptidyl carrier protein and TE: thioesterase domain. (b) The proposed pathway for ambigol biosynthesis. PEP: phosphoenolpyruvate and E4P: erythrose-4-phosphate.

Results and discussion

Ab7 (DAHP synthase)

The chlorinated phenolic building blocks of ambigol biosynthesis suggest a link to primary metabolism, *i.e.* the shikimate pathway. Sequence analysis of the protein Ab7 encoded in the BGC indicates it to be a 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase. Members of this enzyme family catalyse the first step in the biosynthesis of aromatic amino acids, as well as of aromatic cofactors like folate and quinones in microorganisms and plants.[5] Ab7 possesses the highly conserved DAHP synthase signature motifs DxxHxN, KPRT and xGxR (Table S1†) and the four amino acid residues involved in binding metal ions in an octahedral geometry to the active site are present (Table S2†).

DAHP enzymes usually regulate the amount of carbon that enters the shikimate pathway, thereby representing a potential bottleneck in specialized metabolite production. This was for example experimentally proven in the biosynthesis of the glycopeptide antibiotic balhimycin. A knockout of the *dahp* gene resulted in a significant reduction of balhimycin titres, while expressing an additional copy of *dahp* increased its production.[6] Hence, it was assumed that Ab7 initiates the first reaction of ambigol biosynthesis by providing sufficient precursors for the shikimate pathway by the interconnection of phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P). To verify the predicted activity of Ab7, it was heterologously overexpressed in *E. coli* cells, purified (Fig. S1†) and used as biocatalyst for the coupling of PEP and E4P to DAHP. In the ensuing spectrophotometric assay, DAHP activity was observed with a temperature optimum of 30 °C, while no activity remained at 50 °C (Fig. S2a†). The optimal pH was 7, with a broad maximum from 6 to 8 (Fig. S2b†). These conditions, neutral pH and a relatively low temperature optimum, reflect the growth conditions of the original ambigol producer *F. ambigua*, which was isolated from shallow water in Switzerland. Ab7 was active in the same range with different divalent metal cations added to the assay; EDTA treatment, however, abolished enzyme activity (Table S3†). This proved Ab7 to be a metalloenzyme. The K_M and k_{cat} values were determined to be 0.699 ± 0.375 μM and 0.281 ± 0.063 s^{-1}, respectively (Fig. S3†).

The next enzyme involved in precursor synthesis is Ab5. BLASTp alignment of Ab5 revealed several hits to chorismate lyase or 4-HBA synthetase with the highest similarity of 60.11% and a query cover of 86%. Since it was shown that chorismic acid production is promoted by the DAHP synthase

Ab7 and as we could also prove that Ab6 recognizes 4-HBA (**7**) as its substrate (see below), it seems reasonable to assign Ab5 the role of a chorismate lyase catalysing the production of 4-HBA (**7**) from chorismic acid. As the information obtained from BLASTp alignment was considered sufficient, no further *in vitro* characterisation of Ab5 was performed. For conversion of 4-HBA (**7**) to 2,4-dichlorophenol (**6**) the former has to be decarboxylated and chlorinated twice.

To study if the predicted halogenases Ab1 and Ab10 are halogenating 4-HBA (**7**), these enzymes were overexpressed, purified (Fig. S4 and S5†) and treated with 4-HBA (**7**) in the presence of Flavin reductase (Fre) (Fig. S6†). However, no production of mono- or di-chlorinated compounds was observed (Fig. S7†). This indicates that at least one halogenase acts on a carrier protein bound substrate instead of free substrate. A similar behaviour was previously described for example for the halogenases SgcC3[7] and PltA[8] (involved in the chlorination of C-1027 and pyoluteorin, respectively). Therefore, in a next step, the function of Ab6, Ab8 and the domains in Ab9 was analysed.

Ab6 (adenylate-forming enzyme)

Sequence analysis predicted Ab6 to be a freestanding adenylation (A) domain. To verify this prediction, it was tested if Ab6 activates 4-HBA (**7**). Accordingly, the enzyme was heterologously expressed and the purified version (Fig. S8a†) was used for the respective *in vitro* assays. CoA was used as a general mimic of the phosphopantheteine (PPT) arm of the respective carrier protein. This revealed that the activation proceeds by a two-step mechanism. First, 4-HBA reacted with ATP, thereby forming a mixed anhydride of AMP. Then, in a second step, AMP was displaced by coenzyme A (CoA), which could be observed photometrically and verified by LCMS (Fig. 3a–c).

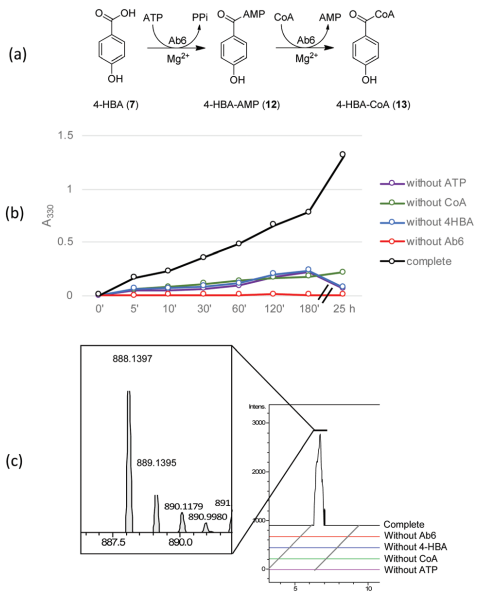


Fig. 3 (a) The reaction scheme of the Ab6 *in vitro* assay, (b) the generation of 4-HBA-CoA (**13**) over time (*x* axis) observed by the intensity of its specific UV absorption at 330 nm (ref. 9 and 10) (*y* axis) and (c) 4-HBA-CoA formation is confirmed by LCMS. The image shows the extracted ion chromatograms of the $[M + H]^+$ peak of 4-HBA-CoA (**13**) as well as a close-up of the respective mass spectrum (m/z = 888.1397; error = 4.3912 ppm). The complete LCMS chromatogram is shown in Fig. S9.†

Ab8 (acyl carrier protein)

Since 4-HBA (**7**) has been verified as the substrate of Ab6, we next addressed the question to which carrier protein the substrate will be attached. There are two annotated carrier proteins in the BGC – the ACP Ab8 and the PCP Ab9. The *in vitro* assay was performed using purified Ab8 or Ab9 (C-PCP) (Fig. S8†) as the substrate receiving protein, which was treated with **7** as substrate in the presence of Ab6 (Fig. 4a). The didomain Ab9 (C-PCP) was used instead the entire Ab9 (C-PCP-TE), since the latter construct could not be obtained in a soluble form in sufficient yield.

As described above, in the first reaction stage, Ab6 activates **7** and attaches it to a carrier protein. To achieve displacement of **7** in form of the corresponding amide from the carrier protein, the solution was treated with cysteamine, thus yielding compound **15** (Fig. 4a). In a first control trial, which was performed without any carrier protein but in the presence of Ab6, the product amide **15** could still be detect. This is attributed to reaction of the reactive mixed anhydride **12** with cysteamine (Fig. S10†). By HPLC purification of the respective carrier protein (Ab8 or Ab9 (C-PCP), respectively) prior to cysteamine treatment, the assay revealed that Ab6 favours the attachment of **7** to the ACP Ab8 over Ab9 (C-PCP) (Fig. 4b).

Ab10 (FAD-dependent halogenase)

In the ambigol BGC two putative halogenases are encoded by *ab1* and *ab10*. The *ab10* gene was annotated as FAD-dependent oxidoreductase, which is in accordance with the fact that most experimentally proven halogenases are $FADH_2$-dependent.[11,12] The proteins that showed the highest sequence identity to Ab10 that were proposed to catalyse halogenation reactions were McnD,[13] AerJ[14] and ApdC[15] (with an identity/query coverage of 64.82/97%, 64.46/96%, and 63.69/96%, respectively). Like those three proteins, Ab10 displays the two highly conserved motifs GxGxxG and WxWxIP characteristic for $FADH_2$-dependent halogenases[16] (Fig. S12†). Moreover, the three previously described halogenases have two main features in common: (i) they are encoded in proximity of NRPS gene clusters, and (ii) they act on free or carrier protein tethered aromatic substrates. Consequently, it was hypothesised that Ab10 catalyses the halogenation of Ab8-bound **7**.

To explore the reactivity of Ab10 *in vitro*, we synthesised the thioester 4-HBA-SNAC (**16**) (Fig. 5), which should mimic the

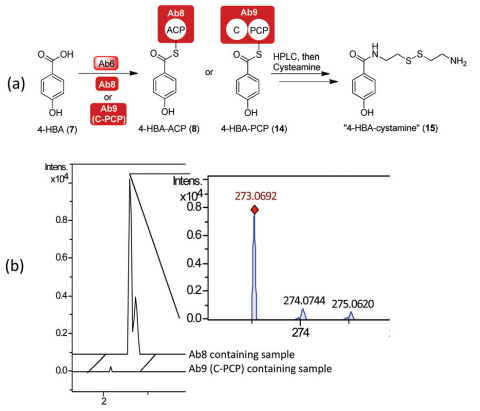


Fig. 4 (a) The reaction scheme underlying the *in vitro* assay testing carrier protein attachment to Ab8 and Ab9. (b) The LCMS data confirm the preferential formation of product **15** in samples employing Ab8 as a carrier protein. The image shows the extracted ion chromatograms of the $[M + H]^+$ peak of **15** as well as a close-up of the respective mass spectrum (m/z = 273.0692; error = 12.45 ppm). The full LCMS chromatogram is shown in Fig. S11.†

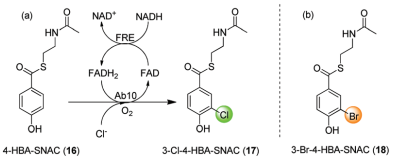


Fig. 5 (a) The proposed reaction catalysed by Ab10 and (b) the brominated product formed, if the reaction is carried out in the presence of bromide ions.

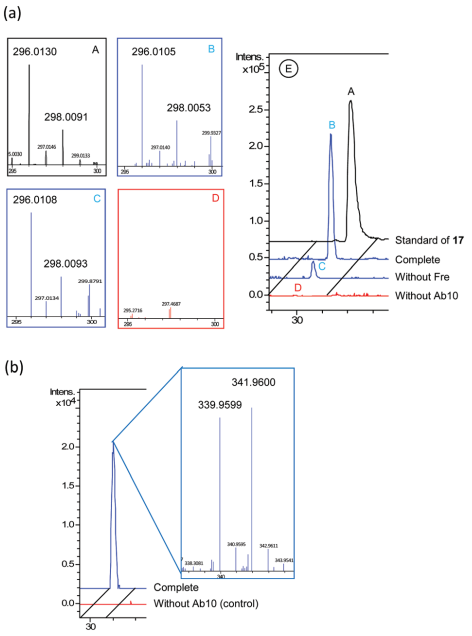


Fig. 6 (a) Images A–D show the HR-mass spectra obtained for different reaction conditions. Image E displays a close-up of the corresponding extracted ion chromatograms (EIC) of the $[M + Na]^+$ peak of the product **17**, which elutes at a retention time of 30.1 min. Observed m/z for $[M + Na]^+$ of *A* = 296.0130 (error = 3.7160 ppm), *B* = 296.0105 (error = 4.7295 ppm) and *C* = 296.0108 (error = 3.7161 ppm). The full LCMS chromatogram is shown in Fig. S13.† (b) The LCMS data confirm the substrate flexibility of Ab10 with respect to bromination. The image shows a close-up of the extracted ion chromatogram corresponding to the $[M + Na]^+$ peak of the expected compound **18**, which elutes at a retention time of 30.7 min. The respective mass spectrum shows the $[M + Na]^+$ peak of **18** at m/z = 339.9599 (error = 4.1181 ppm), also displaying the characteristic isotopic pattern for mono-bromination. The full LCMS chromatogram is shown in Fig. S14.†

natural carrier protein-tethered substrate for the halogenation reaction. This thioester motif resembles the part of the PPT arm that is orientated towards the carrier protein-attached substrate.[17] The other components for the *in vitro* assay, namely Ab10 (Fig. S5†) and the Flavin reductase (Fre) of *E. coli* (Fig. S6†), were obtained by heterologous expression and subsequent purification. The latter was used in the assay to regenerate the co-factor $FADH_2$, which is essential for the reaction.[18]

Using the SNAC thioesters mimicking the Ab8-tethered substrate allowed the halogenation reaction to take place. The need for co-factor regeneration by Fre was obvious, since without Fre, only minor amounts of the product 3-Cl-4-HBA-SNAC (**17**) were detected (Fig. 6a, comparison between B and C). The specificity of Ab10 for the bound substrate makes sense for the organism to prevent random halogenation of **7**, which is also part of primary metabolism and thus a common molecule in the cell.

In a next step, the flexibility of Ab10 to attach other halogens was tested. Therefore, the chloride ions in the assay were exchanged against bromide or iodide ions, respectively. Analysis of the samples by HR-ESI-LCMS revealed that the brominated product **18** was formed (Fig. 6b), while an iodinated derivative could not be detected (Fig. S15†). This is in accordance with the fact that terrestrial organisms mostly produce chlorinated compounds; however, for bacteria, cultivated in the presence of bromide ions instead of chloride ions, it was often shown before that they possess the ability to produce bromo-analogues of the natural chloro-metabolites.[19–22] However, *in vivo* experiments to replace the chlorine substituents in the ambigols with bromine by altering the composition of the cultivation medium were unsuccessful.[2] The reason for this phenomenon is probably attributable to cellular regu-

lation. Yet, the underlying mechanism is still elusive and needs further investigation.

Ab9 (NRPS module)

The dedicated NRPS module Ab9 encoded in the ambigol BGC consists of a condensation (C), a peptide carrier protein (PCP) and a thioesterase (TE) domain. Since there is no amide/peptide bond in the ambigol structures, we suspect that the C domain acts as a gatekeeper towards Ab8-bound 3-Cl-4-HBA (**9**), which is formed as a product upon action of Ab10, and transfers it to Ab9-PCP. This hypothesis was tested in an assay consisting of Ab6, Ab9 (C-PCP) and 4-HBA-SNAC (**16**) or 3-Cl-4-HBA-SNAC (**17**), respectively, as artificial substrate mimics (Fig. 7a). When 4-HBA-SNAC (**16**) was used as substrate, **15** could not be detected (Fig. S16†), indicating that ACP-bound 4-HBA is not the substrate for the C domain of Ab9. However, when 3-Cl-4-HBA-SNAC (**17**) was applied as substrate, the product **19** could be observed (Fig. 7b); thereby, demonstrating that ACP-bound 3-Cl-4-HBA is recognized by the C domain and that the latter acts as a gatekeeper towards the halogenated ACP-bound **9**.

The ensuing TE domain of Ab9 is likely to promote hydrolysis releasing 3-Cl-4-HBA (**11**) from Ab9 (Fig. 2b). This assumption was put to a test by a suitable *in vitro* experiment. As a prerequisite for the intended *in vitro* experiment, the Ab9 di-domain (PCP-TE) was heterologously expressed and purified (Fig. S18†). As substrate 3-Cl-4-HBA-SNAC (**17**) was used, which was synthesised as mimic of the natural substrate. The TE domain was expected to catalyse the hydrolysis of the thioester bond, either in a natural setup between substrate and PPT

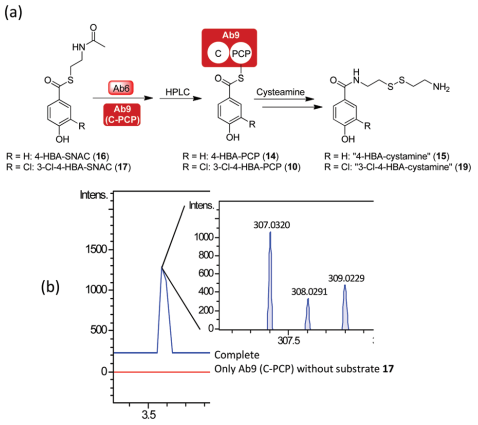


Fig. 7 (a) The reaction scheme underlying the *in vitro* assay investigating the role of the C domain of Ab9. (b) The LCMS chromatogram shows the formation of **19** only in the presence of 3-Cl-4-HBA-SNAC (**17**) proving the transfer of **17** to the PCP domain of Ab9. The image shows the extracted ion chromatograms of the $[M + H]^+$ peak of **19** as well as a close-up of the respective mass spectrum (m/z = 307.0320, error = 5.21). The full LCMS chromatogram is shown in Fig. S17.†

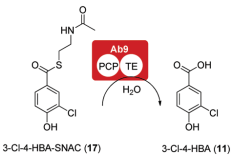


Fig. 8 The proposed reaction of TE domain from Ab9.

arm, or from the synthetic SNAC-substrate 3-Cl-4-HBA-SNAC (**17**) (Fig. 8). To examine this hypothesis, the artificial substrate **17** was incubated with Ab9 (PCP-TE). Unfortunately, LCMS analysis of substrate **17** revealed the presence of minor amounts of the expected hydrolysed product 3-Cl-4-HBA (**11**) (Fig. S19†). It thus is likely that a small amount of the thioester 3-Cl-4-HBA-SNAC (**17**) is spontaneously cleaved under aqueous conditions to the corresponding carboxylic acid 3-Cl-4-HBA (**11**). To prevent an erroneous conclusion, the *in vitro* assay was monitored over time and compared to standards of 3-Cl-4-HBA-SNAC (**17**) and 3-Cl-4-HBA (**11**) as well as to control experiments (without Ab9 (PCP-TE)) to observe the increase of 3-Cl-4-HBA (**11**) and decrease of 3-Cl-4-HBA-SNAC (**17**). Indeed, our LCMS-based assay confirmed that Ab9 (PCP-TE) did catalyse hydrolysis of the thioester bond, resulting in 3-Cl-4-HBA (**11**) (Fig. 9a). In our assay, the enzyme consumed the substrate 3-Cl-4-HBA-SNAC (**17**) in a time-dependent manner over 24 h.

In a next step, the specificity of the TE for the hydrolysis of the non-chlorinated SNAC substrate was investigated. When 4-HBA-SNAC (**16**) was employed as substrate, hydrolysis to 4-HBA (**7**) was observed (Fig. 9b). This indicates that the TE domain from Ab9 does not have strict substrate specificity and is consequently not acting as a gatekeeper to the non-halogenated substrate.

The specificity of Ab6 with respect to the product released by the TE domain, *i.e.* 3-Cl-4-HBA (**11**), was also investigated using our LCMS-based assay. Using the same experimental setup as before, we could not detect the m/z of 3-Cl-4-HBA-CoA, indicating that 3-Cl-4-HBA (**11**) is not recognized by Ab6 as a substrate (Fig. S23†).

Ab1 (putative FAD-dependent halogenase)

An initial BLASTp alignment of Ab1 against the database resulted in tryptophan 7-halogenases and NAD(P)/FAD-dependent oxidoreductases as closest matches, while a further refined *in silico* analysis suggested resemblance of Ab1 to the Flavin-dependent halogenases (FDH). This assignment was additionally supported by the presence of the characteristic GxGxxG and WxWxIP motifs[16] (Fig. S24†). As 3-Cl-4-HBA (**11**) is the hydrolysis product obtained by the action of Ab9, it seems logical to assume, that **11** serves as the substrate of Ab1, thereby producing 2,4-dichlorophenol (**6**). However, to date we were not able to verify this assumption, as treatment of **11** with heterologously expressed Ab1 did not show production of **6** under various conditions (Fig. S25†). As it is possible that also Ab1 acts on carrier protein attached substrates, an *in vitro*

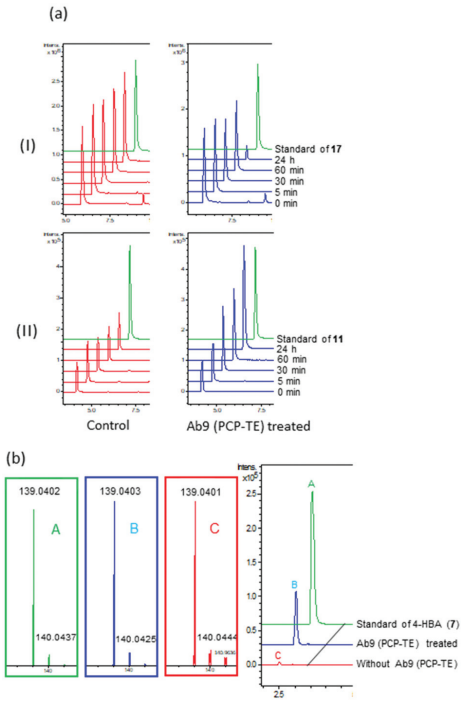


Fig. 9 (a) LCMS analysis of the *in vitro* assay of the Ab9 (PCP-TE). Shown are the extracted ion chromatograms of the $[M - H]^-$ peak of (I) **17** and (II) **11** obtained by LCMS measurement in negative mode. Red lines represent the control reaction (without Ab9 (PCP-TE)), blue lines represent the assay with Ab9 (PCP-TE); the respective standard is given in green. The time-dependent increase of the product **11** and the consumption of the substrate **17** can be observed. The full LCMS chromatograms are shown in Fig. S20 and S21.† (b) Investigation of the specificity of Ab9 (PCP-TE) using 4-HBA-SNAC (**16**) as the substrate. Assay was incubated for 24 hours. The graphic displays the extracted ion chromatograms of the $[M + H]^+$ peak of 4-HBA (**7**) and the corresponding mass spectra obtained by LCMS measured in positive mode. Green, blue and red lines represent the standard of **7**, the product of the Ab9 (PCP-TE) treated reaction and the control sample, respectively. Traces of **7** were also detected in the control without enzyme, indicating a slow uncatalysed hydrolysis reaction. The full LCMS chromatograms are shown in Fig. S22.†

assay similar to the Ab10 assay was done using Ab1 and 3-Cl-4-HBA-SNAC (**17**) as the substrate. However, formation of **6** could not be detected and also addition of Ab9 (PCP-TE) did not result in 2,4-dichlorophenol (**6**) production (Fig. S26†).

Transformation of **11** to **6** involves decarboxylation as well as chlorination. In case Ab1 affects only the chlorination step, while decarboxylation proceeds spontaneously or by an unspecific decarboxylase, 2-chlorophenol should be a possible substrate of Ab1. Yet, this hypothesis still awaits experimental confirmation and preliminary experiments for the incubation of 2-chlorophenol with Ab1 did not result in the expected product **6** (Fig. S27†). Alternatively, decarboxylation and chlorination could take place simultaneously. Thus, further investigation of this reaction is necessary.

Conclusions

Based on previous knowledge on the role of Ab2 and Ab3[4] and our recently obtained results on the function of Ab6, Ab7, Ab8, Ab9, and Ab10, we were able to gain further understanding of ambigol biosynthesis. We therefore propose a biosynthetic pathway for the formation of 3-chloro-4-hydroxybenzoic acid (**11**), which is a precursor of the ambigol building block 2,4-dichlorophenol (**6**), as depicted in Fig. 2b. The results of our *in vitro* assays show that (i) Ab6 activates 4-HBA (**7**) in a two-step procedure, thereby leading to the formation of 4-HBA-CoA (**13**). (ii) The activated substrate is attached to the ACP Ab8 instead of PCP of Ab9. (iii) The Flavin-dependent halogenase Ab10 is responsible for the conversion of 4-HBA-SNAC (**16**) to 3-Cl-4-HBA-SNAC (**17**), thus proving that Ab10 is only acting on carrier protein bound substrates. As seen above, the relevant carrier protein is the ACP Ab8. (iv) Interestingly, Ab10 is flexible enough to enable bromination as well as chlorination. (v) The C domain from Ab9 acts as a gatekeeper towards Ab8-bound 3-Cl-4-HBA. (vi) The TE domain of Ab9 finally accomplishes the release of 3-Cl-4-HBA (**11**), as proven by the hydrolysis of 3-Cl-4-HBA-SNAC (**17**). (vii) The TE domain has no strict substrate specificity towards non-halogenated substrates as was shown by treatment with 4-HBA-SNAC (**16**). (viii) As the released 3-Cl-4-HBA is not a substrate of Ab6, side reactions are avoided. (ix) In spite of initial experiments, the role of Ab1 still remains elusive and requires further investigation.

Material and method

Plasmid transformation, general cultivation and protein purification

Transferring plasmid(s) into a host was always done by electroporation. Electro-competent cells were always freshly prepared and used at the same day. All the steps were carried out at 4 °C. 0.5 mL of an over night grown *E. coli* culture was inoculated in 50 mL of Luria Bertani (LB) medium and incubated at 37 °C, 180 rpm until reaching an OD_{600} of 0.4–0.6. The culture was harvested by centrifugation (10 000 rpm for 3 min) and the cell pellet was washed 3 times with 50 mL of 10% sterile ice-cold glycerol. Finally, the pellet was resuspended in 500 µL of 10% sterile glycerol solution and was ready to use as electro-competent cells. For transferring a plasmid into *E. coli* cells, 5–15 µL of purified plasmid were added to 50 µL of electro-competent cells and loaded into the pre-chilled electroporation cuvette with a diameter of 1 mm. A voltage of 25 kV cm^{-1} was applied in a Biorad MicroPulser™. Cells were immediately recovered by 1 mL LB medium and incubated in 2 mL tubes at 37 °C for 1 h. Cells were pelleted, most of the supernatant dis-

carded and the pellet was resuspended in the last drops. The suspension was spread on agar plates containing appropriate antibiotic(s) for selection and incubated overnight at 37 °C.

1% (v/v) of an overnight culture of *E. coli* cells carrying the corresponding construct(s) and grown in a mixture of LB medium and a suitable antibiotic(s) was used to inoculate 1 L of a mixture of LB medium containing the respective antibiotic(s) and incubated at 37 °C, 180 rpm until reaching OD_{600} of about 0.5. Subsequently, the culture was induced with a final IPTG concentration of 0.3 mM and incubation was continued at 18 °C overnight. Thereafter, the culture was centrifuged at 4 °C. The obtained pellet was lysed under cooling (ice bath) using ultrasound (70% power, 5 × 1 minute). The lysate was separated from the cell debris by centrifugation at maximum speed for about 30 min. The supernatant was further purified by use of Ni-NTA resin (Qiagen), following the manufacture's protocol.

LCMS measurement

UPLC-HRMS measurements were performed on an Agilent Infinity 1290 UPLC system equipped with an Acquity UPLC BEH C18 1.7 μm (2.1 × 100 mm) column and an Acquity UPLC BEH C18 1.7 μm VanGuard Pre-Column (2.1 × 5 mm; both columns were purchased from Waters) setup coupled to a DAD detector and a micrOTOFQ II mass spectrometer (Bruker). The LC part was run using a gradient (A: H_2O, 0.1% FA; B: MeCN, 0.1% FA; Flow: 600 μL min^{-1}): 0 min: 95%A; 0.80 min: 95%A; 18.70 min: 4.75%A; 18.80 min: 0%A; 23.00 min: 0%A; 23.10 min: 95%A; 25.00 min: 95%A and the column oven temperature was set to 45 °C.

In vitro assay of DAHP synthase, Ab7

The sequence encoding Ab7 was amplified by PCR using primers containing the sequence of restriction sites *Eco*RI and *Hind*III. The resulting PCR product was subcloned into a pGEM-T vector and transferred to *E. coli* XL1Blue. After verifying the amplificate by Sanger sequencing, the *ab7*-containing sequence was excised from the pGEM-T vector using *Eco*RI and *Hind*III. The fragment was gel purified and subsequently ligated into the expression vector pET28, which resulted in the construct pET.DAHP.Fa. Afterwards, the expression vector was transferred to *E. coli* BL21 (DE3) cells yielding the strain *E. coli* BL21 (DE3)/pET.DAHP.Fa. The latter strain served subsequently as expression host.

For the assay, a standard reaction mixture containing 80 μM PEP, 350 μM E4P, 100 μM $MnCl_2$ or $MnSO_4$ in Bis-tris propane buffer pH 5–9 was incubated with purified Ab7. Consumption of PEP in the reaction mixture was determined by a continuous spectrophotometric assay at a UV absorption of 232 nm. For temperature and pH optimisation of the assay, the mixture was pre-incubated for 5 minutes at the desired temperature and pH, and then the reaction was started by addition of the enzyme. The reaction was monitored after 5 or 10 minutes incubation. For assay optimisation with respect to the employed divalent metal ion, the mixture used was essentially the same, except for the added metal salt.

In vitro assay to confirm the substrate of Ab6

The sequence encoding Ab6 was amplified by PCR adding the *Bam*HI and *Xho*I restriction sites to the primers used. The PCR product was subcloned into a pGEM-T vector and transferred to *E. coli* XL1Blue. After verifying the amplificate by Sanger sequencing, the *ab6*-containing sequence was excised from the pGEM-T vector using *Bam*HI and *Xho*I. Then, the fragment was gel purified and subsequently ligated into the expression vector pET28, which resulted in the construct pET-Ab6. Afterwards, the vector was transferred to expression host *E. coli* BL21 (DE3) cells, yielding the strain *E. coli* BL21 (DE3)/pET-Ab6.

From a 1 L cultivation, Ab6 was obtained with a yield of 1.38 mg after purification (Fig. S5†). The concentration in the elution buffer was 4.6 mg mL^{-1} as analysed by the Bradford method. For the *in vitro* assay, the standard reaction mixture contained 5 mM $MgCl_2$, 0.5 mM ATP, 0.25 mM CoA, 0.6 mM substrate (stock substrate solution was dissolved in DMSO) and 10 μL purified Ab6 in 50 mM Tris-Cl buffer at pH 9.2. Generation of the product 4-HBA-CoA (**13**) was monitored at a UV absorption of 330 nm.[9,10] To confirm product formation after incubation, the sample was analysed by LCMS measurement.

In vitro assay of Fre

Fre was purified from 1 L cultivation volume (Fig. S6†) and obtained in 3.15 mg yield as determined by Bradford assay. For the *in vitro* assay, besides adding Fre, three different conditions were applied: (1) 25 mM HEPES buffer (pH 7.5), 1 mM NADH, 0.05 mM FAD and 10 mM NaCl;[18] (2) 10 mM phosphate buffer (pH 7.2), 2.4 mM NADH, 0.01 mM FAD and 25 mM NaCl;[23] (3) 75 mM Tris-Cl (pH 7.5), 4 mM NADH, 0.1 mM FAD and 10 mM NaCl.[8] Assays were performed in 2 mL final volume scale. At each time point of measurement, 400 μL were removed from the reaction tube and used in the spectrophotometric measurement. All the assays were incubated at 30 °C for 5, 30, 60, 120 and 180 minutes. The activity of Fre was determined by measuring the decrease of the absorbance at 340 nm due to the oxidation of NADH.[24] The corresponding buffer was used as a blank. The control was the standard reaction mixture without Fre. Enzymatic activity was observed in all buffer systems tested (Table S4 and Fig. S28†). For the initial Ab10 assay, phosphate buffer was selected.

In vitro assay of Flavin-dependent halogenases

The *ab1* gene was amplified by PCR and the *Bam*HI and *Sal*I restriction sites were added by the primers. The PCR product was subcloned into a pGEM-T vector and transferred to *E. coli* XL1Blue cells. After sequence verification, *ab1* was excised by restriction digestion using *Bam*HI and *Sal*I. Then, the gel purified DNA fragment was ligated into the expression vector pET28 that was linearized before using the same restriction enzymes, which resulted in the construct pET-Ab1. The assembled vector was transferred to the expression host *E. coli* BL21 (DE3), yielding the strain *E. coli* BL21 (DE3)/pET-Ab1.

In the case of Ab10, a codon-optimized version of *ab10* was synthesized (GenScript). The codon-optimized version was cloned into pET28a using the restriction enzymes *Eco*RI and *Hind*III. The resulting plasmid was transferred to the expression host *E. coli* BL21 (DE3), containing plasmid pGROe7. pGROe7 is a helper plasmid encoding for the chaperones GroEL and GroEs.

Purification of the cell lysate obtained from 1 L cultivation volume each (Fig. S4 and S5†), yielded 4.2 mg of protein Ab1 and 2.48 mg of protein Ab10 as determined by Bradford assay. For the *in vitro* assay, 10 µL of the Flavin-dependent halogenase (Fre) and a substrate aliquot (stock solution was dissolved in DMSO) corresponding to a final concentration of 0.6 mM was added. The reaction was performed under 3 different conditions, which are as follows: (1) 75 mM Tris-Cl (pH 7.5), 4 mM NADH, 0.1 mM FAD, and 10 mM NaCl;[8] (2) 10 mM phosphate buffer (pH 7.2), 2.4 mM NADH, 0.01 mM FAD, and 25 mM NaCl;[23] (3) 25 mM HEPES buffer (pH 7.5), 1 mM NADH, 0.05 mM FAD, and 10 mM NaCl.[18] The reaction was performed in 50–100 µL final volume. The mixture was vortexed prior to incubation at 30 °C for 24 hours. For a 100 µL reaction, quenching was achieved by adding 7.5 µL MeCN containing 0.25% TFA, and 42.5 µL methanol. The solution was vortexed and centrifuged at maximum speed for 2 minutes. The clear supernatant was subjected to LCMS measurement. The results showed no difference concerning the outcome of the assay with respect to the three different buffer systems (Fig. S29†). Without any preference, phosphate buffer was selected for the further experiments.

In vitro assay of carrier proteins Ab8 and Ab9 (C-PCP)

The *ab8* gene was amplified by PCR and the *Bam*HI and *Sal*I restriction sites were added by the primers. The PCR product was subcloned into a pGEM-T vector and transferred to *E. coli* XL1Blue cells. After sequence verification, *ab8* was excised by restriction digestion using *Bam*HI and *Sal*I. Then, the gel purified DNA fragment was ligated into the expression vector pET28 that was linearized before using the same restriction enzymes, which resulted in the construct pET-Ab8. The assembled vector was transferred to the expression host *E. coli* BAP1 containing plasmid pGROe7, yielding *E. coli* BAP1/pET-Ab8/pGROe7. pGROe7 is a helper plasmid encoding for the chaperones GroEL and GroEs.

For Ab9 (C-PCP), the nucleotide sequence encoding the C and PCP domains from Ab9 was amplified using PCR. *Eco*RI and *Not*I restriction sites were added by the primers. The boundaries of the domains were determined *in silico* using NRPSpredictor2 and PHYRE2. The PCR product was subcloned and transferred to *E. coli* XL1Blue cells. After sequence verification, *ab9* (C-PCP) was excised by restriction digestion and ligated into the beforehand linearized expression vector pET28, yielding pET-Ab9-C-PCP. This vector was transferred to the expression host *E. coli* BAP1 containing plasmid pGROe7, yielding the strain *E. coli* BAP1/pET-Ab9-C-PCP/pGROe7.

Purification of Ab8 and Ab9 (C-PCP) (Fig. S8†) was achieved from the cell lysate grown in 0.5 L cultivation volume each, yielding 1.1 and 1.44 mg of protein Ab8 and Ab9 (C-PCP) respectively as determined by Bradford assay. The concentration in the elution buffer was 2.2 and 3.6 mg mL^{-1} for Ab8 and Ab9 (C-PCP) respectively as analysed by the Bradford method. For the *in vitro* assay, the standard reaction mixture contained 5% v/v DMSO, 1 mM TCEP, 10 mM $MgCl_2$, 1 mM 4-HBA, 1 mM ATP, 15 µL purified Ab6, 15 µL purified carrier protein (Ab8 or Ab9 (C-PCP)) in 100 mM Tric-Cl buffer (pH 7.5). The mixtures were incubate for 4 hours at 30 °C. Subsequently, the reaction mixtures were treated with cysteamine to release the substrate from the attached carrier protein according to a literature procedure.[25] To confirm product formation, samples were analysed by LCMS.

For experiments in which the carrier protein was purified prior to cysteamine treatment, HPLC (Shimadzu) was employed. The column was a Symmetry300™ C4 5 µM, 2.1 × 150 mm (Waters). For elution the following gradient was used (A: H_2O, 0.1% TFA; B: MeCN; Flow: 1 mL min^{-1}): 0 min: 80%A; 7.5 min: 80%A; 27.5 min: 0%A; 35 min: 0%A; 35–40 min: 80% A and the column oven temperature was set to 40 °C. Ab8 was collected at retention time of 16.8–17.2 min, while Ab9 (C-PCP) was collected at 19.0–19.6 min. The collected fractions were dried *in vacuo* using a HT12-II centrifugal concentrator (Genevac,) and redissolved in 800 µL H_2O : DMSO (v/v = 1 : 1). Thereafter, the solutions were treated with cysteamine according to a literature procedure,[25] centrifuged at maximum speed for 5 min and the clear supernatant was analysed by LCMS.

In vitro assay of C domain from Ab9 as a gatekeeper

The standard reaction mixture contained 5% v/v DMSO, 1 mM TCEP, 1 mM substrate **16** or **17**, 15 µL purified Ab9 (C-PCP) from stock solution (3.6 mg mL^{-1}) in 100 mM Tric-Cl buffer (pH 7.5). The mixtures were incubated for 4 hours at 30 °C. Prior to cysteamine treatment, the Ab9 (C-PCP) was purified using HPLC as described in the previous paragraph.

In vitro assay of thioesterase (TE)

The sequence encoding the PCP and TE domains from Ab9 (PCP-TE) was amplified using PCR. *Eco*RI and *Not*I restriction sites were added by the primers. The boundaries of the domains were determined *in silico* using NRPSpredictor2 and PHYRE2. The PCR product was subcloned and transferred to *E. coli* XL1Blue cells. After sequence verification, *ab9* (PCP-TE) was excised by restriction digestion and ligated into the beforehand linearized expression vector pET28, yielding pET-Ab9-PCP-TE. This vector was transferred to the expression host *E. coli* BAP1, yielding the strain *E. coli* BAP1/pET-Ab9-PCP-TE.

Purification of the cell lysate obtained from 1 L cultivation volume (Fig. S18†), yielded 0.7 mg of the protein Ab9 (PCP-TE) as determined by Bradford assay. The standard *in vitro* reaction mixture for a 50 µL reaction consisted of 40 µL of 10 mM phosphate buffer (pH 7.2), containing 0.6 mM substrate and 10 µL of the purified Ab9 (PCP-TE). The solution was vortexed prior to incubation at 30 °C for 0, 5, 30, 60 min and 24 hours. The reaction was quenched by adding 7.5 µL MeCN containing 0.25% TFA, and 42.5 µL methanol. The solution was vortexed

and centrifuged at maximum speed for 2 minutes. The clear supernatant was subjected to LCMS measurement.

Synthesis of *S*-2-acetamidoethyl 4-hydroxybenzothioate (4-HBA-SNAC) (16)

The thioester was prepared according to a published procedure.[26] Triethylamine (825 μL, 599 mg, 5.92 mmol, 2.1 eq.) was added to a Schlenk flask and diluted with dry DCM (21.0 mL, 7.50 mL $mmol^{-1}$). 4-Hydroxybenzoic acid (7) (409 mg, 2.96 mmol, 1.05 eq.), EDC·HCl (567 mg, 2.96 mmol, 1.05 eq.), HOBt (400 mg, 2.96 mmol, 1.05 eq.) and *N*-acetylcysteamine (300 μL, 336 mg, 2.82 mmol, 1.0 eq.) were added and the mixture was stirred for 24 h at room temperature. The organic phase was washed with sat. $NaHCO_3$ solution, 0.5 M HCl solution (2×), brine, dried over $MgSO_4$ and filtered. The solvent was removed and the obtained residue was purified by MPLC (column: Reveleris C18–40 g; solvents: A: H_2O + 0.05% TFA, B: MeCN + 0.05% TFA; flow rate: 36 mL min^{-1}; gradient: 0–2 min 5% B, 2–28 min: to 95% B, 28–30 min 95% B) to give desired product **16** (15%, 103 mg, 0.43 mmol) as colorless oil. Analytical data were in agreement with literature values.[26]

^{1}H-NMR (300 MHz, DMSO-d_6): δ = 8.12 (t, J = 5.7 Hz, 1 H), 7.80 (d, J = 8.8 Hz, 2 H), 6.87 (d, J = 8.7 Hz, 2 H), 3.25 (dt, J = 6.1 Hz, 2 H), 3.06 (t, J = 6.5 Hz, 2 H), 1.80 (s, 3 H) (Fig. S30†). **^{13}C-NMR** (75 MHz, DMSO-d_6): δ = 189.0, 169.5, 162.8, 129.4, 127.7, 115.6, 38.5, 28.0, 22.6 (Fig. S31 and 32†). **MS** (ESI+): m/z = 500.8 $[2M + Na]^+$, 479.1 $[2M + H]^+$, 262.1 $[M + Na]^+$.

Synthesis of *S*-2-acetamidoethyl 3-chloro-4-hydroxybenzothioate (3-Cl-4-HBA-SNAC) (17)

Thioester **17** was prepared in analogous fashion as compound **16** using 3-chloro-4-hydroxybenzoic acid (**11**) (511 mg, 2.96 mmol, 1.05 eq.) instead of 4-hydroxybenzoic acid (7). The desired product (3%, 23.0 mg, 84.0 μmol) was obtained as white solid.

^{1}H-NMR (300 MHz, DMSO-d_6): δ = 11.41 (s, 1 H), 8.11 (t, J = 5.6 Hz, 1 H), 7.84 (d, J = 2.2 Hz, 1 H), 7.77 (d, J = 8.5, 2.2 Hz, 1 H), 7.09 (d, J = 8.5 Hz, 1 H), 3.25 (dt, J = 6.1 Hz, 2 H), 3.08 (t, J = 6.5 Hz, 2 H), 1.80 (s, 3 H) (Fig. S33†). **^{13}C-NMR** (75 MHz, DMSO-d_6): δ = 188.3, 169.4, 158.3, 128.7, 128.4, 127.8, 120.4, 116.6, 38.2, 28.3, 22.6 (Fig. S34 and 35†). **MS** (ESI+): m/z = 547.2 $[2M + H]^+$, 273.8 $[M + H]^+$. **HRMS** (ESI+) m/z calcd for $C_{11}H_{12}ClNNaO_3S$ $[M + Na]^+$: 296.0119, found: 296.0130.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Tilman Weber (Technical University of Denmark, DTU) for providing the Fre expression vector. I D. M. Kresna thanks to the Deutscher Akademischer Austauschdienst (DAAD) for the Ph.D. scholarship. Luis Linares Otoya thanks Fondo para la Innovación, la Ciencia y la Tecnología (FINCyT-Innovate Peru) for his fellowship. E. R. D. thanks the Deutsche Bundesstiftung Umwelt (DBU) for her Ph.D. scholarship and CIPSM-Women for funding. T. M. M. thanks the Stiftung der Deutschen Wirtschaft (sdw) for funding. T. A. M. G. thanks the DFG for generous financial support of the work in our laboratory (Emmy Noether program (GU 1233/1-1).

References

1 B. S. Falch, G. M. König, A. D. Wright and O. Sticher, *J. Org. Chem.*, 1993, **58**, 6570–6575.
2 A. D. Wright, O. Papendorf and G. M. König, *J. Nat. Prod.*, 2005, **68**, 459–461.
3 T. Chilczuk, R. Monson, P. Schmieder, V. Christov, H. Enke, G. Salmond and T. H. J. Niedermeyer, *ACS Chem. Biol.*, 2020, **15**, 2929–2936.
4 E. R. Duell, T. M. Milzarek, M. E. Omari, L. J. Linares-Otoya, T. F. Schäberle, G. M. König and T. A. M. Gulder, *Org. Chem. Front.*, 2020, **7**, 3193–3201.
5 K. M. Hermann, *Plant Physiol.*, 1995, **107**, 7–12.
6 J. Thykaer, J. Nielsen, W. Wohlleben, T. Weber, M. Gutknecht, A. E. Lantz and E. Stegmann, *Metab. Eng.*, 2010, **12**(5), 455–461.
7 S. Lin, S. G. V. Lanen and B. Shen, *J. Am. Chem. Soc.*, 2007, **129**, 12432–12438.
8 P. C. Dorrestein, E. Yeh, S. Garneau-Tsodikova, N. L. Kelleher and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**(39), 13843–13848.
9 J. Gibson, M. Dispensa, G. C. Fogg, D. T. Evans and C. S. Harwood, *J. Bacteriol.*, 1994, **176**, 634–641.
10 L. T. Webster Jr., J. J. Mieyal and U. A. Siddiqui, *J. Biol. Chem.*, 1974, **249**(8), 2641–2645.
11 F. H. Vaillancourt, E. Yeh, D. A. Vosburg, S. Garneau-Tsodikova and C. T. Walsh, *Chem. Rev.*, 2006, **106**(8), 3364–3378.
12 K. H. van Pée and E. P. Patallo, *Appl. Microbiol. Biotechnol.*, 2006, **70**, 631–641.
13 A. Tooming-Klunderud, T. Rohrlack, K. Shalchian-Tabrizi, T. Kristensen and K. S. Jakobsen, *Microbiology*, 2007, **153**, 1382–1393.
14 S. Cadel-Six, C. Dauga, A. M. Castets, R. Rippka, C. Bouchier, N. T. de Marsac and M. Welker, *Mol. Biol. Evol.*, 2008, **25**(9), 2031–2041.
15 L. Rouhiainen, L. Paulin, S. Suomalainen, H. Hyytiäinen, W. Buikema, R. Haselkorn and K. Sivonen, *Mol. Microbiol.*, 2000, **37**(1), 156–167.
16 K. H. van Pée and S. Zehner, *Handb. Environ. Chem.*, 2003, **3**, 171–199; K. H. van Pée and S. Zehner, *Handb. Environ. Chem.*, 2003, **3**, 171–199.
17 J. Franke and C. Hertweck, *Cell Chem. Biol.*, 2016, **23**, 1179–1192.
18 E. Yeh, S. Garneau and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**(11), 3960–3965.
19 C. G. Smith, *J. Bacteriol.*, 1958, **75**(5), 577–583.

20 A. P. Doerschuk, J. R. D. McCormick, J. J. Goodman, S. A. Szumski, J. A. Growich, P. A. Miller, B. A. Bitler, E. R. Jensen, M. Matrishin, M. A. Petty and A. S. Phelps, *J. Am. Chem. Soc.*, 1959, **81**(12), 3069–3075.
21 K. H. van Pée, O. Salcher, P. Fischer, M. Bokel and F. Lingens, *J. Antibiot.*, 1983, **36**, 1735–1742.
22 B. Bister, D. Bischoff, G. J. Nicholson, S. Stockert, J. Wink, C. Brunati, S. Donadio, S. Pelzer, W. Wohlleben and R. D. Süssmuth, *ChemBioChem*, 2003, **4**, 658–662.
23 S. Zehner, A. Kotzsch, B. Bister, R. D. Süssmuth, C. Méndez, J. A. Salas and K. H. van Pée, *Chem. Biol.*, 2005, **12**, 445–452.
24 J. K. Lee and H. Zhao, *J. Bacteriol.*, 2007, **189**(23), 8556–8563.
25 H. P. Mguyen and K. Yokoyama, *Biochemistry*, 2019, **58**, 2804–2808.
26 G. Prasad, L. S. Borketey, T.-Y. Lin and N. A. Schnarr, *Org. Biomol. Chem.*, 2012, **10**, 6717–6723.

Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.

In vitro Characterization of 3-Chloro-4-hydroxybenzoic Acid Building Block Formation in Ambigol Biosynthesis

I Dewa Made Kresna,[a,b] Luis Linares-Otoya,[a,b] Tobias Milzarek,[c] Elke R. Duell,[c,d] Mahsa Mir Moseni,[e,f] Ute Mettal,[a,b] Gabriele M. König,[e] Tobias A. M. Gulder[c,g] and Till F. Schäberle[a,b,*]

Supporting Information

Table S1. Conserved motifs of DAHP synthase Ab7.

Organism	Motif			Accession Number
	DxxHxN	KPRT	xGxR	
S. cerevisiae	I**D**YS**H**G**N**S 285	**KPRT** 117	I**G**A**R** 180	GI: 48425087
C. albicans	V**D**CS**H**G**N**S 283	**KPRT** 108	I**G**A**R** 170	GI: 3647668
E. coli	I**D**FS**H**A**N**S 271	**KPRT** 102	I**G**A**R** 165	GI: 12932715
H. influenzae	V**D**FS**H**A**N**S 279	**KPRT** 109	I**G**A**R** 172	GI: 950411
S. typhimurium	V**D**CS**H**G**N**S 272	**KPRT** 103	I**G**A**R** 166	GI: 1252280
F. ambigua	I**D**CS**H**G**N**S 273	**KPRT** 103	I**G**A**R** 166	This study

Table S2. Multiple sequence alignment of the metal binding amino acid (AA) residues of DAHP synthase Ab7.

Organism	AA residue binding to the metal ion				Accession Number
	C	H	E	D	
S. cerevisiae	GP**C**SI 78	S**H**GN 284	I**E**SN 318	T**D**AC 344	GI: 48425087
C. albicans	GP**C**SI 72	S**H**GN 282	I**E**SN 316	T**D**AC 345	GI: 3647668
E. coli	GP**C**SI 63	S**H**AN 270	V**E**SH 304	T**D**AC 328	GI: 12932715
H. influenzae	GP**C**SI 70	S**H**AN 278	V**E**SH 312	T**D**AC 336	GI: 950411
S. typhimurium	GP**C**SI 64	S**H**GN 271	I**E**SN 305	T**D**AC 330	GI: 1252280
F. ambigua	GP**C**SI 64	S**H**GN 272	L**E**SN 306	T**D**KC 331	This study

Table S3. Influence of metal salt addition on the enzymatic activity of Ab7. One unit of enzyme activity is defined as the consumption of 1 µmol PEP per minute.

Metal Salt	Activity Unit
$MnCl_2$	7.5±0.5
$MgCl_2$	7.0±1.0
$ZnCl_2$	7.5±0.5
$CuSO_4$	2.0±1.0
$CdSO_4$	5.0±1.0
EDTA	0.0±0.0

Table S4. The results of the Fre assay for measurement of UV absorbance at 340 nm. Consumption of NADH by Fre was calculated by the difference of $A_{340\ nm}{}^{1}$ minus $A_{340\ nm}{}^{2}$.

Reaction in HEPES Buffer (a)

Incubation time (min)	$A_{340\ nm}{}^{1}$ (Control)	$A_{340\ nm}{}^{2}$ (with Fre)	NADH consumption by Fre $A_{340\ nm}{}^{1} - A_{340\ nm}{}^{2}$
5	0,582	0,322	0,26
30	0,505	0,038	0,467
60	0,431	0,035	0,396
120	0,317	0,035	0,282
180	0,254	0,034	0,22

Reaction in Phosphate Buffer (b)

Incubation time (min)	$A_{340\ nm}{}^{1}$ (Control)	$A_{340\ nm}{}^{2}$ (with Fre)	NADH consumption by Fre $A_{340\ nm}{}^{1} - A_{340\ nm}{}^{2}$
5	1,332	1,105	0,227
30	1,269	0,602	0,667
60	1,203	0,455	0,748
120	1,102	0,422	0,68
180	1,021	0,411	0,61

Reaction in Tris-Cl Buffer (c)

Incubation time (min)	$A_{340\ nm}{}^{1}$ (Control)	$A_{340\ nm}{}^{2}$ (with Fre)	NADH consumption by Fre $A_{340\ nm}{}^{1} - A_{340\ nm}{}^{2}$
5	2,169	2,029	0,14
30	1,995	1,58	0,415
60	1,912	1,519	0,393
120	1,775	1,394	0,381
180	1,688	1,308	0,38

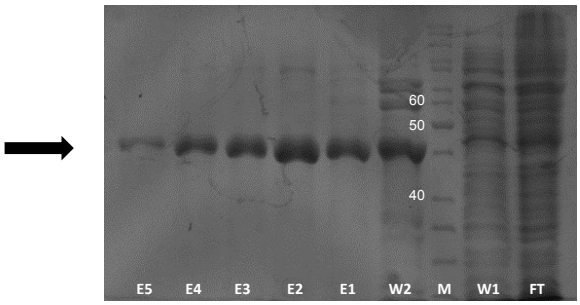


Figure S1. SDS page analysis of Ab7 after purification. FT, flow through; W1 and W2, washing steps (with 30 and 50 mM imidazole, respectively); E1 – E5, elution fractions (100- 350 mM imidazole); M, marker (kDa). The arrow indicates the band corresponding to Ab7. Elution fractions 1 – 5 were pooled, concentrated and redissolved in Tris HCl buffer (pH 7). The protein concentration was determined spectrophotometrically to be around 1 mg/mL, using Thermo Scientific NanoDrop.

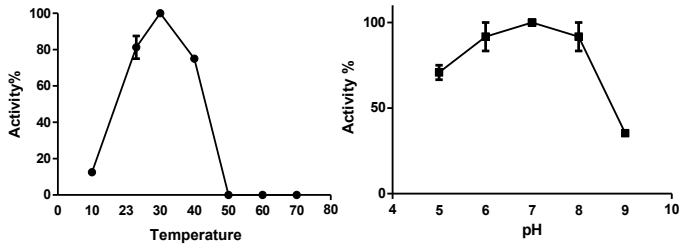
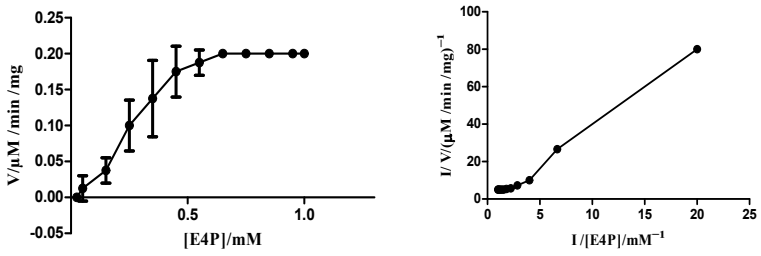


Figure S2. Effect of (a) incubation temperature and (b) pH on Ab7 activity.



(a) (b)

Figure S3. Lineweaver-Burk plot of Ab7 with various E4P concentrations (0.025 - 1 mM) and a constant concentration of PEP. Using Graf Pad Prism 5, kinetic constants of Ab7 were determined according to double-reciprocal curves. A double-reciprocal plot of Ab7 was calculated against the substrate concentration. The concentration of PEP was always preserved at 80 µM while the concentration of E4P was varied between 0.025 and 1 mM. The maximum velocity achieved by the system, at saturating substrate concentrations (v_{max}), was determined to be 0.3725 ± 0.083 (U/mg^{-1}) and the obtained K_m and k_{cat} values were shown to be 0.6993 ± 0.3746 mM and 0.0281 ± 0.063 s^{-1}, respectively.

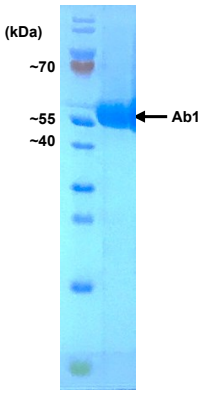


Figure S4. Purification of Ab1 as analyzed by SDS-Page.

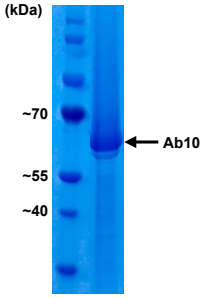


Figure S5. Purification of Ab10 as analyzed by SDS-Page.

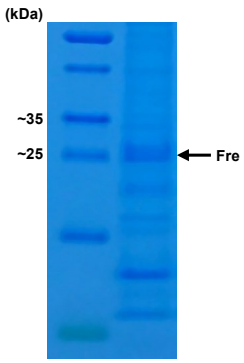


Figure S6. Purification of Fre as analyzed by SDS-Page.

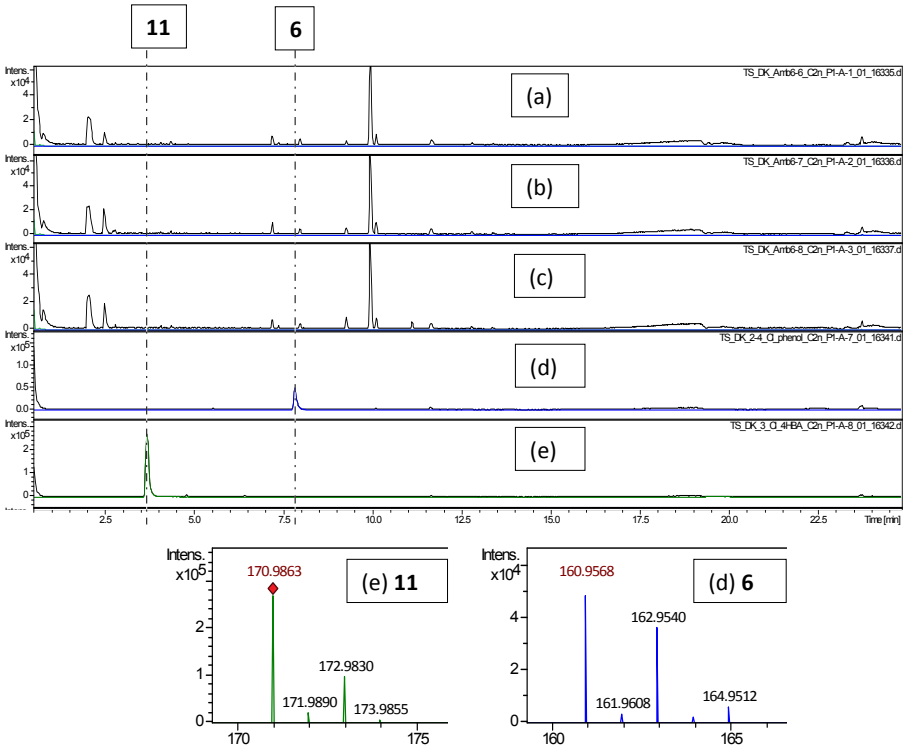


Figure S7. LCMS data for the in vitro assay using 4-HBA (**7**) as the substrate of Ab1, Ab10, or both. The assay was conducted (a) only with Ab10, (b) only with Ab1, (c) with Ab1 and Ab10. The (d) standard of 2,4-dichlorophenol (**6**), and the (e) standard of 3-Cl-4-HBA (**11**) served for comparison. Mass spectra were measured in negative mode. The upper part of the figure shows the respective chromatograms, while the lower part provides for (d) and (e) close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Displayed in green are the extracted ion chromatograms (EIC) of the $[M-H]^-$ peak of **11** (theoretical m/z = 170.9843). Shown in blue are the extracted ion chromatograms of the $[M-H]^-$ peak of **6** (theoretical m/z = 160.9555). As apparent from the images, neither substance **6** nor **11** could be detected.

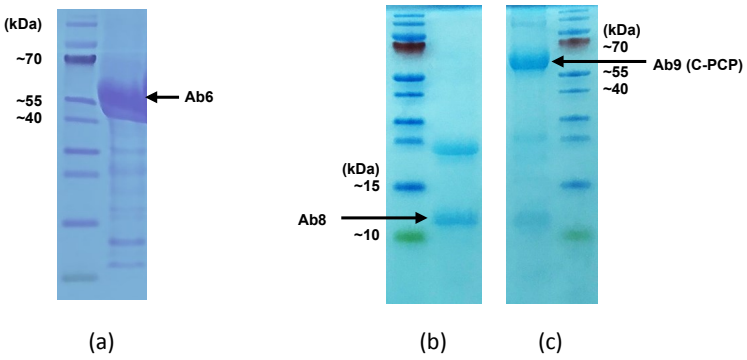


Figure S8. Purification of (a) Ab6, (b) Ab8, and (c) Ab9 (C-PCP) as analyzed by SDS-Page.

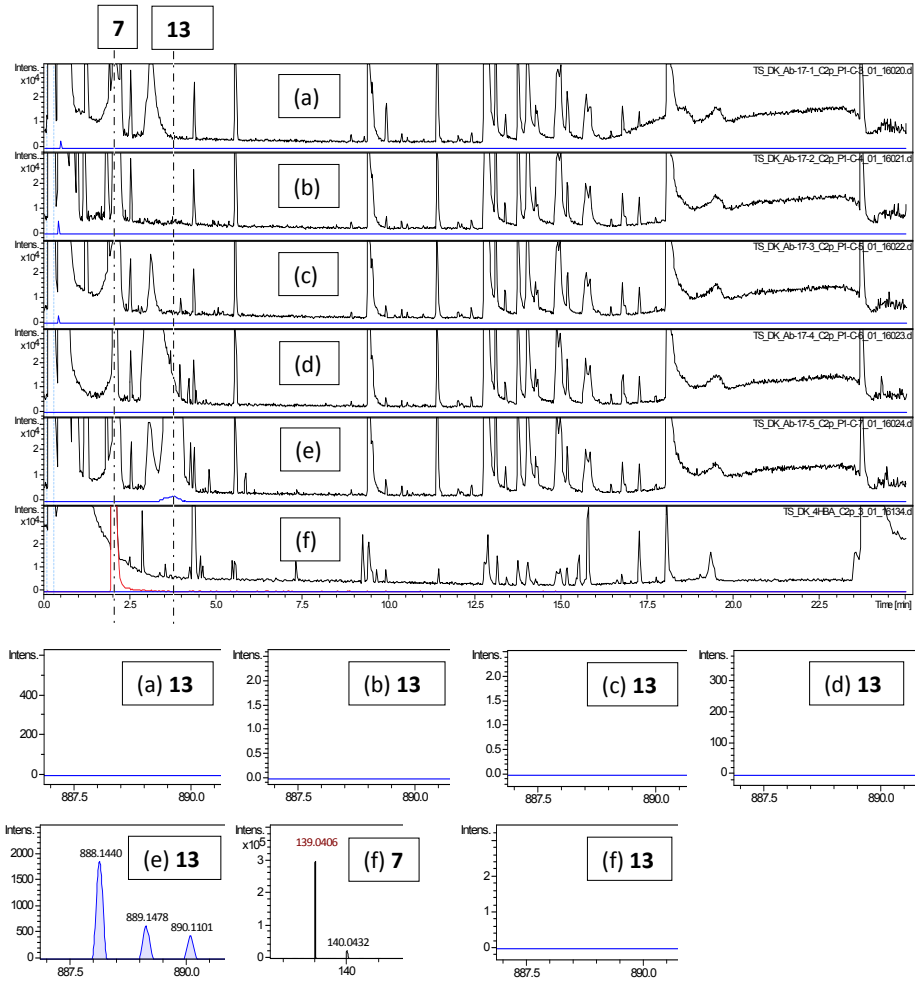


Figure S9. LCMS data of the Ab6 *in vitro* assay. The assay was performed (a) without ATP, (b) without CoA**,** (c) without the substrate 4-HBA (**7**), (d) without Ab6, and as (e) complete reaction. For comparison a (f) standard of **7** was used. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. The red chromatogram in (f) illustrates the extracted ion chromatogram (EIC) of the $[M+H]^+$ peak of the substrate standard **7** (m/z theoretical = 139.0395). The blue chromatogram shows the EIC of the $[M+H]^+$ peak of the product 4-HBA-CoA (**13**) (m/z theoretical = 888.1436).

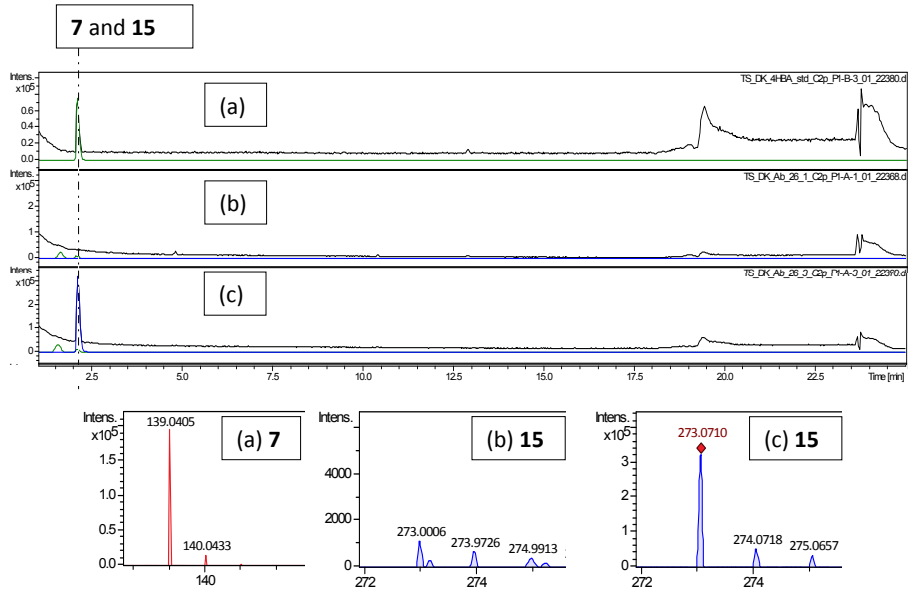


Figure S10. LCMS analysis of the control experiment for the *in vitro* carrier protein assay. For this control trial the substrate 4-HBA (**7**) was incubated with cysteamine in the presence or absence of Ab6. The (a) standard of substrate **7** served for comparison. The assay was performed (b) without Ab6, and (c) with Ab6. Mass spectra were measured in positive mode. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Displayed in green are the extracted ion chromatograms (EIC) of the $[M+H]^+$ peak of the substrate **7**. The blue chromatogram shows the EIC of the $[M+H]^+$ peak of the product "4-HBA-cystamine" (**15**) (theoretical m/z = 273.0726). As apparent from the images, product **15** could be detected in sample (c).

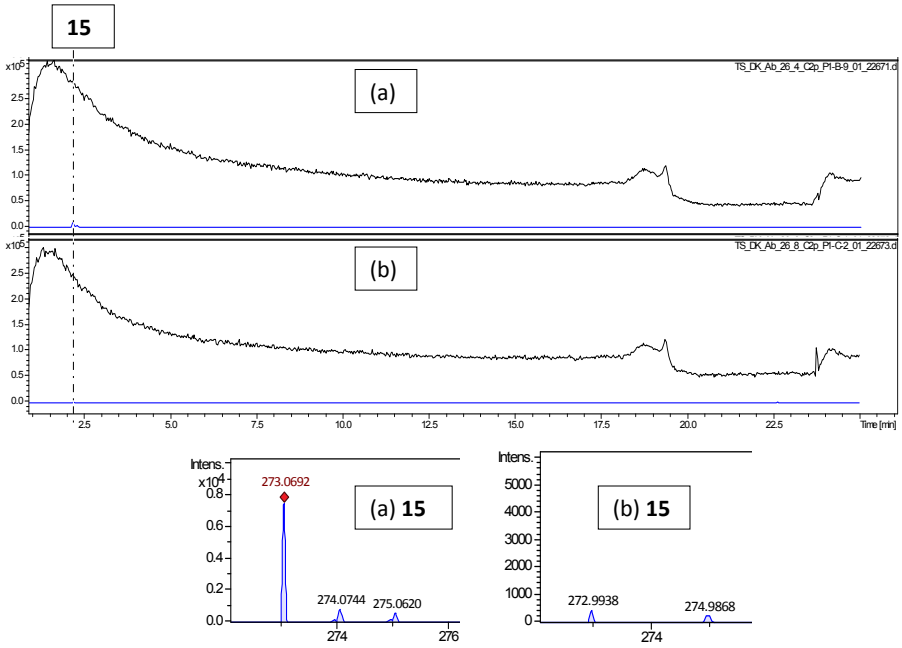


Figure S11. LCMS analysis of the Ab8 and Ab9 (C-PCP) carrier protein *in vitro* assay. The assay was performed in the presence of Ab6 using 4-HBA (**7**) as substrate. Prior to cysteamine treatment, Ab8 and Ab9 (C-PCP) were purified by HPLC. Mass spectra were measured in positive mode. In the assay either (a) Ab8 or, (b) Ab9 (C-PCP) served as a carrier protein. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Displayed in blue are the extracted ion chromatograms (EIC) of the $[M+H]^+$ peak of the product "4-HBA-cystamine" (**15**) (theoretical m/z = 273.0726). As apparent from the images, product **15** could be detected only when Ab8 was used as a carrier protein.

```
CLUSTAL O(1.2.4) multiple sequence alignment

Ab10 ------------------------------------------------------------
McnD MKTVEFLAYLNSLQINLWAENDKLRYRSPQGVMTPELLGKLKERKEQLIALLRQKAEDLG 60
ApdC MKTVNFLSHLNDLGINVWVENDKLRYRSPKGVIIPELLQELKERKEELIAFLRQQAEDLN 60
AerJ MKTVEFLSDLNHLGVTIWMEGDKLRYRSPQGVMTPDLLEQLKEHKEELIVLLREQADNFS 60


Ab10 MKNIYDVAIC**G**S**G**LA**G**LTLARQLKLKMPDISVVVLDRLARPLPEAGFKVGESSVEVGAFY 60
McnD EAEVYDVVIC**G**G**G**LA**G**LTLARQLKLQKLNISVIVLDKIARPLPEASFKVGESTVEVGAFY 120
ApdC QAETYDVVIC**G**G**G**LA**G**LTLARQLKLQKPNMAIAVLDKMSGLSPEASFKVGESTVEVGAFY 120
AerJ SETDYDVAIC**G**G**G**LA**G**LTLGRQLKLKQPNLSVVVLDKMARPLPEAGFKVGESTVEVGAFY 120
         ***.***.*******.*****:  :::: ***:::   ***.******:*******
               GxGxxG motif

Ab10 LAHIVQLEDYLEKQHLHKLGLRYFLGDTKGPFHKRPEIGLSKYHFPNSYQLDRGKLENDL 120
McnD LANTLQLTDYFEEQHLVKLGLRYFFNNSATNFQDRPELGLSEFHLPNSYQIDRGKLENDL 180
ApdC LANTLQLTDYFEEQHLVKLGLRYFFNNSATNFQERPELGLSEFHAPNSYQIDRGKLENDL 180
AerJ LANTLQLTDYFDHQHLPKLGLRYFFKPQETEFHKRPELGLSEFHAPKSYQIDRGKLENDL 180
     **: :** **::.*** *******:      *:.***:***::* *:***:*********

Ab10 RSINTEAGVELLEGCLVKDIELGDP-QQLHQIIYTQEN--NKATQAIQARWVVDSMGYRR 177
McnD RAFNVEAGVELRENCLVNEIELAVGLQQHHKIVYTQDKGDHKKTNVIQARWVVDAMGRRR 240
ApdC RQFNMEAGVELREGCLVNEIALAEGLQQHHKVVYTQGDGDNRKNKIIKARWVVDAMGRRR 240
AerJ RQFNIEAGIDLKENCSVKDIEFAEGLQQQHKIIYTQGSGANQKTHCIKSRWVVDAMGRRR 240
     * :* ***::* *.* *::* :.   ** *:::*** .  :: .: *::*****:** **

Ab10 FLQRKLGLAKPKNSQFSAVWFRVEGRFDVSDFVPSTEIEWHERVPHNNRYYSTNHLCGEG 237
McnD FLQKKLGLDKPNNAQFGAVWFRVEGRFDISDFVPSSEEKWHNRVPNKNRYYSTNHLCGEG 300
ApdC FLQRKLGLDKPNNDNFGAVWFRVNGRFDIGDFVPSSEEKWHNRVPNKNRYYSTNHLCGEG 300
AerJ FIQKKLGLAKPNHNNYSAVWFRVEGRFDVSNFVPASEEKWHRRVPNNNRYYSTNHLCGEG 300
     *:*:**** **:: ::.******:****:.:***::* :**.***::**************

Ab10 Y**W**V**W**L**IP**LSTGYTSIGIVTNEEIHPFGTYHTYEKAFQWLEKHEPVVAFHLKSNPPVDFMK 297
McnD Y**W**V**W**T**IP**LSTGHTSIGIVARQDIHPLKTYYNYELAYQWLQKNEPTLAFHLADKQPEDFRK 360
ApdC Y**W**V**W**T**IP**LSTGHTSIGIVARQDIHPLKNYYNYELAYQWLQKNEPVLACHLKDKEPEDFRK 360
AerJ Y**W**V**W**L**IP**LSTGYTSIGIVARQDIHPLKNYHNYELAFQWLRENEPVLAAHLEGKSPEDFRK 360
     **** ******:******:.::***: .*:.** *:***.::**.:* ** .: * ** *
      WxWxIP motif

Ab10 IPQYSYSSNQVFSINRWACVGVAGVFADPFYSPGTDLIGFGNSLITQMVELDRENKLTPE 357
McnD MPKYSYSSKQVFSYNRWACVGEAGTVPDPFYSPGTDNIGFGNSLTTQLITLDLEGKLTQE 420
ApdC MPKYSYSSKQVFSSNRWACVGEAGTFPDPFYSPGTDNIGFGNSLTTQLIALDLEDKLTKE 420
AerJ MPKYSYSSKQVFSFNRWACVGEAGLFPDPFYSPGSDSIGFGNSLTTQMIELDLKGQLTPE 420
     :*:*****:**** ******* ** . *******:* ******* **:: ** :.:** *

Ab10 IVNEANRFLITYNESVTSNIHNAYLCFGNETVMVMKYIWDVLSAWAFSAPMMFNSLFLDS 417
McnD KVKDANHFYLSYSDGVTLNIQNAYNCLGNGMVMATKFIWDTLSGWTFSGLMMFNSIYLDQ 480
ApdC KVEDANYFYISYSDGVTLNIQNAYNCLGNGIVMATKFIWDTLSGWTFSGLMMFNSIFLDQ 480
AerJ RVDDANHFYLTYHDGLTFNIQNSYNCMGNGIVMATKMIWDTLAGWTFGCLMMFNSIFLDP 480
      *.:** * ::* :.:* **:*:* *:**  **. * ***.*:.*:*.  *****::**

Ab10 DKRAKVRKGTGQFFLLAQRMNQLFRDWAVQSQRRTSFEFIDYLQIPFVRELRARNLKTNK 477
McnD EFRTKVQQISSKFFPLSYRMQQLFKDWANQSLHRVNFEFIDYLAIPFVEELRSRNLRYNQ 540
ApdC DFRTRVQEINSKFFPLSYRMQQLFRDWANKSLNRVNFEFIDYLAIPFVDELRTRNLKSNK 540
AerJ ELKMKVQQINAEFFPLSYRIQQLFRDWANQSLGRVSFEFIDYLAIPFVNELRTRNLQSNK 540
     : : :*:: ..:** *: *::***:*** :*  *..******* **** ***:***: *:

Ab10 TEQELIDDHLASIKLFEELAQVIFLLALEDTMPEKSADFPSPVWLNAWVVSLDDKRWEID 537
McnD TESEIIENYLSSLKLLEEVAQVMFHLALEDIMPEMLPKVSSNSWLNAWAISLNASKWESD 600
ApdC TKSEIIENYLSSLKLLEEVAQVMFHLALEDTMPDMLSKVNSNSWLNAWAISLDISKWEAD 600
```

```
AerJ TEAEIIEAYRASLKLLEEFAQVIFQIALEDTSPELLTKINEHPWLNAWAISLDASKWETD 600
     *: *:*: : :*:**:**.***:* :****  *:   .. .  *****.:**: .:** *

Ab10 GLFRPTSKPRDLRPMMEQLWQNIHFRAADRDSSLITA 574
McnD GLFAPKSEPRNLDLIKQQLWEAIGK------------ 625
ApdC GLFTPKSEPRNLNLIKQQLWEAIGK------------ 625
AerJ GLFSPKNEPRNLHSIKKQYLEAINR------------ 625
     *** *..:**:*  : :*  : *
```

Figure S12. Alignment of Ab10 with McnD, ApdC and AerJ using CLUSTALO (1.2.4) multiple sequence alignment (https://www.uniprot.org/align/). GxGxxG and WxWxIP motifs are highlighted in bold letters.

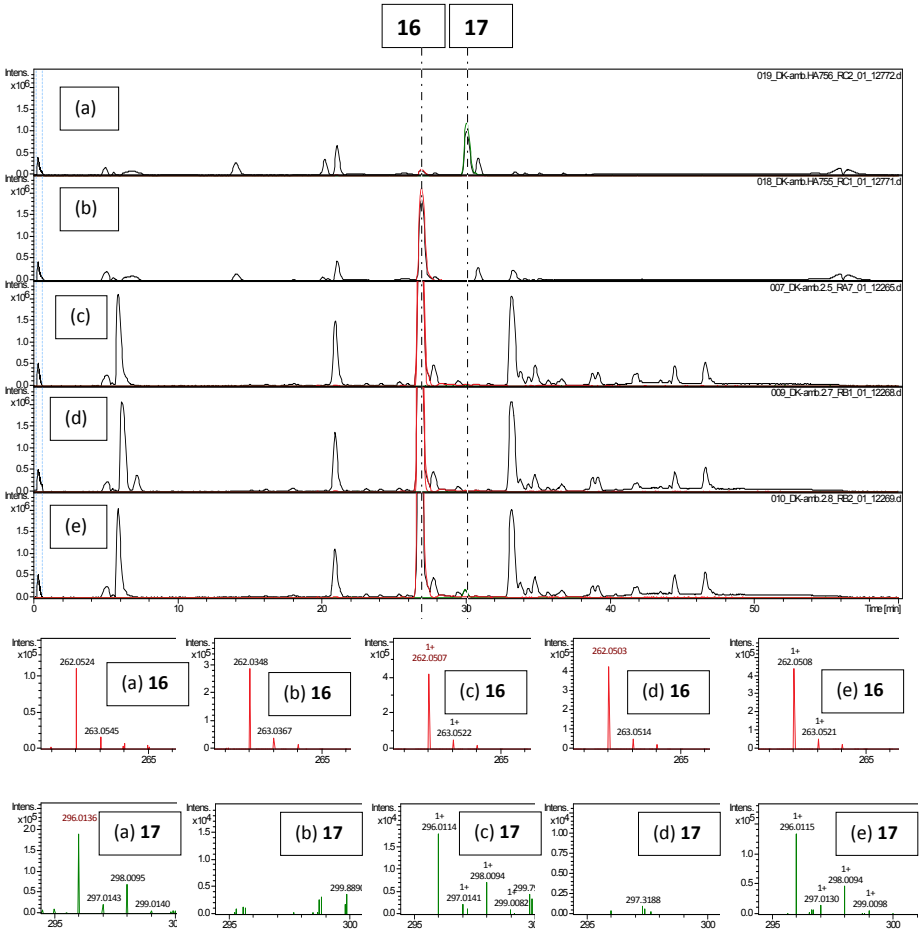


Figure S13. LCMS analysis of the Ab10 *in vitro* assay. For comparison a (a) standard of the product 3-Cl-4-HBA-SNAC (**17**), and a (b) standard of substrate 4-HBA-SNAC (**16**) were used. The assay itself was performed (c) without Fre, (d) without Ab10, and as (d) complete reaction. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Displayed in red are the extracted ion chromatograms (EIC) of the $[M+Na]^+$ of the substrate **16** (m/z theoretical = 262.0508). The green chromatogram shows the EIC to $[M+Na]^+$ of the product **17** (m/z theoretical = 296.0119).

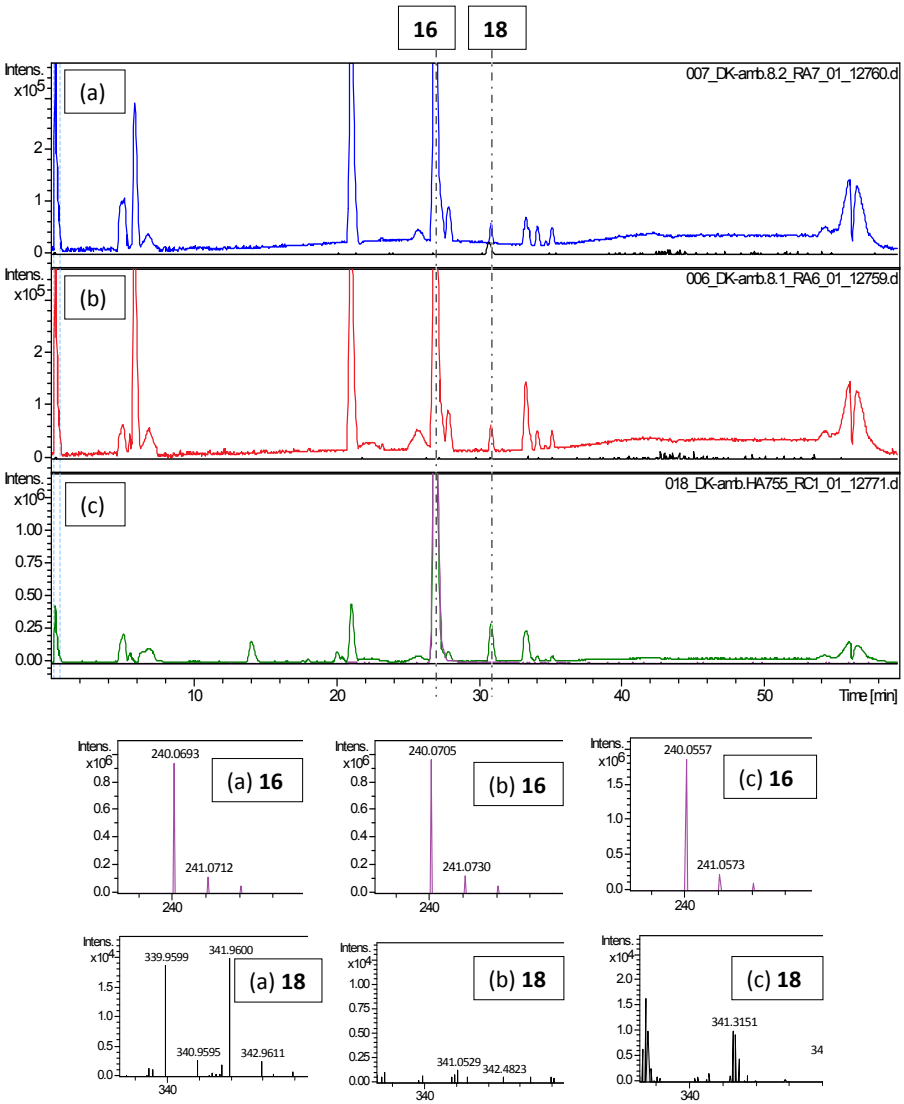


Figure S14. LCMS analysis of the *in vitro* assay of Ab10 using bromide as the halogen source for investigating Ab10 substrate specificity. The assay was performed as a (a) complete reaction, and (b) without Ab10. A (c) standard of the substrate 4-HBA-SNAC (**16**) was used for comparison. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) are shown in blue, red and green, respectively. The purple chromatogram in (c) illustrates the extracted ion chromatogram (EIC) of the $[M+H]^+$ peak of the substrate **16** (m/z theoretical = 240.0694). Black chromatograms in "a", "b" and "c" show the EIC of the $[M+H]^+$ peak of the product 3-Br-4-HBA-SNAC **18** (m/z theoretical = 339.9619).

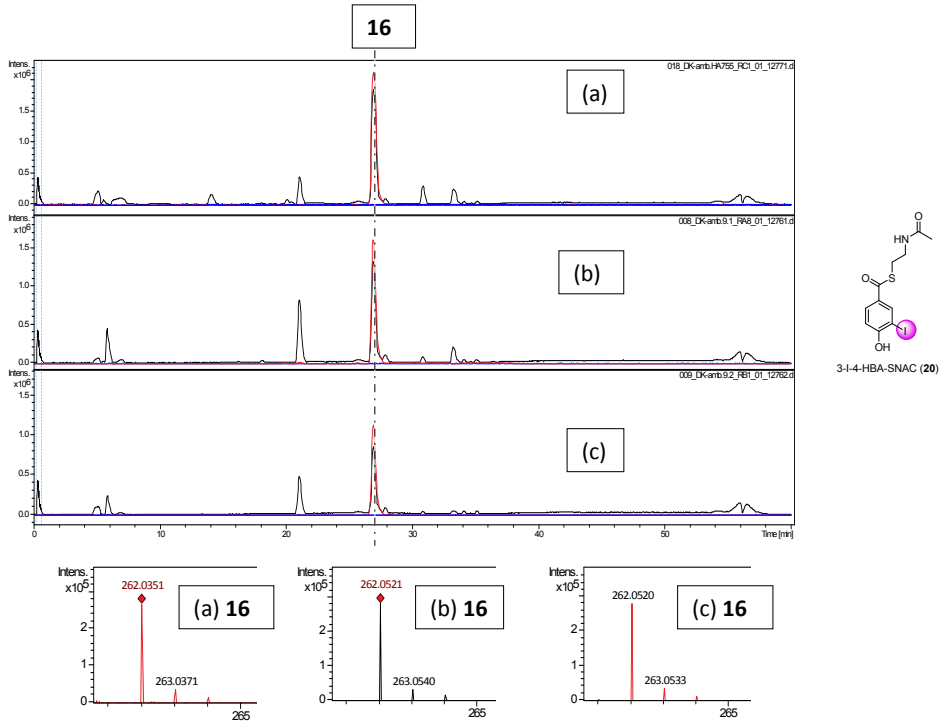


Figure S15. LCMS analysis of the *in vitro* assay of Ab10 using iodide as the halogen source. Mass spectra were measured in positive mode. For comparison the (a) standard of the substrate 4-HBA-SNAC (**16**) was used. The assay was performed (b) without Ab10, and as a (c) complete reaction. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Red chromatograms show the extracted ion chromatogram (EIC) of the $[M+Na]^+$ peak of **16** (theoretical m/z = 262.0508). Blue chromatograms show the EIC of the $[M+H]^+$, $[M+Na]^+$, and $[M+K]^+$ peaks (theoretical m/z = 365.9655, 387.9457, 403.9214 respectively) of the expected product 3-I-4-HBA-SNAC (**20**). As apparent from the images, no peak of the expected product **20** could be observed.

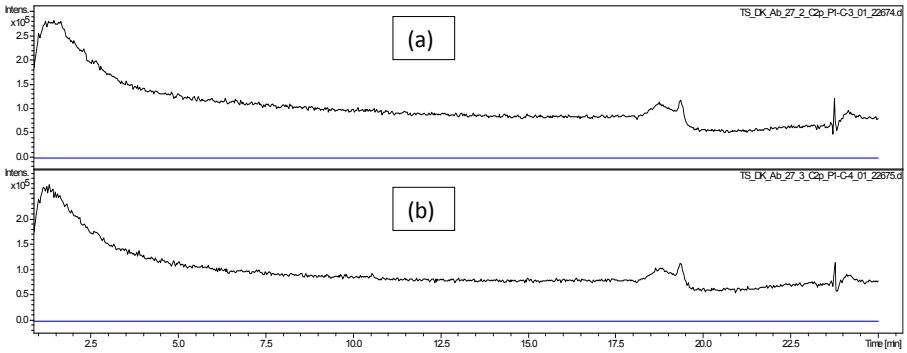


Figure S16. LCMS analysis of the *in vitro* assay of the C domain from Ab9 using 4-HBA-SNAC (**16**) as the substrate. The assay was performed using (a) only Ab9 (C-PCP) without substrate **16**, and as (b) complete reaction. Mass spectra were measured in positive mode. Base peak chromatograms (BPC) of each sample are shown in black. Blue chromatograms show the extracted ion chromatogram (EIC) of the $[M+H]^+$ peak of the product "4-HBA-cystamine" (**15**). As apparent from the images, product **15** could not be detected, neither in the control sample (a) nor in the complete reaction (b).

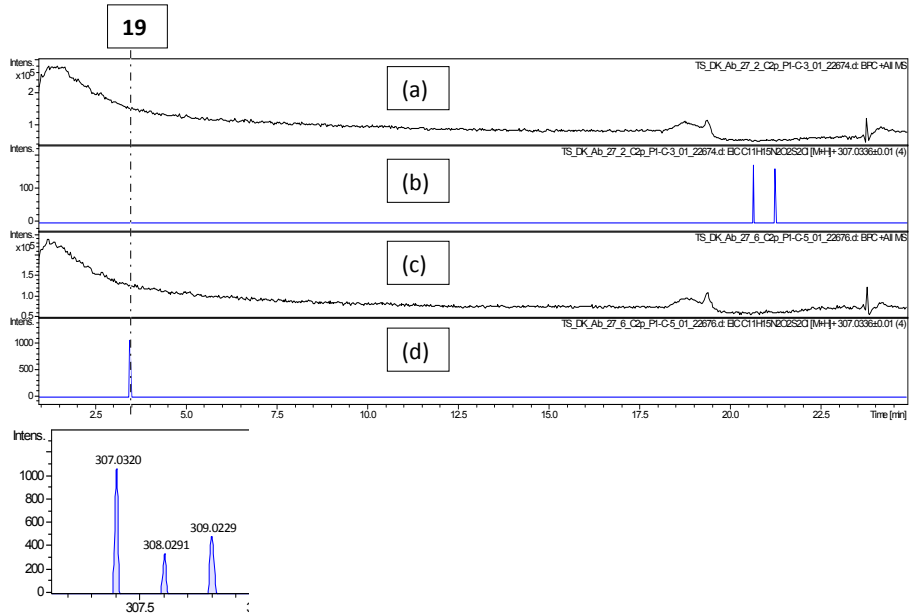


Figure S17. LCMS analysis of the *in vitro* assay of the C domain from Ab9 using 3-Cl-4-HBA-SNAC (**17**) as the substrate. The assay was performed using (a, b) only Ab9 (C-PCP) without substrate **17**, and as a (c, d) complete reaction. Mass spectra were measured in positive mode. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Blue chromatograms show the extracted ion chromatogram (EIC) of the $[M+H]^+$ peak of the product "3-Cl-4-HBA-cystamine" **19** (theoretical m/z = 307.0336). As apparent from the images, product **19** could only be detected in the complete reaction sample.

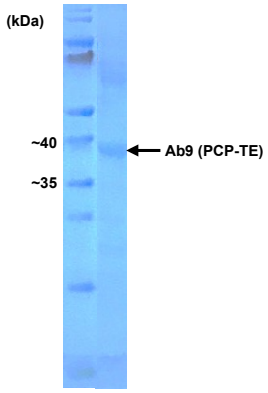


Figure S18. Purification of Ab9 (PCP-TE) as analyzed by SDS-Page.

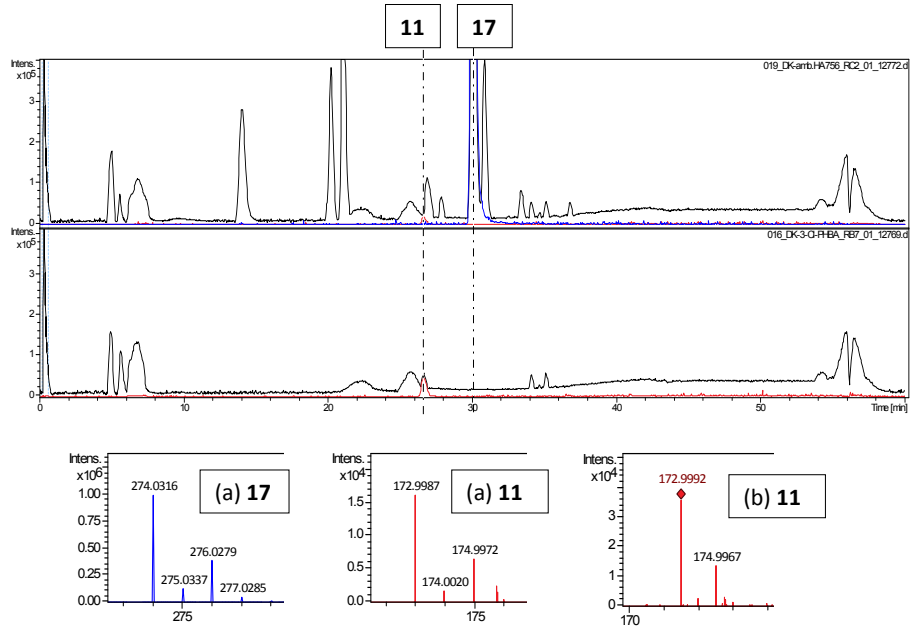
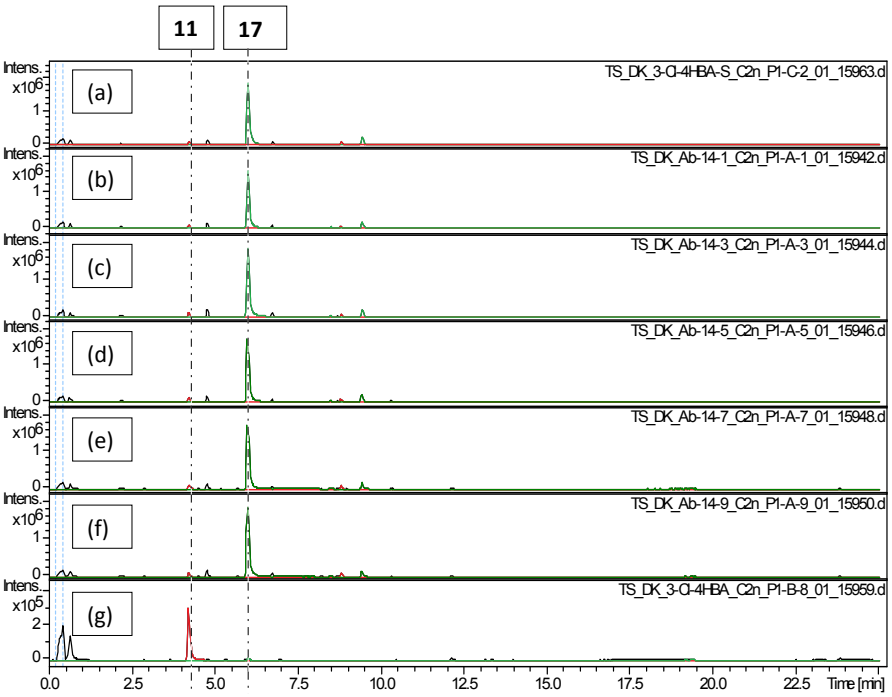
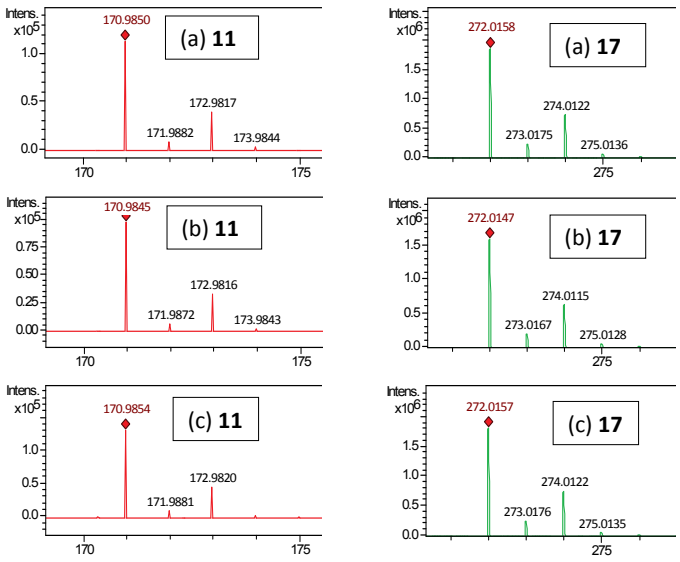


Figure S19. LCMS analysis of the (a) standard 3-Cl-4-HBA-SNAC (**17**) in comparison to the (b) standard of 3-Cl-4-HBA (**11**). Mass spectra were measured in positive mode. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Red chromatograms show the extracted ion chromatogram (EIC) of the $[M+H]^+$ peak of **11** (theoretical m/z = 173.0000). Blue chromatograms show the EIC of the $[M+H]^+$ peak of **17** (theoretical m/z = 274.0299). As apparent from the images, there is a contamination of **11** in the standard of **17**.


11
17
(a)
(b)
(c)
(d)
(e)
(f)
(g)
TS_DK_3-Cl-4HBA-S_C2n_P1-C-2_01_15963.d
TS_DK_Ab-14-1_C2n_P1-A-1_01_15942.d
TS_DK_Ab-14-3_C2n_P1-A-3_01_15944.d
TS_DK_Ab-14-5_C2n_P1-A-5_01_15946.d
TS_DK_Ab-14-7_C2n_P1-A-7_01_15948.d
TS_DK_Ab-14-9_C2n_P1-A-9_01_15950.d
TS_DK_3-Cl-4HBA_C2n_P1-B-8_01_15959.d
Intens.
Time [min]


(a) 11
170.9850
172.9817
171.9882
173.9844
(a) 17
272.0158
274.0122
273.0175
275.0136
(b) 11
170.9845
172.9816
171.9872
173.9843
(b) 17
272.0147
274.0115
273.0167
275.0128
(c) 11
170.9854
172.9820
171.9881
(c) 17
272.0157
274.0122
273.0176
275.0135
Intens.

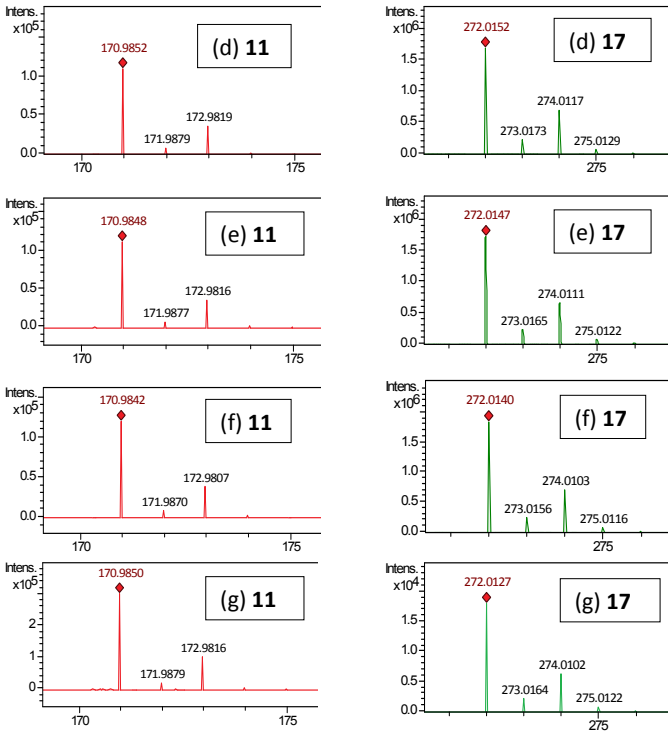
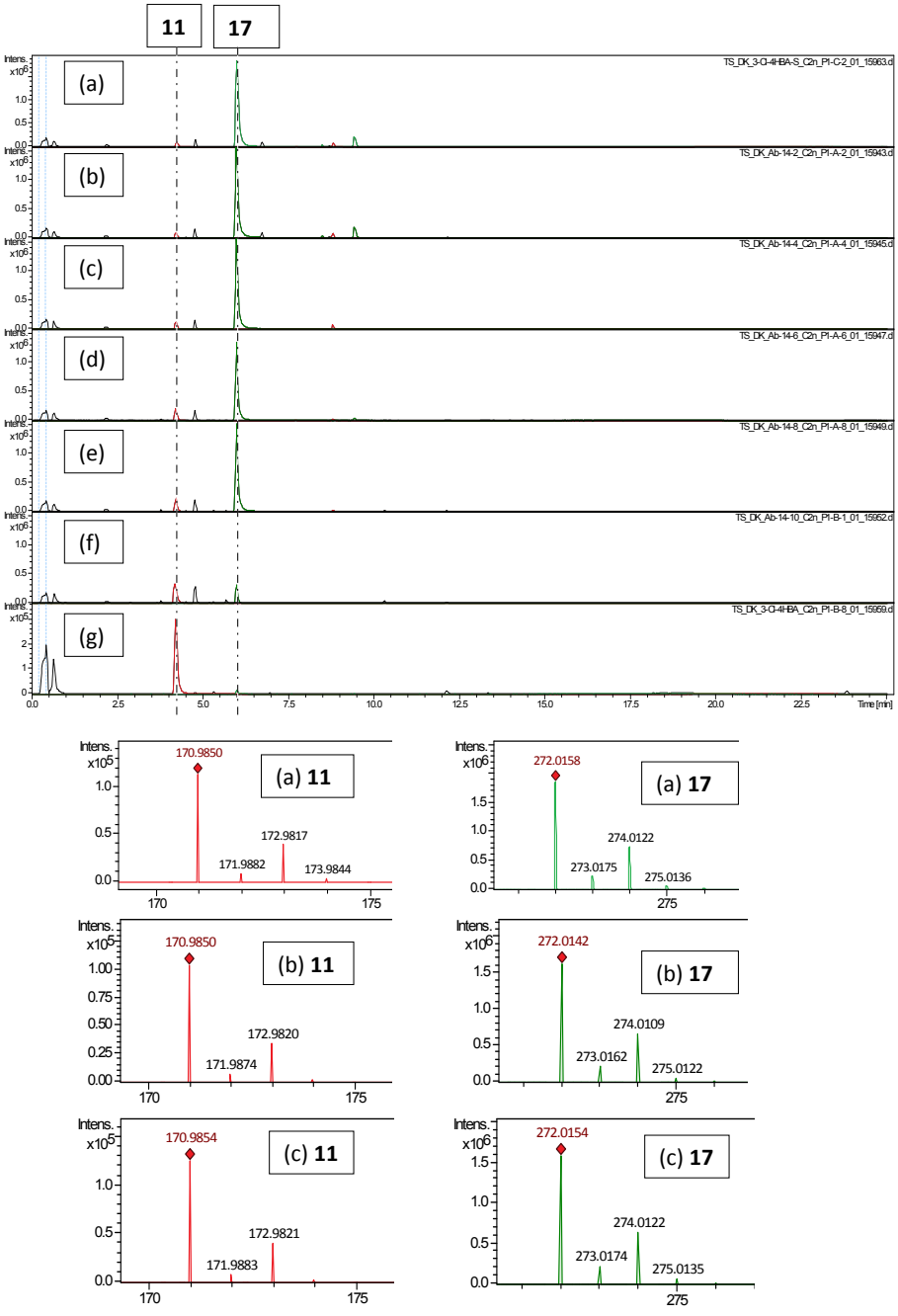


Figure S20. LCMS analysis of the control reaction (without Ab9 (PCP-TE)) of the time dependent Ab9 (PCP-TE) *in vitro* assay. For comparison a (a) standard of substrate 3-Cl-4-HBA-SNAC (**17**) and a (g) standard of the product 3-Cl-4-HBA (**11**) were used. The reactions (b), (c), (d), (e), (f) were incubated for 0, 5, 30, 60 min and 24 hours respectively. Mass spectra were measured in negative mode. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Green chromatograms show the extracted ion chromatogram (EIC) of the $[M-H]^-$ peak of the substrate **17** (m/z theoretical = 272.0148). Red chromatograms show the EIC of the $[M-H]^-$ peak of the product **11** (m/z theoretical = 170.9848).


11
17
(a)
(b)
(c)
(d)
(e)
(f)
(g)
Time [min]
(a) 11
170.9850
172.9817
171.9882
173.9844
(a) 17
272.0158
274.0122
273.0175
275.0136
(b) 11
170.9850
172.9820
171.9874
(b) 17
272.0142
274.0109
273.0162
275.0122
(c) 11
170.9854
172.9821
171.9883
(c) 17
272.0154
274.0122
273.0174
275.0135

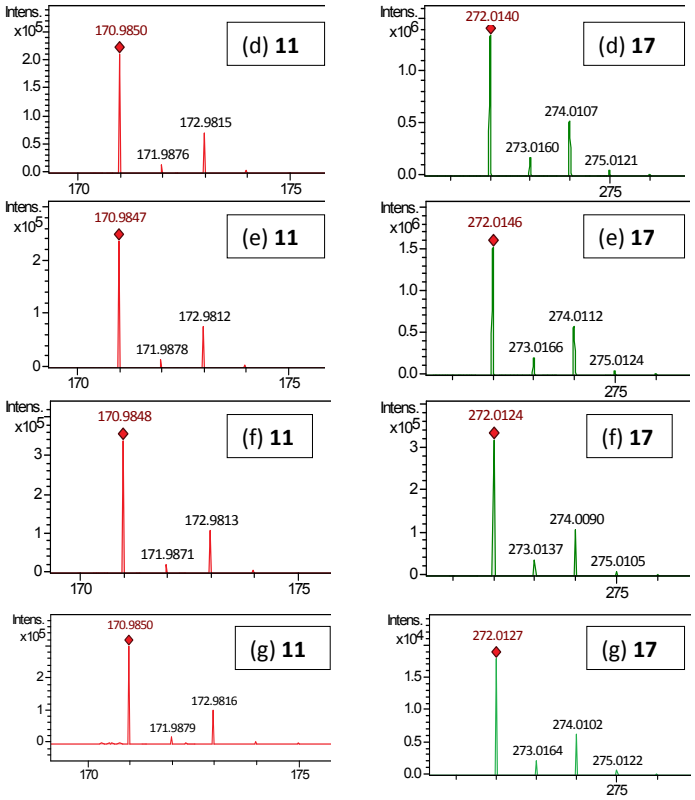


Figure S21. LCMS analysis of the complete Ab9 (PCP-TE) treated reaction of the time dependent Ab9 (PCP-TE) *in vitro* assay. For comparison a (a) standard of substrate 3-Cl-4-HBA-SNAC (**17**) and a (g) standard of the product 3-Cl-4-HBA (**11**) were used. The reactions (b), (c), (d), (e), (f) were incubated for 0, 5, 30, 60 min and 24 hours respectively. Mass spectra were measured in negative mode. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Green chromatograms show the extracted ion chromatogram (EIC) of the $[M-H]^-$ peak of the substrate **17** (m/z theoretical = 272.0148). Red chromatograms show the EIC of the $[M-H]^-$ peak of the product **11** (m/z theoretical = 170.9848).

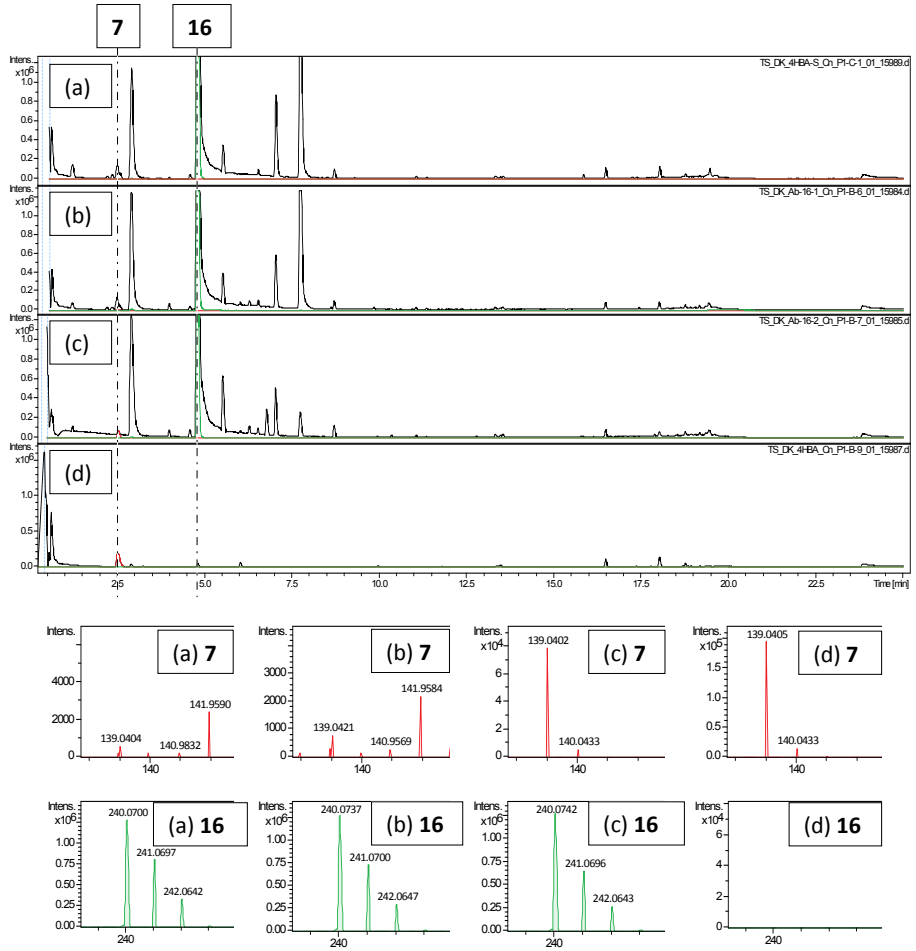


Figure S22. LCMS analysis of Ab9 (PCP-TE) substrate specificity experiment with respect to the substrate 4-HBA-SNAC (**16**). For comparison a (a) standard of the substrate **16**, and a (d) standard of the product 4-HBA (**7**) were used. The assay was carried out (b) without Ab9 (PCP-TE) (control), and as a (c) complete reaction. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Green chromatograms show the extracted ion chromatogram (EIC) of the $[M+H]^+$ peak of the substrate **16** (m/z theoretical = 240.0694). Red chromatograms show the EIC of the $[M+H]^+$ peak of the product **7** (m/z theoretical = 139.0395).

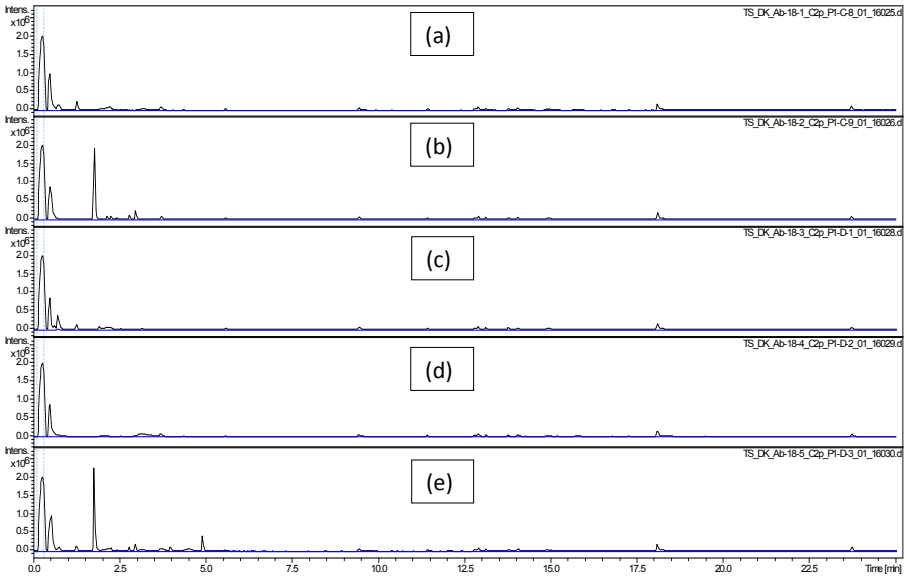


Figure S23. LCMS data of the *in vitro* investigation of Ab6 substrate specificity with respect to the substrate 3-Cl-4-HBA (**11**). The assay was carried out (a) without ATP, (b) without CoA, (c) without substrate **11**, (d) without Ab6, and as (e) complete reaction. Mass spectra were measured in positive mode. Base peak chromatograms (BPC) of each sample are shown in black. Blue chromatograms show the extracted ion chromatogram (EIC) of the $[M+H]^+$ and $[M+Na]^+$ peaks of 3-Cl-4-HBA-CoA (theoretical m/z = 922.1046 and 944.0866 respectively). As apparent from the EIC, the expected product could not be detected.

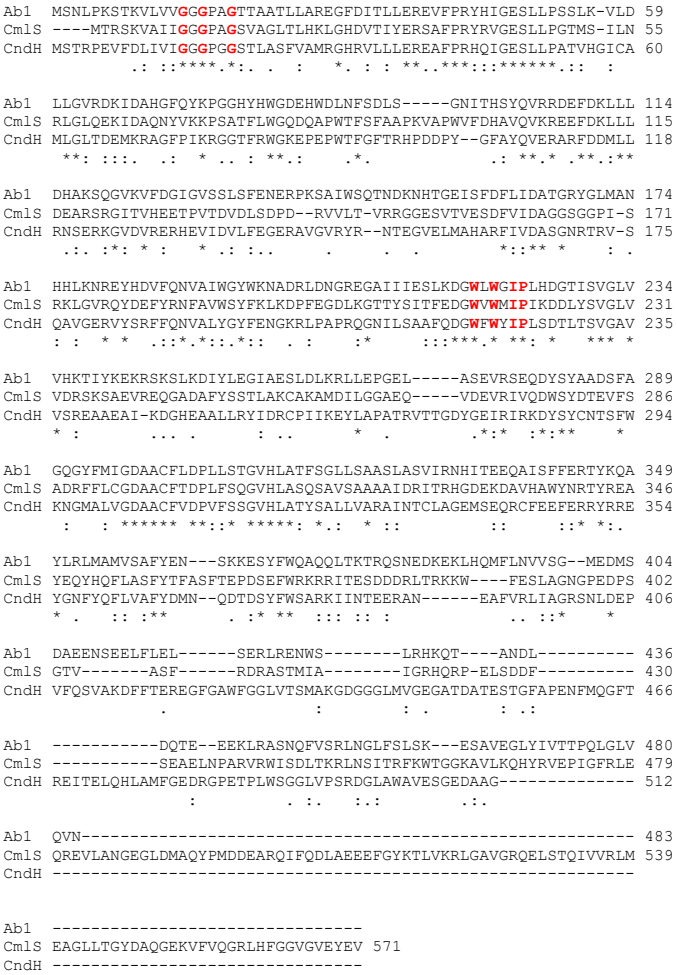


Figure S24. Alignment of Ab1 against ClmS and CndH with GxGxxG and WxWxIP motifs being highlighted in bold red letters. BLASTp alignment of Ab1 against the database shows, that almost all of the 100 first hits are tryptophan 7-halogenases or NAD(P)/FAD-dependent oxidoreductases with the highest percentage of identity being 72.52%. However, none of those hits is so far experimentally characterized. A refined BLASTp search using the pdb database and including only experimentally characterized enzymes, revealed the first two hits to be ClmS and CndH. However, as can be concluded from the query cover, both of them do not share AA sequence in their C-terminal region. Differences in the C-terminal region of Flavin-dependent halogenases (FDH) are the origin of their regio- and substrate specificity, as described by the characterization of PyrH[1]. The N-terminus usually displays the region which forms the FAD binding site[2].

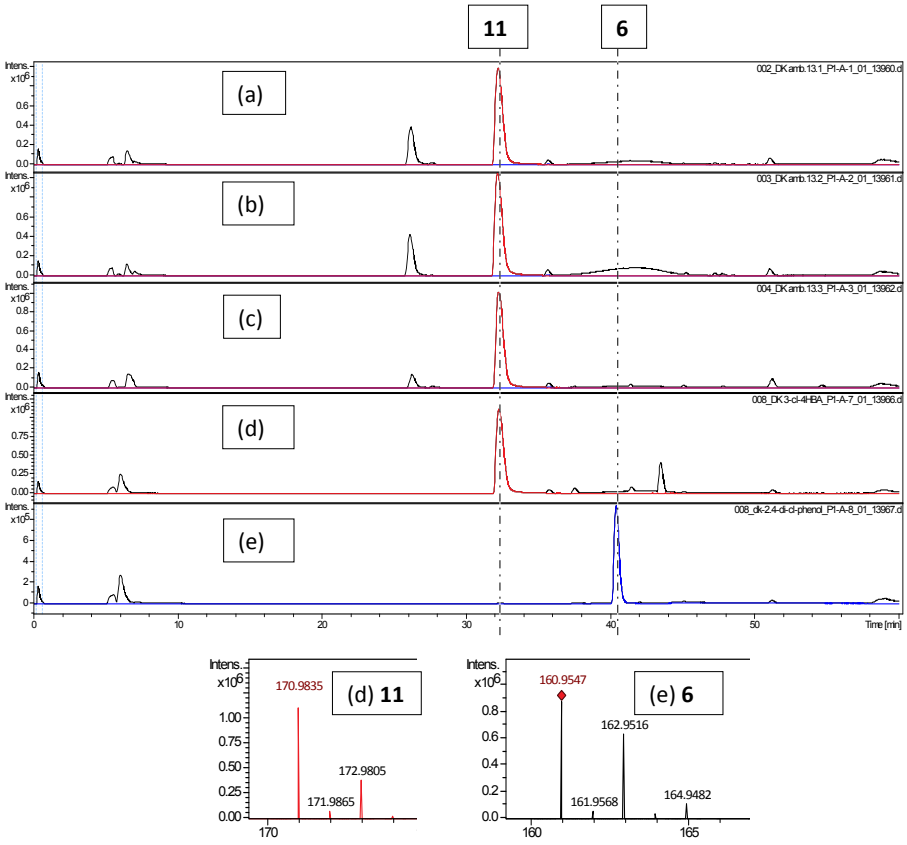


Figure S25. LCMS analysis of the *in vitro* assay using 3-Cl-4-HBA (**11**) as the substrate of Ab1. The assay was carried out (a) without Fre, as (b) complete reaction, (c) without Ab1. The (d) standard of substrate **11**, and the (e) standard of the product 2,4-dichlorophenol (**6**) served for comparison. Mass spectra were measured in negative mode. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Red chromatograms show the extracted ion chromatogram (EIC) of the $[M-H]^-$ peak of **11** (theoretical m/z = 170.9843). Blue chromatograms show the EIC of the $[M-H]^-$ peak of **6** (theoretical m/z = 160.9555). As apparent from the images, the expected product **6** could not be observed.

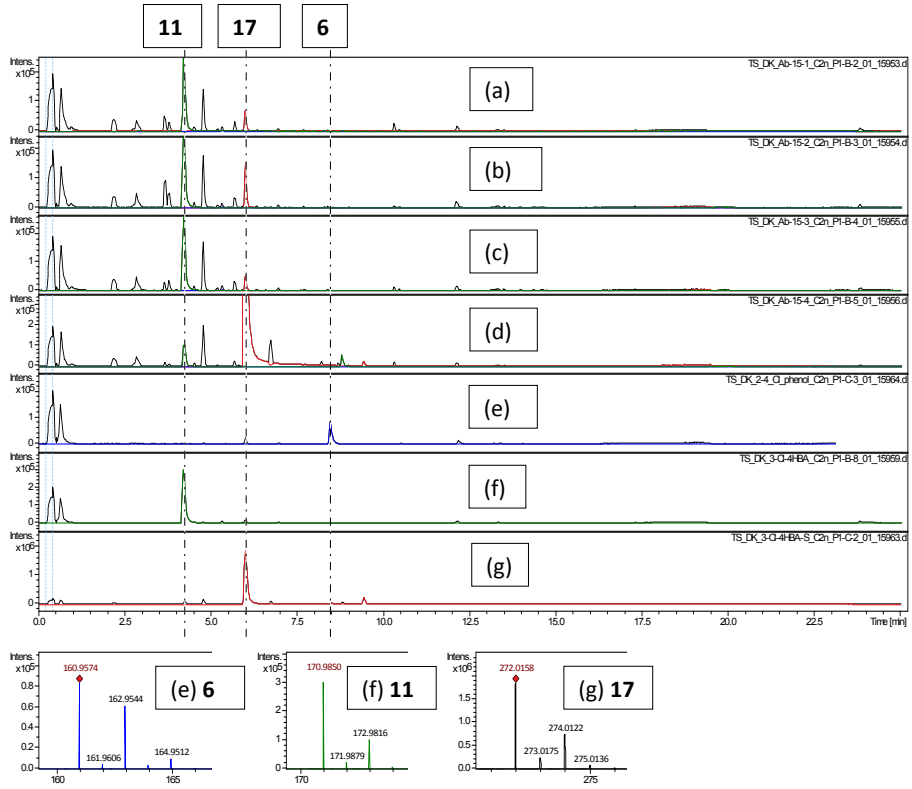


Figure S26. LCMS analysis of the *in vitro* assay using 3-Cl-4-HBA-SNAC (**17**) as the substrate of Ab1 in the presence of Ab9 (PCP-TE). The assay was carried out (a) without Ab1, as (b) complete reaction (with Ab1 and Ab9 (PCP-TE)), (c) without Fre, and (d) without Ab9 (PCP-TE). The (e) standard of 2,4-dichlorophenol (**6**), the (f) standard of 3-Cl-4-HBA (**11**), and the (g) standard of substrate **17** served for comparison. Mass spectra were measured in negative mode. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Blue chromatograms show the extracted ion chromatogram (EIC) of the $[M-H]^-$ peak of **6** (theoretical m/z = 160.9555). Green chromatograms show the EIC of the $[M-H]^-$ peak of **11** (theoretical m/z = 170.9843), while the red chromatograms display the EIC of the $[M-H]^-$ peak of **17** (theoretical m/z = 272.0143). The significant decrease of **17** and concomitant increase of **11** in the reactions (a)-(c) relative to (d) can be attributed to the activity of Ab9 (PCP-TE) which was not present in reaction (d). However, as apparent from the images, under various condition, the expected product **6** could not be observed.

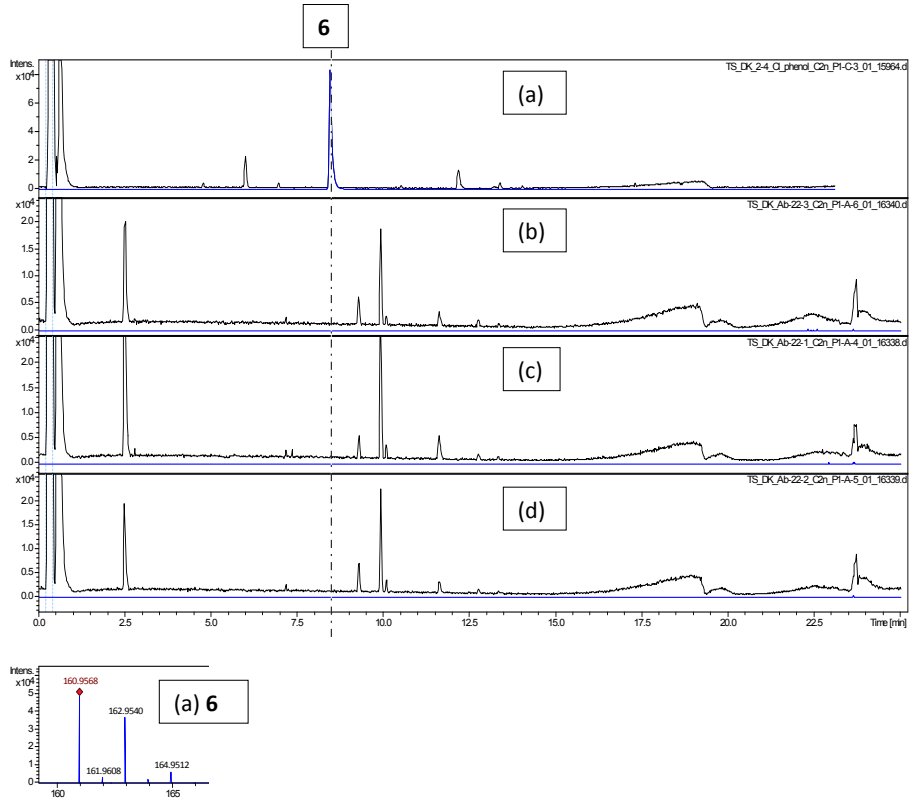


Figure S27. LCMS analysis of the *in vitro* assay using 2-chlorophenol as the substrate of Ab1. For comparison a (a) standard of the product 2,4-dichlorophenol (**6**) was employed. The assay was carried out as (b) complete reaction, (c) without Fre, and (d) without Ab1. Mass spectra were measured in negative mode. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Blue chromatograms show the extracted ion chromatogram (EIC) of the $[M\text{-}H]^-$ peak of **6** (theoretical m/z = 160.9555). As apparent from the images, the expected product **6** could not be detected.

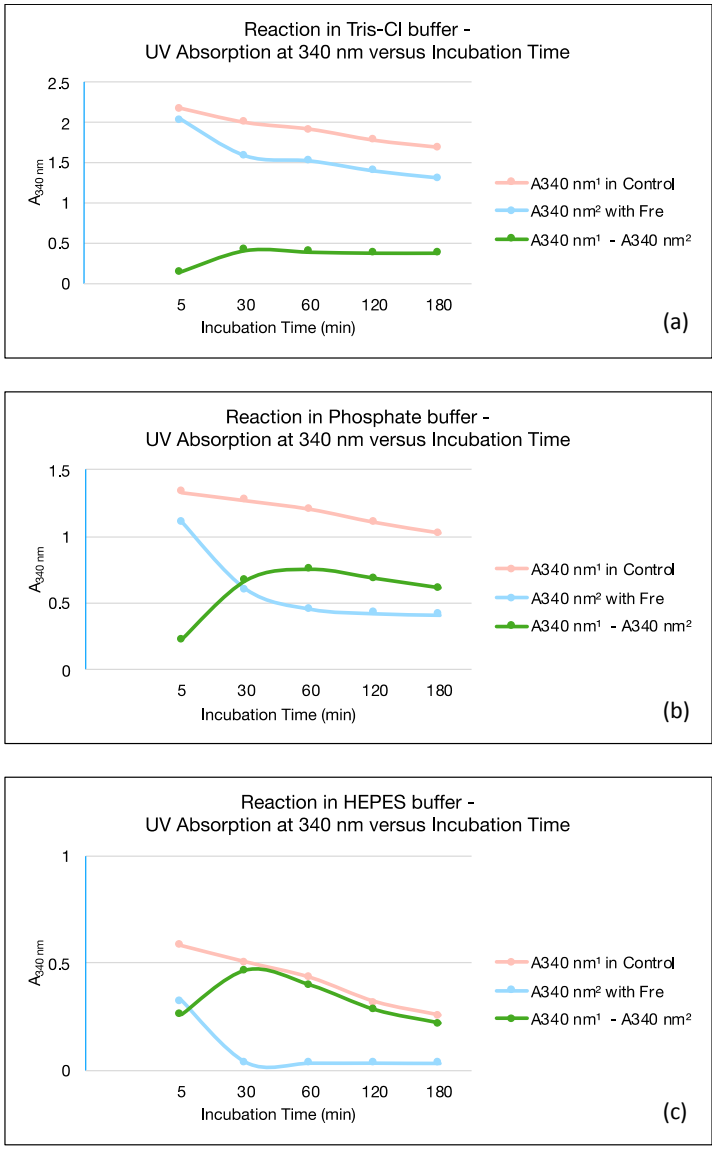


Figure S28. Graphic representation of Table S5 showing NADH consumption (green line) by Fre in (a) Tris-Cl, (b) Phosphate, and (c) HEPES buffer. Without Fre, the absorption ($A_{340\ nm}$) was decreasing over time, too. Therefore, it is likely that NADH naturally degraded over time. However, when Fre was present, the decrease of $A_{340\ nm}$ was accelerated, indicating that NADH as substrate of Fre was consumed. According to our results, Fre is active in all buffers tested.

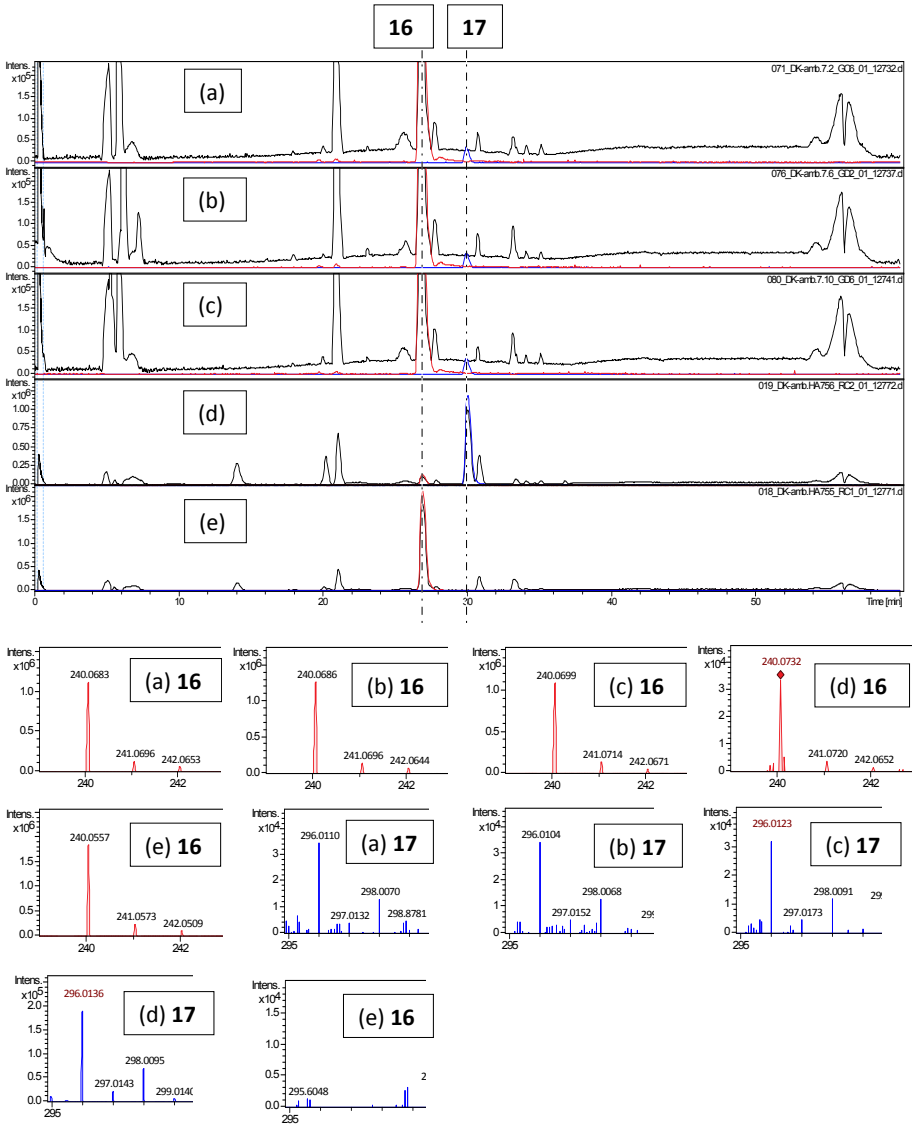


Figure S29. LCMS analysis of the *in vitro* assay of Ab10 in three different buffers. The utilized buffers were (a) Tris-Cl, (b) phosphate, and (c) HEPES, respectively. The (d) standard of the product 3-Cl-4-HBA-SNAC (**17**) and the (e) standard of the substrate 4-HBA-SNAC (**16**) served for comparison. Mass spectra were measured in positive mode. The upper part of the figure shows the respective chromatograms, while the lower part provides close-ups of the corresponding mass spectra. Base peak chromatograms (BPC) of each sample are shown in black. Blue chromatograms show the extracted ion chromatogram (EIC) of the $[M+Na]^+$ peak of **17** (theoretical m/z = 296.0119), while red chromatograms display the EIC of the $[M+H]^+$

peak of the substrate **16** (theoretical m/z = 240.0689). As shown in the images, the reaction product **17** was produced in comparable amounts in the three different buffers used (judging by the peak intensity), thus indicating that Ab10 works equally well in all three buffer systems.

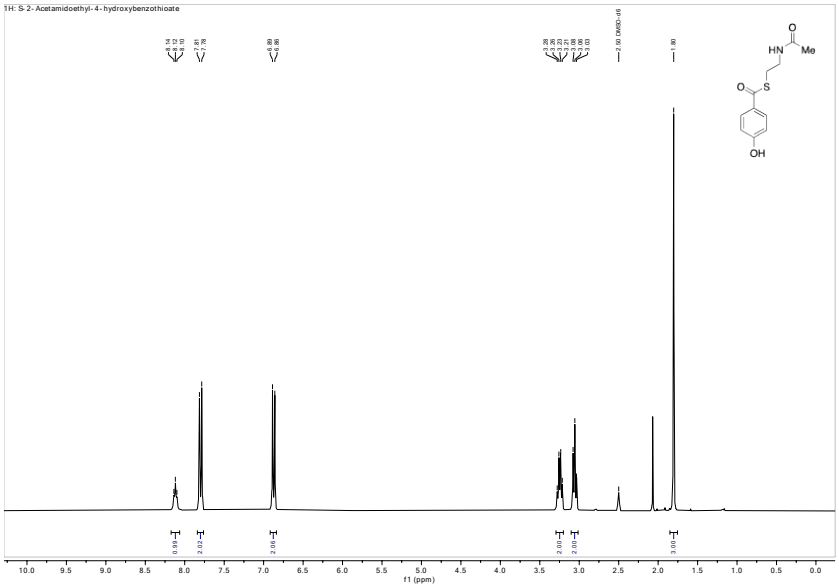


Figure S30. ^{1}H spectrum of *S*-2-acetamidoethyl 4-hydroxybenzothioate (**16**).

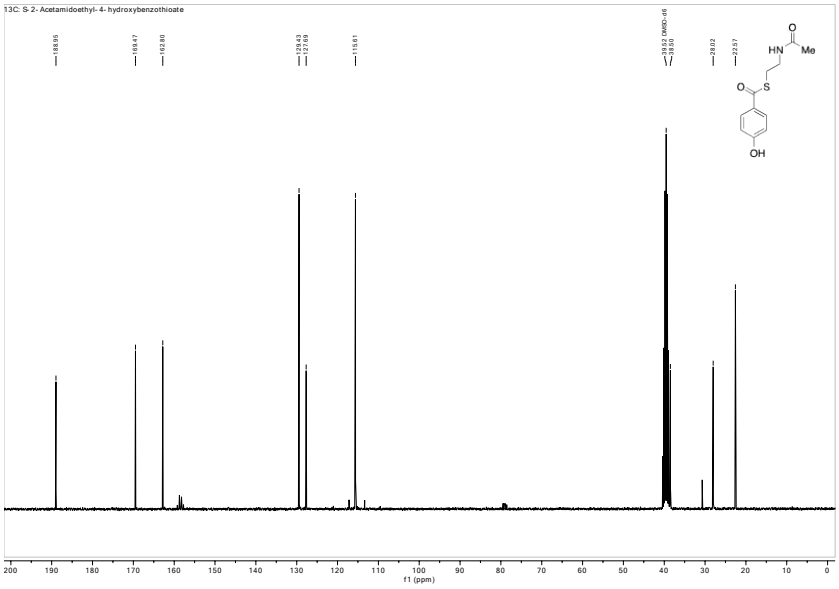


Figure S31. ^{13}C spectrum of *S*-2-acetamidoethyl 4-hydroxybenzothioate (**16**).

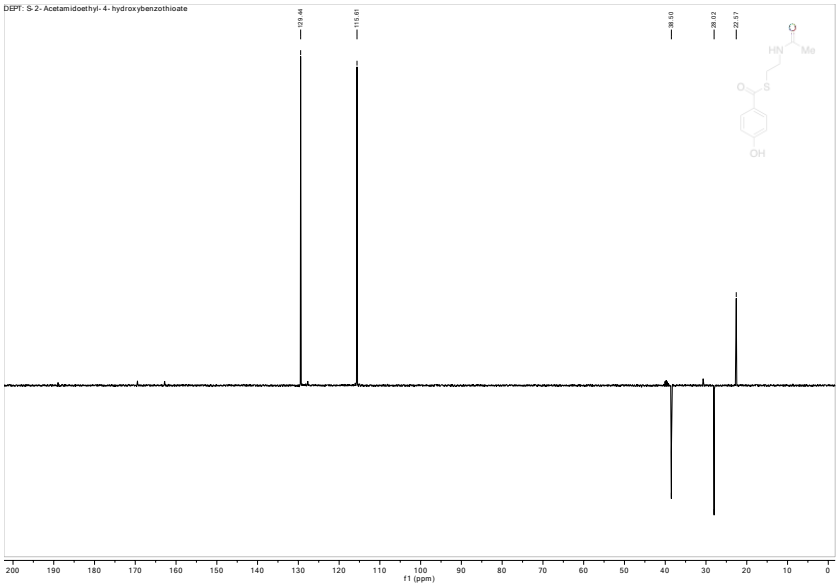


Figure S32. DEPT spectrum of *S*-2-acetamidoethyl 4-hydroxybenzothioate (**16**).

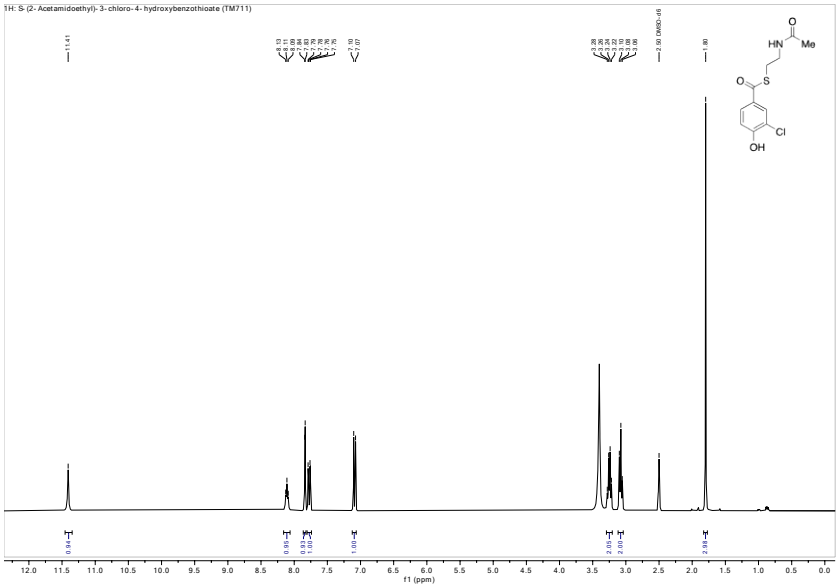


Figure S33. ^{1}H spectrum of *S*-2-acetamidoethyl 3-chloro-4-hydroxybenzothioate (**17**).

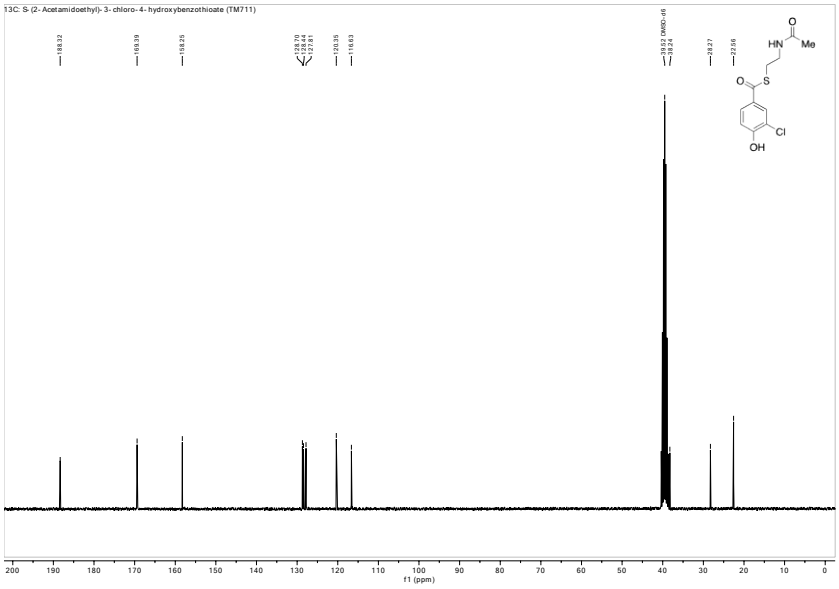


Figure S34. ^{13}C spectrum of *S*-2-acetamidoethyl 3-chloro-4-hydroxybenzothioate (**17**).

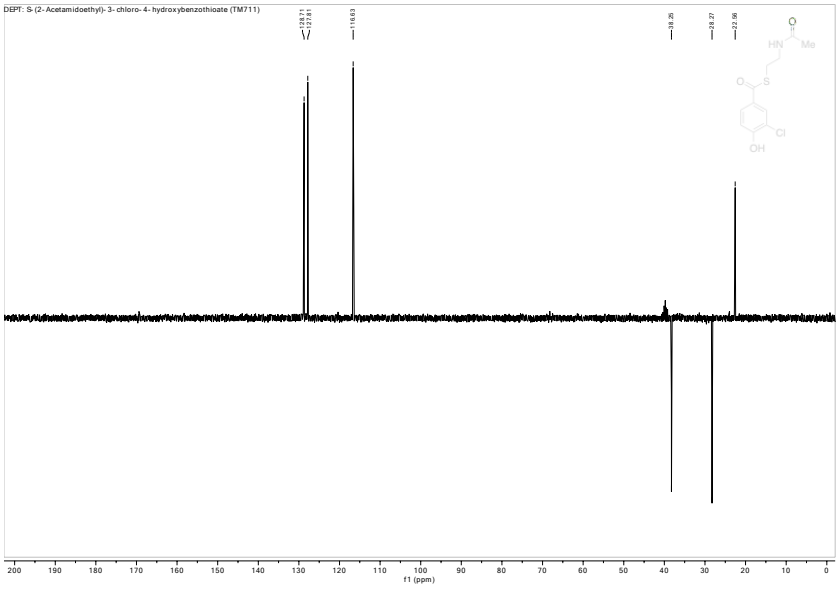


Figure S35. DEPT spectrum of *S*-2-acetamidoethyl 3-chloro-4-hydroxybenzothioate (**17**).

References

[1] X. Zhu, W. De Laurentis, K. Leang, J. Herrmann, K. Ihlefeld, K. H. van Pée, J. H. Naismith, Journal of Molecular Biology, 2009, 391, 74-85

[2] K. Podzelinska, R. Latimer, A. Bhattacharya, L. C. Vining, D. L. Zechel, Z. Jia, Journal of Molecular Biology, 2010, 397, 316-331.

3. Chapter II

Genome-mining-guided discovery and characterization of the PKS-NRPS-hybrid polyoxyperuin produced by a marine-derived Streptomycete

Kresna, I D. M., Wuisan, Z. G., Pohl, J. M., Mettal, U., Otoya, V. L., Gand, M., Marner, M., Otoya, L. L., Böhringer, N., Vilcinskas, A., Schäberle, T. F.

Published manuscript: *Journal of Natural Products* **85**, 888-898 (2022). doi: 10.1021/acs.jnatprod.1c01018

Summary: The manuscript presents the discovery and characterization of a PKS-NRPS-hybrid compound from a marine-derived Streptomycete by genome mining. The strain showed activity in an initial antimicrobial screening. The strain was genome sequenced and annotated for BGC identification. As a result, a BGC that we named *pop* was identified. DNA recombination was performed on one hand to overexpress the BGC by introducing a synthetic strong promoter upstream of a putative activator and on the other hand to knock out (KO) a structural gene in the BGC. LCMS analysis of the extracts revealed the absence of two ions in the KO strain, thereby confirming the products of the *pop* BGC. Furthermore, the concentrations of these compounds were significantly higher in the overexpression strain compared to the wild-type (WT). The overexpression strain was cultivated to enable purification of the products. Structure elucidation was performed, confirming new compounds named polyoxyperuin A seco acid and polyoxyperuin A. The latter possesses strong antibiotic activity against gram-positive pathogens.

Contribution: I Dewa Made Kresna designed and planned the cloning strategies and performed these experiments. He analyzed the results and wrote the most parts of the initial manuscript draft.

JOURNAL OF NATURAL PRODUCTS

pubs.acs.org/jnp Article

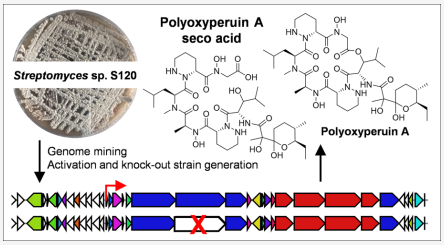
Genome-Mining-Guided Discovery and Characterization of the PKS-NRPS-Hybrid Polyoxyperuin Produced by a Marine-Derived Streptomycete

I Dewa M. Kresna, Zerlina G. Wuisan, Jean-Marie Pohl, Ute Mettal,* Virginia Linares Otoya, Martin Gand, Michael Marner, Luis Linares Otoya, Nils Böhringer, Andreas Vilcinskas, and Till F. Schäberle*

Cite This: *J. Nat. Prod.* 2022, 85, 888–898 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The azinothricin family comprises several cyclic hexadepsipeptides with diverse pharmacological bioactivities, including antimicrobial, antitumoral, and apoptosis induction. In this work, using a genome mining approach, a biosynthetic gene cluster encoding an azinothricin-like compound was identified from the *Streptomyces* sp. s120 genome sequence (*pop* BGC). Comparative MS analysis of extracts from the native producer and a knockout mutant led to the identification of metabolites corresponding to the *pop* BGC. Furthermore, regulatory elements of the BGC were identified. By overexpression of an LmbU-like transcriptional activator, the production yield of **1** and **2** was increased, enabling isolation and structure elucidation of polyoxyperuin A seco acid (**1**) and polyoxyperuin A (**2**) using high-resolution mass spectrometry and NMR spectroscopy. Compound **1** exhibited a low antibiotic effect against *Micrococcus luteus*, while **2** showed a strong Gram-positive antibiotic effect in a micro-broth-dilution assay.



The fact that pathogens are becoming resistant to available antibiotics implicates that continuous discovery of new compounds with antibacterial activity is necessary to fuel the antibiotic development pipeline. Natural products are a proven and valuable source for bioactive molecules, and activity-guided fractionation of extracts is a classic approach to identify compounds with desired bioactivity. However, a main drawback of this classical approach is that frequently detected antibiotic hits are caused by already known compounds, thus hampering the discovery of truly novel structures. Therefore, in the past decade additional data analysis became more and more important to prioritize samples before laborious isolation and profiling. An increase in publicly available sequence data in combination with great advances in bioinformatic tools for the detection of biosynthetic gene clusters (BGCs) and improvements of molecular biology techniques today represent an option to apply a genome mining approach.[1] The genome sequence of a given organism of interest can be analyzed with respect to BGCs encoding for the biosynthesis of specialized metabolites that might possess the desired bioactivity. However, it is still most challenging to predict the bioactivity of an *in silico* detected compound. Therefore, screening assays remain an essential part of discovery projects.

During the continuing screening efforts of our lab to detect and identify compounds with antibacterial activities, the strain *Streptomyces* sp. s120 (isolated from the Peruvian coast) showed inhibition of both clinically relevant methicillin-resistant *Staphylococcus aureus* (MRSA) and enterohemorrhagic *Escherichia coli* (EHEC) strains.[2] Therefore, the genome of this strain was sequenced and its metabolome was analyzed to dereplicate known compounds. Indeed, the known compounds naphthyridinomycin and resistomycin could be dereplicated using molecular networking, as well as *in silico* analysis, which enabled detection of BGCs corresponding to these compounds.[2] In addition, using antiSMASH,[3] a BGC was detected that shares high similarity to the BGCs of the peptidic natural products polyoxypeptin[4] and aurantimycin.[5] These molecules belong to the azinothricin family that consists of cyclic hexadepsipeptides (Figure 1). These hexadepsipeptides are biosynthesized by a nonribosomal peptide synthetase (NRPS) system, which starts synthesis of the core peptide after the polyketide synthase (PKS)-derived acyl side chain has been attached to the first amino acid residue.[4,5] For the various members of this cyclic peptide class, many different

Received: October 27, 2021
Published: March 3, 2022

https://doi.org/10.1021/acs.jnatprod.1c01018
J. Nat. Prod. 2022, 85, 888–898

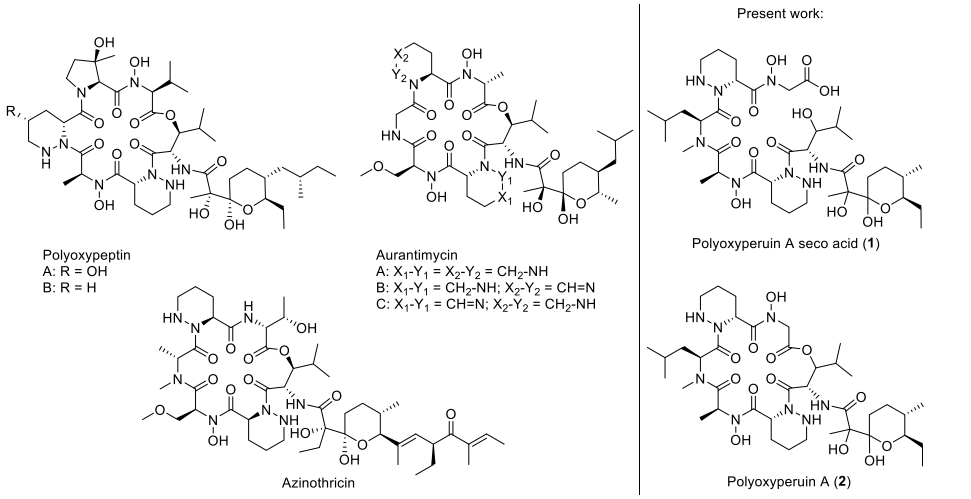


Figure 1. Chemical structures of selected compounds from the azinothricin family (left) and structures of polyoxyperuin A seco acid (**1**) and polyoxyperuin A (**2**) (right).

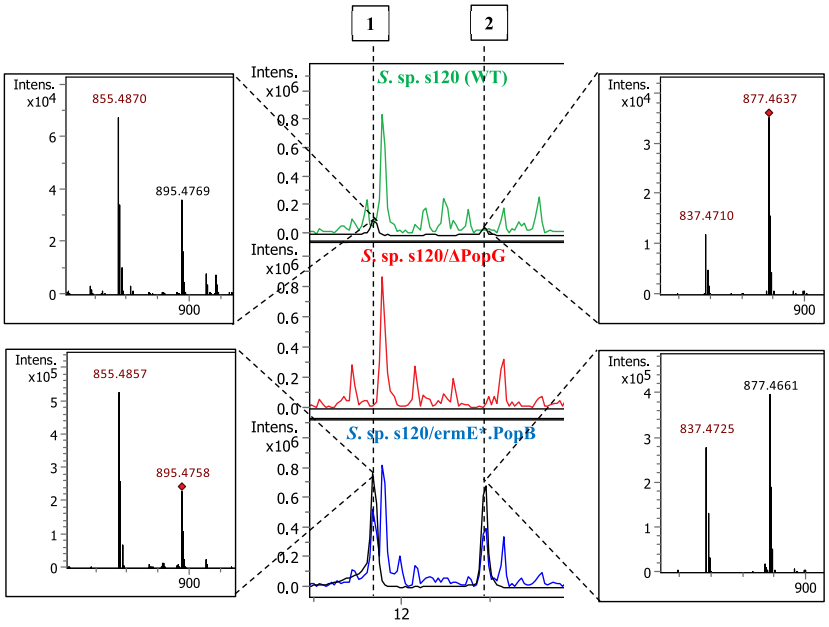


Figure 2. LCMS analysis of the base peak chromatogram (BPC) of the extract from *Streptomyces* sp. s120 (WT) (green), *Streptomyces* sp. s120/ΔPopG (red), and *Streptomyces* sp. s120/ermE*.PopB (blue). The extracted ion chromatograms of the $[M + Na]^+$ peak of compounds **1** and **2** are shown in black. The theoretical values calculated for $[M + Na]^+$ and $[M - H_2O + H]^+$ are m/z 895.4747 and 855.4822 for compound **1** and m/z 877.4642 and 837.4716 for compound **2**.

bioactivities such as apoptosis-inducing activity (polyoxypeptins A and B),[6] antitumor activity (diperamycin,[7] citropeptin,[8] and kettapeptin[9]), and activity against Gram-positive bacteria (IC101,[10] diperamycin,[7] variapeptin,[11] aurantimycin,[12] kettapeptin,[9] azinothricin,[13] and A83586C[14]) were reported. Among these substances, only the PKS-NRPS hybrid BGCs

https://doi.org/10.1021/acs.jnatprod.1c01018
J. Nat. Prod. 2022, 85, 888−898

of polyoxypeptin and aurantimycin have been described so far.[4,5] The antibacterial target is not described; however it was reported that a multidrug resistance ABC transporter is mediating aurantimycin A resistance.[15] Here, we report the *in silico* identification of a putative BGC and its experimental proof by generation of knockout (KO) and activation mutants. This enabled the isolation of the compounds and their structure elucidation, which was found to be in accordance with the proposed biosynthetic hypothesis.

RESULTS AND DISCUSSION

Analysis of the Polyoxyperuin BGC and Identification of Compounds 1 and 2. Scanning the assembled genome sequencing data of *Streptomyces* sp. s120 for putative BGCs encoding for the biosynthesis of natural products with antibacterial activity, a BGC (accession number: OK625705) was identified with 48% similarity to the polyoxypeptin and 51% to the aurantimycin BGC, respectively (Figure S1). To identify the product corresponding to this BGC, a gene putatively encoding an NRPS module (i.e., *popG*) was knocked-out by inserting an apramycin resistance gene cassette (Figure S2), resulting in strain *Streptomyces* sp. s120/ΔPopG. This modification led to the identification of two $[M + Na]^+$ ions at m/z 895.47 (compound **1**, named polyoxyperuin A seco acid) and 877.46 (compound **2**, named polyoxyperuin A) in the extract of the wild-type strain, which were absent in the extract of the knockout strain (Figure 2). Based on the HRMS data, the molecular formulas of compounds **1** and **2** were established as $C_{39}H_{68}N_8O_{14}$ and $C_{39}H_{66}N_8O_{13}$. The increase in the number of ring and double-bond equivalents (RDBE) for compound **2** (RDBE = 10 for compound **1**; RDBE = 11 for compound **2**) and the similar fragment ions observed in the MS/MS spectra (Figure 6) for both compounds suggest that compound **1** represents a linear peptide, while compound **2** contains a closed ring system. However, the production was so low that these ions could only be observed in the respective extracted ion chromatograms (EIC), while they could not be detected in the base peak chromatogram. Therefore, in a next step it was planned to induce the expression of the biosynthetic genes. Analyzing the predicted open reading frames (ORFs) of the BGC, two LmbU-like regulatory genes were detected (i.e., *popб* and *popB*) (Figure 4a and Table S1). LmbU is regarded as a transcriptional regulator that promotes biosynthesis, as it was for instance shown in the biosynthesis of lincomycin,[16] hormaomycin,[17] and himastatin.[18] The eight base pair palindromic sequence (5′-CGCCGGCG-3′), which forms the LmbU binding site,[16] was identified upstream of several genes of the BGC, e.g., *popE*, *popJ*, *popK*, and *popR* (Figure 4a and Figure S3). Hence, one of the two LmbU-like regulators, *popB*, was overexpressed by introduction of the strong constitutive synthetic promoter *ermE** upstream of *popB* (Figure S4). This yielded the transgenic strain *Streptomyces* sp. s120/ermE*.PopB. Subsequently, the wild type and the overexpression strain were fermented. Analysis of the resulting extracts by HRESI/MS^2 showed that the peak area of compounds **1** and **2** was 8 and 16 times higher, respectively, for the strain overexpressing *popB* compared to the wild-type control (Figure 2). This result showed that the BGC is functional and that PopB is indeed an activator.

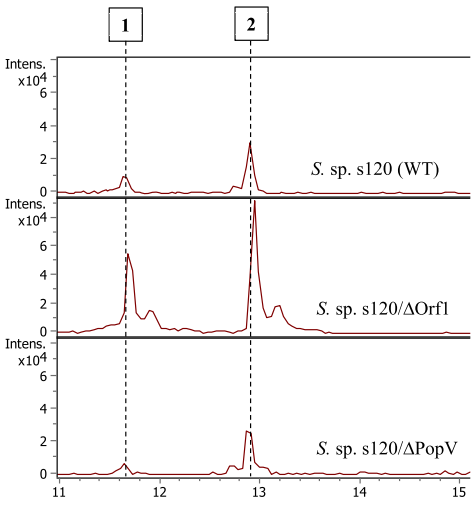


Figure 3. Comparison of LCMS chromatograms for the extract of *Streptomyces* sp. s120 (WT), *Streptomyces* sp. s120/ΔOrf1, and *Streptomyces* sp. s120/ΔPopV. Displayed are the extracted ion chromatograms of the $[M + Na]^+$ peak for compounds **1** and **2**.

The KO and overexpression strains already linked the *in silico* identified BGC with polyoxyperuin biosynthesis, and it was aimed to obtain further insights into the minimum BGC. Therefore, the genes *orf1* and *popV*, which are annotated as DNA-binding protein and acyl-CoA carboxylase, respectively, were inactivated by insertion of an apramycin resistance cassette, yielding the transgenic strains *Streptomyces* sp. s120/ΔOrf1 and *Streptomyces* sp. s120/ΔPopV (Figures S5 and S6). Analysis of extracts by HR-UPLC-MS revealed that compounds **1** and **2** could still be produced by the transgenic strains (Figure 3). Hence, Orf1 and PopV are not essential for polyoxyperuin biosynthesis; however they are involved in regulation and precursor supply. The increase in the production of compounds **1** and **2** that was observed when *orf1* was disrupted indicated that Orf1 serves as a repressor in polyoxyperuin biosynthesis. On the other hand, PopV, an acyl-CoA carboxylase, is predicted to be involved in the synthesis of malonyl-CoA and methylmalonyl-CoA as the extender units for the PKS system. We assume that deletion of *popV* did not affect the production of compounds **1** and **2**, since these common extender units can be complemented in trans.

Biosynthesis of Polyoxyperuins. In general, the putative polyoxyperuin biosynthesis can be divided into several steps. The acyl side chain is built up by an assembly-line-like polyketide synthase type 1 system. ATSignature and Minowa algorithms, which are integrated in antiSMASH,[3] predicted that (2*S*)-methylmalonyl-CoA (L-configuration) serves as the substrate for the AT domains of PopN, O, and Q, while for PopP malonyl-CoA acts as substrate (Table S2). The conserved motifs GH**S**VG and HAF**H** (catalytic residues are in bold) in PopP (Figure S7) and the presence of GH**S**QG and YAS**H** motifs in PopN, O, and Q are also indicating PopP to be specific for malonyl-CoA and the other ones for (2*S*)-methylmalonyl-CoA (Figure S7).[19]

In the PKS PopN, the catalytic cysteine residue for *trans*-thioesterification in ketosynthase (KS) domains is substituted with glutamine (Figure S8). Such a mutation in KS is often

https://doi.org/10.1021/acs.jnatprod.1c01018
J. Nat. Prod. 2022, 85, 888−898

Journal of Natural Products pubs.acs.org/jnp Article

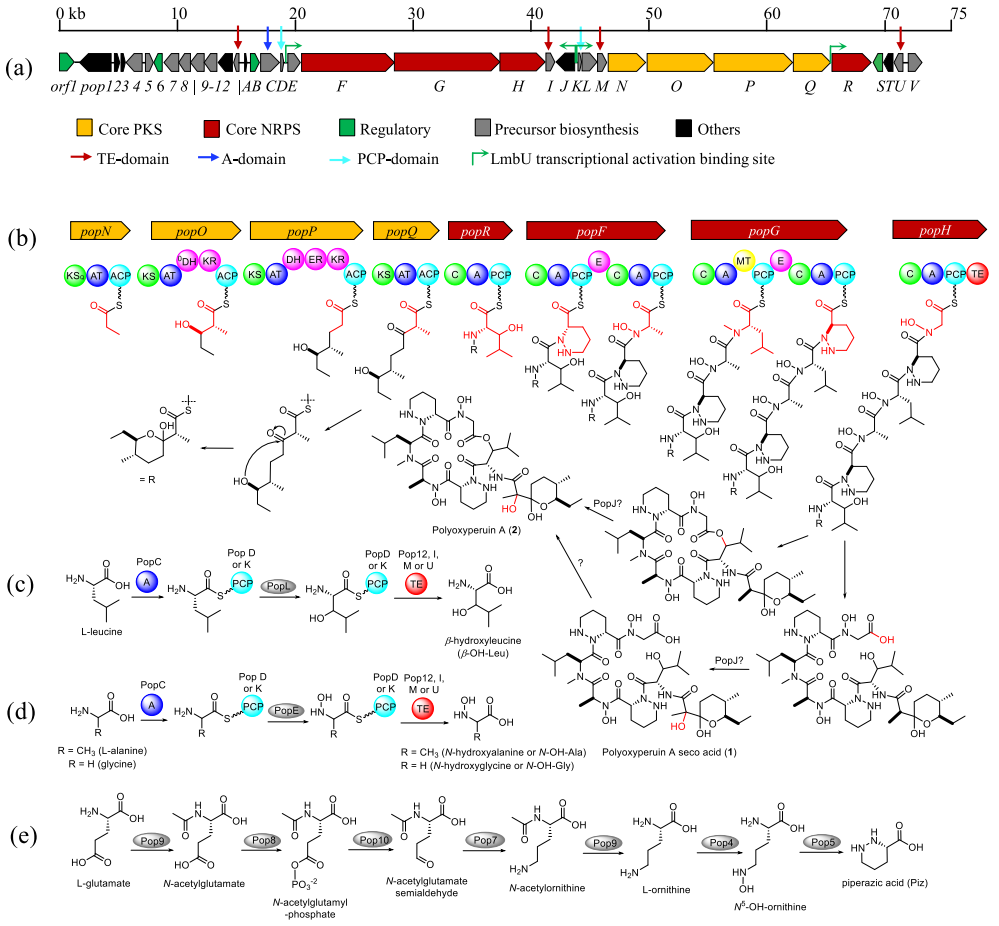


Figure 4. (a) BGC and (b) proposed biosynthetic pathway of compounds **1** and **2**. Proposed biosynthetic pathways of the nonproteinogenic amino acids (c) β-hydroxy-L-leucine (β-OH-L-Leu), (d) *N*-hydroxy-L-alanine (*N*-OH-L-Ala) and *N*-hydroxyglycine (*N*-OH-Gly), as well as (e) L-piperazic acid (L-Piz).

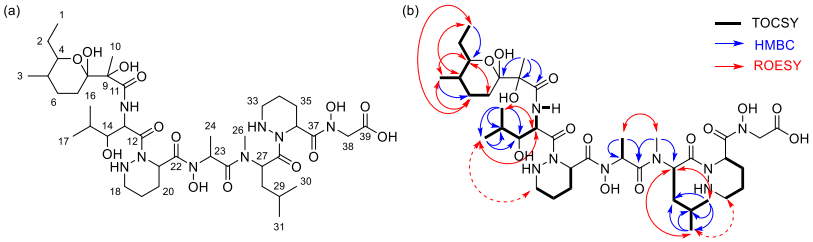


Figure 5. (a) Proposed structure of **1** and (b) the key correlations from HMBC (blue arrows) and ROESY experiments (red arrows). The spin systems derived from TOCSY experiments are displayed as bold black bonds. Dashed lines represent weak correlations.

https://doi.org/10.1021/acs.jnatprod.1c01018
J. Nat. Prod. 2022, 85, 888−898

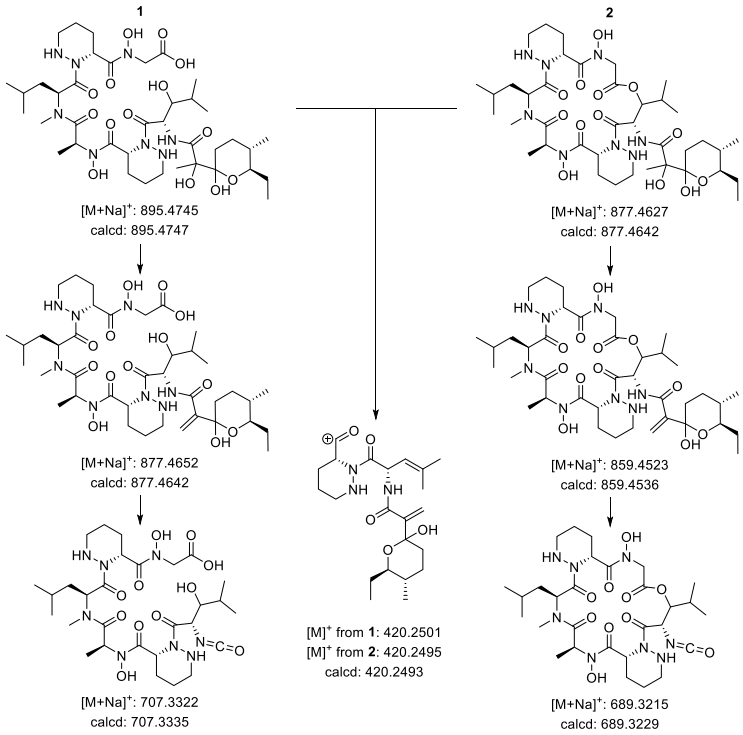


Figure 6. Proposed fragmentation pattern of **1** and **2** obtained from HRMS analysis (more details are shown in Figures S12−S14).

found in the loading modules of modular type I PKSs and is commonly referred to as "KS_Q".[19,20] The glutamine residue in this position has been reported to organize the KS for decarboxylation in the chain initiation reaction.[19,20] For the putative loading module PopN, decarboxylation of the ACP-tethered methylmalonyl-CoA results in a propionate residue as first unit of the PKS chain. The following PopO harbors, beside the three basic domains KS, AT, and ACP, an additional dehydratase (DH) and ketoreductase (KR) domain, whereas PopP carries a DH, enoylreductase (ER), and KR domain, indicating conversion of the preceding building block to a saturated alkyl moiety. Amino acid sequence analysis of the ER domain of PopP suggested that it belongs to the D-type ERs, which do not possess the fingerprint residue tyrosine, known to influence the stereochemistry of reduction[19] (Figure S9). The domain architectures for the PopO homologous enzymes, ArtP in the aurantimycin[5] and PlyU in the polyoxypeptin[4] BGC, are the same. However, the DH domains of ArtP and PlyU are proposed to be nonfunctional.[5,4] While the inactivity of the DH from PlyU is due to a mutation of a key amino acid residue from histidine to glutamine in the conserved motif HxxxGxxxxP,[4] it remains elusive why the DH of ArtP is inactive, as the key amino acids are conserved[5] (Figure S10). Based on this observation, the DH domain from PopO, even though showing nonmutated conserved motifs (Figure S10), was predicted to be nonfunctional, which is verified by the collinearity of the structures **1** and **2**. Moreover, the presence of the LDD motif in PopO(KR) and the absence of proline two positions downstream of the catalytic tyrosine indicate that PopO(KR) belongs to the group of B1-type KRs.[19] The PopP(KR) on the other hand can be classified as an A1-type KR, due to the absence of the LDD motif as well as the absence of histidine three positions upstream of the catalytic tyrosine[19] (Figure S11).

After the assembly of the PKS-derived acyl side chain and the biosynthesis of the nonproteinogenic amino acid precursors, these building blocks are assembled by the NRPS assembly line. Among the proposed amino acid building blocks of the polyoxyperuins **1** and **2**, four are nonproteinogenic amino acids (i.e., *β*-OH-L-Leu, L-Piz, *N*-OH-L-Ala, and *N*-OH-Gly). Genes that are putatively encoding the enzymes catalyzing precursor biosynthesis were identified (Figure 4 and Table S1). For instance, the enzymes Pop4-5 and Pop7-10 could catalyze L-Piz formation starting from glutamate (Figure 4e). Furthermore, the free-standing A-domain PopC is proposed to be involved in the biosynthesis of *β*-OH-L-Leu, *N*-OH-L-Ala, and *N*-OH-Gly (Figure 4c and d). PopC shares 100% Stachelhaus signature sequence similarity to PlyC and ArtC (Table S3). For the homologue ArtC, it has been shown experimentally that L-alanine, L-leucine, L-serine, L-valine, and glycine are accepted as substrates.[5] In analogy, first, the stand-alone A-domain (PopC) activates L-leucine, L-alanine, or glycine and attaches it to a stand-alone PCP (PopD or PopK). The attachment of the hydroxy group in *β*-OH-L-Leu is

https://doi.org/10.1021/acs.jnatprod.1c01018
J. Nat. Prod. 2022, 85, 888−898

Journal of Natural Products pubs.acs.org/jnp Article

Table 1. NMR Analysis of Compound 1 in $CDCl_3$ (^1H at 600 MHz, ^{13}C at 150 MHz)

position	δ_C,[a] type	δ_H,[b] mult. (J in Hz)	HMBC	ROESY
1	9.8, CH_3	0.81, t (7.2), obs[c]	4	5, 6
2	29.8, CH_2	1.25, m		
3	17.8, CH_3	0.81, d (6.5)	5, 6, 7	4, 5, 6
4	77.1, CH	3.44, t (9.7)		7
5	34.5, CH	1.25, m		1, 3
6	25.6, CH_2	1.61, m		1, 3
		1.24, m		1, 3
7	27.7, CH_2	1.56, m		
8	98.9, C			
9	77.7, C			
10	21.1, CH_3	1.56, s	8, 9, 11	
11	177.2, C=O			
11-NH		7.97, d (8.8)		
12	ND[d], C=O			
13	47.1, CH	5.94, d (7.5)		16, 17
14	79.7, CH	3.46, m		16, 17
15	32.1, CH	1.70, m		
16	18.6, CH_3	0.99, d (7.1), obs[c]	14,15,17	14
17	19.1, CH_3	0.98, d (7.4), obs[c]	14, 15, 16	14
18	47.3, CH_2	3.16, d (13.7)		
		2.76, d (13.7)		16, 17
19	21.4, CH_2	1.57, m, obs[c]		
		1.53, m, obs[c]		
20	24.9, CH_2	2.22, d (14.3)		
		1.95, m		
21	49.3, CH	5.58, d (5.6)		20
22	ND,[d] C=O			
23	51.7, CH	5.36, q (7.1)		24, 26
24	14.1, CH_3	1.46, d (7.1)	23, 25	23, 26
25	175.9, C=O			
26	32.4, CH_3	3.10, s	25, 27	24
27	54.3, CH	5.61, d (11.0)		30, 31
28	37.0, CH_2	1.72, m		30, 31
		1.67, m		
29	26.2, CH	1.45, m, obs[c]		
30	23.5, CH_3	0.95, d (6.5)	28, 29, 31	28
31	20.9, CH_3	0.90, d (6.5)	28, 29, 30	28
32	ND,[d] C=O			
33	47.1, CH_2	3.09, m, obs[c]		30, 31
		2.70, d (14.0)		/
34	20.2, CH_2	1.78, m, obs[c]		
		1.58, m, obs[c]		
35	25.5, CH_2	2.18, d (14.2)		
		1.89, m		
36	47.0, CH	5.47, d (6.0)		35
37	ND,[d] C=O			
38	51.0, CH_2	4.67, d (17.2)		
		4.04, d (17.2)		
39	ND,[d] C=O			

[a] ^{13}C chemical shifts assigned by HSQC and HMBC correlations. [b] ^1H chemical shifts partially assigned by COSY, TOCSY, and ROESY correlations. [c] The abbreviation "obs" is used for signals partially obscured by a different compound signal. [d] Not detectable due to low concentration.

proposed to be catalyzed by the putative cytochrome PopL. This is in accordance with the activity proposed for the homologue ArtM.[5] Moreover, the introduction of the hydroxy group in *N*-OH-L-Ala and *N*-OH-Gly is proposed to be catalyzed by the oxidoreductase/hydrolase PopE, which is in accordance with the activity proposed for the homologue ArtE.[5] Thereafter, the modified *β*-OH-L-Leu, *N*-OH-L-Ala, and *N*-OH-Gly are subsequently released by catalytic activity of a stand-alone thioesterase (TE) (e.g., Pop12, I, M, or U).

Furthermore, the substrate preferences of most of the A-domains in the NRPS modules could be predicted by *in silico* analysis (Stachelhaus signature sequence) and are in accordance with the final molecules **1** and **2**, except for the A_1-domain of PopG and the A-domain of PopH. The first A-

https://doi.org/10.1021/acs.jnatprod.1c01018
J. Nat. Prod. 2022, 85, 888−898

domain of the NRPS assembly line, PopR, is predicted to activate β-OH-L-Leu, which is as well in accordance with the highly similar signature sequence of known A-domains recognizing β-OH-L-Leu (Table S4). The starter condensation (C) domain of PopR connects the acyl side chain to β-OH-L-Leu, thus initiating the assembly of the peptide core architecture. Likewise, it was found that the A_1-domain of PopF and the A_2-domain of PopG recognize L-Piz (Table S5), while the A_2-domain of PopF recognizes *N*-OH-L-Ala (Table S6). For the A_1 domain of PopG we propose that it activates L-leucine, which is modified by the methyl transferase (MT) present in this module. Furthermore, *N*-OH-Gly should be the substrate of the PopH A-domain. This is supported by the fact that the substrate signature sequence shares 70% similarity to the respective A_1-domain of ArtG in aurantimycin biosynthesis, which activates glycine (Table S7). Based on the high similarity of the polyoxyperuin BGC and structure to aurantimycin and polyoxypeptin, it is assumed that the epimerase (E) domains in PopF and PopG likewise recognize the two L-Piz moieties and thereupon convert them to the D-configuration.

Finally, the nascent chain is released from the NRPS system by the catalytic activity of the TE domain (PopH). It can be expected that the cyclized form of the molecule **2** represents the final natural product. However, hydrolysis can also occur without cyclization, yielding the linear peptide **1**. This phenomenon has been previously observed among NRPS-derived products; for instance for the GameXPeptides it was reported that the speed of the peptide synthesis might be too fast to allow for correct thioesterase-catalyzed cyclization, which was probably competing with faster hydrolytic release.[21] Even though from our point of view it is most likely that cyclization and release from the NRPS system take place simultaneously by the action of PopH, it cannot be ruled out that **1** might be converted to **2** by the action of a suitable, so far unidentified, stand-alone enzyme. The final oxidation step, resulting in hydroxylation of the acyl side chain and thus yielding the final products **1** and **2**, is proposed to be catalyzed by the putative FAD-dependent monooxygenase PopJ. However, the time point of this tailoring reaction remains elusive.

Structure Elucidation. In order to isolate both molecules, a 9 L cultivation of *Streptomyces* sp. s120/ermE*.PopB was carried out. HPLC-based purification yielded 1.4 mg of pure compound **1** and 1.4 mg of compound **2**, although not entirely pure. Further purification of compound **2** was performed, but it could not be recovered in sufficient amount after these additional purification steps.

NMR experiments of **1** identified in total six amino acids, which are β-hydroxyleucine (β-OH-Leu), *N*-OH-alanine (*N*-OH-Ala), *N*-methyl-leucine (*N*-Me-Leu), *N*-OH-glycine (*N*-OH-Gly), and two piperazic acid (Piz) moieties as well as one tetrahydropyran (THP)-based residue. The presence of *N*-OH-Gly and *N*-OH-Ala was derived from the chemical shift of the respective α-CH/α-CH_2 (δ_C 51.7 and δ_H 5.36 for *N*-OH-Ala; δ_C 51.0 and δ_H 4.67 and 4.04 for *N*-OH-Gly), which is more deshielded compared to the unhydroxylated amino acids.[7,22] Correlations in HMBC and ROESY experiments showed that the amide nitrogen between the alanine and leucine residues is methylated (position 26; Table 1). According to the COSY experiment, β-OH-Leu is the sole amino acid bearing an unsubstituted NH group in the system (position 11-NH; Table 1). As the ^{13}C chemical shift of the carbonyl group for the THP-based residue (δ_C 177.2) does not match the shift expected for a heteroatom-substituted amide, whereas it fits quite well for a secondary amide,[7] it can be ruled out that the THP moiety is bound to either of the piperazic acid residues or any of the *N*-hydroxylated amino acids. Furthermore, HMBC and ROESY experiments showed that the amino function of *N*-Me-Leu is attached via an amide bond to *N*-OH-Ala. Consequently, it was concluded that the THP moiety forms a secondary amide bond with the amino function of β-OH-Leu. The further sequence of the amino acids was determined based on the MS/MS fragmentation and the observed ROESY correlations, which gave the order β-OH-Leu-Piz-*N*-OH-Ala-*N*-Me-Leu-Piz. The amino acid glycine showed no correlations to another fragment in any 2D spectra and was therefore determined to be located at the C-terminus of the peptide using the method of exclusion. Concerning the noncyclic nature of peptide **1**, we found that the signal for the proton at the β-carbon of the β-OH-Leu (position 14; Table 1) experiences a considerable shielding in comparison to the expected value of the alkoxy moiety of an ester. This directly indicates it to be a hydroxy moiety, leading to the conclusion that there is no ring closure between β-OH-Leu and glycine. On the basis of these data, we propose the structure of **1** in Figure 5.

The ^1H NMR spectrum of the isolated compound **2** showed a significant degree of similarity to the spectrum of **1**. Similar observations were made during MS/MS experiments, in which **1** and **2** showed similar fragmentation patterns (Figure 6 and Figures S12−S14). The proposed fragmentation patterns for both samples are given in Figure 6. This was a further indication that both compounds are of similar makeup. Consequently, we propose the structure of **2** to be the cyclic depsipeptide. Yet, even though refermentation was carried out, we were not able to obtain a sufficient quantity of pure **2** for conclusive NMR analysis to confirm the open/closed relationship between **1** and **2**. However, in order to investigate the open/closed relationship between compounds **1** and **2**, a sample of compound **2** was hydrolyzed using a solution of potassium carbonate. Even though the hydrolysate still contained unreacted educt and the reaction was accompanied by the formation of various side products, LCMS analysis (Figures S15 and S16) showed the formation of a substance eluting at the same retention time (t_R = 11.7 min) as compound **1**, displaying peaks at *m*/*z* 855.4793 ($[M - H_2O + H]^+$ of **1**) and *m*/*z* 895.4744 ($[M + Na]^+$ of **1**), thereby further supporting the open/closed relationship between compounds **1** and **2**.

The proposed structures are in accordance with the information obtained from the investigation of the BGC. The absolute configurations of the amino acids and the THP ring were proposed based on the analysis of the BGC and similarities to known azinothricin derivatives. The configurations at C-8, C-9, and C-14 are unknown because the gene cluster analysis did not provide strong trends to predict the configuration of those atoms, as the selectivity of the cyclization/hydroxylation reactions creating these stereocenters remains elusive.

Biological Activities of Compounds 1 and 2. Compounds **1** and **2** were tested for their antimicrobial activity even though **2** was not completely pure. In the employed test panel, **1** showed moderate activity against *Micrococcus luteus*, while **2** showed strong activity against nearly all tested Gram-positive pathogens (Table 2). As expected

https://doi.org/10.1021/acs.jnatprod.1c01018
J. Nat. Prod. 2022, 85, 888−898

Table 2. Minimum Inhibitory Concentration in μg/mL of Compounds 1 and 2

test organism	compound 1	compound 2[a]
Escherichia coli ATCC35218	>64	>64
Staphylococcus aureus ATCC25923	>64	0.125
Mycobacterium smegmatis ATCC607	>64	>64
Bacillus subtilis DSM10	>64	0.125−0.25
Micrococcus luteus DSM20030	32	<0.031
Listeria monocytogenes DSM20600	>64	<0.031

[a]Compound is not completely pure.

from the comparison to other azinothricin-like compounds IC101,[10] diperamycin,[7] variapeptin,[11] aurantimycin,[12] kettapeptin,[9] azinothricin,[13] and A83586C,[14] which exhibited antibiotic activity against Gram-positive pathogens, the cyclic depsipeptide **2** showed superior activity relative to the linear compound **1**. Even though compound **2** was not entirely pure, the increase in activity is so pronounced that it should not be caused solely by residual impurities.

CONCLUSION

Driven by genome mining, a BGC belonging to the azinothricin family was identified for *Streptomyces* sp. s120. Comparative MS-profiling of the extracts produced by the WT and the transgenic knockout strain (ΔPopG) showed two missing peaks at *m*/*z* 895.47 Da and *m*/*z* 877.46 Da, which correspond to the polyoxyperuins **1** and **2**, respectively. Disruption of the genes *orf1* and *popV* did not abolish the production of **1** and **2**, indicating those genes are not essential for polyoxyperuin biosynthesis. Moreover, the strong constitutive promoter *ermE** was introduced upstream of PopB. Using the later mutant strain, production of peptides **1** and **2** was significantly enhanced compared to the WT strain, confirming PopB as an activator. This overexpression strain, *Streptomyces* sp. s120/ermE*.PopB, was used to produce polyoxyperuin for NMR analysis and activity profiling. An effort to optimize the production yielded 1.4 mg of the pure linear peptide **1** and 1.4 mg of the not completely pure cyclic peptide **2** from a total cultivation volume of 9 L. Structure elucidation of the polyoxyperuins by NMR and HRMS proved the viability of the chosen BGC-based discovery approach. Compound **1** showed moderate activity against *Micrococcus luteus*, while **2** showed strong activity to nearly all tested Gram-positive pathogens.

EXPERIMENTAL SECTION

General Experimental Procedures. ^{1}H NMR spectra (600 MHz) and ^{13}C NMR spectra with proton noise decoupling (^{13}C{^{1}H} NMR, 150 MHz) were recorded on a Bruker "Avance III HD 600 MHz" spectrometer. Chemical shifts are denoted in ppm relative to the residual $CHCl_3$ signal at 7.26 ppm (^{1}H) and 77.16 ppm (^{13}C).[23] NMR spectra were analyzed using the software TopSpin 4.1.1 (Bruker BioSpin GmbH) and MestReNova 14.2.1 (Mestrelab Research, S.L.).

Standard ultra-high-performance liquid chromatography high-resolution mass spectrometry (UHPLC-HRMS) measurements were carried out on an Agilent Infinity 1290 UPLC system equipped with an Acquity UPLC BEH C18 1.7 μm (2.1 × 100 mm) column and an Acquity UPLC BEH C18 1.7 μm VanGuard pre-column (2.1 × 5 mm) (Waters) coupled to a DAD detector and a Bruker "micrOTOFQ II" mass spectrometer (Bruker Daltonics) with an electrospray ionization source. MS/MS fragmentation was performed using the auto-MS/MS setting. Calibration of mass spectra was achieved using 10 mM sodium formate in H_2O/^{i}PrOH (1:1) as internal standard.

High-accuracy UHPLC-HRMS measurements were performed on a 1290 UHPLC system (Agilent) equipped with a DAD, ELSD, and maXis II (Bruker) ESI-qTOF-UHRMS. The employed column was an Acquity UPLC BEH C18 1.7 μm (2.1 × 100 mm) column with an Acquity UPLC BEH C18 1.7 μm VanGuard pre-column (2.1 × 5 mm). The column oven temperature was maintained at 45 °C, and the following gradient was used: 0 min: 95% A; 0.30 min: 95% A; 18.00 min: 4.75% A; 18.10 min: 0% A; 22.50 min: 0% A; 22.60 min: 95% A; 25.00 min: 95% A (A: H_2O, 0.1% formic acid; B: MeCN, 0.1% formic acid; flow: 600 μL/min). The injection volume was either 1 or 2 μL. MS data were acquired over a range from *m*/*z* 50 to 2000 in positive mode. Auto MS/MS fragmentation was achieved with rising collision energy (for single charged ions: 35−70 eV over a gradient from *m*/*z* 500 to 2000; for double charged ions: 25−50 eV over a gradient from *m*/*z* 500 to 2000). Calibration of mass spectra was achieved using sodium formate in H_2O/^{i}PrOH (1:1) as internal standard.

Analysis of mass spectra was accomplished using the software Data Analysis 4.2 (Bruker Daltonics).

Bacterial Strain. The *Streptomyces* sp. s120 strain was previously isolated from the Peruvian coast.[2] Its 16S rRNA sequence was deposited at GenBank with the accession number MF796625.[2] The strain grew optimally at a temperature of 40 °C and a NaCl concentration below 1%.[2]

Genetic Manipulation of *Streptomyces* sp. s120. DNA manipulation of *Streptomyces* sp. s120 (WT) relies on a double-crossover recombination event. Plasmids pCAP.ΔPopG, pCAP.ermE*.PopB, pCAP.ΔOrf1, and pCAP.ΔPopV were used to generate the transgenic strains *Streptomyces* sp. s120/ΔPopG, *Streptomyces* sp. s120/ermE*.PopB, *Streptomyces* sp. s120/ΔPopOrf1, and *Streptomyces* sp. s120/ΔPopV, respectively.

The plasmid pCAP.ΔPopG was constructed by assembling four DNA fragments from PCR using Q5 polymerase (NEB Biolabs). The vector containing a kanamycin resistance gene (*aph*) and *ori* was derived from the template pCAP03-*acc*(3)IV.[24] The template was first digested with the restriction enzymes *Xho*I and *Nde*I, and the resulting 10551 bp linear DNA fragment was purified and used as a template for PCR using primers pCAP03.part.F and pCAP03.part.R. The later PCR product, a 3877 bp DNA fragment, was purified and used as a vector. The apramycin resistance gene (*acc*(3)IV) and *oriT* (1431 bp) were obtained by PCR amplification from the template of the apramycin resistance cassette (sequence S1) with the primers cassette.F and pIJ_tcp830_rev. The homologous regions 1 and 2 (HR1 and -2) were amplified from genomic DNA of *Streptomyces* sp. s120 (WT) with the primer pair KO.HR1.F/KO.HR1.R for HR1 and New.KO.HR2.F/KO.HR2.R for HR2. The four DNA fragments (vector, apramycin resistance gene with *oriT*, HR1, and HR2) were assembled using the Gibson assembly (GA) method.[25] The assembled product (circular plasmid DNA) was transformed to *E. coli* top10 cells (for preservation) and *E. coli* ET12567[26] cells (for conjugation) using LB medium containing kanamycin and apramycin as antibiotics. The correct construct was confirmed by PCR and a DNA restriction pattern.

As described above for pCAP.ΔPopG, the plasmid pCAP.ermE*.PopB was constructed by assembling four DNA fragments. Consequently, assembly of the four DNA fragments, transformation, and plasmid product verification were performed similar to the procedure outlined for pCAP.ΔPopG construction. The promoter ermE* as well as *oriT-acc*(3)IV genes were obtained by PCR amplification of the template of the ermE* cassette (sequence S2) using the primers cassette.F and cassette.R. HR1 and -2 were amplified from genomic DNA of *Streptomyces* sp. s120 (WT) with the primer pair C1.HR1.F/C1.HR1.R for HR1 and C1.HR2.F/C1.HR2.R for HR2.

Similarly, the construction (assembly of DNA fragments, transformation, and plasmid product verification) of pCAP.ΔOrf1 and pCAP.ΔPopV was performed as described for the construction of pCAP.ΔPopG. For pCAP.ΔOrf1, HR1 and -2 were amplified by the

https://doi.org/10.1021/acs.jnatprod.1c01018
J. Nat. Prod. 2022, 85, 888−898

primer pair KO.g20.HR1.F/KO.g20.HR1.R and KO.g20.HR2.F/KO.g20.HR2.R, respectively. For pCAP.ΔPopV, HR1 and -2 were amplified by the primer pair KO53-F/KO53-R and KO55-F/KO55-R, respectively. All primers employed in this study are listed in Table S8.

For conjugation to the *Streptomyces* sp. s120 WT strain, the plasmid was transferred by triparental conjugation.[27] The positive event of double crossover was verified by PCR using genomic DNA of candidate positive colonies as templates (respective schemes are shown in Figures S2, S4, S5, and S6).

Standard Protocol for Small-Scale Cultivation and Extract Preparation for LCMS. A 5 μL aliquot of a cryoculture of a particular strain was used to inoculate 20 mL of Med-3 (Table S9) in a 100 mL Erlenmeyer flask, and the resulting culture was incubated for 4 days, at 30 °C and 180 rpm. Thereafter, the culture was transferred to a 50 mL Falcon tube. An equal amount of 20 mL of EtOAc was added to the Falcon tube, shaken vigorously, and centrifuged at maximum speed for 1 min to facilitate phase separation. The EtOAc layer was collected via pipet and dried using a rotary evaporator. The dried extract was dissolved in 1 mL of MeOH, centrifuged, and subjected to LCMS measurement.

Production Optimization and Purification of Compounds 1 and 2. Medium optimization was done by cultivating the selected strain in eight different media (Table S9) containing apramycin (50 μg/mL). Several media were modified compared to the original references. Cultivation was performed in 10 mL of medium in a 50 mL Erlenmeyer flask. Each medium was inoculated with 0.05% v/v of a cryoculture and incubated for 4 days at 30 °C and 180 rpm. Extraction and preparation of LCMS samples was carried out following the above-described standard protocol. Interestingly, LCMS analysis showed that polyoxyperuins were produced in good amounts only in media containing soybean powder (Figure S17). Furthermore, Med-4 exhibits the best production of **1**, with an insignificant difference to Med-3, while for production of **2**, Med-3 proved most suitable. To obtain both compounds, Med-3 was selected for further optimization. Next, the influence of the flask size (selected sizes: 100, 300, 500, 1000, and 2000 mL) on production was investigated. The ratio (v/v) of medium volume to flask volume was maintained at 15−25%. After extraction with EtOAc (EtOAc/aqueous phase = 3:1) and drying, the extract was dissolved in MeOH, with the volume of MeOH amounting to 10% of the cultivation medium volume. The methanolic extracts were analyzed by LCMS. It turned out that production of **1** was best when cultivation was carried out in a 300 mL flask, while for **2**, production proceeds equally well in a 100 and 300 mL flask (Figure S18). Interestingly, while generally an increase in flask size from 100 to 300 mL either led to an increased production or did not alter production, a further increase in flask size above 300 mL dramatically decreased production yield. It was suspected that this behavior was due to different aeration conditions caused by different flask volumes as the diameter of rotation of the shaker was constant. It was reasoned that in case this observation was attributable to the aeration conditions, then increasing the medium volume while at the same time maintaining flask volume constant should reduce the production since this leads to a decreased aeration. This assumption was tested by performing cultivation in a constant flask size of 300 mL while varying the medium volume (100, 150, and 200 mL, respectively). As expected, the larger the medium cultivation volume was, the less production was observed, so that aeration was verified to impact production (Figure S19). Moreover, the kinetics of polyoxyperuin production was investigated by drawing a sample for LCMS analysis every day for 10 days in total. After 1 day of incubation, no polyoxyperuin production was detectable (Figure S20). On the second day, polyoxyperuin A seco acid (**1**) reached the highest yield, while for polyoxyperuin A (**2**) the yield on the third day was slightly increased compared to the second day. The production of both polyoxyperuins decreased again on the fourth day (Figure S20). Overall, the highest production amounts were observed for cultivation in Med-3 for 2 days using either 20 mL of medium in a 100 mL Erlenmeyer flask or 50 mL of medium in a 300 mL Erlenmeyer flask, thus constituting the optimized production conditions.

After optimization, large-scale production of a total volume of 9 L of Med-3 was achieved using a combination of 300 mL flasks containing 50 mL of medium and 100 mL flasks with 20 mL of medium. Each flask containing Med-3 was inoculated with 5 μL of the *Streptomyces* sp. s120/ermE* cryoculture and incubated for 2 days at 37 °C and 180 rpm. Thereafter, the culture was extracted with three times as much EtOAc. The organic extract was separated using a separatory funnel, dried over Na_2SO_4 to remove any traces of water, and filtered. The filtrate was subsequently evaporated to dryness using a rotary evaporator, yielding 1.2 g of the organic extract. The extract was purified by reversed-phase HPLC employing a Shimadzu HPLC system (Shimadzu Deutschland GmbH) equipped with a VP250/10 Nucleodur C18 Gravity-SB, 5 μm (250 × 10 mm) column (Macherey-Nagel GmbH & Co. KG). The following gradient was used for separation: (solvent A: H_2O + 0.1% TFA; solvent B: MeCN + 0.1% TFA; column temperature: 40 °C; flow rate: 3 mL/min): 0−10 min: 10% solvent B, 10−40 min: 10−100% solvent B, 40−50 min: 100% solvent B, 50−50.01 min: 100−10% solvent B, and 50.01−60 min: 10% solvent B. Compound **1** was collected in the retention time (t_R) interval 36.5−37.2 min, and **2** was collected at 40.75−41.5 min, yielding 2.8 and 1.4 mg, respectively. The respective fractions were dried *in vacuo* using a HT12-II centrifugal concentrator (Genevac, Ips). The combined fractions containing **1** were further purified by HPLC using a Lux 5 μm Cellulose-1, 250 × 4.60 mm (Phenomenex) column. The employed gradient was as follows: (solvent A: H_2O + 0.1% TFA; solvent B: MeCN + 0.1% TFA; column temperature: 40 °C; flow rate: 1 mL/min): 0−10 min: 35% solvent B, 10−35 min: 35−75% solvent B, 35−35.01 min: 75−100% solvent B, 35.01−42.5 min: 100% solvent B, 42.5−42.51 min: 100−35% solvent B, and 42.51−50 min: 35% solvent B. Compound **1** was collected in the retention time interval (t_R) 22.5−26 min, thereby affording 1.4 mg of the purified compound. Compound **2** could not be further purified using this column, because the impurity eluted at a similar t_R.

*Polyoxyperuin A seco acid (**1**):* reddish-brown solid; ^1H and ^{13}C NMR data, see Table 1; HRESIMS $[M + Na]^+$ *m/z* 895.4745 (calcd for $C_{39}H_{68}N_8O_{14}Na^+$, 895.4747).

*Polyoxyperuin A (**2**):* reddish-brown solid; ^1H NMR (600 MHz, $CDCl_3$) δ 7.92 (s), 5.99 (br s), 5.55 (d, *J* = 5.7 Hz), 5.46 (br s), 5.40−5.37 (m), 5.37−5.32 (m), 5.32−5.27 (m), 5.17 (d, *J* = 3.4 Hz), 5.15 (d, *J* = 2.9 Hz), 4.05−4.01 (m), 3.99−3.95 (m), 3.66 (d, *J* = 5.5 Hz, overlay), 3.64 (s), 3.44 (t, *J* = 9.6 Hz), 3.18−3.13 (m), 3.06 (s). 3.05 (s), 2.79−2.74 (m), 2.24−2.19 (m), 1.99−1.86 (m), 1.66−1.52 (m), 1.48 (t, *J* = 6.8 Hz), 1.37−1.27 (m), 1.25 (s), 0.99−0.95 (m), 0.94−0.90 (m), 0.90−0.88 (m), 0.878 (s), 0.86−0.81 (m). (Only the main compound peaks as derived from comparison with compound **1** are given. Owing to the impurity of compound **2**, integrals are not indicated.) HRESIMS $[M + Na]^+$ *m/z* 877.4627 (calcd for $C_{39}H_{66}N_8O_{13}Na^+$, 877,4642).

Hydrolysis of Compound 2. Based on a literature procedure,[28] a sample solution of polyoxyperuin A (**2**) in MeOH (0.34 μmol, 3.5 mM) was added to a solution of K_2CO_3 (3.7 mg, 26.8 μmol, 79 equiv) in 57 μL of Milli-Q H_2O. The concentration of compound **2** in the reaction mixture was adjusted with MeOH to 2 mM. The reaction mixture was agitated for 8 h at room temperature. Afterward, 200 μL of H_2O was added and the reaction mixture was extracted four times with 300 μL of $CHCl_3$/PrOH (2:1). The combined organic phases were washed once with 400 μL of H_2O. Subsequently, the organic phase was evaporated to dryness. The obtained residue was redissolved in 75 μL of MeOH and analyzed by high-accuracy LCMS.

Minimum Inhibitory Concentrations. The antibacterial properties of compounds **1** and **2** were determined by micro-broth-dilution assays in 96-well microtiter plates. Each compound was dissolved in a 2:7 MeOH/DMSO mixture and tested in a 12-step dilution series ranging from 64 to 0.03 μg/mL. All concentrations were tested in triplicate.

The indicator strains *Escherichia coli* ATCC35218, *Staphylococcus aureus* ATCC25923, *Bacillus subtilis* DSM10, and *Micrococcus luteus* DSM20030 were incubated overnight at 37 °C and agitated at 180 rpm in 100 mL Erlenmeyer flasks. After incubation the cell density was diluted to 5×10^5 cells/mL in cation-adjusted Mueller Hinton II

https://doi.org/10.1021/acs.jnatprod.1c01018
J. Nat. Prod. 2022, 85, 888−898

medium. *Listeria monocytogenes* DSM20600 was incubated for 2 days before the cell density was adjusted. Rifampicin, tetracycline, and gentamycin were used as positive controls at concentrations identical to compounds **1** and **2** (64−0.03 μg/mL). Bacterial suspensions without compounds **1** and **2** or positive control were used as negative controls. Medium background was averaged from five replicates. Assay incubation at 37 °C, 180 rpm, and 85% relative humidity was carried out for 18 h (2 days for *Listeria*) before the turbidity of each well was measured with a spectrophotometer at 600 nm (LUMIstar Omega, BMG Labtech GmbH). Growth inhibition was calculated relative to the absorption of the controls. Cell viability of *L. monocytogenes* was assessed by ATP quantification (BacTiter-Glo, Promega GmbH) according to the manufacturer's instructions. The preculture of *Mycobacterium smegmatis* ATCC607 was incubated in brain−heart infusion broth (48 h, 37 °C, 180 rpm) before the cell concentration was adjusted in Mueller Hinton II medium. Isoniazid was used instead of gentamycin as a third positive control. Cell viability was evaluated after 48 h (37 °C, 180 rpm, 85% relative humidity) via ATP quantification.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jnatprod.1c01018.

NMR and MS/MS spectra for compounds **1** and **2**; LCMS data for the cultivation optimization experiments and for the hydrolysis experiment of compound **2**; data on cloning and primers used in this study; bioinformatics data on the whole BGC as well as on the AT, KS, ER, DH, and KR domains and on the substrate signature of the NRPS A-domains; kinetic analysis of the polyoxyperuin production; composition of media used in this study (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Ute Mettal − *Institute for Insect Biotechnology, Justus-Liebig-University Giessen, 35392 Giessen, Germany; Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany*; Phone: +49 64197219142; Email: Ute.Mettal@chemie.uni-giessen.de

Till F. Schäberle − *Institute for Insect Biotechnology, Justus-Liebig-University Giessen, 35392 Giessen, Germany; Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany; German Center of Infection Research (DZIF), 35392 Giessen, Germany*; orcid.org/0000-0001-9947-8079; Phone: +49 64197219140; Email: Till.F.Schaeberle@agrar.uni-giessen.de

Authors

I Dewa M. Kresna − *Institute for Insect Biotechnology, Justus-Liebig-University Giessen, 35392 Giessen, Germany; Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany*

Zerlina G. Wuisan − *Institute for Insect Biotechnology, Justus-Liebig-University Giessen, 35392 Giessen, Germany; Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany; German Center of Infection Research (DZIF), 35392 Giessen, Germany*

Jean-Marie Pohl − *Institute for Insect Biotechnology, Justus-Liebig-University Giessen, 35392 Giessen, Germany; Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany*

Virginia Linares Otoya − *Unidad de Posgrado en Farmacia y Bioquímica, Facultad de Farmacia y Bioquímica, Universidad Nacional de Trujillo, 13011 Trujillo, Peru*

Martin Gand − *Institute of Food Chemistry and Food Biotechnology, Faculty of Biology and Chemistry, Justus-Liebig-University Giessen, 35392 Giessen, Germany*; orcid.org/0000-0001-8211-691X

Michael Marner − *Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany*; orcid.org/0000-0002-1024-1567

Luis Linares Otoya − *Institute for Insect Biotechnology, Justus-Liebig-University Giessen, 35392 Giessen, Germany*

Nils Böhringer − *Institute for Insect Biotechnology, Justus-Liebig-University Giessen, 35392 Giessen, Germany; Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany; German Center of Infection Research (DZIF), 35392 Giessen, Germany*

Andreas Vilcinskas − *Institute for Insect Biotechnology, Justus-Liebig-University Giessen, 35392 Giessen, Germany; Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany*; orcid.org/0000-0001-8276-4968

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jnatprod.1c01018

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank the Deutscher Akademischer Austauschdienst (DAAD) for funding under the framework of PROPERU and especially the Ph.D. scholarship to I.D.M.K. and Z.G.W. Research in the Schäberle lab was further supported by the German Center of Infection Research (DZIF) and the Hessen State Ministry of Higher Education, Research and the Arts (HMWK) via the LOEWE Center for Insect Biotechnology and Bioresources. The authors would like to thank Dr. H. Hausmann for the NMR measurements.

■ REFERENCES

(1) Albarano, L.; Esposito, R.; Ruocco, N.; Costantini, M. *Mar. Drugs* **2020**, *18*, 1−17.

(2) Linares-Otoya, L.; Linares-Otoya, V.; Armas-Mantilla, L.; Blanco-Olano, C.; Crüsemann, M.; Ganoza-Yupanqui, M. L.; Campos-Florian, J.; König, G. M.; Schäberle, T. F. *Mar. Drugs* **2017**, *15*, 1−14.

(3) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2019**, *47*, 81−87.

(4) Du, Y.; Wang, Y.; Huang, T.; Tao, M.; Deng, Z.; Lin, S. *BMC Microbiol.* **2014**, *14*, 1−12.

(5) Zhao, H.; Wang, L.; Wan, D.; Qi, J.; Gong, R.; Deng, Z.; Chen, W. *Microbial Cell Factories* **2016**, *15*, 1−12.

(6) Umezawa, K.; Nakazawa, K.; Ikeda, Y.; Naganawa, H.; Kondo, S. *J. Org. Chem.* **1999**, *64*, 3034−3038.

(7) Matsumoto, N.; Momose, I.; Umekita, M.; Kinoshita, N.; Chino, M.; IInuma, H.; Sawa, T.; Hamada, M.; Takeuchi, T. *Journal of Antibiotics* **1998**, *51*, 1087−1092.

(8) Hayakawa, Y.; Nakagawa, M.; Toda, Y.; Seto, H. *Agric. Biol. Chem.* **1990**, *54*, 1007−1011.
(9) Maskey, R. P.; Fotso, S.; Sevvana, M.; Usón, I.; Grün-Wollny, I.; Laatsch, H. *J. Antibiot.* **2006**, *59*, 309−314.
(10) Ueno, M.; Amemiya, M.; Someno, T.; Masuda, T.; Iinuma, H.; Naganawa, H.; Hamada, M.; Ishizuka, M.; Takeuchi, T. *Journal of Antibiotics* **1993**, *46*, 1658−1665.
(11) Nakagawa, M.; Hayakawa, Y.; Adachi, K.; Seto, H. *Agric. Biol. Chem.* **1990**, *54*, 791−794.
(12) Gräfe, U.; Schlegel, R.; Ritzau, M.; Ihn, W.; Dornberger, K.; Stengel, C.; Fleck, W. F.; Gutsche, W.; Härtl, A.; Paulus, E. F. *Journal of Antibiotics* **1995**, *48*, 119−125.
(13) Maehr, H.; Liu, C. M.; Palleroni, N. J.; Smallheer, J.; Todaro, L.; Williams, T. H.; Blount, J. F. *Journal of Antibiotics* **1986**, *39*, 17−25.
(14) Smitka, T. A.; Deeter, J. B.; Hunt, A. H.; Mertz, F. P.; Ellis, R. M.; Boeck, L. D.; Yao, R. C. *Journal of Antibiotics* **1988**, *41*, 726−733.
(15) Hauf, S.; Engelgeh, T.; Halbedel, S. *J. Bacteriol.* **2021**, *203*, 1−16.
(16) Hou, B.; Lin, Y.; Wu, H.; Guo, M.; Petkovic, H.; Tao, L.; Zhu, X.; Ye, J.; Zhang, H. *J. Bacteriol.* **2018**, *200*, 1−16.
(17) Cai, X.; Teta, R.; Kohlhaas, C.; Crüsemann, M.; Ueoka, R.; Mangoni, A.; Freeman, M. F.; Piel, J. *Chemistry & Biology* **2013**, *20*, 839−846.
(18) Xie, Y.; Li, Q.; Qin, X.; Ju, J.; Ma, J. *Metabolic Engineering Commun.* **2019**, *8*, 1−6.
(19) Keatinge-Clay, A. T. *Nat. Prod. Rep.* **2012**, *29*, 1050−1073.
(20) Long, P. F.; Wilkinson, C. J.; Bisang, C. P.; Cortés, J.; Dunster, N.; Oliynyk, M.; McCormick, E.; McArthur, H.; Mendez, C.; Salas, J. A.; Staunton, J.; Leadlay, P. F. *Mol. Microbiol.* **2002**, *43*, 1215−1225.
(21) Bode, E.; Brachmann, A. O.; Kegler, C.; Simsek, R.; Dauth, C.; Zhou, Q.; Kaiser, M.; Klemmt, P.; Bode, H. B. *ChemBioChem.* **2015**, *16*, 1115−1119.
(22) Wyche, T. P.; Ruzzini, A. C.; Beemelmanns, C.; Kim, K. H.; Klassen, J. L.; Cao, S.; Poulsen, M.; Bugni, T. S.; Currie, C. R.; Clardy, J. *Org. Lett.* **2017**, *19*, 1772−1775.
(23) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. *Journal of Organic Chemistry* **1997**, *62*, 7512−7515.
(24) Tang, X.; Li, J.; Millán-Aguiñaga, N.; Zhang, J. J.; O'Neill, E. C.; Ugalde, J. A.; Jensen, P. R.; Mantovani, S. M.; Moore, B. S. *ACS Chem. Biol.* **2015**, *10*, 2841−2849.
(25) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A., III; Smith, H. O. *Nat. Methods* **2009**, *6*, 343−345.
(26) Flett, F.; Mersinias, V.; Smith, C. P. *FEMS Microbiology Letters* **1997**, *155*, 223−229.
(27) Kieser, T.; Bibb, M. J.; Butter, M. J.; Chater, K. F.; Hopwood, D. A. *Practical Streptomyces Genetics: A Laboratory Manual*; Crowes: Norwich, England, 2000.
(28) Shilabin, A. G.; Kasanah, N.; Wedge, D. E.; Hamann, M. T. *J. Med. Chem.* **2007**, *50*, 4340−4350.

https://doi.org/10.1021/acs.jnatprod.1c01018
J. Nat. Prod. 2022, 85, 888−898

Supporting Information

Genome Mining-Guided Discovery and Characterization of the PKS-NRPS-Hybrid Polyoxyperuin produced by a Marine-Derived Streptomycete

I Dewa M. Kresna,[†,‡] Zerlina G. Wuisan,[†,‡] Jean-Marie Pohl,[†,‡] Ute Mettal,[†,‡,] Virginia Linares Otoya,[§] Martin Gand,[^] Michael Marner,[‡] Luis Linares Otoya,[†] Nils Böhringer,[†,‡] Andreas Vilcinskas,[†,‡] Till F. Schäberle[†,‡,‖,*]*

[†]Institute for Insect Biotechnology, Justus-Liebig-University Giessen, 35392 Giessen, Germany

[‡]Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch for Bioresources, 35392 Giessen, Germany

[§]Unidad de Posgrado en Farmacia y Bioquímica, Facultad de Farmacia y Bioquímica, Universidad Nacional de Trujillo, 13011. Trujillo, Peru

[^]Institute of Food Chemistry and Food Biotechnology, Faculty of Biology and Chemistry, Justus-Liebig-University Giessen, 35392 Giessen, Germany

[‖]German Center of Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, 35392 Giessen, Germany

Table of contents

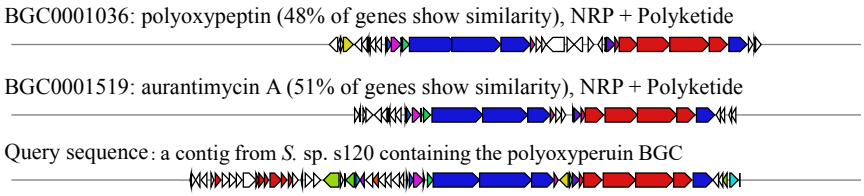


Figure S1. Similarity between the BGCs for polyoxypeptin, aurantimycin, and polyoxyperuin. Similarity analysis was accomplished using antiSMASH.[1]

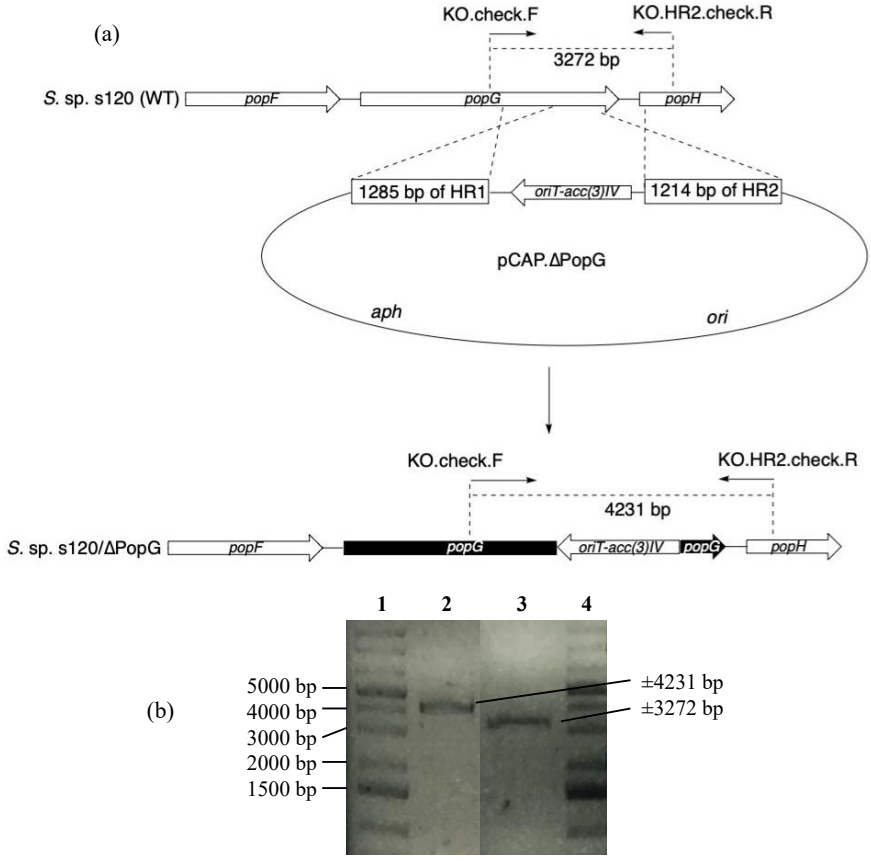


Figure S2. (a) Construction of the *S.* sp. s120/ΔPopG mutant by the insertion of *oriT-acc*(3)IV, thereby disrupting *popG*. (b) Confirmation of *PopG* disruption due to the insertion of *oriT-acc*(3)IV by PCR using the primers KO.check.F and KO.HR2.check.R (listed in Table S8). Lane 1 and 4: molecular weight marker; Lane 2: PCR analysis of the ΔPopG mutant strain; Lane 3: PCR analysis of the wild type strain.

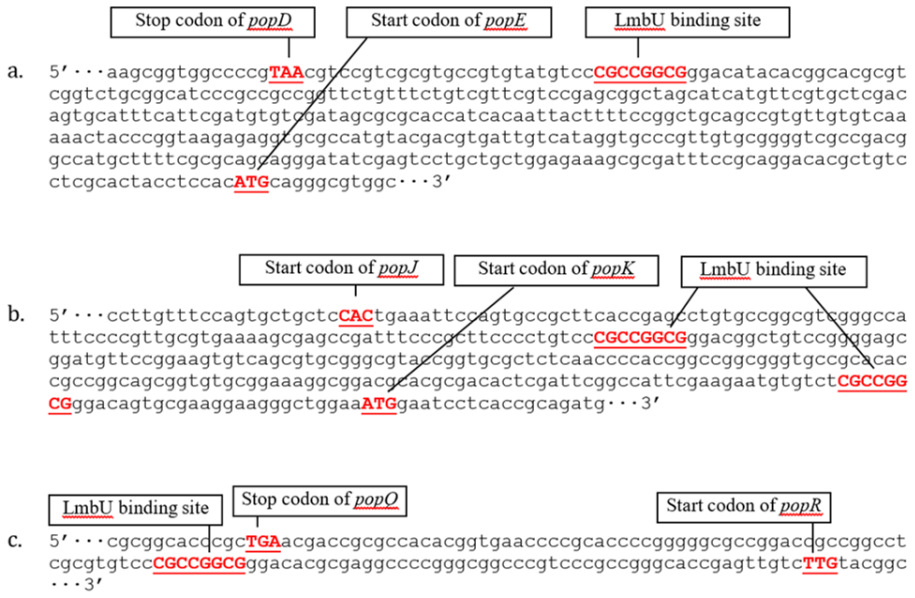


Figure S3. Nucleotide sequence of three areas (a, b and c) containing the predicted LmbU binding sites (5'-CGCCGGCG-3').[2]

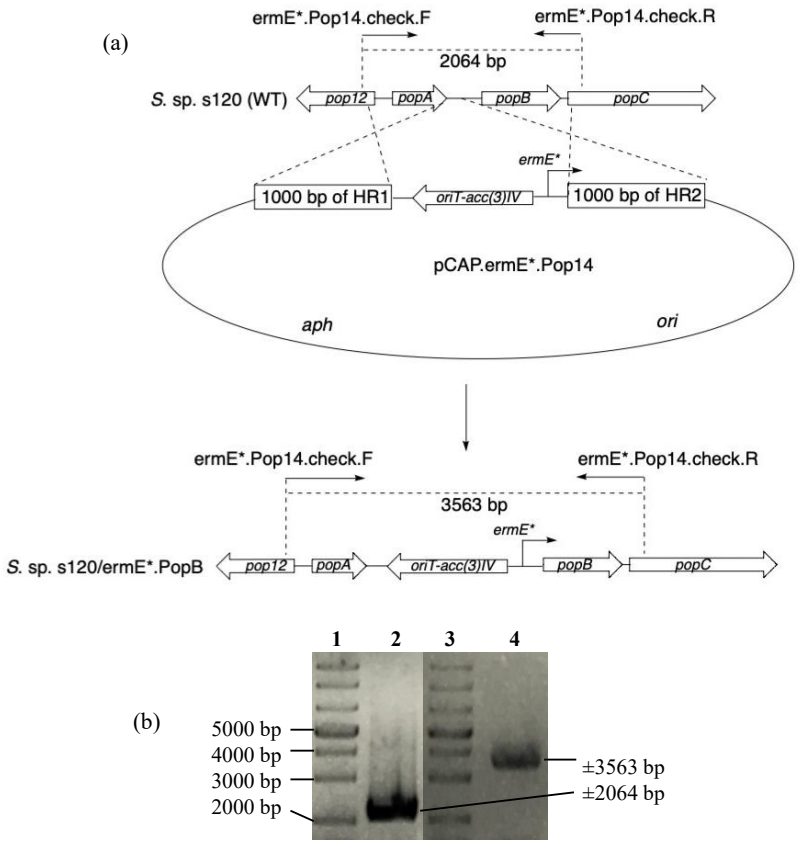


Figure S4. (a) Construction of the *S.* sp. s120/ermE*.PopB mutant by insertion of *oriT-acc*(3)IV-*ermE** upstream of *popB*. (b) Confirmation of the insertion of *oriT-acc*(3)IV-*ermE** upstream of *popB* by PCR using the primers ermE*Pop14.check.F and ermE*.Pop14.check.R (listed in Table S8). Lane 1 and 3: molecular weight marker; Lane 2: PCR analysis of wild type strain; Lane 4: PCR analysis of ermE*.PopB mutant strain.

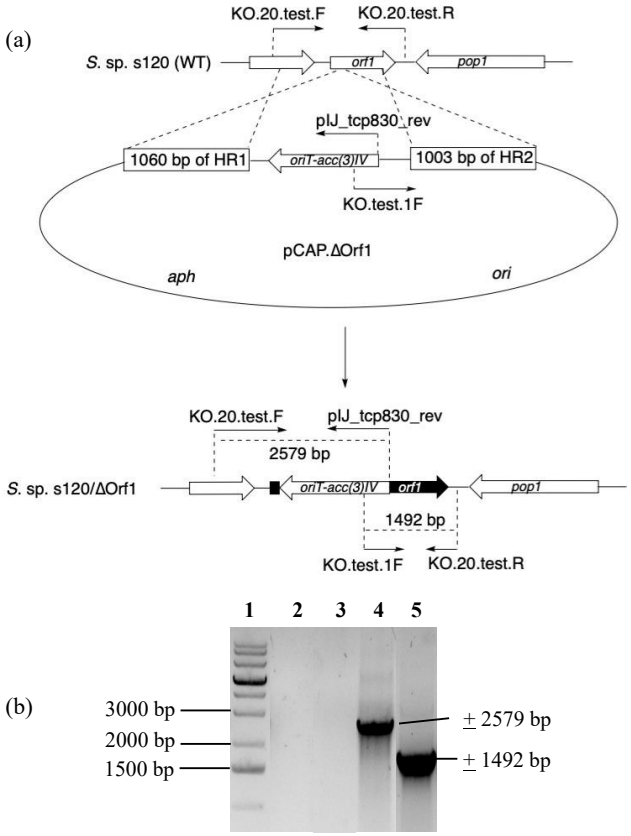


Figure S5. (a) Construction of the *S.* sp. s120/ΔOrf1 mutant by insertion of *oriT-acc*(3)IV in the middle of *orf1*. (b) Confirmation of the *oriT-acc*(3)IV insertion in the middle of *orf1* by PCR. Lane 1: molecular weight marker; Lane 2 and 3: PCR analysis of the wild type strain; Lane 4 and 5: PCR analysis of the ΔOrf1 mutant strain; Lane 2 and 4: primers KO.20.test.F and pIJ_tcp830_rev were used for PCR analysis; Lane 3 and 5: primers KO.test.1F and KO.20.test.R were used for PCR analysis. All primers are listed in Table S8.

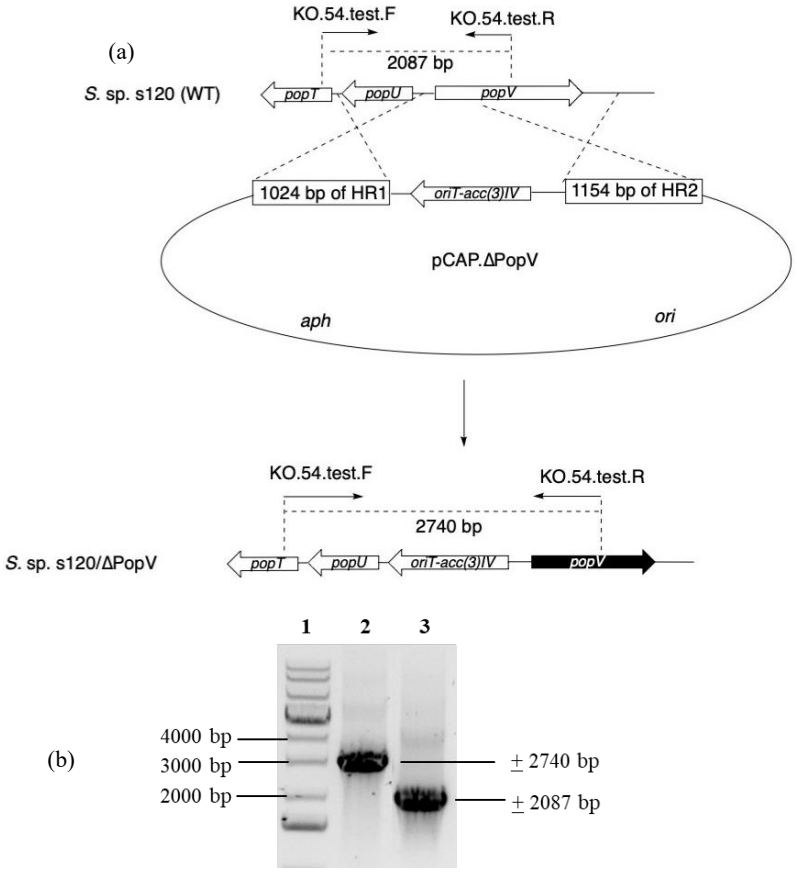
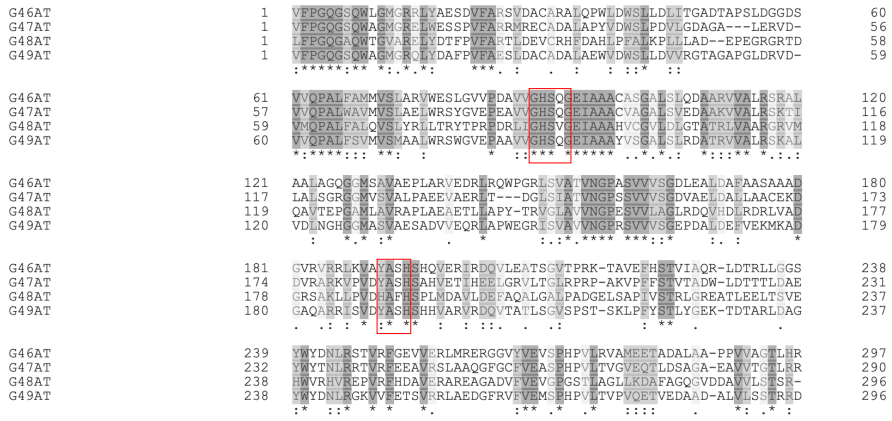


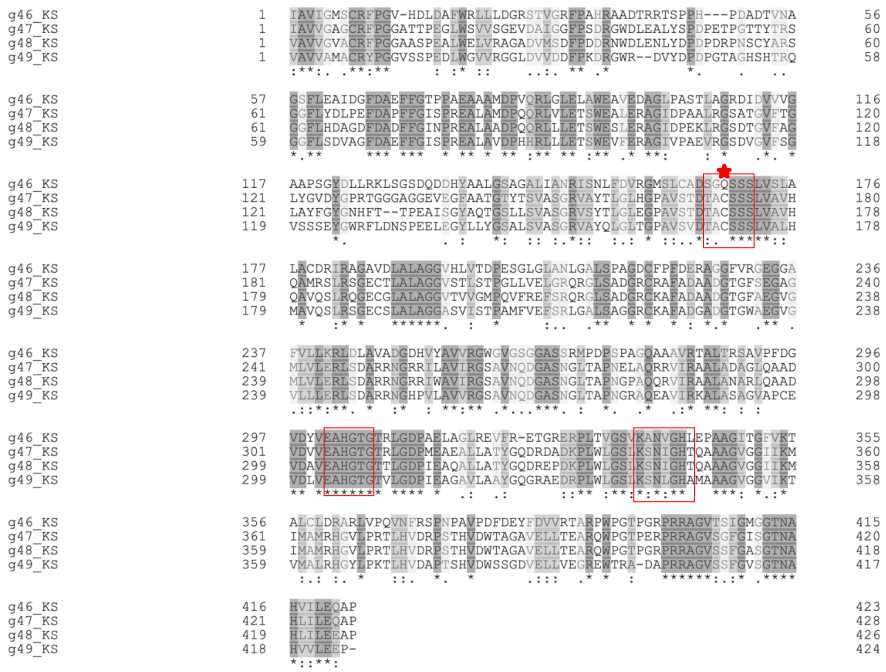
Figure S6. (a) Generation of *S.* sp. s120/ΔPopV by insertion of *oriT-acc*(3)IV in the middle of *popV*. (b) Confirmation of *PopV* disruption by insertion of *oriT-acc*(3)IV by PCR using primers KO.54.test.F and KO.54.test.R (listed in Table S8); Lane 1: molecular weight marker; Lane 2: PCR analysis of the ΔPopV mutant strain; Lane 3; PCR analysis of the wild type strain.



Description:

G46AT = PopN(AT) G47AT = PopO(AT) G48AT = PopP(AT) G49AT = PopQ(AT)

Figure S7. Alignment of AT domains from PopN-Q. The red rectangles represent the conserved motifs, which are also related to substrate specificity. The AT domains of PopN, O, and Q possess GHSQG and YASH motifs, implying that their substrate is (*2S*)-methylmalonyl-CoA[3], while PopP(AT) bears GHSVG and HAFH motifs implying that its substrate is malonyl-CoA.[3]



Description:

g46_KS = PopN(KS) g47_KS = PopO(KS) g48_KS = PopP(KS) g49_KS = PopQ(KS)

Figure S8. Alignment of the four KS domains of PopN-Q. The sequences in the red rectangles indicate the conserved sites (1. TACSSS; 2. EAHGTG; 3. KSNIGHT) of the KS domains.[3] The red star highlights the substitution of the cysteine (C) residue involved in *trans*-thioesterification by glutamine (Q) in PopN(KS).

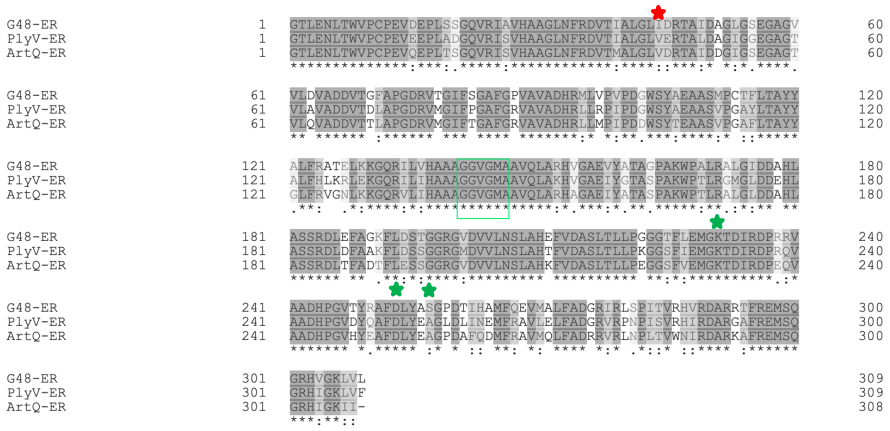


Figure S9. Alignment of the ER domains of PopP, PlyV, and ArtQ. The green rectangle shows the conserved pyrophosphate-binding motif where NADPH binds.[3] PopP(ER) belongs to the class of D-type ERs, due to the absence of the catalytic tyrosine in the position marked by the red star (finger print residue).[3] Green starts indicate the catalytic residues.[3]

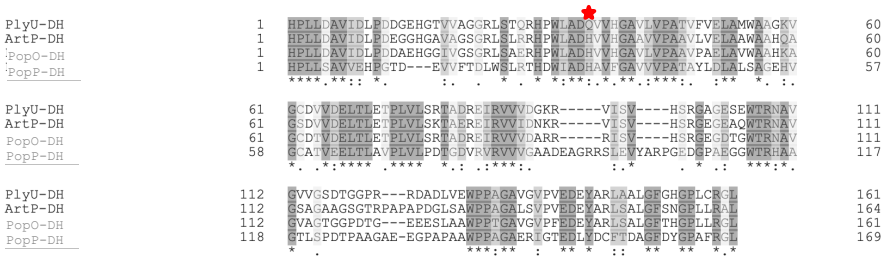


Figure S10. Alignment of DH domains from PlyU, ArtP, PopO, and PopP. The red star marks the catalytic histidine residue of the HxxxGxxxxP motif.[3]

Description:

G47KR = PopO(KR) G48KR = PopP(KR)

Figure S11. Alignment of the KR domains from PopO and PopP. The substrate fingerprint of the KR domains is highlighted by red rectangles, while the catalytic residues are marked by red stars.[3] The green rectangle designates the NADPH binding domain TGGTGxLG.[3] The presence of the LDD motif in PopO(KR) and the absence of proline two positions downstream of the catalytic tyrosine indicates that PopO(KR) belongs to the group of B1-type KRs.[3] The PopP(KR) on the other hand can be classified as A1-type KR, due to the absence of the LDD motif as well as the absence of histidine three positions upstream of the catalytic tyrosine.[3]

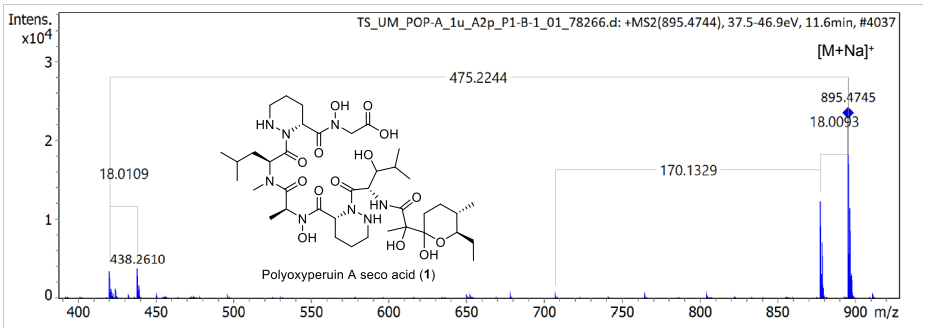


Figure S12. MS/MS spectrum of compound **1**. The parent ion at *m/z* 895.4745 represents the $[M+Na]^+$ peak of compound **1**. Fragment structures are displayed in Figure 6.

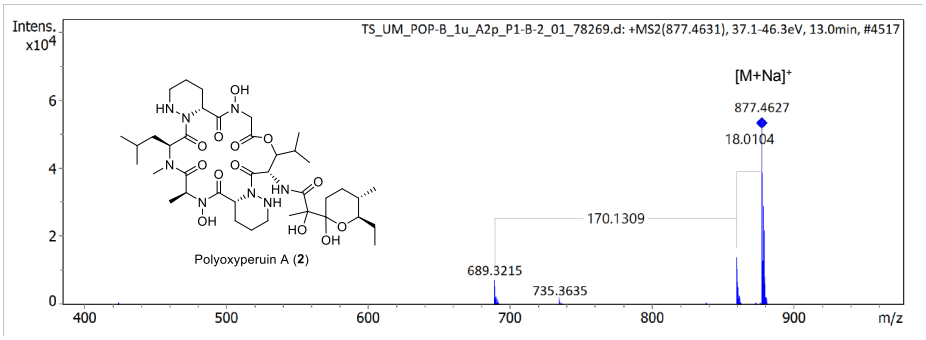


Figure S13. MS/MS spectrum of compound **2**. The parent ion at *m/z* 877.4627 represents the $[M+Na]^+$ peak of compound **2**. Fragment structures are displayed in Figure 6.

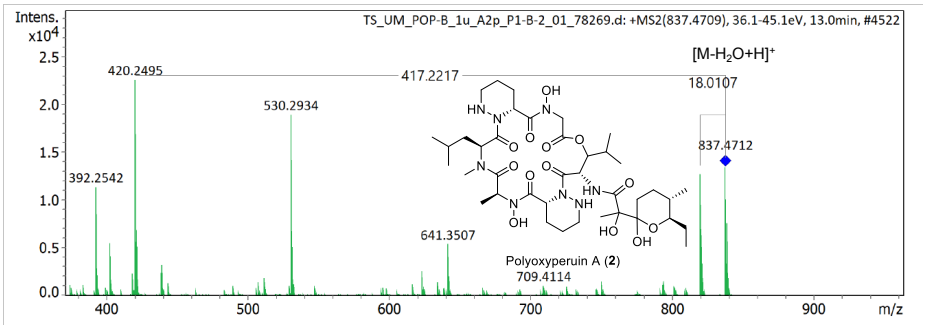


Figure S14. MS/MS spectrum of compound **2**. The parent ion at *m/z* 837.4712 represents the $[M\text{-}H_2O+H]^+$ peak of compound **2**. Fragment structures are displayed in Figure 6.

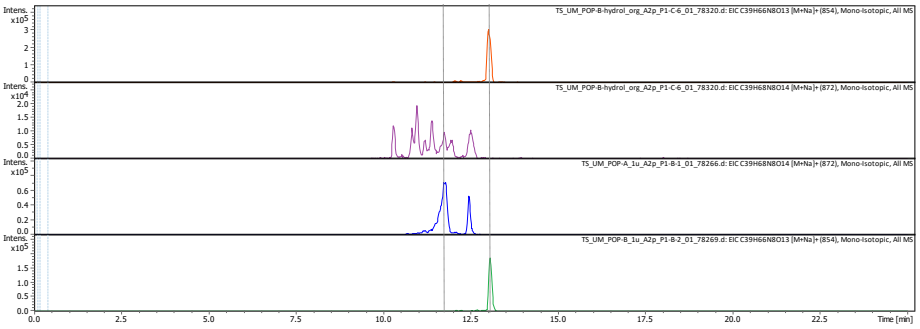


Figure S15. LCMS analysis for the hydrolysis of compound **2**. Orange: extracted ion chromatogram (EIC) of the $[M+Na]^+$ peak of compound **2** ($C_{39}H_{66}N_8O_{13}$; educt) in the hydrolysate; purple: EIC of the $[M+Na]^+$ peak of compound **1** ($C_{39}H_{68}N_8O_{14}$; product) in the hydrolysate; blue: EIC of the of the $[M+Na]^+$ peak of the reference compound **1** ($C_{39}H_{68}N_8O_{14}$; polyoxyperuin A seco acid); green: EIC of the of the $[M+Na]^+$ peak of the reference compound **2** ($C_{39}H_{66}N_8O_{13}$; polyoxyperuin A).

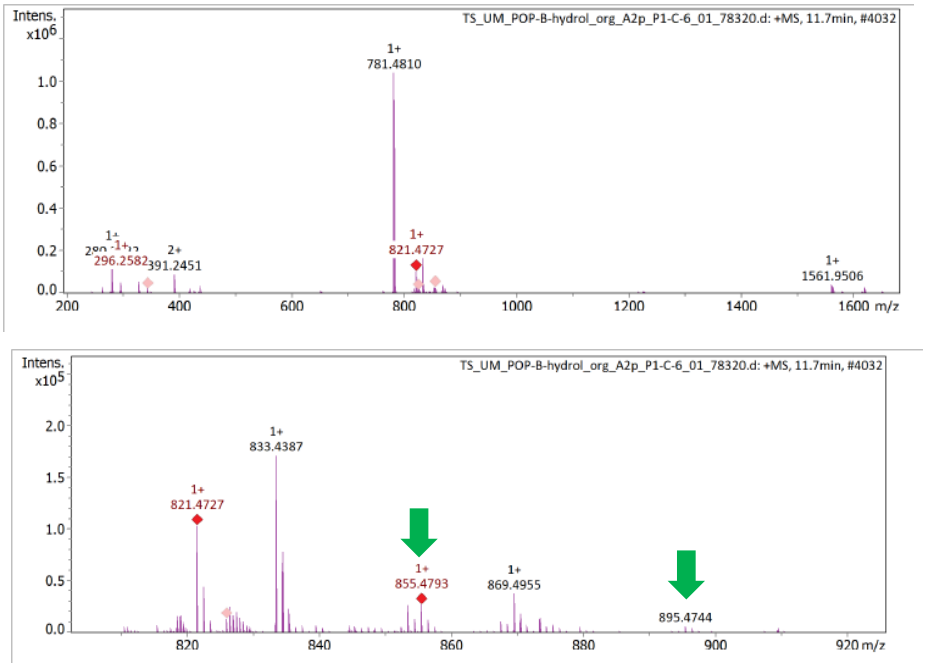


Figure S16. MS spectrum for the hydrolysis of compound **2**. The spectrum at the top shows the complete spectrum, while the spectrum at the bottom depicts a close-up. Peaks indicating the formation of compound **1** upon hydrolysis of compound **2** are marked with green arrows and represent the $[M+Na]^+$ (*m/z* 895.4744; calcd for $C_{39}H_{68}N_8O_{14}Na^+$, 895.4747) and $[M\text{-}H_2O+H]^+$ (*m/z* 855.4793; calcd for $C_{39}H_{67}N_8O_{13}^+$, 855,4822) peaks respectively.

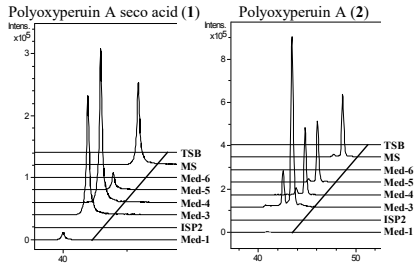


Figure S17. LCMS analysis of the medium optimization experiments for polyoxyperuin production. Shown are the extracted ion chromatograms of the $[M+Na]^+$ peak of compounds **1** (left) and **2** (right).

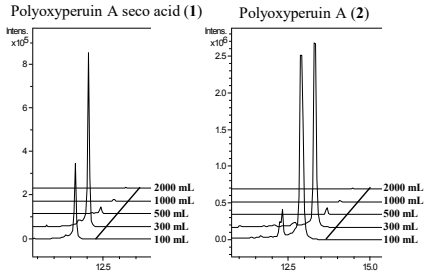


Figure S18. LCMS analysis of the optimization experiments testing the influence of the flask size on polyoxypeptin production. Shown are the extracted ion chromatograms of the $[M+Na]^+$ peak of compounds **1** (left) and **2** (right).

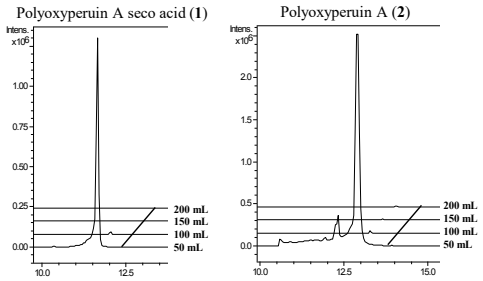


Figure S19. LCMS analysis of the cultivation experiments for investigation of the aeration conditions. The medium volumes were varied, while the flask size was maintained constant at 300 mL for polyoxyperuin production. Displayed are the extracted ion chromatograms of the $[M+Na]^+$ peak of compounds **1** (left) and **2** (right).

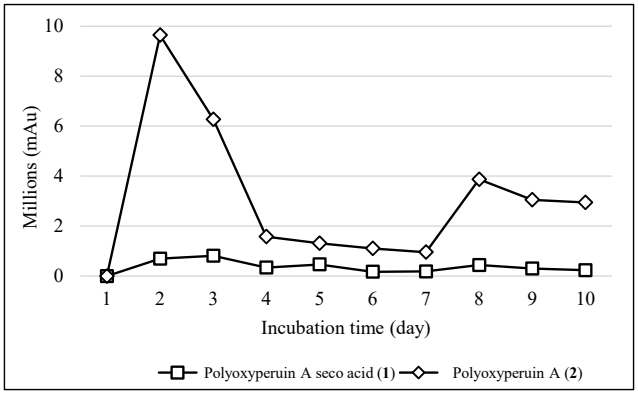


Figure S20. Kinetic analysis of the polyoxyperuin production. The *y*-axis represents the peak area (in mAu) of the $[M+Na]^+$ peak of compounds **1** and **2**, while the *x*-axis represents the incubation time (in days) for cultivation up to 10 days.

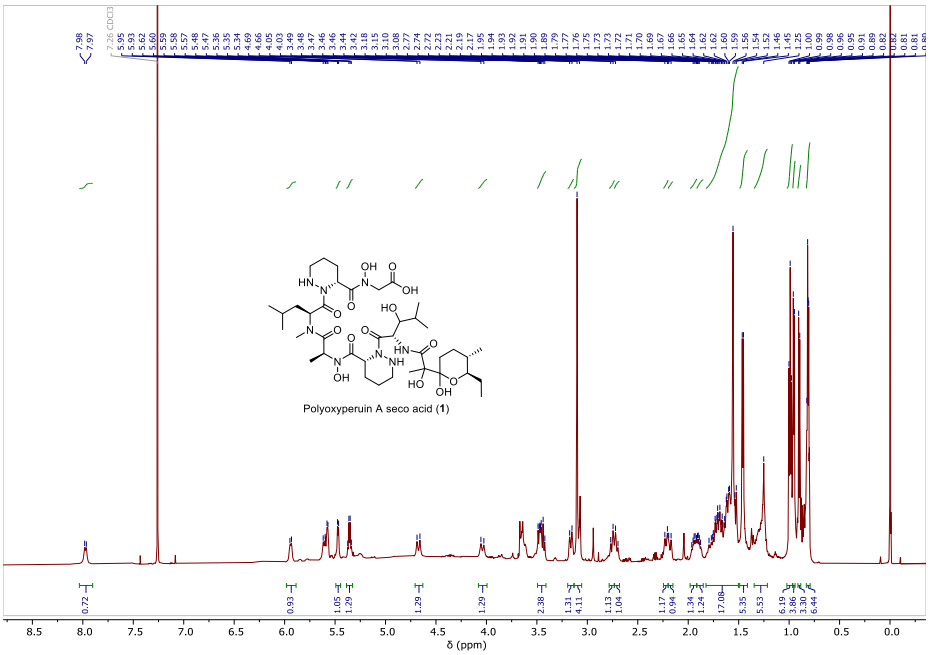


Figure S21. ^{1}H-NMR spectrum (600 MHz, $CDCl_3$) of compound **1**.

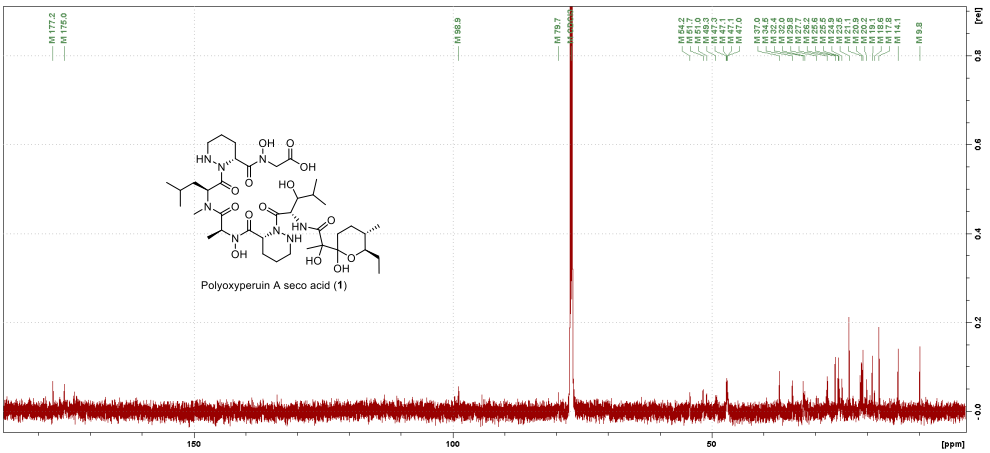


Figure S22. $^{13}C\{^{1}H\}$-NMR spectrum (150 MHz, $CDCl_3$) of compound **1**.

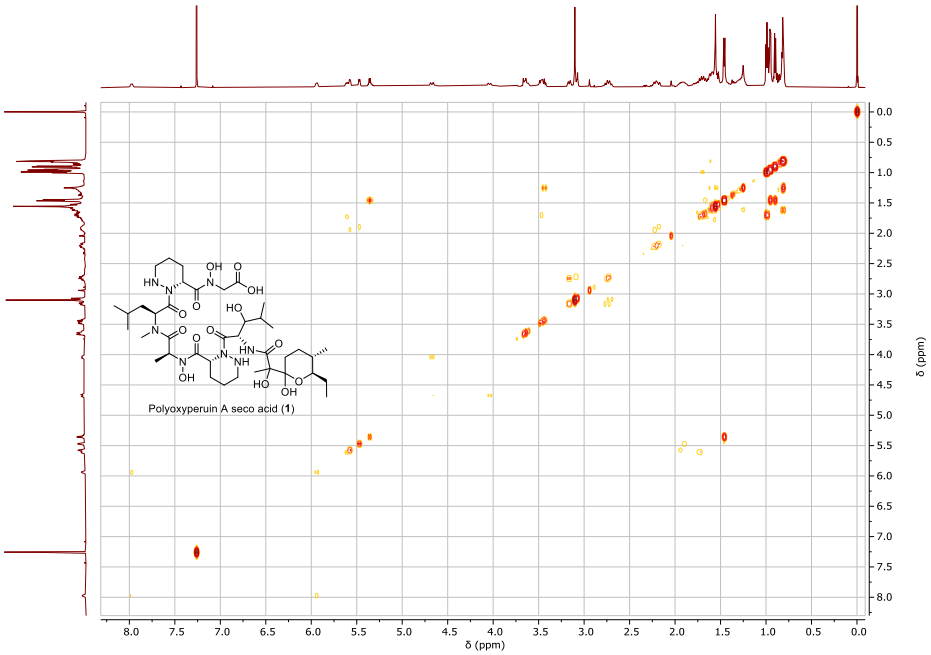


Figure S23. COSY spectrum of compound **1** (in $CDCl_3$).

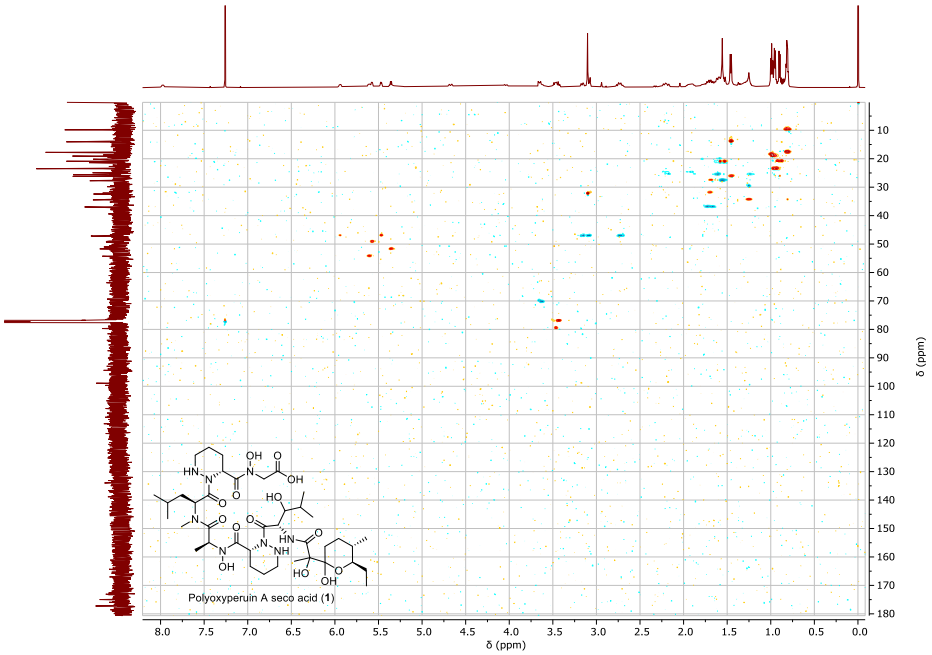


Figure S24. HSQC spectrum of compound **1** (in $CDCl_3$).

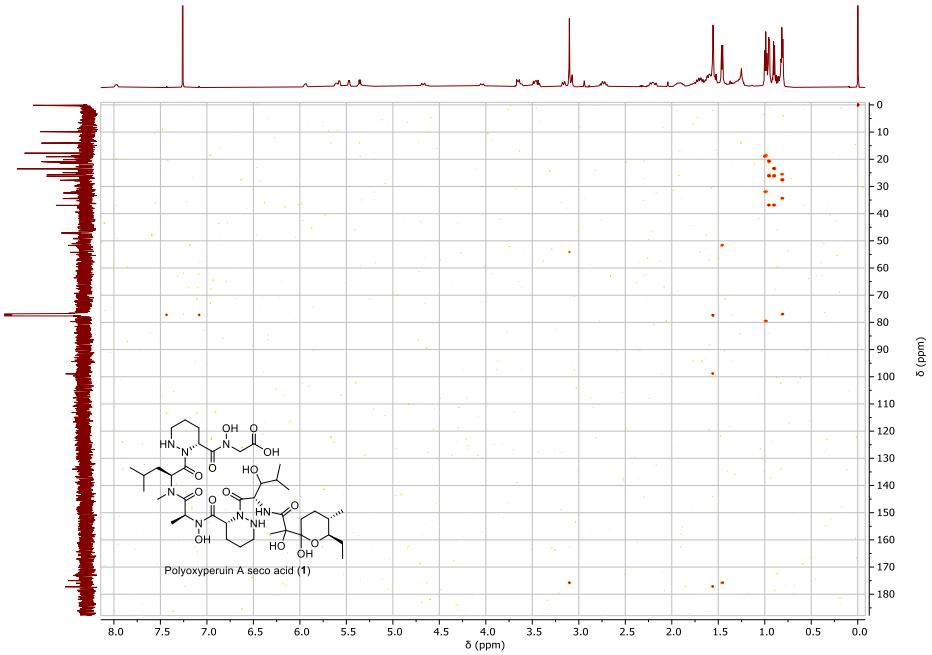


Figure S25. HMBC spectrum of compound **1** (in $CDCl_3$).

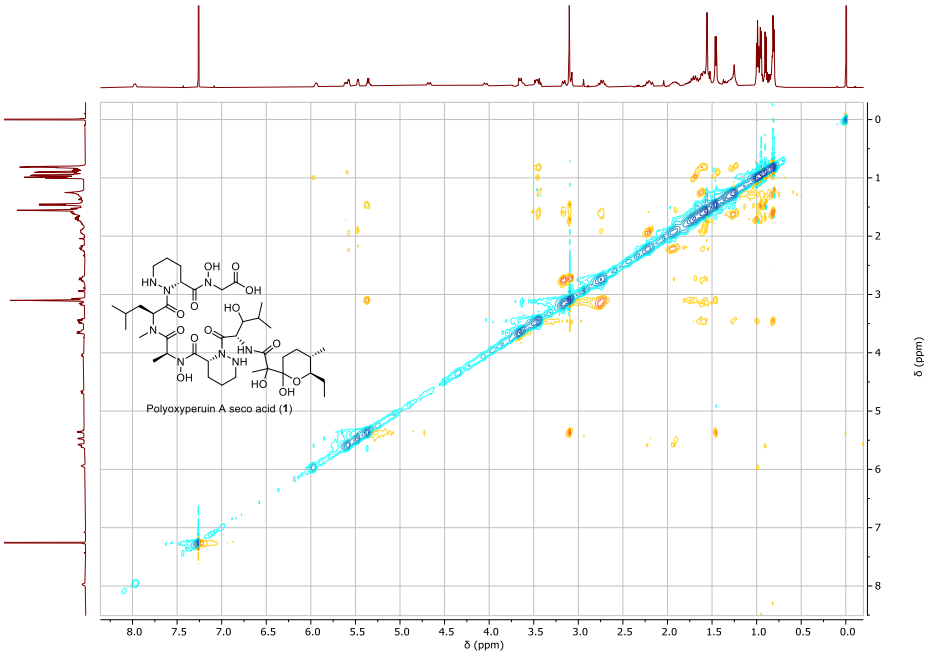


Figure S26. ROESY spectrum of compound **1** (in $CDCl_3$).

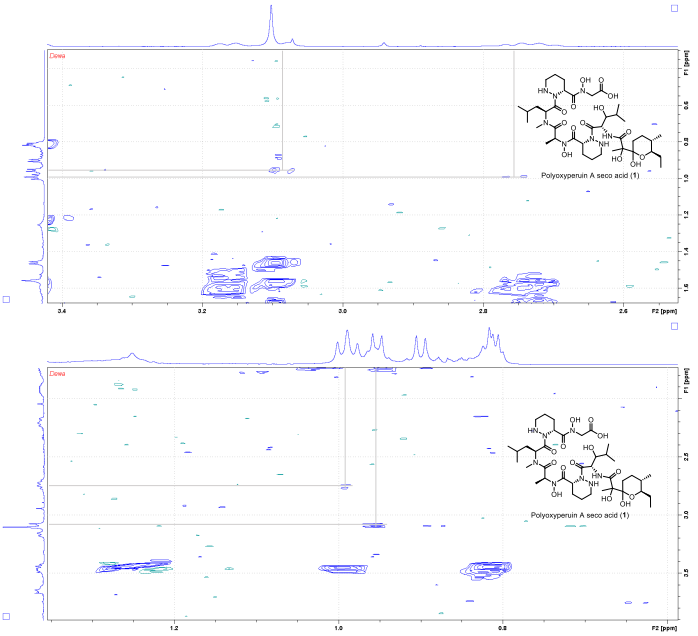


Figure S27. Close-ups of the ROESY spectrum of compound **1** (in $CDCl_3$).

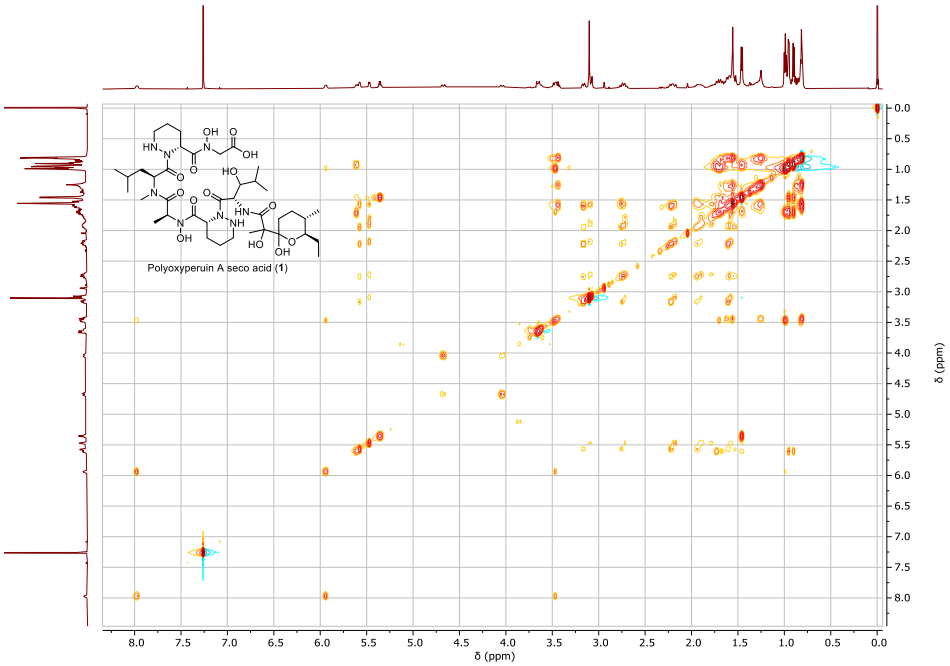


Figure S28. TOCSY spectrum of compound **1** (in $CDCl_3$).

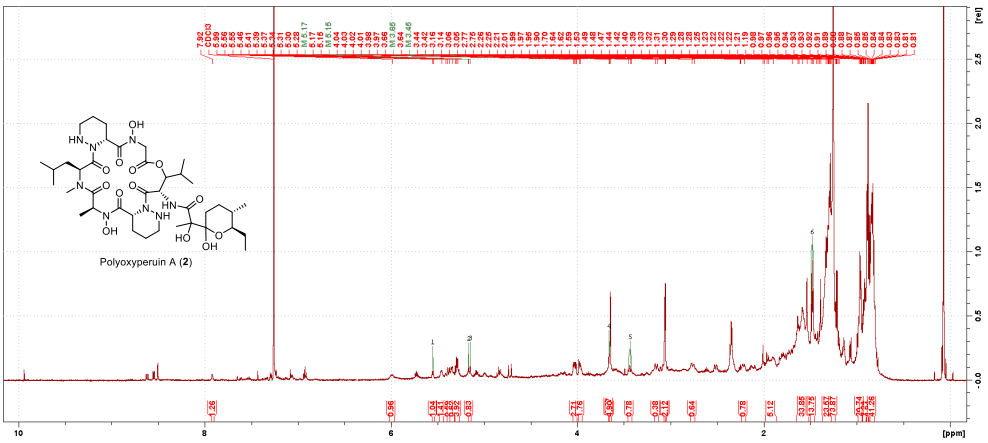


Figure S29. ^{1}H-NMR spectrum (600 MHz, $CDCl_3$) of compound **2**.

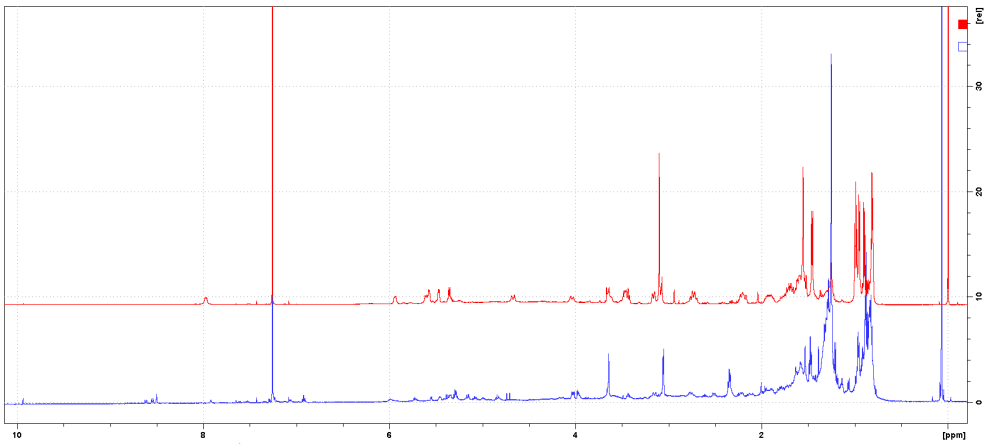


Figure S30. Comparison of the ^{1}H-NMR spectra of compound **1** (red) and compound **2** (blue). Both spectra were recorded in $CDCl_3$ at 600 MHz.

Table S1. Deduced Functions of ORFs in the Biosynthetic Gene Cluster of Polyoxyperuin.

Protein	Size[a]	Protein homolog[c]	Origin	Proposed function	Identity/ Similarity (%)	Ref. or acc. number
Orf1	412	-	*Streptomyces incarnatus*	DNA-binding protein	84/88	AKJ15799.1
Pop1	884	-	-	MT-A (NRPS)[b]	-	-
Pop2	153	-	*Streptomyces* sp. HNS054	Nuclear transport factor 2 family protein	99/100	WP_048460357.1
Pop3	120	-	*Streptomyces griseoincarnatus*	Hypothetical protein (unknown)	99/100	WP_189463054.1
Pop4	451	KtzI	*Kutzneria* sp. 744	Lysine or ornithine *N*-hydroxylase	59/73	[4]
Pop5	214	KtzT (Orf4)	*Kutzneria* sp. 744	Heme-dependent enzyme	57/66	[5, 6]
Pop6	228	LmbU	*Streptomyces lincolnensis*	LmbU transcriptional regulator	49/63	[2]
Pop7	409	QchE	*Streptomyces* sp. MBT76	*N*-acetylornithine aminotransferase	73/81	[7]
Pop8	325	QchD	*Streptomyces* sp. MBT76	Acetylglutamate kinase	67/80	[7]
Pop9	383	QchC	*Streptomyces* sp. MBT76	Bifunctional glutamate *N*-acetyltransferase/ amino acid acetyltransferase	79/86	[7]
Pop10	342	QchA	*Streptomyces* sp. MBT76	*N*-acetyl-γ-glutamyl-phosphate reductase	74/81	[7]
Pop11	396	-	*Streptomyces* sp. ZS0098	Serine hydrolase	99/99	WP_183153332.1
Pop12	248	-	*Streptomyces* sp. NRRL B-1140	TE-domain	88/91	WP_053672919.1
PopA	71	ArtA	*Streptomyces aurantiacus* JA 4570	MbtH-like protein	89/95	[8]
PopB	249	ArtB	*Streptomyces aurantiacus* JA 4570	LmbU-like transcriptional regulator	73/80	[8]
PopC	535	ArtC	*Streptomyces aurantiacus* JA 4570	A-domain	79/87	[8]
PopD	76	ArtD	*Streptomyces aurantiacus* JA 4570	PCP-domain	83/86	[8]
PopE	349	ArtE	*Streptomyces aurantiacus* JA 4570	FAD-dependent oxidoreductase	79/89	[8]
PopF	2592	ArtF	*Streptomyces aurantiacus* JA 4570	C-A-PCP-E-C-A-PCP	77/85	[8]

PopG	2948	ArtG	*Streptomyces aurantiacus* JA 4570	C-A-MT-PCP-E-C-A-PCP	74/83	[8]
PopH	1272	ArtH	*Streptomyces aurantiacus* JA 4570	C-A-PCP-TE	73/84	[8]
PopI	247	ArtI	*Streptomyces aurantiacus* JA 4570	TE-domain	80/84	[8]
PopJ	516	-	*Streptomyces* sp. E1N211	FAD-dependent monooxygenase	99/99	WP_114875797.1
PopK	104	ArtL	*Streptomyces aurantiacus* JA 4570	PCP-domain	67/78	[8]
PopL	416	ArtM	*Streptomyces aurantiacus* JA 4570	Cytochrome	86/92	[8]
PopM	246	ArtN	*Streptomyces aurantiacus* JA 4570	TE-domain	72/86	[8]
PopN	1046	ArtO	*Streptomyces aurantiacus* JA 4570	KS_Q-AT-ACP	74/81	[8]
PopO	1844	ArtP	*Streptomyces aurantiacus* JA 4570	KS-AT-DH-KR-ACP	67/76	[8]
PopP	2203	ArtQ	*Streptomyces aurantiacus* JA 4570	KS-AT-DH-ER-KR-ACP	73/81	[8]
PopQ	1044	ArtR	*Streptomyces aurantiacus* JA 4570	KS-AT-ACP	75/84	[8]
PopR	3303	ArtT	*Streptomyces aurantiacus* JA 4570	C-A-PCP	70/79	[8]
PopS	258	-	*Streptomyces* sp. E2N171	MerR family transcriptional regulator	99/99	WP_121721408.1
PopT	266	-	*Streptomyces* sp. ZS0098	Alpha/beta hydrolase	99/99	RMI88881.1
PopU	253	-	*Streptomyces* sp. ZS0098	TE-domain	99/99	WP_122216010.1
PopV	536	-	*Streptomyces* sp. E1N211	Acyl-CoA carboxylase subunit beta	99/99	WP_114875795.1

[a] Number of amino acids.

[b] Domain was predicted by antiSMASH.[1]

[c] Protein homologs were chosen based on the their identity and similarity to the proteins within the aurantimycin BGC. The ones, which display low identity or similarity to any proteins within the aurantimycin BGC were submitted to a BLASTp search, and hit proteins which have been

previously described were selected. If there is no hit for described proteins, then the highest identity protein was selected.

FAD: flavin adenine dinucleotide, MT: methyl transferase, A: adenylation domain, TE: thioesterase, PCP: peptidyl carrier protein, C: condensation domain, E: epimeration domain, KS: ketosynthase, AT: acyltransferase, ACP: acyl carrier protein, DH: dehydratase, KR: ketoreductase, ER: enoylreductase.

Table S2. The score of AT domain prediction by ATSignature and Minowa from AntiSMASH.[1]

Protein	Predicted substrate	ATSignature score	Minowa Score
PopN(AT)	(2*S*)-methylmalonyl-CoA	83.3%	186.7
PopO(AT)	(2*S*)-methylmalonyl-CoA	100%	209.9
PopP(AT)	malonyl-CoA	75%	156.7
PopQ(AT)	(2*S*)-methylmalonyl-CoA	100%	218

Table S3. Comparison of the substrate signature of PopC, PlyC, and ArtC.

Protein	Stachelhaus signature*	Substrate (L-configuration)	BGC of	Origin	Accession number
PlyC	DLRHLGQDVK	Leucine, valine, alanine	Polyoxypeptin	*S.* sp. MK498-98F14	AGZ15455.1
ArtC	DLRHLGQDVK	Alanine, leucine, serine, valine and glycine	Aurantimycin	*S. aurantiacus* JA 4570	WP_016638466.1
PopC	DLRHLGQDVK	Leucine, valine, glycine	Polyoxyperuin	*S.* sp. s120	This study

* Predicted by NRPSPredictor2 incorporated in AntiSMASH version 5.2.0.[1]

Table S4. Comparison of the substrate signature of the NRPS A-domains recognizing *β*-hydroxy-L-leucine (*β*-OH-L-Leu) as substrate.

Protein	Stachelhaus signature*	Substrate	BGC of	Origin	Accession number
PlyX	DTLWWGGVFK	*β*-OH-L-Leu	Polyoxypeptin	*S.* sp. MK498-98F14	AGZ15476.1
ArtT	DALWWGGVLK	*β*-OH-L-Leu	Aurantimycin	*S. aurantiacus* JA 4570	WP_078620912.1
PopR	DALWWGGVLK	*β*-OH-L-Leu	Polyoxyperuin	*S.* sp. s120	This study

* Predicted by NRPSPredictor2 incorporated in AntiSMASH version 5.2.0.[1]

Table S5. Comparison of the substrate signature of the NRPS A-domains recognizing L-piperazic acid (L-Piz) as substrate.

Protein	Stachelhaus signature*	Substrate	BGC of	Origin	Accession number
PlyF(A_1)	DVFSVASYAK	L-Piz	Polyoxypeptin	*S.* sp. MK498-98F14	AGZ15458.1
PlyG(A_1)	DVFSIAAYAK	L-Piz	Polyoxypeptin	*S. sp. MK498-98F14*	AGZ15459.1
ArtF(A_1)	DVFSVASYAK	L-Piz	Aurantimycin	*S. aurantiacus* JA 4570	WP_016638469.1
ArtG(A_2)	DVFTVAAYAK	L-Piz	Aurantimycin	*S. aurantiacus* JA 4570	WP_016638470.1
PopF(A_1)	DVFSVASYAK	L-Piz	Polyoxyperuin	*S.* sp. s120	This study
PopG(A_2)	DVFSVAAYAK	L-Piz	Polyoxyperuin	*S.* sp. s120	This study

* Predicted by NRPSPredictor2 incorporated in AntiSMASH version 5.2.0.[1]

Table S6. Comparison of the substrate signature of the NRPS A-domains recognizing *N*-hydroxy-L-alanine (*N*-OH-L-Ala) as substrate.

Protein	Stachelhaus signature*	Substrate	BGC of	Origin	Accession number
PlyF(A_2)	DVPDEALVEK	*N*-OH-L-Ala	Polyoxypeptin	*S.* sp. MK498-98F14	AGZ15458.1
PopF(A_2)	DVPDEALVEK	*N*-OH-L-Ala	Polyoxyperuin	*S.* sp. s120	This study

* Predicted by NRPSPredictor2 incorporated in AntiSMASH version 5.2.0.[1]

Table S7. Comparison of the substrate signature of the NRPS A-domains recognizing glycine and *N*-hydroxyglycine (*N*-OH-Gly) as substrates.

Protein	Stachelhaus signature*	Substrate	BGC of	Origin	Accession number
ArtG(A_1)	DILQLGVIWK	Glycine	Aurantimycin	*S. aurantiacus* JA 4570	WP_016638470.1
PopH(A)	DITQMGWIWK	*N*-OH-Gly	Polyoxyperuin	*S.* sp. s120	This study

* Predicted by NRPSPredictor2 incorporated in AntiSMASH version 5.2.0.[1]

Table S8. Listed primers used in this study.

No	Name	Sequence
1	pCAP03.part.F	GACCGAGATAGGGTTGAGTG
2	pCAP03.part.R	TACGTCGCGGTGAGTTCAGG
3	cassette.F	ATTCCGGGGATCCGTCGACC
4	cassette.R	CCTCCTACCCGCTGGATCCT
5	C1.HR1.F	CCTGAACTCACCGCGACGTAGGCCAGTGGCGGTACAGCAC
6	C1.HR1.R	GGTCGACGGATCCCCGGAATTCAGAGTGTCTTCCTGGGTG
7	C1.HR2.F	AGGATCCAGCGGGTAGGAGGGTTGGGAGATGAAGGTTGTG
8	C1.HR2.R	CACTCAACCCTATCTCGGTCCGCGCGTCCAGCTCGGCGTA
9	pIJ_tcp830_rev	CCTCCCAGATCTCTATCACTG
10	KO.HR1.F	CCTGAACTCACCGCGACGTATGCAGTCCCTGCTGCTGTTC
11	KO.HR1.R	GGTCGACGGATCCCCGGAATAGGTCGAGCCTGTCGAGAAG
12	New-KO.HR2.F	AGTGATAGAGATCTGGGAGGGAGTCGTTCTACGCCACCAC
13	KO.HR2.R	CACTCAACCCTATCTCGGTCAATCCGTGCGCGACGACTTC
14	KO53-F	CCTGAACTCACCGCGACGTAGACCCCAGGTCAGGGTCAAC

15	KO53-R	GGTCGACGGATCCCCGGAATTGACGTCATGAGCTCCGGCA
16	KO55-F	AGTGATAGAGATCTGGGAGGGGATCAGCCAGGACGACCTG
17	KO55-R	CACTCAACCCTATCTCGGTCGCAGCGTGGAGAGGTCGTTG
18	KO.check.F	TGGCAGTACCTGGATCTGAC
19	KO.HR2.check.R	GTGGCGATGTGGTGAATGAC
20	ermE*Pop14. check.F	AGCAGCCGCACGTCCC
21	ermE*Pop14. check.R	GGTGAGGAAGTGCGCGAAACG
22	KO.54.test.F	GTCATGTACGTCAGCGTCCA
23	KO.54.test.R	CGACGAACTCCTCGATGACG
24	KO.g20.HR1.F	ACGCATCGGCTGCCAGC
25	KO.g20.HR1.R	TATCCGGTGGACATGCCGTCA
26	KO.g20.HR2.F	TGCAAGCGGCGGACTTCG
27	KO.g20.HR2.R	ATGCGCCAGATGTCGGCG
28	KO.20.test.F	AGGACATCGGCCGCATCCTC
29	KO.test.1F	GCAGCTCCAGCCTACAGCTG
30	KO.20.test.R	TGACGACGAGCCCTTCTGGA

Table S9. Composition of eight different media used in this study.

Chemicals	Med-1	ISP2	Med-3	Med-4	Med-5	Med-6	MS	TSB
Glucose (g/L)*	10	4	10		25			2.5
Starch (g/L)	10							
Glycerol (%v/v)	1			1		2		
Dextrin (g/L)				10				
Peptone (g/L)	5							20
Meat extract (g/L)	5							
NaCl (g/L)	5		5			0.5		5
$CaCO_3$ (g/L)	3.2		3	1	4	2		
Malt extract (g/L)		10						
Yeast extract (g/L)		4		1.5	2			
Soybean powder (g/L)			20	5	15		20	
$(NH_4)_2SO_4$ (g/L)				1		2		
K_2HPO_4 (g/L)								2.5
KH_2PO_4 (g/L)						0.5		
$MgSO_4$ (g/L)						2		
Mannitol (g/L)							20	
Ref.	[9]	-	[10]	[11]	[12]	[13]	-	-

* = autoclaved separately (stock solution 500 g/L) and admixed prior to use.

The abbreviation “Med” denotes “medium”.

ATTCCGGGGATCCGTCGACCTGCAGTTCGAAGTTCCTATTCTCTAGAAAGTATAGG

AACTTCGAAGTTCCCGCCAGCCTCGCAGAGCAGGATTCCCGTTGAGCACCGCCAGGT

GCGAATAAGGGACAGTGAAGAAGGAACACCCGCTCGCGGGTGGGCCTACTTCACCT

ATCCTGCCCGGCTGACGCCGTTGGATACACCAAGGAAAGTCTACACGAACCCTTTGG

CAAAATCCTGTATATCGTGCGAAAAAGGATGGATATACCGAAAAAATCGCTATAAT

GACCCCGAAGCAGGGTTATGCAGCGGAAAATGCAGCTCACGGTAACTGATGCCGTA

TTTGCAGTACCAGCGTACGGCCCACAGAATGATGTCACGCTGAAAATGCCGGCCTTT

GAATGGGTTCATGTGCAGCTCCATCAGCAAAAGGGGATGATAAGTTTATCACCACC

GACTATTTGCAACAGTGCCGTTGATCGTGCTATGATCGACTGATGTCATCAGCGGTG

GAGTGCAATGTCGTGCAATACGAATGGCGAAAAGCCGAGCTCATCGGTCAGCTTCT

CAACCTTGGGGTTACCCCCGGCGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAGCGT

CCGGCCCCTCGAAGATGGGCCACTTGGACTGATCGAGGCCCTGCGTGCTGCGCTGG

GTCCGGGAGGGACGCTCGTCATGCCCTCGTGGTCAGGTCTGGACGACGAGCCGTTC

GATCCTGCCACGTCGCCCGTTACACCGGACCTTGGAGTTGTCTCTGACACATTCTGG

CGCCTGCCAAATGTAAAGCGCAGCGCCCATCCATTTGCCTTTGCGGCAGCGGGGCCA

CAGGCAGAGCAGATCATCTCTGATCCATTGCCCCTGCCACCTCACTCGCCTGCAAGC

CCGGTCGCCCGTGTCCATGAACTCGATGGGCAGGTACTTCTCCTCGGCGTGGGACAC

GATGCCAACACGACGCTGCATCTTGCCGAGTTGATGGCAAAGGTTCCCTATGGGGTG

CCGAGACACTGCACCATTCTTCAGGATGGCAAGTTGGTACGCGTCGATTATCTCGAG

AATGACCACTGCTGTGAGCGCTTTGCCTTGGCGGACAGGTGGCTCAAGGAGAAGAG

CCTTCAGAAGGAAGGTCCAGTCGGTCATGCCTTTGCTCGGTTGATCCGCTCCCGCGA

CATTGTGGCGACAGCCCTGGGTCAACTGGGCCGAGATCCGTTGATCTTCCTGCATCC

GCCAGAGGCGGGATGCGAAGAATGCGATGCCGCTCGCCAGTCGATTGGCTGAGCTC

ATAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGAAGCAGC
TCCAGCCTACAGCTGTTGGCTACTCTATCATTGATAGTGGTAGGATCCCTATCAGTG
ATAGAGATCTGGGAGG

Sequence S1. Linear DNA sequence of the apramycin resistance cassette (*oriT*-*acc*(3)IV).

ATTCCGGGGATCCGTCGACCTGCAGTTCGAAGTTCCTATTCTCTAGAAAGTATAGG
AACTTCGAAGTTCCCGCCAGCCTCGCAGAGCAGGATTCCCGTTGAGCACCGCCAGGT
GCGAATAAGGGACAGTGAAGAAGGAACACCCGCTCGCGGGTGGGCCTACTTCACCT
ATCCTGCCCGGCTGACGCCGTTGGATACACCAAGGAAAGTCTACACGAACCCTTTGG
CAAAATCCTGTATATCGTGCGAAAAAGGATGGATATACCGAAAAAATCGCTATAAT
GACCCCGAAGCAGGGTTATGCAGCGGAAAATGCAGCTCACGGTAACTGATGCCGTA
TTTGCAGTACCAGCGTACGGCCCACAGAATGATGTCACGCTGAAAATGCCGGCCTTT
GAATGGGTTCATGTGCAGCTCCATCAGCAAAAGGGGATGATAAGTTTATCACCACC
GACTATTTGCAACAGTGCCGTTGATCGTGCTATGATCGACTGATGTCATCAGCGGTG
GAGTGCAATGTCGTGCAATACGAATGGCGAAAAGCCGAGCTCATCGGTCAGCTTCT
CAACCTTGGGGTTACCCCCGGCGGTGTGCTGCTGGTCCACAGCTCCTTCCGTAGCGT
CCGGCCCCTCGAAGATGGGCCACTTGGACTGATCGAGGCCCTGCGTGCTGCGCTGG
GTCCGGGAGGGACGCTCGTCATGCCCTCGTGGTCAGGTCTGGACGACGAGCCGTTC
GATCCTGCCACGTCGCCCGTTACACCGGACCTTGGAGTTGTCTCTGACACATTCTGG
CGCCTGCCAAATGTAAAGCGCAGCGCCCATCCATTTGCCTTTGCGGCAGCGGGGCCA
CAGGCAGAGCAGATCATCTCTGATCCATTGCCCCTGCCACCTCACTCGCCTGCAAGC
CCGGTCGCCCGTGTCCATGAACTCGATGGGCAGGTACTTCTCCTCGGCGTGGGACAC
GATGCCAACACGACGCTGCATCTTGCCGAGTTGATGGCAAAGGTTCCCTATGGGGTG

CCGAGACACTGCACCATTCTTCAGGATGGCAAGTTGGTACGCGTCGATTATCTCGAG

AATGACCACTGCTGTGAGCGCTTTGCCTTGGCGGACAGGTGGCTCAAGGAGAAGAG

CCTTCAGAAGGAAGGTCCAGTCGGTCATGCCTTTGCTCGGTTGATCCGCTCCCGCGA

CATTGTGGCGACAGCCCTGGGTCAACTGGGCCGAGATCCGTTGATCTTCCTGCATCC

GCCAGAGGCGGGATGCGAAGAATGCGATGCCGCTCGCCAGTCGATTGGCTGAGCTC

ATAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGAAGCAGC

TCCAGCCTACAGCTAGCCGCGGTCGATCTTGACGGCTGGCGAGAGGTGCGGGGAGG

ATCTGACCGACGCGGTCCACACGTGGCACCGCGATGCTGTTGTGGGCACAATCGTGC

CGGTTGGTAGGATCCAGCGGGTAGGAGG

Sequence S2. Linear DNA sequence of the *ermE** cassette (*oriT-acc*(3)IV-*ermE**).

References

1. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. *Nucleic Acids Research* **2019**, *47*, 81-87.
2. Hou, B.; Lin, Y.; Wu, H.; Guo, M.; Petkovic, H.; Tao, L.; Zhu, X.; Ye, J.; Zhang, H. *Journal of Bacteriology* **2018**, *200*, 1-16.
3. Keatinge-Clay, A. T. *Nat. Prod. Rep.* **2012**, *29*, 1050-1073.
4. Neumann, C. S.; Jiang, W.; Heemstra, Jr J. R.; Gontang, E. A.; Kolter, R.; Walsh, C. T. *ChemBioChem* **2021***, 13*, 972-976.
5. Fujimori, D. G.; Hrvatin, S.; Neumann, C. S.; Strieker, M.; Marahiel, M. A.; Walsh, C. T. *PNAS* **2007***, 104*, 16498-16503.
6. Du, Y.-L.; He, H.-Y.; Higgins, M. A.; Ryan, K. S. *Nature Chemical Biology* **2017***, 13*, 836-838.
7. Gubbens, J.; Wu, C.; Zhu, H.; Filippov, D. V.; Florea, B. I.; Rigali, S; Overkleeft, H. S.; Wezel, G. P. v. *ACS Chem. Biol.* **2017***, 12*, 2756-2766.
8. Zhao, H. ; Wang, L.; Wan, D.; Qi, J.; Gong, R.; Z. Deng and W. Chen, *Microbial Cell Factories* **2016***,* 15, 1-12.
9. Du, Y.; Wang, Y.; Huang, T.; Tao, M.; Deng, Z.; Lin, S. *BMC Microbiology* **2014***, 14*, 1-12.
10. Zhao, H.; Wang, L.; Wan, D.; Qi, J.; Gong, R.; Deng, Z.; Chen, W. *Microbial Cell Factories* **2016***, 15*, 1-12.
11. Matsumoto, N.; Momose, I.; Umekita, M.; Kinoshita, N.; Chino, M.; Iinuma, H.; Sawa, T.; Hamada, M.; Takeuchi, T. *The Journal of Antibiotics* **1998***, 51*, 1087-1092.
12. Hayakawa, Y.; Nakagawa, M.; Toda, Y.; Seto, H. *Agric. Biol. Chem.* **1990***, 54*, 1007-1011.
13. Leet, J. E.; Schroeder, D. R.; Golik, J.; Matson, J. A.; Doyle, T. W.; Lam, K. S.; Hill, S. E.; Lee, M. S.; Whitney, J. L.; Krishnan, B. S. *The Journal of Antibiotics* **1996***, 49*, 299-311.

4. Chapter III

Production of antimicrobial compounds by homologous and heterologous expression

Kresna, I D. M., Wuisan, Z. G., Schäberle, T. F.

This manuscript is an invited contribution to the book entitled 'Antibiotics: Methods and Protocols, Second Edition'. The book is currently in preparation.

Summary: The chapters present routinely used methodology in a basic standard protocol format. The manuscript presents a detailed protocol of DNA recombination techniques for homologous expression in *Streptomyces* and heterologous expression using *Escherichia coli* as a host.

Contribution: I Dewa Made Kresna designed and wrote the part about the homologous expression in *Streptomyces*. He did the literature research, wrote the initial draft including generation of the figure.

Production of antimicrobial compounds by homologous and heterologous expression

I Dewa M. Kresna[a,b], Zerlina G. Wuisan[a,b,c] and Till F. Schäberle[a,b,c]

[a] Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, Ohlebergsweg 12, 35392 Giessen, Germany.

[b] Department of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Ohlebergsweg 12, 35392 Giessen, Germany.

[c] German Center for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Giessen, Germany.

Summary

Natural product discovery campaigns aim to identify compounds with the desired bioactivity, *e.g*. metabolites with antibiotic activity. The major driver of many projects is still the finding of bioactive extracts, which will be followed up to isolate the activity-causing agent as pure compound. However, nowadays also additional strategies can be used to increase the probability of success. Metabolomics approaches indicate chemical novelty and genomics allow identification of putative biosynthetic gene clusters (BGCs) of interest, even though the corresponding metabolite is unknown. Whatever the entry to the campaign is, at one point the scientists need to have the real compound in hand to analyze it detail. Hence, expression must be achieved to yield in the compound of interest; either to link it to the corresponding putative BGC, or to overcome the bottleneck of sparse compound supply. Therefore, homologous and heterologous expression approaches are a feasible way forward to increase production yield, shorten fermentation time, or to get BGCs expressed at all for which no suitable fermentation condition was identified. In this chapter, expression approaches in bacteria are described to biosynthesize compounds of interest. Homologous expression, by genetic manipulation of the original *Streptomyces* producer strain, as well as heterologous expression in the microbial workhorse *Escherichia coli* are exemplified.

Keywords: Heterologous expression, Homologous expression, *E. coli, Streptomyces*

1. Introduction

The initial production of antimicrobial compounds is a prerequisite for research and development projects, starting with first bioactivity screening, and advancing to identification, purification, and structure elucidation of hits, to enable rational based decisions about if an active compound should be moved on to the next development steps. Considering natural products, the producer organism has to biosynthesize the compound of interest in sufficient amounts and therefore, suitable conditions must exist. This can be achieved by fermentation of the original producer using different conditions, as described in the first edition of this book[1]. However, if no suitable condition is identifiable, there are still further options. The expression of the biosynthetic gene cluster (BGC) within the producer strain could be switched on, *e.g.* by the insertion of a, or several, promoter(s) upstream of the genes of interest. Increasing the expression yield of a gene, or a complete BGC, in a system from where it originates (*i.e.*, the natural producer) is called homologous expression. However, to do so, first knowledge about the producer strain must be available, *e.g.* which promoter regions will increase expression. Second, genetic manipulation of the strain must be possible. Another option would be to transfer the genetic information to biosynthesize the compound of interest into a heterologous host to reach heterologous expression therein. Several bacterial strains optimized for expression of single genes and BGCs are available. In addition, other types of host organisms beside bacteria exist and are established, *e.g.* fungal strains, and cell lines of mammalian or plant origin. A close phylogenetic relationship between original and heterologous host is often productive. However, the more exotic a natural producer is, the less is known about which hosts might be suitable. Due to the fact that the degenerated genetic code comprises more than one triplet codon for one specific amino acid to be added to the growing peptide chain during translation, different bacteria have their own preferences. Therefore, it might be useful to not work with the original DNA sequence for heterologous expression. Instead, a codon-optimized version of the gene(s) can be synthesized, adapted to the envisaged heterologous host. Several commercial suppliers offer to synthesize the DNA sequence of interest and directly provide the option to optimize the codon usage for the chosen host. Once, the expression is reached, this will open new possibilities to further tune the expression yield or to even modify the compound by biotechnological approaches.

In this chapter, a workflow is presented that can be easily adapted to your lab. It is exemplified by (i) the heterologous expression of a ribosomally and posttranslational modified peptide

(RiPP) BGC in *Escherichia coli* cells[2], and by (ii) the homologous expression of a modular non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) hybrid BGC in a *Streptomyces* strain[3]. The latter strain was isolated from an environmental sample[4].

2. Materials

2.1. General

2.1.1. Instrumentation

1. PCR machine
2. Centrifuge for 1.5-2 mL tubes
3. Centrifuge for 15-50 mL tubes
4. Sterile bench
5. Vortex
6. Shaking device with temperature control
7. Incubator
8. Ice generator machine
9. Nanodrop or any instrument to measure DNA concentration and OD_{600}
10. MicroPulser (including 0.2 cm cuvettes) for electroporation
11. Complete module for agarose gel electrophoresis including gel documentation instrument
12. Autoclave
13. -80 °C freezer for longtime storage of microorganisms
14. -20 °C freezer for short time storage
15. 4 °C fridge
16. Temperature controllable shaker or water bath
17. Hotplate with magnetic stirrer
18. pH meter
19. Micro pipette 100-1000 μL
20. Micro pipette 10-100 μL
21. Micro pipette 1-10 μL
22. 60 °C incubator

2.1.2. Kits

1. Agarose purification for large fragment kit
2. Bacterial DNA isolation kit
3. Plasmid purification kit

2.1.3. Strains, plasmids and primers

1. Plasmid pCAP03-acc(3)IV (kanamycin and apramycin resistance)[5]
2. *E. coli* ET12567/pUB307 (chloramphenicol and kanamycin resistance)[6]
3. *E. coli* ET12567 (chloramphenicol resistance)[6]
4. Apramycin resistance cassette with *oriT* and *ermE** promoter. The cassette could be purchased and the DNA sequence is shown in Sequence S1.
5. Primers for the experiment as described in section 3.2 and 3.3.
6. Target *Streptomyces* strain which is genetically accessible
7. *E. coli* host strain for heterologous expression
8. pRSFduet™-1

2.1.4. Chemicals

1. Tryptone
2. Yeast extract
3. NaCl
4. Agar
5. Mannitol
6. Soya flour
7. Soytone
8. Glucose
9. K_2HPO_4
10. Isopropyl β-D-1-thiogalactopyranoside (IPTG): store at -20 °C freezer.
11. Kanamycin: store at 4 °C fridge.
12. Apramycin: store at 4 °C fridge.
13. Chloramphenicol: store at 4 °C fridge.

14. Nalidixid acid: store at 4 °C fridge.
15. ddH_2O
16. Glycerol
17. Tris
18. HCl
19. $MgCl_2$
20. dNTPs: store at -20 °C.
21. Dithiothreitol (DTT): store at -20 °C freezer.
22. Polyethylene glycol (PEG) 8000
23. Nicotinamide adenine dinucleotide (NAD): store at -20 °C freezer
24. T5 exonuclease: store at -20 °C freezer
25. DNA polymerase for PCR (Phusion): store at -20°C freezer
26. Taq ligase: store at -20°C freezer
27. TE buffer: 10 mM Tris-HCl pH8, 1 mM EDTA; store at room temperature
28. Lysozyme: 10 mg/mL stock solution in TE buffer; store at -20 °C in 1 mL aliquots
29. High-Fidelity DNA polymerase for PCR (Hot start): store at -20 °C
30. Restriction enzymes: store at -20°C
31. Agarose: stock solution at 1% (w/v) concentration in TAE buffer; store at 60 °C
32. TAE buffer, 50x stock solution: 242.2 g Tris base, 18.612 g EDTA disodium salt dihydrate, 57.1 mL acetic acid, dissolved in 1 L ddH_2O. Working solution is prepared by diluting 40 mL of 50x stock solution into 2 L ddH_2O; store at room temperature.
33. Midori Green DNA stain: store at 4 °C.

2.1.5. Others

1. Mixed cellulose esters (MCE) membrane, hydrophilic, 0.025 μm pore size
2. 15 mL falcon tubes
3. 50 mL falcon tubes
4. Syringes

5. Sterile filters (0.2 μm pore size)
6. Petri dishes
7. Schott bottles
8. PCR tubes
9. 1.5-2 mL tubes
10. 2 mL cryo tubes
11. Cotton
12. Scalpel

2.2. Medium

Before use, medium has to be sterilized by using an autoclave and cooled to room temperature (RT). Initially, weight all the necessary materials and pour them into a Schott bottle. Thereafter, add ddH_2O up to the required volume and homogenize the solution by shacking or stirring. Then, autoclave the solution and let it cool to RT (*see* **Note 1**). In the following, the recipes of the media used in this experiment are described.

1. Lysogeny broth (LB) medium: 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl and 15 g/L agar (agar is only added for solid medium)
2. 2 x WT broth medium: 16 g/L tryptone, 10 g/L yeast extract, 5g/L NaCl
3. MS medium: 20 g/L mannitol, 20 g/L soya flour, 20 g/L agar
4. TSB medium: 17 g/L tryptone, 3 g/L soytone, 2.5 g/L glucose, 5 g/L NaCl, 2.5 g/L K_2HPO_4

2.3. Antibiotic and IPTG stock solutions

The antibiotic stock solutions are prepared in a 1000x concentration (in relation to the final working concentration). Concentration of the IPTG stock solution is 1 M. Filters with a pore size of 0.2 μm are used to sterilize the solution, to prevent degradation of the antibiotic and IPTG by autoclaving. Thereafter, store the solutions at -20 °C in a freezer (*see* **Note 2**).

1. Kanamycin (Kan) 50 mg/mL in ddH_2O
2. Apramycin (Apra) 50 mg/mL in ddH_2O
3. Chloramphenicol (Cm) 25 mg/mL in EtOH

4. Nalidixic acid (NA) 20 mg/mL in ddH_2O (*see* **Note 3**)
5. IPTG 1 M in ddH_2O

3. Methods

3.1. General

3.1.1. Generating electro-competent cells

1. Grow the pre-culture of the desired *E. coli* strain in 5 mL LB with necessary antibiotics in 15 mL falcon tubes or other sterile containers. Incubate at 37 °C in a shaker (180 rpm) overnight.
2. Transfer an inoculum of the pre-culture to new 50 mL LB medium (this will be the main culture) with necessary antibiotics. Of the pre-culture 1% is used to inoculate the main culture.
3. Incubate the main culture in a shaker (180 rpm) at 37 °C until it reaches an OD_{600} of 0.4 - 0.6.
4. Transfer the main culture to a 50 mL falcon tube and pellet the cells by centrifugation at 15,000 g, 2 min, at 4 °C.
5. Discard the supernatant. All following steps should be performed on ice.
6. Add 30 mL of ice-cold sterile ddH_2O and re-dissolve the pellet by shacking and vortex vigorously.
7. Centrifuge the suspension at 15,000 g, 2 min, at 4 °C.
8. Discard the supernatant.
9. Repeat step 6-8 to re-wash the cell.
10. Add 30 mL of ice-cold sterile 10% (w/v) glycerol and re-dissolve the pellet by shacking and vortex vigorously.
11. Repeat step 7-8.
12. Add 400 μL of ice-cold sterile 10% (w/v) glycerol and re-suspend cells until the solution is homogenous. After this step, the cells have become electro-competent and can be directly used or stored at -80 °C.
13. Prepare aliquots of 50 μL of the competent cells in 1.5 mL or 2 mL sterile tubes until all the competent cells are consumed.
14. Store the ready-to-use electro-competent cells in a -80 °C freezer.

3.1.2. Preparing the ready-to-use Gibson Assembly (GA) master mix[7]

1. Add the following ingredients to a 1.5 mL tube: 250 μL 1 M Tris-HCl pH 7.5, 25 μL 1 M $MgCl_2$, 50 μL 10 mM dNTP, 25 μL 1 M DTT, 125 mg PEG-8000 and 1.7 mg NAD.
2. Dissolve the ingredients completely by pipetting up and down (*see* **Note 4**).
3. Add ddH_2O to the final volume of 500 μL and resuspend. This solution represents the 5x isothermal buffer.
4. Transfer 320 μL of the isothermal buffer solution from the previous step to a new 1.5 mL or 2 mL tube.
5. From now on, each step has to be performed on ice. To the 320 μL 5x isothermal buffer solution add 1.2 μL of T5 exonuclease, 20 μL Phusion polymerase (not Hotstart), 160 μL Taq ligase and 700 μL ddH_2O.
6. Homogenize the solution by pipetting up and down and vortexing.
7. Prepare 15 μL aliquots of the solution in PCR tubes until all the solution is consumed. Each of the tubes contains now the ready-to-use Gibson assembly (GA) master mix solution needed for one experiment.
8. Store the aliquots at -20 °C in a freezer.

3.1.3. Transformation of *E. coli* by electroporation[8]

1. All steps should be performed on ice-cold temperature.
2. Prepare the ice-cold ready-to-use electro-competent cells (*see* Method 3.1.1), ice-cold LB, the salt-free-DNA to be transferred and ice-cold 0.2 cm cuvette(s).
3. Turn the MicroPulser on for electroporation.
4. Set the pre-programmed setting to Ec2 (2.5 kV) and the parameters to "time constant" (msec).
5. Add 2-4 μL of the salt-free-DNA (*see* **Note 5**) to a ready-to-use 50 μL of electro-competent cells (*see* Method 3.1.1).

6. Transfer the electro-competent cell-DNA mix to the 0.2 cm cuvette (make sure the solution is touching both electrodes).
7. Place the cuvette in the chamber slide.
8. Release the pulse once and thereafter directly re-suspend the mix by adding 1 mL of ice-cold LB to the cuvette and mix by pipetting until homogeny (*see* **Note 6**).
9. Transfer the solution to a 1.5 mL or 2 mL tube and incubate at 37 °C and shaking (180 rpm) for 45-60 min.
10. Centrifuge the culture at 15,000 x g, 1 min and discard the supernatant until only around 100 μL of supernatant are remaining.
11. Re-suspend the solution by pipetting and spread it to LB agar plates containing the corresponding antibiotics for selection.
12. Incubate the plate at 37 °C overnight.

3.1.4. Generation of *E. coli* cryo-cultures for storage

1. Prepare a sufficient amount of sterile 50% glycerol (w/v) in ddH_2O by autoclaving.
2. Grow the *E. coli* strain in 3-5 mL of LB medium with corresponding antibiotic(s) for selection overnight at 37 °C in a shaker (180 rpm) as pre-culture.
3. Transfer 1 mL of the pre-culture to a sterile cryo-tube and add an equal amount of 50% sterile glycerol (w/v).
4. Vortex the cryo-culture and store at -80 °C in a freezer.

3.1.5. Generation of Streptomycetes spore suspension for storage[9]

1. Spread a desired *Streptomyces* strain on MS agar medium with corresponding antibiotics at its optimal growth temperature (usually 28-30 °C are used) for several days until sporulation can be observed. Sporulation can be recognized by a greyish coloration, since the spores are pigmented.
2. Add 4 mL of sterile ddH_2O on top of the agar plate and gently scrape off the spores using sterile cotton.

3. Take a sterile syringe and suck the spore-containing water through the cotton (which will serve as a filter for bigger particles) and collect the ddH_2O containing the spores in a sterile 15 mL falcon tube.
4. Centrifuge the spore suspension at 15,000 x g for 5 min and thereafter discard the supernatant.
5. Add 1-2 mL of sterile 20% glycerol (w/v) to the spore pellet and re-suspend the spores therein.
6. Transfer the spore suspension to a sterile cryo tube and store it at -80 °C freezer.

3.1.6. Restriction analysis of plasmid DNA

1. Inoculate the clone tested positive by colony PCR in 5 mL LB medium containing necessary antibiotics (*see* **Note 7**).
2. Incubate the culture at 37 °C, on a shaker (180 rpm), overnight.
3. Isolate the plasmid from the culture using a plasmid isolation kit.
4. For restriction analysis take an aliquot of the plasmid DNA, add the selected restriction enzymes and their respective buffer according to the restriction enzyme manufacturers protocol. Total volume of the restriction mix should be 20-50 μL.
5. Incubate the restriction mix overnight at the recommended temperature optimum of the restriction enzymes (usually 37 °C).
6. For analysis, add loading dye and run the mix over a 1% agarose gel.
7. Analyze and document the DNA restriction pattern.

3.2. Construction of the plasmid to be used for conjugation of Streptomyces and homologous recombination

The DNA manipulation explained herein relies on a double crossover recombination. The homologous regions necessary for the recombination event to take place are introduced into the bacterium as part of a designed plasmid. This plasmid has several features: (i) Two DNA regions homologous to the sequences flanking the desired site of modification, (ii) resistance genes to enable selection in *E. coli* cells during vector construction and in *Streptomyces* cells after the recombination event took place, (iii) an origin of transcription (*ori*) for propagation of the plasmid in *E. coli*, (iv) an origin of transfer (*oriT*) to enable transfer of the DNA by

conjugating from the donor *E. coli* to the recipient *Streptomyces* cell, and (v) if necessary a promoter (*ermE** in our example) (Figure 1). The plasmid can replicate in *E. coli*, which enables construction of the plasmid in this host. However, the plasmid is used to transfer the desired DNA sequence, but cannot be propagated by the recipient *Streptomyces* strain. The initial (single) cross over could happen in either homologous region 1 or 2. Regardless in which region the recombination took place, the whole plasmid would be integrated, resulting in a strain that possesses the genes for kanamycin and apramycin resistance. When the second cross over takes place, it could happen at each of the homologous region, either 1 or 2. If the second cross over happen in the same region as the initial cross over, then the chromosome is reverted back to the original wild type status, which does not have any antibiotic resistance gene – and thus will die during antibiotic selection pressure. If the second cross over happens in the region different from the initial cross over event, the resulting strain would carry only the apramycin and not the kanamycin resistant cassette. To identify the latter case, the clones are screened using two replica plates with supplemented with a different antibiotic.

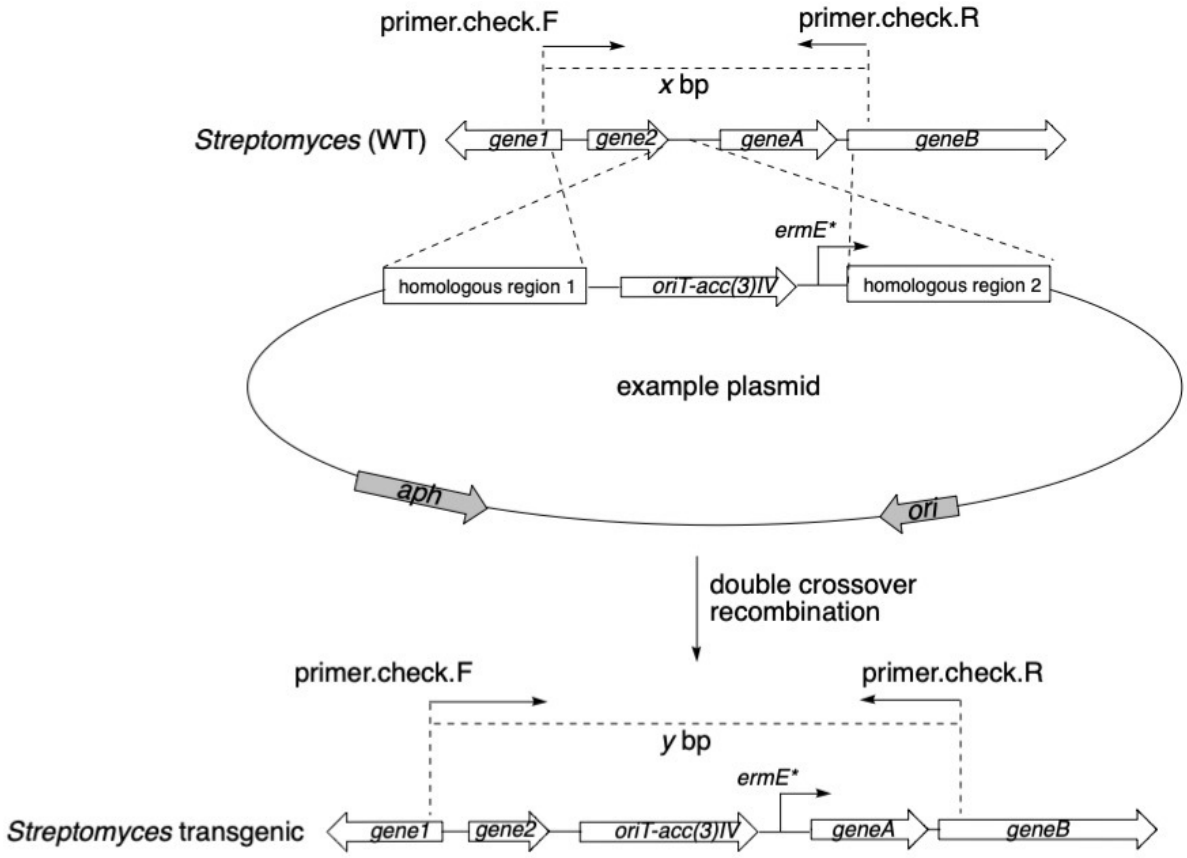


Figure 1. Scheme of the introduction of the constitutive ermE* promoter by double crossover. To introduce the promoter upstream of the target gene A, a plasmid is designed that carries two sequence regions, which are homologous to the genome of the *Streptomyces* wild type (WT) host. Here, the example plasmid carries the antibiotic resistance cassettes for apramycin (encoded by the gene *acc*(3)IV)) and kanamycin (encoded by the *aph* gene). The genetic organization of the WT and of the mutant strain after successful integration of the apramycin resistance cassette in combination with the ermE* promoter are given. The location of a primer pair to analyze if the targeted integration took place is indicated.

3.2.1. DNA fragment of homologous regions 1 and 2

1. Determine the objective of the experiment. Is an over expression or a knock out desired?
2. Decide which manipulation should be performed, e.g. deletion, insertion (introducing *ermE**), or both (replacement).
3. Determine the homologous regions (HR) 1 and 2 (each ±1000 bps in length, example in Figure 1).
4. Design a primer that can amplify the HRs 1 and 2 with annealing temperature (Tm) of 65-66 °C (to determine the Tm, use the online tool provided by the company from which the DNA polymerase is purchased). Next, add additional extension nucleotides 5'-cctgaactcaccgcgacgta-3' to the 5'-forward primer of HR1, 5'-ggtcgacggatccccggaat-3' to the 5'-reverse primer of HR1, 5'-aggatccagcgggtaggagg-3' to the 5'-forward primer of HR2 and 5'-cactcaaccctatctcggtc-3' to the 5' of the reverse primer of HR2 (*see* **Note 8**).
5. Using the designed primers and genomic DNA of your *Streptomyces* WT strain as a template, amplify the homologous regions 1 and 2 by PCR. The annealing temperature is chosen according to the primers (in our example 65 °C).
6. Isolate the desired PCR products by agarose gel and subsequent large fragment purification kit.
7. Determine the DNA concentration of HR1 and HR2 in pmol/μL (*see* **Note 9**).

3.2.2. DNA fragment of the vector

1. Digest the vector pCAP03-acc(3)IV with the restriction enzymes *Xho*I and *Nde*I by overnight incubation at 37 °C.
2. Purify the 10551 bps linear DNA fragment by agarose gel and subsequent large fragment purification kit.
3. Used the pure DNA fragment as a template for PCR using primers pCAP03.part.F (gaccgagatagggttgagtg) and pCAP03.part.R (tacgtcgcggtgagttcagg). The annealing temperature for the PCR program is 55 °C. The desired PCR amplificate (fragment length is 3877 bps) carries a kanamycin resistance gene (*aph*) and an *ori*.
4. Purify the desired PCR product by agarose gel and subsequent large fragment purification kit.
5. Determine the DNA concentration in pmol/μL (*see* **Note 9**).

3.2.3. DNA fragment of the apramycin resistance gene (*acc*(3)IV) and *oriT*

1. Amplified the apramycin resistance cassette containing *oriT* using primer cassette.F (attccggggatccgtcgacc) and cassette.R (tgtaggctggagctgcttcgaa) for a knock out experiment or using primer cassette.F and ermE*.R (cctcctacccgctggatcct) for an over expression experiment (the latter primer pair will result in an amplificate that includes the *ermE** promoter). The annealing temperature is 65 °C.
2. Purify the resulting 1369-1499 bps desired PCR product by agarose gel and subsequent large fragment purification kit.
3. Determine the DNA concentration in pmol/μL (*see* **Note 9**).

3.2.4. Fusion of the four DNA fragments using Gibson assembly

1. Calculate how many μL of each DNA fragment (from section 3.2.1-3) should be used to reach a molar ratio of vector:HR1:HR2:cassette of 1:1:1:1. The final volume should be 5 μL.
2. Add the 5 μL mix of the DNA fragments to the ready-to-use 15 μL Gibson assembly (GA) master mix (*see* Method 3.1.2).

3. Incubate the GA reaction at 50 °C for two hours (use PCR machine if possible).
4. During waiting for step 3, pour 20 mL of ddH_2O into a petri dish and place a 0.025 μm pore size MCE membrane on the water.
5. After the isothermal incubation (step 3), transfer the complete 20 μL of the GA reaction carefully using a micropipette onto the prepared 0.025 μm pore size MCE membrane (step 4).
6. Dialyze the GA reaction for 15-30 minutes.
7. Transfer the desalted GA mix to *E. coli* ET12567 cells by electroporation (*see* Method 3.1.3), and thereafter spread it on LB agar supplemented with the three antibiotics Cm/Kan/Apra.
8. Verify colonies growing on the LB $agar_{Cm/Kan/Apra}$ selection plates by colony PCR. Replicate the colony to a new LB $agar_{Cm/Kan/Apra}$ plate for maintaining the colony and incubate it at 37 °C overnight.
9. Isolate the plasmid DNA from the clones tested positive by colony PCR to analyze the DNA restriction pattern (*see* Method 3.1.6).
10. As a final verification, you can submit the plasmid DNA tested positive by colony PCR and by DNA restriction analysis for Sanger sequencing (plasmid DNA is obtained by Method 3.1.6).
11. Pick one validated colony carrying the correct construct from the agar plate and prepare a cryo-culture for storage (*see* Method 3.1.4).

3.2.5. Transfer of the generated plasmid to the *Streptomyces* strain by tri-parental conjugation[9]

1. Inoculate *E. coli* ET12567/pUB307 (carrying the helper plasmid necessary for tri-parental conjugation) and *E. coli* ET12567 (carrying the plasmid generated before, *see* Method 3.2.4) each separately into 10 mL LB medium supplemented with Kan and Cm for *E. coli* ET12567/pUB307 and additional Apra for the strain carrying the generated plasmid for transfer (*see* Method 3.2.4).
2. Grow them overnight at 37 °C in a shaker (180 rpm).
3. Take an inoculum of 100 μL of each of the overnight cultures into separate 10 mL of fresh LB medium with the corresponding antibiotics.

Incubate at 37 °C in a shaker (180 rpm) until an OD_{600} of 0.4-0.6 is reached (*see* **Note 10**).

4. Wash the cells twice with 10 mL of LB medium to remove the antibiotics that might inhibit the *Streptomyces* strain.
5. Re-suspend the cell pellet of strain *E. coli* ET12567/pUB307 in 0.5-1 mL of LB medium and the pellet of *E. coli* ET12567 carrying the plasmid to be transferred in 0.2-0.4 mL of LB medium. Now, the *E. coli* cells are ready for conjugation.
6. For the *Streptomyces*, add 10-20 µL of the *Streptomyces* wild-type spore suspension (previously generated in the way explained in Method 3.1.5) to 200 µL of 2 x YT broth medium. For each conjugation reaction, a separate tube should be prepared. Heat shock the spore suspension at 50 °C for 10 min and then allow it to cool down to room temperature (RT).
7. Mix 0.1 mL of each of the two *E. coli* cell suspensions and 0.2 mL of the heat shocked spores by pipetting up and down.
8. Plate out 50 µL and 350 µL of the mixture on two MS agar plates containing 10 mM $MgCl_2$ (without antibiotics) and incubate at 30 °C for 16-20 hours.
9. Overlay the plates with 1.5 mL of sterile ddH_2O containing 0.5 mg of NA (equivalent to 20 µL of NA stock solution) and 2-4 mg Kan (40 µL of Kan stock solution). Use a spreader to lightly distribute the antibiotic solution evenly.
10. Continue incubation at the optimum growth temperature of the used *Streptomyces* strain until sporulation can be observed.
11. Pick up several single colonies and strike them as separated colonies on an MS agar plate supplemented with NA and Apra. By this plating, single colonies should be obtained and you should get rid of remaining *E. coli* cells.
12. Incubate the plates at the optimum growth temperature of the *Streptomyces* strain until sporulation can be observed.
13. To verify the positive double crossover recombination event, pick and replica streak several single colonies on MS agar supplemented with

Kan and onto MS agar supplemented with Apra to maintain the colonies. The inability of the picked colonies to grow under Kan selection confirms the double cross recombination event (*see* **Note 11**).

14. To verify if the recombination took place at the desired location, do a PCR test using two primers that bind outside of the recombination event border (primer.check.F and primer.check.R in Figure 1). The length of the PCR product should be different (in case there is a deletion or insertion of nucleotides) in comparison to the WT strain (*see* **Note 12**).
15. Once a transgenic colony has been verified, prepare a spore suspension for storage (*see* Method 3.1.5).
16. The transgenic strain is ready to be used for cultivation.

3.3. Heterologous expression of a BGC in Escherichia coli

In contrast to the homologous expression approach, the heterologous expression system described here is plasmid-based, which means that the target BGC is cloned into an expression vector and then introduced into the heterologous expression host, in which it is propagated. The plasmid pRSFDuet™ -1 was chosen as the heterologous expression vector due to several reasons: (i) the presence of two multiple cloning sites; (ii) the presence of T7-*lac* operon that facilitates high gene expression upon IPTG induction; (iii) high plasmid copy number; (iv) kanamycin resistance cassette that does not interfere with the resistance gene of the selected heterologous expression host, *E. coli* Rosetta™ (DE3) that has chloramphenicol resistance. *E. coli* Rosetta™ (DE3) was chosen as the heterologous host because it carries the pRARE plasmid that provides tRNAs for rare codons in *E. coli*, (*i.e.* AGG, AGA, AUA, CUA, CCC, GGA). The cloning of the target BGC into the heterologous expression vector is done by Gibson Assembly. Therefore, 20-40 bp overlaps at each end of the fragments are needed (Figure 2). pRSFDuet™ -1 as the heterologous expression vector is linearized by *Nde*I and *Avr*II to remove the S-tag in the second multiple cloning site. The digestion using both restriction enzymes will result in sticky ends. However, during the Gibson Assembly reaction, the T5 exonuclease will chew the 5' ends of the fragments. Therefore, the overlapping regions that should be introduced to the fragments by using PCR are as following: (i) in the forward primer: 5'-GTATAAGAAGGAGATATACA-3' and (ii) in the reverse primer: 5'-TGCTCAGCGGTGGCAGCAGC -3'.

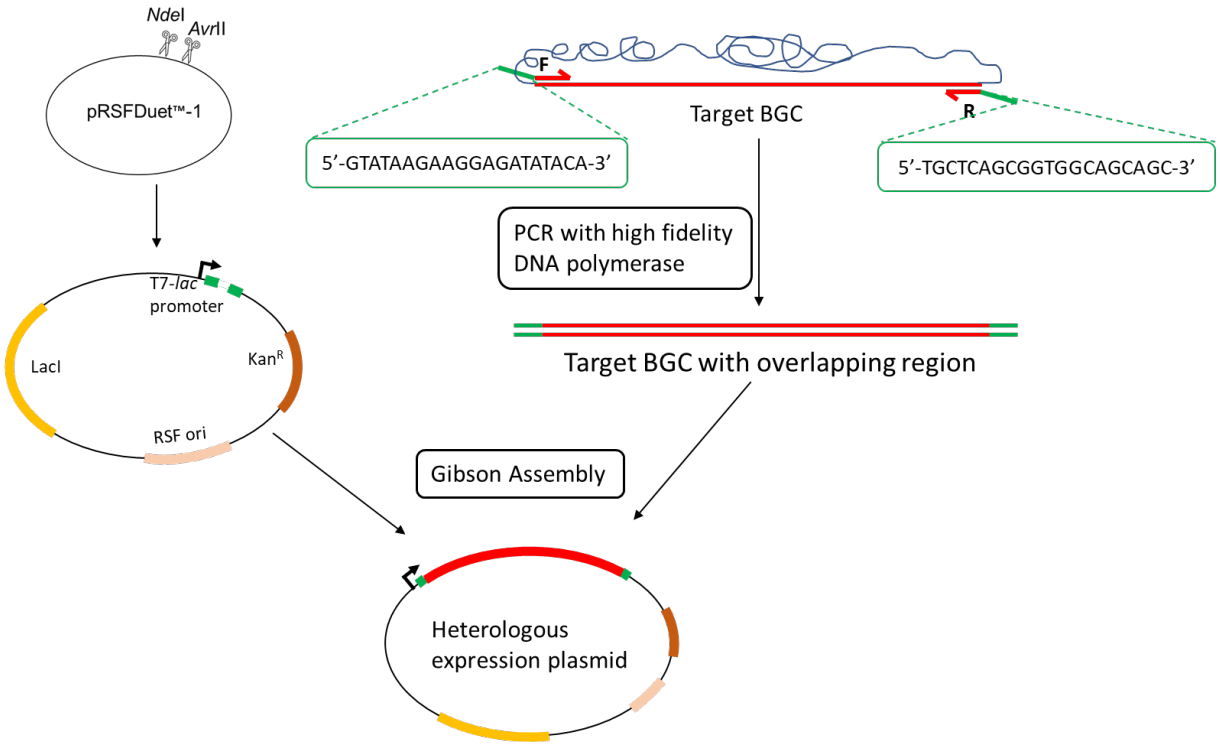


Figure 2. Scheme for cloning the target BGC into the heterologous expression vector. *Nde*I and *Avr*II are used to linearize pRSFDuet™ -1. The forward (**F**) and reverse (**R**) primers that are used for amplification of the target BGC from the genomic DNA of the original strain should have at least a 20 bp region that is homologous to the ends of the linearized plasmid (highlighted in green). The target BGC is inserted into the heterologous expression vector by Gibson Assembly (therefore, the homologous regions between the DNA fragments are needed).

3.3.1 Preparation of Gibson assembly fragments

1. Design a PCR primer pair for amplification of the target BGC. Each primer should have a minimum length of 40 bp. 20 nt of the 5' end of the primers should contain overlapping regions with the heterologous expression vector, and 20 nt of the 3' end of the primers should bind to target BGC. (*See* **Note 13 and 14**)
2. Amplify target BGC using High-Fidelity DNA polymerase according to manufacturer's protocol.
3. Linearize pRSFDuet™ -1 as the heterologous expression vector using the following reaction at 37°C, overnight: (*See* **Note 15**)

16 μL	**pRSFDuet™-1***
1 μL	*Nde*I (20 U/μL)
1 μL	*Avr*II (20 U/μL)
2 μL	10x restriction enzyme buffer

*Plasmid was isolated from its maintenance host using plasmid isolation kit. The optimal plasmid concentration should be around 100 ng/μL. (*See* **Note 16**)

4. Run a 1% agarose gel for visualization of the Gibson assembly fragments (amplified target BGC and linearized heterologous expression vector). To prevent DNA damage from UV, use Midori Green instead of Ethidium Bromide.
5. Cut the DNA fragments with correct size using a scalpel, and collect the fragments in different 2 mL tubes.
6. Purify DNA from agarose gel using "Agarose purification for large fragment kit". Elute the pure DNA using ≤10 μL autoclaved ddH_2O, pre-warmed in a 60°C oven.

3.3.2. Plasmid construction by Gibson Assembly[7] and transferring the heterologous expression construct into a heterologous host.

1. Take 1 aliquot (15 μL) of the ready-to-use GA master mix *(see* Method 3.1.2), add a total volume of 5 μL of the purified DNA fragments in equimolar amounts (*see* **Note 17**).
2. Incubate at 50°C for 1 hour (*see* **Note 18**).
3. Pour 20 mL ddH_2O into a petri dish and place a 0.025 µm nitrocellulose membrane on the water.
4. After the isothermal incubation, take all 20 μL of the Gibson Assembly reaction carefully using a micropipette, and place it on the 0.025 µm nitrocellulose membrane.
5. Dialyze the Gibson Assembly reaction for 15-30 minutes.
6. Transfer 5 μL the GA reaction into a ready-to-use electrocompetent cell of *E. coli* previously prepared (*see* Method 3.1.1) and do transformation as described in Method 3.1.3.

7. Plate transformed cells on LB agar medium supplemented with 50 μg/mL kanamycin (for selection of resistant cells harboring the desired plasmid) and incubate at 37°C (*see* **Note 19**).

3.3.3. Cultivation of heterologous expression host

1. Take a single colony carrying the constructed plasmid that was proven to be assembled correctly by PCR test and restriction enzyme digest (*see* Method 3.1.6). This colony should be preserved as cryoculture following the method described in Method 3.1.4.
2. Inoculate the single colony into 5 mL LB supplemented with kanamycin (50 μg/mL) and chloramphenicol (25 μg/mL), incubate at 37°C with 180 rpm shaking overnight.
3. On the next day, take 500 μL of the overnight culture to inoculate 50 mL LB supplemented with kanamycin (50 μg/mL) and chloramphenicol (25 μg/mL), incubate at 37°C with 180 rpm shaking until an OD_{600} of 0.4 – 0.6 is reached.
4. Induce with 25 μL of 1M IPTG (final concentration of 0.5 mM). Incubate at 37°C with 180 rpm shaking.
5. Take 1 mL samples for LC-MS analysis to detect/follow compound production after at least one day cultivation (*see* **Note 20 and 21**).

4. Notes

1. Hot medium in glassware can be cooled faster using normal tap water. Put the hot bottle into a vessel containing room temperature tap water for several minutes. Do not completely sink the bottle to prevent contamination at the bottleneck and lid area.

2. It might be useful to divide the antibiotic stock solution into aliquots of smaller volume 1-1.5 mL in sterile 1.5-2 mL tubes. This will reduce the time needed for thawing the frozen solution before use.

3. Nalidixic acid is not easily soluble in water. To increase the solubility, you could add several drops of 1 M NaOH.

4. If necessary, heat the solution to 60 °C to increase the solubility.

5. Pipetting DNA has to be done slow and carefully to prevent breaking due to shearing forces.

6. If a short circuit occurs, repeat step 4 with less volume of DNA solution. There might be remaining impurities like salts.

7. For picking the colony from the agar plate, you can use a sterile pipette tip or tooth pick. Take only half of the colony so that you still have another half left for further steps. If most of the colony is used up, you can also just re-incubate the plate at 37 °C overnight. Then, the remaining cells will multiply again and can be used for the next steps.

8. The additional nucleotide in the primers are meant to be the homologous regions for Gibson assembly.

9. If a Nanodrop instrument is used, the DNA concentration will be given μL/mL. To adjust the concentration for Gibson Assembly you must convert it to pmol/mL. (You could use online tools like the Promega biomath calculator (https://www.promega.de/en/resources/tools/biomath/) to do so. Subsequently, you can calculate that the fragments are used in the same molarity.

10. Normally, *E. coli* ET12567/pUB307 will grow faster than *E. coli* ET12567 carrying the designed plasmid.

11. It might happen that only a single crossover event took place from all the picked colonies, which would result in clones that carry apramycin and kanamycin resistance genes. If necessary, you can verify this by PCR using primer.check.F and primer.check.R (remember that in case of only single cross over event, whole plasmid is integrated meaning that you need to re-calculate the extension time for PCR). As a solution, you can grow these single crossover clones in TSB medium for several days; thereafter, take 100 μL of the culture and spread it on MS agar supplemented with Apra to get single clones. Then, continue with step 12 in section 3.3.

12. Keep in mind that the length of the PCR amplificate is only an indication and small changes in the DNA length (between mutant and

WT) might not be observable in the agarose gel after PCR. In this case, further verification by Sanger sequencing might be necessary.

13. Overlapping regions between the DNA fragments can be extended up to 40 bp to increase the efficiency of Gibson assembly.

14. If the target BGC is larger than 10 kb, it can be amplified as two fragments. The overlapping region of 20-40 bp between the fragments needs to be taken into account during primer design. Increasing amount of fragments could decrease Gibson Assembly efficiency. It is advised to have not more than 5 fragments.

15. Linearized plasmid can also be generated by amplification using a High-Fidelity DNA polymerase according to manufacturer's protocol. After the final elongation step of the PCR, 1 μL of *Dpn*I is added into the PCR reaction tube and incubated at 37°C for 1 hour. This step reduces false positives.

16. Lower plasmid concentration could lead to a lower number of colonies after Gibson assembly.

17. The concentration of the fragments to be assembled should be between 0.05–0.5 pmols for a GA using 2-3 fragments. If any fragments are smaller than 200 bp, use a 5-fold excess of these fragments.

18. During the isothermal incubation for Gibson Assembly, prepare electro-competent cells to be transformed with the Gibson Assembly result.

19. Corroborate the correct assemblies by doing PCR tests and checking the restriction pattern using suitable restriction enzyme(s).

20. Samples can be taken at certain time points to find the optimum cultivation period for compound production. Samples taken before one-day cultivation after IPTG induction might produce too low amount of the target compound to be unambiguously detected by LC-MS analysis.

21. As negative control, prepare a clone of the heterologous expression host carrying the empty expression vector, and cultivate it using the exactly same conditions as for the heterologous expression culture.

References

1. Harms H, König GM, Schäberle TF (2016) Production of antimicrobial compounds by fermentation. In: Sass P. (eds) Antibiotics. Methods in Molecular Biology, vol 1520. Humana Press, New York, NY, pp 49-61. https://doi.org/10.1007/978-1-4939-6634-9_3
2. Wuisan ZG, Kresna IDM, Böhringer N, Lewis K, Schäberle TF (2021) Optimization of heterologous Darobactin A expression and identification of the minimal biosynthetic gene cluster. Metabolic Engineering 66: 123-136. doi: 10.1016/j.ymben.2021.04.007
3. Kresna IDM, Wuisan ZG, Pohl J-M, Mettal U, Otoya VL, Gand M, Marner M, Otoya LL, Böhringer N, Vilcinskas A, Schäberle TF (2022) Genome Mining-Guided Discovery and Characterization of the PKS-NRPS-Hybrid Polyoxyperuin Produced by a Marine Streptomycete. Under revision.
4. Linares-Otoya L, Linares-Otoya V, Armas-Mantilla L, Blanco-Olano C, Crüsemann M, Ganoza-Yupanqui ML, Campos-Florian J, König GM, Schäberle TF (2017) Diversity and Antimicrobial Potential of Predatory Bacteria from the Peruvian Coastline. Mar. Drugs 15(10) 308: 1-14. doi: 10.3390/md15100308
5. Tang X, Li J, Millán-Aguiñaga N, Zhang JJ, O'Neill EC, Ugalde JA, Jensen PR, Mantovani SM, Moore, BS (2015) Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. ACS Chem. Biol. 10(12): 2841–2849. doi: 10.1021/acschembio.5b00658
6. Flett F, Mersinias V, Smith CP (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from Escherichia coli to methyl DNA-restricting streptomycetes. FEMS Microbiology Letters 155: 223-229. doi: 10.1111/j.1574-6968.1997.tb13882.x
7. Gibson DG, Young L, Chuan R-Y *et al*. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods 6:343-345. doi: 10.1038/nmeth.1318
8. BIO-RAD (-) MicroPulser ™ Electroporation Apparatus Operating Instruction and Application Guide. BIO-RAD. Available vio DIALOG. https://www.bio-rad.com/webroot/web/pdf/lsr/literature/4006174B.pdf. Accessed 24 Dec 2021
9. Tang X (-) Direct Cloning and Heterologous Expression of Biosynthetic gene cluster. The Moore Lab. Available via DIALOG. https://media.addgene.org/data/plasmids/69/69862/69862-attachment_sMTMGzzwy5aw.pdf. Accessed 24 Dec 2021

5. Chapter IV:

Optimization of heterologous Darobactin A expression and identification of the minimal biosynthetic gene cluster

Wuisan, Z. G., Kresna, I D. M., Böhringer, N., Lewis, K. and Schäberle, T. F.

Published manuscript: *Metabolic Engineering* **66**, 123-136 (2021). doi: 10.1016/j.ymben.2021.04.007

Summary: In this manuscript is the improvement of the darobactin A production yield in heterologous expression systems detailed. Optimization parameters include different heterologous producer strains, different origin of the darobactin biosynthetic gene clusters (BGCs) and modification of these. Furthermore, the regulation of the biosynthetic pathway was investigated and the minimum BGC was identified. The effect of the *darA* to *darE* transcript ratio on production yield was tested.

Contribution: I Dewa Made Kresna contributed the section “A dedicated peptidase is not part of the BGC”. Therefore, he designed and planed the strategies and experimental set ups. He performed experiments, analyzed results for that particular section producing the Figure 9 in the manuscript.

Metabolic Engineering 66 (2021) 123–136

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/meteng

Optimization of heterologous Darobactin A expression and identification of the minimal biosynthetic gene cluster

Zerlina G. Wuisan [a], I Dewa M. Kresna [a], Nils Böhringer [a,d], Kim Lewis [b], Till F. Schäberle [a,c,d,*]

[a] Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, 35392, Giessen, Germany
[b] Antimicrobial Discovery Center, Northeastern University, Department of Biology, Boston, MA, USA, 02115
[c] Branch for Bioresources of the Fraunhofer Institute for Molecular Biology and Applied Ecology, 35394, Giessen, Germany
[d] German Center for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Giessen, Germany

ARTICLE INFO

Keywords:
Antibiotic
Heterologous expression
Biosynthesis
RiPP

ABSTRACT

Darobactin A (DAR) is a ribosomally synthesized and post-translationally modified peptide (RiPP) antibiotic, which was initially identified from bacteria belonging to the genus *Photorhabdus*. In addition, the corresponding biosynthetic gene cluster (BGC) was identified and subsequently detected in several bacteria genera. DAR represents a highly promising lead structure for the development of novel antibacterial therapeutic agents. It targets the outer membrane protein BamA and is therefore specific for Gram-negative bacteria. This, together with the convincing *in vivo* activities in mouse infection models, makes it a particular promising candidate for further research. To improve compound supply for further investigation of DAR and to enable production of novel derivatives, establishment of an efficient and versatile microbial production platform for these class of RiPP antibiotics is highly desirable. Here we describe design and construction of a heterologous production and engineering platform for DAR, which will ensure production yield and facilitates structure modification approaches. The known Gram-negative workhorses *Escherichia coli* and *Vibrio natriegens* were tested as heterologous hosts. In addition to that, DAR producer strains were generated and optimization of the expression constructs yielded production titers of DAR showing around 10-fold increase and 5-fold decrease in fermentation time compared to the original product description. We also report the identification of the minimal DAR BGC, since only two genes were necessary for heterologous production of the RiPP.

1. Introduction

Antibiotic resistance has become one of the major threats on global health, as infectious diseases are becoming more and more difficult to be treated by the market available antibiotics. The world health organization (WHO) puts critical priority to the research and development of new antibiotics against the multi-drug-resistant (MDR) Gram-negative pathogens, e.g. *Acinetobacter baumanii*, *Pseudomonas aeruginosa*, *Escherichia coli, Klebsiella pneumoniae* ("WHO publishes list of bacteria for which new antibiotics are urgently needed," 2017). The outer membrane (OM) of Gram-negative bacteria plays a major role in their resistance against certain drugs, due to the fact that the OM functions as a permeability barrier, which restricts drug penetration into the cell (Silhavy et al., 2010). Hence, there is an obvious need to fill the antibiotic development pipeline with candidate molecules that can be evolved into medicinal drugs for the treatment of infectious diseases caused by Gram-negative bacteria.

Darobactin A (DAR), a novel antibiotic that selectively kills Gram-negative pathogens, *e.g. Acinetobacter baumannii* (MIC, 8 μg/ml), *Pseudomonas aeruginosa* PAO1 (MIC, 2 μg/ml), *Escherichia coli* (MIC, 4 μg/ml), and *Klebsiella pneumoniae* (MIC, 2 μg/ml), was discovered in *Photorhabdus khanii* HGB1456. Experimental proof was provided that DAR binds to BamA, which is the central component of the OM β-barrel assembly machinery. It helps to fold and insert β-barrel proteins such as porins into the OM. If this chaperone-like function is impaired, it will result in the disruption of OM formation. In addition to its good *in vitro* activity, DAR showed promising efficacy in mouse septicemia and thigh infection models without cytotoxic effects (Imai et al., 2019). Therefore, DAR has emerged as a promising drug lead.

Heterologous expression is often used to express a silent gene cluster (Hegemann et al., 2013; Yamanaka et al., 2014), express BGCs from unculturable bacteria (Crüsemann et al., 2018; Ongley et al., 2013),

* Corresponding author. Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, 35392, Giessen, Germany.
E-mail address: till.f.schaeberle@agrar.uni-giessen.de (T.F. Schäberle).

https://doi.org/10.1016/j.ymben.2021.04.007
Received 11 February 2021; Received in revised form 23 March 2021; Accepted 11 April 2021
Available online 16 April 2021

increase yield (Flinspach et al., 2010; Sucipto et al., 2017), as well as to proof the functionality of a putative BGC (Linares-Otoya et al., 2017). In addition to that, the genetic tools for modification of the BGC in a heterologous expression system are more available to study the biosynthetic pathway (Bouhired et al., 2014; Crüsemann et al., 2018) and to produce new derivatives of the natural product (Paulus et al., 2018).

In our previous study, we identified a 6.2 kb biosynthetic gene cluster (BGC) that was responsible for DAR production (Fig. 1b). Knocking out the BGC abolished DAR production, but reinstatement of the BGC into *Photorhabdus khanii* DSM3369 could not restore DAR production, indicating that the regulatory framework in *P. khanii* DSM3369 does not allow interference. However, the final confirmation of the BGC was achieved by heterologous expression in *E. coli* BW25113, which resulted in production of DAR. Even though the production yield was low (<1 mg/L), it clearly indicated the general suitability of *E. coli* as expression host for DAR.

A long fermentation period (10–14 days) was required to obtain DAR from its native *Photorhabdus* producer strains, with a maximum yield of only 3 mg/L (Imai et al., 2019). Although the heterologous expression in *E. coli* BW25113 could shorten the fermentation period to three days, the yield was still lower than the one of the native producer strains.

The present study aims at overcoming the major limitation for further investigation of DAR, *i.e.* the compound supply. Therefore, the goal was improvement of DAR production by optimization of the biosynthesis using a heterologous production system. We choose several different heterologous hosts, cloned the respective DAR BGC from different species, including upstream regions of the BGC, and lastly created a DAR-resistant heterologous host to boost DAR production. Furthermore, *in silico* predicted DAR BGCs were experimentally confirmed and the minimal BGC necessary for heterologous production was identified.

2. Material and methods

2.1. Plasmid and strain construction

Several constructs were generated during this project for heterologous expression of DAR (Fig. 2). The vector used for the expression constructs was in all cases pRSFDuet™-1 (Merck KGaA, Darmstadt, Germany). Chromosomal DNA used as template for amplification of the DAR BGC was isolated using the innuPREP Bacteria DNA Kit (Analytik Jena AG, Jena, Germany). In general, fragments were amplified using Q5 DNA polymerase (New England Biolabs, Ipswich, USA) and purified

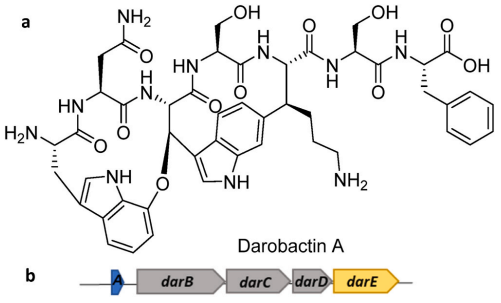


Fig. 1. Darobactin (DAR) is a ribosomally synthesized and post-translationally modified peptide (RiPP), encoded by the *dar* operon. a, DAR is a modified heptapeptide consisting of the seven amino acids W^1-N^2-W^3-S^4-K^5-S^6-F^7, with an ether bond between W^1 and W^3, and a *C-C* bond between W^3 and K^5. **b**, The *dar* BGC (in total 6.2 kb in length) consists of *darA* that encodes the precursor peptide; *darBCD* that encode subunits of an ABC transporter; and *darE* that encodes a radical *S*-adenosylmethionine enzyme (RaS).

from the agarose gel using Large Fragment DNA Recovery Kit (Zymo Research, Irvine, USA). The polymerase chain reaction (PCR) was performed in a Biometra TRIO thermocycler (Analytik Jena AG) using the following program: 95 °C for 2 min; 34 cycles of 95 °C for 45 s, 60–70 °C for 45 s (applied annealing temperature was depending on the primer sequence), 72 °C for 30 s/kb (extension time was varied depending on the length of the fragment to be amplified), followed by a final extension step at 72 °C for 5 min. All primer sequences can be found in Table S2.

Plasmid pZW-ADC3 carries the genes of the DAR BGC without the intergenic region between *darA* and *darB*. Therefore, *darA* was amplified from *Photorhabdus khanii* HGB1456 (Table S1, GenBank accession number WHZZ00000000) using the primer pair ADC-F7 and ADC-R1; and *darB* to *darE* was amplified using the primers ADC-F2 and ADC-R7. pRSFDuet™-1 was linearized using *Nde*I and *Avr*II restriction enzymes (New England Biolabs) to insert the two purified fragments into the second multiple cloning site of the vector under the control of the T7-*lac* promoter. To do this, the one-step isothermal DNA assembly protocol described by Gibson was followed (Gibson et al., 2009), with the minor modification that 1.2 µl of 10 U/µL T5 exonuclease was added instead of 0.64 µL. Therefore, the final concentration of the Gibson reaction mix was the following: 100 mM Tris-HCl pH7.5, 10 mM $MgCl_2$, 0.2 mM each dNTP, 10 mM DTT, 5% PEG-8000, 1 mM NAD, 7.5 U/mL T5 exonuclease, 25 U/mL Phusion polymerase, 4 U/µL Taq DNA ligase, 0.02–0.5 pmol DNA fragments. This reaction mix was then incubated at 50 °C for 1 h. After the isothermal assembly, the reaction was dialyzed using a 0.025 µm nitrocellulose membrane (Merck™ MF-Millipore™, Ireland), and subsequently transferred to *E. coli* TOP10 cells as a plasmid maintenance host by electroporation using Micropulser Electroporator (Bio-Rad, California, USA) in a 0.2 cm electroporation cuvette at a voltage of 2.5 kV.

The second and third construct, carries the native DAR BGC from *Photorhabdus khanii* HGB1456 (pZW-ADC5) and *Photorhabdus khanii* DSM3369 (GenBank accession number AYSJ00000000) (pZW-ADC6), respectively. For both constructs, the respective BGC was amplified using the primer pair ADC-F7 and ADC-R7; therefore, the respective bacterial genomic DNA was used as template. Gibson Assembly was performed as described above.

Plasmids pZW-ADC3.2 and pZW-ADC5.2 were created by restriction-digest of pZW-ADC3 and pZW-ADC5. Each plasmid was restricted using *Nco*I and *Not*I (New England Biolabs). Then, an additional codon optimized version of *Photorhabdus namnaonensis darA* (Sequence S1) was inserted into the first multiple cloning site of the pRSFDuet™-1 based vectors by Gibson Assembly.

The plasmids pZW-ADC7 and pZW-ADC8 carry the native DAR BGC from *P. khanii* HGB1456 with an additional 200 bp and 605 bp upstream region of *darA*, respectively. The BGC were amplified from the bacterial genomic DNA using the primer pair ADC-F9 and ADC-R7 for the pZW-ADC7 insert, and the primer pair ADC-F10 and ADC-R7 for the pZW-ADC8 insert. The respective inserts were assembled to the *Nde*I-*Avr*II linearized pRSFDuet™-1 using Gibson Assembly.

The plasmid pZW-ADC9 carries the DAR BGC from *P. khanii* DSM3369 without the transporter genes, *i.e.* only *darA* and *darE*. *darA* was amplified from the bacterial genomic DNA using the primer pair ADC-F7 and ADC9-R and *darE* was amplified using the primer pair ADC9-F and ADC-R7. Both fragments were assembled to the *Nde*I-*Avr*II linearized pRSFDuet™-1 using Gibson Assembly.

Synthetic codon-optimized DNA of *darA* and *darE* from *P. namnaonensis* (Sequence S1 and S2) were amplified using the primer pairs pro.CO.F – pro.CO.R and RS.CO.F and RS.CO.R, respectively. Both amplicons were assembled to the *Nde*I-*Avr*II linearized pRSFDuet™-1 vector by Gibson Assembly, creating pZW-ADC12.

pZW-YerA4 carries the DAR BGC derived from *Yersinia frederiksenii* ATCC 33641 (GenBank accession number NZ_JPPS00000000) that was amplified using the primer pair YerA-F4 and YerA-R4. The amplicon was also assembled to the *Nde*I-*Avr*II linearized pRSFDuet™-1 by using Gibson assembly.

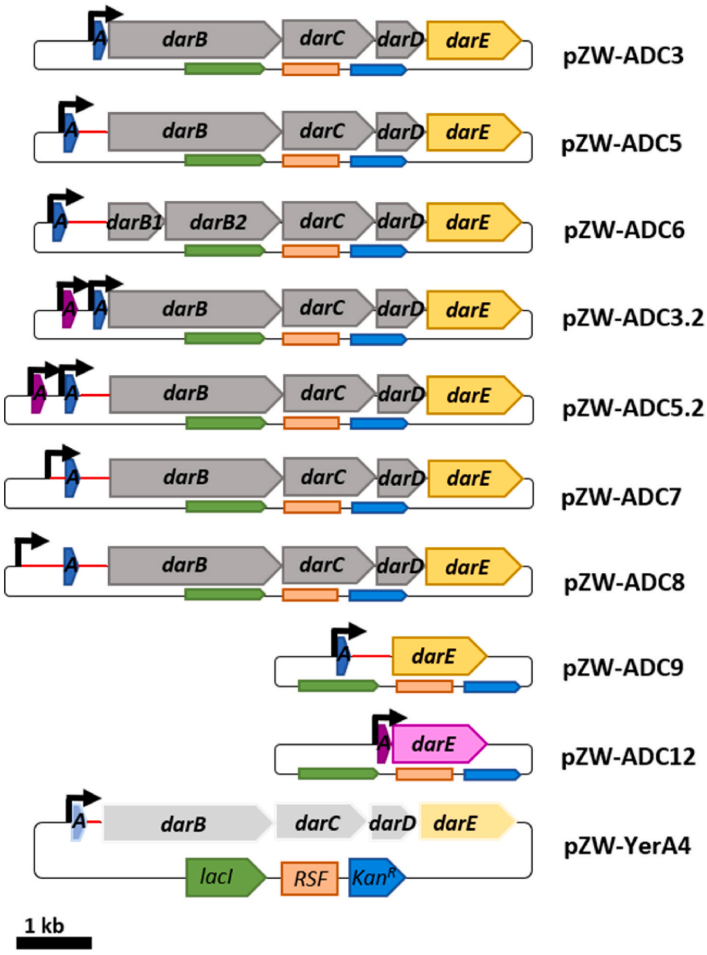


Fig. 2. Expression constructs created in this study. The DAR BGC was cloned into the pRSFDuet™-1 vector under control of the T7-*lac* promoter. pZW-ADC3 has a 'streamlined' DAR BGC from *P. khanii* HGB1456, where all intergenic regions were removed, while pZW-ADC5 and pZW-ADC6 carry the native cluster from *P. khanii* HGB1456 and *P. khanii* DSM3369, respectively. Addition of a codon optimized *darA* from *P. namnaonensis* (in purple) to pZW-ADC3 and pZW-ADC5 created pZW-ADC3.2 and pZW-ADC5.2. Plasmid pZW-ADC7 carries the *P. khanii* HGB1456-derived DAR BGC with an additional 200 bp upstream region of *darA* and pZW-ADC8 harbors a 605 bp upstream region. Plasmid pZW-ADC9 harbors the *P. khanii* DSM3369-derived DAR BGC, omitting *darBCD*, while pZW-ADC12 carries codon optimized *darA* and *darE* (in pink) from *P. namnaonensis* and pZW-YerA4 carries the DAR BGC from *Yersinia frederiksenii* ATCC 33641. The black arrow indicates the T7-*lac* promoter and the color code for *lacI*, RSF and the kanamycin resistance cassette (Kan^R) is kept constant. A red line indicates sequence derived from the BGC, while a black line indicates the vector backbone. The *lacI* gene encodes for the *lac* operon repressor and RSF is an origin of replication that was derived from RSF1030, which allows the plasmid to be maintained at a high copy number in the cell.

Following assembly and propagation in *E. coli* TOP10 cells, all constructs were checked by test PCR and by their restriction pattern.

2.2. Qualitative analysis of different promoters

The qualitative analysis of different promoters was done by visualization using green fluorescence protein (GFP) as reporter gene, and by testing the inhibition of *E. coli* MG1655 BamA6 (Ruiz et al., 2006). To create the construct pZW-ADC5-GFP, *gfpmut3.1* was amplified from pFU95 (GenBank accession number: JF796092.1) using the primer pair GFP-F and GFP-R2 and pZW-ADC5 was amplified using the primer pair ADC5-GFP-F2 and ADC5-GFP-R. Then, these two fragments were assembled by Gibson Assembly. Thereafter, plasmid pZW-ADC5-GFP was amplified using the primer pair T5-F and T5-R. Thereafter, one-fragment-Gibson-Assembly was performed to replace the T7-*lac* with the T5 promoter (Shibui et al., 1988).

The J23101 promoter (Davis et al., 2011) was amplified using overlap extension PCR. The first primer pair, *i.e.* J23101-F and J23101-R, in which the primers complement each other, has a 20 bp 5' overhang region on each primer. The first PCR was performed to fill the overhang using the following conditions: 95 °C for 2 min; 15 cycles of 95 °C for 45 s, 60 °C for 45 s, 72 °C for 15 s. Subsequently, the resulting double stranded DNA template was amplified using the second primer pair J23101-F2 and J23101-R2. Hence, this primer pair was added and the following PCR program was applied: 95 °C for 2 min; 30 cycles of 95 °C for 45 s, 55 °C for 45 s, 72 °C for 15 s, followed by a final extension step at 72 °C for 5 min. The therefrom-resulting 125 bp fragment was assembled with pZW-ADC5-GFP (Fig. S1) that was amplified using the primer pair ADC-F7 and J23101-ADC5-GFP-R.

The prpB promoter and its regulatory gene *prpR* were amplified from *E. coli* MG1655 genomic DNA (GenBank accession number NC_000913.3) with the primer pair pPro-F and pPro-R, and the primer pair pPro-A5-GFP-F and pProA5-GFP-R was used to amplify pZW-ADC5-GFP (Fig. S1). The two fragments were assembled using Gibson Assembly.

Overnight cultures of *E. coli* Rosetta™(DE3) that carry the respective

constructs with different promoters were grown in LB with kanamycin and chloramphenicol at 37 °C. 5 μL of the culture was spotted on LB agar with necessary antibiotics and promoter inducer. The constructs with T5 and J23101 promoter did not need any inducer, T7 promoter was induced with 1 mM IPTG, and prpB promoter was induced with 20 mM sodium propionate. The agar plates were incubated at 30 °C for three days, and GFP-depending fluorescence was observed by a blue-light transilluminator at 470 nm. Subsequently, the plates were overlaid with LB-agar containing 1% overnight culture of *E. coli* MG1655 BamA6, kanamycin and necessary inducer. The plates were incubated at 30 °C, and the inhibition zones were documented on the next day.

2.3. Generation of E. coli mutant strains as heterologous hosts

Two *E. coli* mutant strains were created in this study: (i) *E. coli* Rosetta™(DE3)(Dar^R) and (ii) *E. coli* BAP1 Δ*tolC*:: *aac(3)-IV* (Apr^R); by means of λ-Red recombination. The DAR resistant strain was generated by introduction of three point mutation into the *bamA* gene, which are 1300A>G, 1334A>C, and 2113G>A. That these three point mutations result in a DAR resistant phenotype was confirmed by a previous study, whereby DAR resistance increased to 128 μg/mL (Imai et al., 2019). A linear PCR product of the *bamA* gene with the three point mutation was amplified from this *E. coli* DAR-resistant mutant (strain3) (Imai et al., 2019) using the bamA-recF and bamA-recR primer pair. On the other hand, an apramycin resistance cassette (*aac(3)-IV*) flanked with FRT regions was amplified using the TolCKO-F and TolCKO-R primer pair to be introduced to *E. coli* BAP1 to create the *tolC* knockout mutant.

The λ-Red recombination to create the mutant strain was done according to Datsenko and Wanner (2000), with some modifications. The target *E. coli* strains were transformed with pKD46, a heat-sensitive plasmid that carries λ red genes under the control of the araBAD promoter. They were grown in lysogeny broth (LB) with carbenicillin (50 μg/mL) at 30 °C for 1 h, supplemented with 20 mM L-arabinose, then further incubated to reach the OD_{600} of ≈0.6. Thereafter, the cells were made electrocompetent by three times washing with ice-cold 10% glycerol. The linear PCR amplicon (~100 ng) was added into 50 μL of electrocompetent *E. coli* cell solution and electroporation was performed as previously described in section 2.1. After 1 h incubation for cell recovery, the *E. coli* Rosetta™(DE3)(Dar^R) strain was plated on LB with DAR (32 μg/mL) and the *E. coli* BAP1 Δ*tolC*:: *aac(3)-IV* (Apr^R) was plated on LB with apramycin (50 μg/mL) as selection markers, and incubated at 37 °C to promote the loss pf pKD46. The DAR resistant strain was confirmed by sequencing of the *bamA* gene and the *tolC* knockout strain was confirmed by PCR.

2.4. Heterologous host strains, media and maintenance

All bacterial strains used in this study were maintained in glycerol stocks (25% (v/v)) at −80 °C. *E. coli* and *P. khanii* strains were grown in LB medium at 37 °C and 30 °C, respectively, unless stated otherwise. The LB medium was prepared with 10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl, and supplemented with 50 mg/L kanamycin and 25 mg/L chloramphenicol when required for the purpose of selection. Kanamycin supplementation was done for *E. coli* strains carrying the expression constructs and chloramphenicol was supplemented for *E. coli* Rosetta™(DE3) cultivation.

To generate a growth curve, 50 mL of LB medium were supplemented with required antibiotics, inoculated with 500 μL of an overnight culture, and cultivated at 30 °C with 180 rpm shaking. Samples for OD_{600} measurement were taken at 0, 2, 4, 6, 9, 12, 16, 25, 54, 78, and 102 h after inoculation.

Vibrio natriegens Vmax™ was grown in LB medium with artificial seawater (LB-ASW) at 30 °C. The artificial seawater was prepared with 0.1 g/L KBr, 23.48 g/L NaCl, 10.61 g/L $MgCl_2.6H_2O$, 1.47 g/L $CaCl_2.2H_2O$, 0.66 g/L KCl, 0.04 g/L $SrCl_2.6H_2O$, 3.92 g/L Na_2SO4, 0.19 g/L $NaHCO_3$, 0.03 g/L H_3BO_3. 200 mg/L kanamycin was supplemented to the LB-ASW for the growth of *V. natriegens* Vmax™ that carries the expression constructs.

To make electrocompetent cells for electroporation, *E. coli* strains and *V. natriegens* Vmax™ were grown in LB and LB-ASW, respectively, to OD_{600} between 0.4 and 0.7. Then, the cells were washed two times with electroporation buffer (10% glycerol for *E. coli* strains, and 680 mM sucrose, 7 mM K_2HPO_4, pH 7 for *V. natriegens* Vmax™)(Weinstock et al., 2016). Subsequently, the cells were resuspended in 1:100 volume of electroporation buffer compared to the initial culture volume, and aliquoted into chilled tubes. The electrocompetent cells were always made fresh prior to electroporation.

2.5. Heterologous expression of DAR

The expression constructs were transferred from the maintenance host, *E. coli* TOP10, into different expression hosts by electroporation. The transformation of *E. coli* expression hosts was performed as described above for *E. coli* TOP10, while the transformation of *V. natriegens* Vmax™ was done in a 0.1 cm electroporation cuvette at a voltage of 900 V. After the electroporation, the strains were incubated for 1 h in their respective growth medium and temperature. Then, they were plated on LB(-ASW) plates containing kanamycin as the selective agent with the aforementioned concentration, with the addition of chloramphenicol when *E. coli* Rosetta™(DE3) was used as the host. Single colonies were picked from the selective plates and the presence of the respective expression plasmid was confirmed by PCR. Colonies with correct constructs were then inoculated in 3 mL LB(-ASW) containing required antibiotics and incubated at 37 °C or 30 °C overnight (see point 2.2). 500 μL of this pre-culture were used to inoculate 50 mL of fresh LB (-ASW) medium containing kanamycin, incubated at 37 °C or 30 °C until an OD_{600} of 0.4–0.6 was reached, and then induced with IPTG (final concentration of 0.5 mM). After IPTG induction, the cultures were incubated at 30 °C with 180 rpm shaking.

2.6. UPLC-HRMS analysis

DAR production was analyzed by UPLC-HRMS. From the expression culture, a 1 mL aliquot was taken and centrifuged to separate the medium and the bacterial cell. The medium was lyophilized, 1 mL methanol was added, the mixture was sonicated in a Bandelin Sonorex RK255 ultrasonic bath (Berlin, Germany) for 30 min, and centrifuged at 10,000×*g* for 5 min. The methanol was removed, and the pellet was resuspended in 1 mL deionized water. After a final centrifugation at 10,000×*g* for 5 min, the sample was ready to be injected to the UPLC-HRMS system. To prepare the sample from the cell pellet, 500 μL of methanol was added prior to sonication for 30 min. Then, 500 μL of deionized water was added, and the sonication was continued for another 15 min. Thereafter, the solution was centrifuged to pellet the insoluble part and the supernatant was injected to the UPLC-HRMS system.

The UPLC-HRMS system was an Agilent Infinity 1290 UPLC system equipped with an Acquity UPLC BEH C18 1.7 μm (2.1 × 100 mm) column (Waters, Eschborn, Germany) and an Acquity UPLC BEH C18 1.7 μm VanGuard Pre-Column (2.1 × 5 mm; Waters, Eschborn, Germany) setup coupled to a DAD detector and a micrOTOFQ II mass spectrometer (Bruker, Bremen, Germany). The LC part was run using a gradient (A: H_2O, 0.1% FA; B: MeCN, 0.1% FA; Flow: 600 μL/min): 0 min: 95%A; 0.80 min: 95%A; 18.70 min: 4.75%A; 18.80 min: 0%A; 23.00 min: 0%A; 23.10 min: 95%A; 25.00 min: 95%A and the column oven temperature was set to 45 °C. MS parameters were as follows: nebulizer gas 1.6 bar; gas temperature, 200 °C; gas flow, 8 L/min; capillary voltage, 4500 V; endplate offset, 500 V; measurement was done in positive ion mode.

A DAR standard curve was generated by plotting the peak area of DAR from the extracted ion chromatogram (EIC) (for the m/z of 483.7089 and 475.1956 ± 0.01) to a series of DAR concentrations (2, 3, 4, 5, 10, 15, 20, 30, 40 mg/L). The DAR concentration from a

heterologous expression culture was quantified by calculating the peak area and interpolating it to the DAR standard curve. The linear range for this quantification method was 3 μg/mL to 30 μg/mL. Therefore, the peak area below the border was not converted to concentration. The standard curve was measured with all batches that were analyzed by UPLC-HRMS to exclude technical differences between measurements.

2.7. Determination of E. coli survival rate in the presence of DAR

Overnight cultures of *E. coli* strains were diluted 1:100 with LB-chloramphenicol and cultivated at 30 °C with 180 rpm shaking. After reaching an optical density at 600 nm (OD_{600}) of 0.6–0.7, these cultures were further diluted to OD_{600} of 0.001, 0.1, and 0.5. The diluted cultures were pipetted into eight wells in a 96-well plate, 100 μL final volume in each well. Different concentrations of antibiotics were adjusted. The first three wells: 6 μg/mL DAR, the second three wells: 30 μg/mL DAR, the seventh well: 128 μg/mL gentamycin as negative control, and the eighth well was not treated with an antibiotic as a positive control of the bacterial growth. The plate was incubated overnight at 30 °C with 180 rpm shaking, and OD_{600} was measured using a Tecan infinite®200 plate reader (Tecan Group Ltd., Männedorf, Switzerland). The survival rate was calculated by the following equation:

$$\frac{OD_{600}\ of\ sample - OD_{600}\ of\ negative\ control}{OD_{600}\ of\ positive\ control - OD_{600}\ of\ negative\ control} x\ 100\%$$

2.8. Expression of DarA

Our first approach to heterologously express *darA* as a His-tagged version was unsuccessful. Therefore, a Maltose Binding Protein (MBP)-encoding sequence was fused to the upstream region. First, pET24c was digested with *Nde*I (New England Biolabs) overnight and dephosphorylated (Thermo Scientific) for 1 h. Then, the solution was directly purified using the Large fragment Zymo-research KIT yielding pET24c-*Nde*I as the 1st DNA fragment. The MBP sequence was amplified by PCR using pMAL-c5x (New England Biolabs) as template and primers Daro.MBP.Xa.F and Daro.MBP.Xa.R. Subsequently, it was purified by agarose gel purification, yielding an 1181 bp fragment. The gene *darA* was amplified by PCR using a synthetic codon-optimized DNA version (Sequence S1) with primers Daro.Xa.pro.F and Daro.Xa.pro.R, yielding a 217 bp DNA fragment. Then, these three fragments were ligated via Gibson Assembly (GA), yielding pET24c.MBP.Xa.darA. Thereafter, the GA reaction was transformed to *E.coli* top10 cells via electroporation and colonies were selected on LB agar plates containing 50 μg/mL kanamycin. Clones were picked and tested by PCR using primers Daro.MBP.Xa.F and Daro.Xa.pro.R, yielding a 1378 bp fragment consisting of *MBP* and *darA*. Verified plasmid DNA was isolated from an overnight *E. coli* TOP10/pET24c.MBP.Xa.darA in LB + kanamycin culture and subsequently transferred to *E.coli* BL21(DE3) as expression host.

The expression of the MBP-DarA fusion protein was performed following the manufacturer's instruction (NEB protein expression manual) with the following modifications. After induction, the culture was incubated at 18 °C at 180 rpm overnight. Purification was done using amylose resin (New England Biolabs). Then, the protein was concentrated using Amicon ultra-15 columns (Merck).

2.9. Trypsin digestion

Trypsin digestion of MBP-DarA was performed following manufacturer's instruction (Promega). Protein solution (50 μL) was mixed with the same volume of 100 mM ammonium bicarbonate. Cysteine reduction was done by addition of 1 μL 500 mM DTT and incubation at 56 °C for 30 min. Thereafter, the sample was cooled down to room temperature and 2 μL of 750 mM iodoacetamide were added. Incubation was done in the dark at room temperature for 30 min. 5 μL of trypsin (1 μg/μL) were added, and the sample was incubate at 37 °C overnight. Thereafter, 0.5% (vol/vol) TFA was added to the solution and 1 μL of the sample was applied onto pH paper to verify a pH < 2. The sample was then centrifuged in a tabletop centrifuge at max speed for 2 min. Clear supernatant was transferred to a new 1.5 mL Eppendorf tube. Further, C18 material was added to a 200 μL pipette tip (stage tip) and the column was equilibrated with 20 μL MeOH (centrifugation at 1000×*g*, 2 min). Then, 20 μL solution B (0.5% formic acid in 80/20% MeCN/water) was added to the stage tip and centrifugation was done at 1000×*g*, 2 min. 20 μL solution A (0.5% formic acid in 5/95% MeCN/water) were added to the stage tip and centrifugation was done at 1000×*g*, 2 min. The last step was repeated). Next, the trypsin-digested sample was added to the stage tip and centrifuged at 1000×*g* for 4 min. The stage tip was then washed with 20 μL solution A and centrifuged again at 1000×*g*, 2 min. The stage tip was then moved to a new 1.5 mL Eppendorf tube and 100 μL of solution B were added, centrifuged at 1000×*g*, 2 min; Then the resulting elution fraction was sent for LCMS analysis.

2.10. Quantification of darA and darE expression level

Overnight cultures of *E. coli* Rosetta™(DE3) strains carrying pZW-ADC3, pZW-ADC5, pZW-ADC7, and pZW-ADC8 were diluted 100-fold with fresh LB supplemented with kanamycin and chloramphenicol, and cultivated at 30 °C at 180 rpm shaking to reach an OD_{600} of 0.4 ± 0.05. Then, IPTG was added (final concentration 0.5 mM) and the cultures were incubated at 30 °C at 180 rpm shaking overnight.

7.5 mL samples were taken from the induced overnight culture for RNA isolation using the Quick-RNA Fungal/Bacterial Miniprep Kit (Zymo Research) according to the manufacturer protocol with modifications as following. The 7.5 mL samples were distributed to 5 tubes, each containing 1.5 mL culture for better cell lysis. The samples were combined again during the Zymo-Spin™IICRColumn centrifugation step. As the final step, RNA was eluted in 50 μL of nuclease-free water.

25 μL of isolated RNA was then treated with 5 units of DNase I (Zymo Research) in 50 μL reaction volume (40 mM Tris-HCl, pH 8.0, 10 mM NaCl, 6 mM $MgCl_2$, and 10 mM $CaCl_2$) for 45 min at room temperature, and the reaction was stopped by incubation at 65 °C for 10 min. Thereafter, 50 μL of 5 M ammonium acetate was added and mixed well before 300 μL of 99% ethanol were added. The mixture was incubated overnight at −20 °C for RNA precipitation. Then, it was centrifuged at 12000×*g*, 15 min, 4 °C and the supernatant was removed. The pellet was washed with 99% ethanol, 70% ethanol and subsequently air-dried. The RNA pellet was dissolved in 40 μL of nuclease-free water, then the concentration and purity was measured using a microvolume spectrophotometer (Eppendorf BioSpectrometer®).

cDNA was synthesized using LunaScript® RT SuperMix Kit (New England Biolabs) from 1 μg RNA for each sample according to manufacturer's protocol. The synthesis reaction was incubated as follows: 25 °C for 2 min, 55 °C for 15 min, 95 °C for 1 min. cDNA was directly used in the qPCR reaction and the same procedure was also done for the no-reverse transcriptase (no-RT) control.

For qPCR, the primer pair A-F3 and A-R3 was designed to target *darA* and the primer pair E-F3 and E-R3 was designed for *darE*. Primer pairs amplified 108 bp and 212 bp fragments, respectively. *rrsA* was used as reference gene, amplified by the primer pair rrsA-F3 and rrsA-R3 to generate 105 bp amplicons (Zhou et al., 2011). Each sample was analyzed for *darA, darE, and rrsA* in 5-fold dilution series in triplicates. The qPCR reaction (10 μL volume) contained 1x Luna Universal qPCR Master Mix (New England Biolabs), 0.25 μM forward primer, 0.25 μM reverse primer, and 1 μL of cDNA. No-RT control was also done in 5-fold dilution series, and no template control (NTC) was analyzed in triplicates (qPCR plate scheme design is shown in Table S3). The qPCR was performed in StepOnePlus™instrument (Applied Biosystems) with SYBR®Green detection, using the following thermocycler run: 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C and 1 min at 60 °C. At the end of the run, a melt curve was measured by the following run: 15 s at 95 °C, 1 min at 60 °C, then increasing the temperature from 60 °C to

95 °C with a thermal ramp rate of 0.3 °C/s.

StepOne™Software v2.3 was used to analyze the qPCR data and to determine quantification cycle (C_q). C_q was then plotted against the cDNA dilution series to create a standard curve for relative quantification of each gene. The qPCR efficiency (E) was calculated by the following equation:

$$E = 10^{\left(-\frac{1}{slope}\right)}$$

Relative amount of each gene was then calculated by the following equation:

$$Relative\ amount = E^{-Cq}$$

The *darA:darE* transcript ratio was calculated from the relative amount of each gene.

3. Results

The starting point of this project was the functional prove of the DAR BGC in *E. coli* (Imai et al., 2019). At that time, the heterologous expression of the DAR BGC was done by cloning the genes *darABCDE* from *P. khanii* HGB1456, thereby removing all of the noncoding DNA, under the control of the *araC*-P_{BAD} in *E. coli* BW25113. Although this system allows tight regulation and high-level protein expression with inexpensive induction by L-arabinose (Guzman et al., 1995), the resulting DAR yield with this system was lower than that of the native producer strains *P. khanii* HGB1456 and *P. khanii* DSM3369. Therefore, we tried other promoters, including inducible and constitutive ones. The *gfp* reporter gene was cloned downstream of the DAR BGC to enable a fast and easy readout. It was observed that green fluorescence correlated with the inhibition zone against a DAR-sensitive test strain (Fig. S1). Out of four promoters, *i.e.* T5, J23101, T7-*lac* and prpB (the latter propionate-inducible), the T7-*lac* promoter proved to produce the highest DAR levels (Fig. S1). The T7 RNA polymerase system under control of the *lac* operator generally has a higher level of protein expression compared to the other systems, due to the selective target of the T7 RNA polymerase for the T7 promoter sequence (Studier and Moffatt, 1986).

The heterologous expression of DAR was quantified by HR-LCMS analysis. The HR-LCMS chromatogram of DAR showed two main peaks: (i) *m/z* 483.7089 (calcd.) that corresponds to $[M+2H]^{2+}$ and (ii) *m/z* 475.1956 (calcd.) that corresponds to $[M+H\text{-}NH_2]^{2+}$ (Fig. 3). Therefore, an EIC was generated from the combination of these two masses, and the peak area was integrated. The DAR concentration was calculated by interpolating the peak area on the DAR standard curve.

3.1. Selection of the heterologous producer strain

As a first step, the expression plasmid pZW-ADC3 was generated. The genes of the BGC were transferred into a 'streamlined' BGC, meaning all intergenic regions were removed, genes have the same orientation and are under the control of the same promoter, *i.e.* T7-*lac*. The construct was transferred into different common *E. coli* and one *Vibrio* strain as expression host: (i) *E. coli* BL21(DE3), (ii) *E. coli* BAP1, (iii) *E. coli* Rosetta™(DE3), and (iv) *Vibrio natriegens* Vmax™Express. *E. coli* BL21 (DE3) is the most widely used strain for protein expression that has a prophage carrying a chromosomally encoded T7 RNA polymerase under the control of *lac*UV5 (Studier and Moffatt, 1986). *E. coli* BAP1 is a derivative of *E. coli* BL21(DE3) that has been genetically modified to produce complex polyketides and non-ribosomal peptides (Pfeifer et al., 2001). Even though DAR is not synthetized by such a system, this BL21 (DE3) derivative was tested as well. *E. coli* Rosetta™(DE3) is a further derivative of *E. coli* BL21(DE3), which carries a plasmid that encodes tRNA genes for rare codons. *Vibrio natriegens* Vmax™Express is an optimized *Vibrio natriegens* strain for protein expression of genes regulated by the T7 promoter. It should be mentioned that *in silico* analysis revealed the presence of the DAR BGC in several *Vibrio* strains (Imai et al., 2019).

Fermentation vessels of all heterologous hosts were sampled from 1 to 8 days. The strain *E. coli* BL21(DE3) produced the least amount of DAR during the complete fermentation time, on average 5 mg/L were detected from the cell pellet and 4.3 mg/L from the medium (Fig. 4). On the other hand, DAR could barely be detected from the cell pellet of *V. natriegens* and could only be observed from the medium. The DAR yield from *V. natriegens* showed a clear increase from day 1 to day 4, with no further increase thereafter. Compared to the other two *E. coli* strains, the DAR yield from *V. natriegens* was the least, with a maximum yield of 11.6 mg/L that was reached on the fourth day of fermentation. *E. coli* BAP1 and *E. coli* Rosetta™(DE3) produced the similar amount of DAR on the first day, but the yield increment of *E. coli* Rosetta™(DE3) was higher. When the yield of DAR in the medium and the cell pellet was combined, we achieved the highest production by *E. coli* Rosetta™(DE3) (26.1 mg/L) on the seventh day of fermentation. Therefore, *E. coli* Rosetta™(DE3) was selected as the heterologous expression host for the DAR BGC in the following experiments.

It should be mentioned that in later experiments it was investigated if the difference in production yield between *E. coli* BL21(DE3) and Rosetta™(DE3) might be due to the rare codons present in the DAR BGC. Therefore, the construct pZW-ADC12 that carries codon-optimized versions of *darA* and *darE* (Sequence S1 and 2) was generated. This plasmid was introduced into the heterologous hosts *E. coli* BL21(DE3)

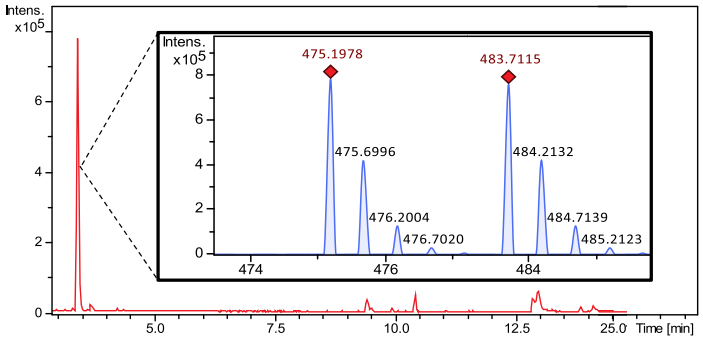


Fig. 3. DAR mass spectra. DAR peak in HR-LCMS ionized into two masses: *m/z* 483.7115 (theoretical value: *m/z* 483.7089, difference of 0.0026) that corresponds to $[M+2H]^{2+}$ and *m/z* 475.1978 (theoretical value: *m/z* 475.1956, difference 0.0022) that corresponds to $[M+H\text{-}NH_2]^{2+}$. For quantification, the EIC of these two masses was integrated.

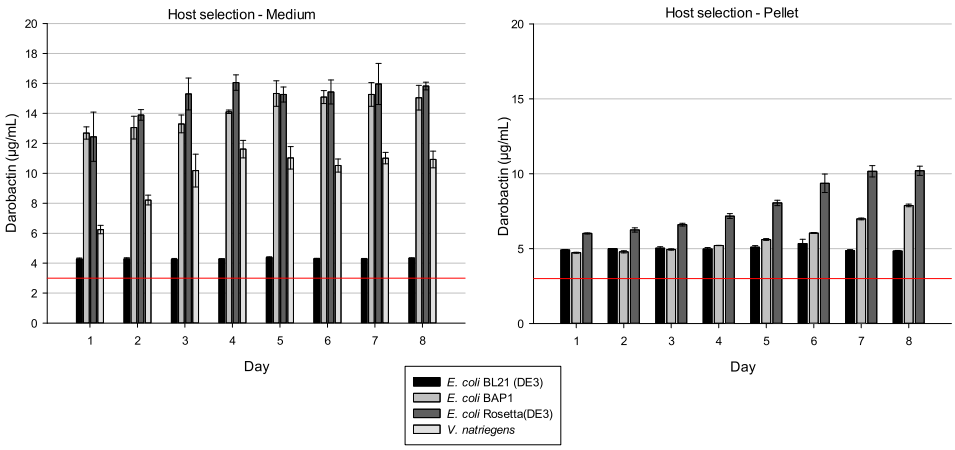


Fig. 4. DAR expression level from different heterologous hosts. Expression plasmid pZW-ADC3 was introduced to four different heterologous host, *i.e. E. coli* BL21(DE3), *E. coli* BAP1, *E. coli* Rosetta™(DE3), and *Vibrio natriegens* Vmax™Express. *E. coli* Rosetta™(DE3) was producing the highest yield in the medium and in the cell pellet. The red line marks the border of the linear range of the calibration curve. Therefore, values below the border are not shown. Data were collected from three biological replicates at day 1 to 8; error bars show standard deviation.

and *E. coli* Rosetta™(DE3), and DAR production was measured. Using this construct, a comparable amount of DAR could be detected in both strains, in the medium as well as in the cell pellet (Fig. S2).

3.2. Generation of an E. coli strain with increased DAR resistance level

Considering that *E. coli* cells are sensitive against the antibiotic to be biosynthesized, a producer strain with an increased resistance level towards DAR was generated. As observed by passaging experiments, DAR resistant strains could be generated. These strains accumulated mutations, *i.e.* more than one, in the *bamA* gene. A high-level resistant strain possessed three point mutations in *bamA*. By inserting these three point mutations into a clean background *E. coli* strain it was proven that these create a high-level resistant phenotype (Imai et al., 2019). Therefore, the mutations were introduced to *E. coli* Rosetta™(DE3) to create *E. coli* Rosetta™(DE3)(Dar^R). The DAR minimum inhibitory concentration (MIC) was determined for both strains to validate the results. Indeed, growth of the original *E. coli* Rosetta™(DE3) strain was inhibited at 2 µg/mL and the *E. coli* Rosetta™(DE3)(Dar^R) strain was able to grow at DAR concentrations >64 µg/mL. Here, it has to be noted that the MIC determination was done for *E. coli* strains under standard conditions, using an OD_{600} of 0.001, while the heterologous expression of DAR was induced when the OD_{600} has reached 0.4–0.6. Therefore, the parent strain without an increased DAR resistance level only had a survival rate of 18% (no growth was visible by eye), when 6 µg/mL DAR were introduced at an OD_{600} of 0.001. However, the strain could maintain 57% growth when the same amount of DAR was added to a culture with an OD_{600} of 0.5 (Fig. S3).

Plasmid pZW-ADC3 was then introduced to *E. coli* Rosetta™(DE3) (Dar^R), and the DAR production was compared to *E. coli* Rosetta™(DE3). However, the original *E. coli* Rosetta™(DE3) strain could produce almost two-fold DAR compared to the strain that has increased resistance to DAR (Fig. 5).

We observed that during the cultivation period the strain *E. coli* Rosetta™(DE3)(Dar^R) was not growing as well as *E. coli* Rosetta™(DE3). To test if the difference in DAR production is due to impaired growth, growth curves of these strains were generated. This revealed that the *E. coli* strain with increased resistance to DAR had an impaired growth compared to its parent strain (Fig. 6). The *E. coli* Rosetta™(DE3) strain had a significantly longer log phase (up to 54 h after inoculation), while the DAR resistant strain has already started its stationary phase at 25 h after inoculation. This means that the parent strain reached a cell density twice as high as the DAR resistant strain. In addition to that, the *E. coli* strains that carry the empty pRSFDuet™-1 vector control were cultivated, and the growth curve showed a similar trend, i.e. growth of *E. coli* Rosetta™(DE3)(Dar^R) was more hindered compared to *E. coli* Rosetta™(DE3) (Fig. S4).

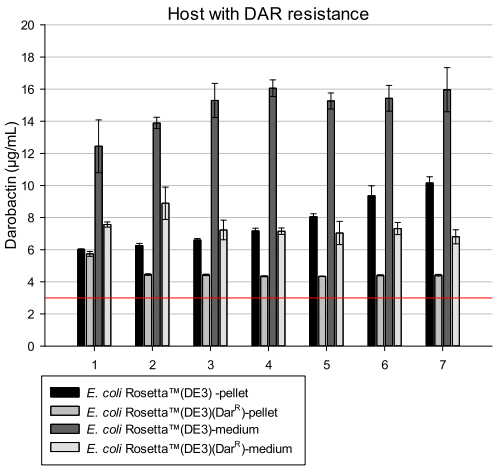


Fig. 5. DAR production in *E. coli* Rosetta™(DE3) with DAR resistance. Heterologous expression of the streamlined DAR BGC in *E. coli* Rosetta™(DE3)/pZW-ADC3 and *E. coli* Rosetta™(DE3)(Dar^R)/pZW-ADC3. The original *E. coli* Rosetta™(DE3) strain produced about two-fold more DAR as *E. coli* Rosetta™(DE3)(Dar^R). Data were collected from three biological replicates at day 1 to 7; error bars show standard deviation. The red line is the linearity limit of the DAR standard curve.

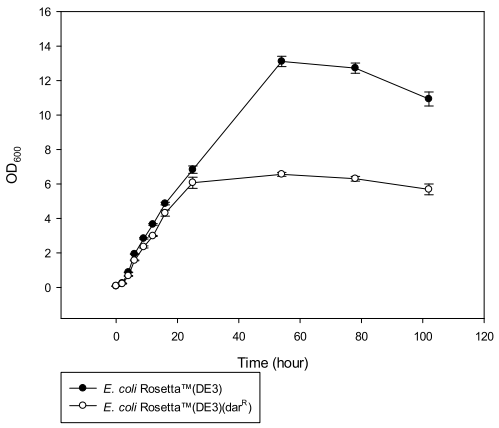


Fig. 6. Growth curve of *E. coli* Rosetta™(DE3) and *E. coli* Rosetta™(DE3) (Dar^R). Optical density (OD) was measured at 600 nm to compare the growth of *E. coli* Rosetta™(DE3) and *E. coli* Rosetta™(DE3)(Dar^R). The *E. coli* strain with increased resistance to DAR has a shorter log phase compared to its parent strain and does not reach the same OD. Data were collected from three biological replicates at the time points given; error bars show the standard deviation.

3.3. Comparison of different DAR BGCs

It was observed before that DAR production yields of *Photorhabdus khanii* DSM3369 were two-fold higher compared to *P. khanii* HGB1456. Furthermore, *in silico* analysis revealed the presence of DAR BGCs in *Yersinia* strains (Imai et al., 2019). Therefore, we cloned the native DAR BGC from *P. khanii* HGB1456, *P. khanii* DSM3369 and *Yersinia frederiksenii* ATCC33641, yielding pZW-ADC5, pZW-ADC6 and pZW-YerA4 respectively. All constructs were transferred into the selected heterologous host, *E. coli* Rosetta™(DE3).

In the heterologous expression system, the production yield difference between the DAR BGCs from both *Photorhabdus* strains were not so prominent. Both strains produced over 20 mg/L already after one-day cultivation and the strain carrying pZW-ADC5 reached the highest yield on the fourth cultivation day, with a total of 27.7 mg/L. Meanwhile, DAR production in the strain carrying pZW-ADC6 was in the same range and remained steady on average of 24.9 mg/L throughout the third to seventh cultivation day (Fig. 7). On the other hand, *E. coli* Rosetta™(DE3) carrying pZW-YerA4 produced only a total of 10.3 mg/L DAR on the second day of cultivation, which was 2.7-fold less, compared to the maximum yield of pZW-ADC5.

A comparison of the DAR BGCs from both *Photorhabdus* strains on DNA sequence level revealed 99% identity between both BGCs. The most evident difference between the two BGCs is an additional 136 bp intergenic region between the *darA* and *darB* genes in *P. khanii* DSM3369 (Fig. S5). To test if this intergenic region will have an impact on DAR production levels, we compared the native DAR BGC from *P. khanii* HGB1456 (pZW-ADC5) to the streamlined BGC, in which the intergenic region was removed (pZW-ADC3). In general, DAR yield is higher in medium than in pellet (Fig. 7). In the cell pellet, DAR production from the host with pZW-ADC3 was accumulating throughout the cultivation period with the most significant difference to pZW-ADC5 occurred on the seventh day. In medium, pZW-ADC5 produced better than pZW-ADC3 and achieved its highest production peak on the fourth day. Overall, pZW-ADC5 could produce more DAR in a shorter period compared to pZW-ADC3 and was used for further experiments.

3.4. Influence of the transporter-encoding genes darBCD

The three genes *darB, darC* and *darD* are coding for subunits of an ABC transporter. To answer the question whether these transporter genes play an additional role to the biosynthesis of DAR and to define the minimum DAR BGC, these genes were removed from the expression construct. Therefore, pZW-ADC9, a construct that carries only *darA* and *darE* from *P. khanii* DSM3369, was created. This experiment showed that without *darBCD*, DAR could still be produced. DAR accumulation in the medium of pZW-ADC6 was higher (±50%) than in pZW-ADC9. However, even though pZW-ADC6 possessed the transporter genes, the accumulation of DAR in the pellet was slightly higher than in the strain carrying pZW-ADC9 (Fig. 8).

The genes *darBCD* encode for proteins that compose a tripartite efflux pump (TEP), which is suspected to work together with the outer

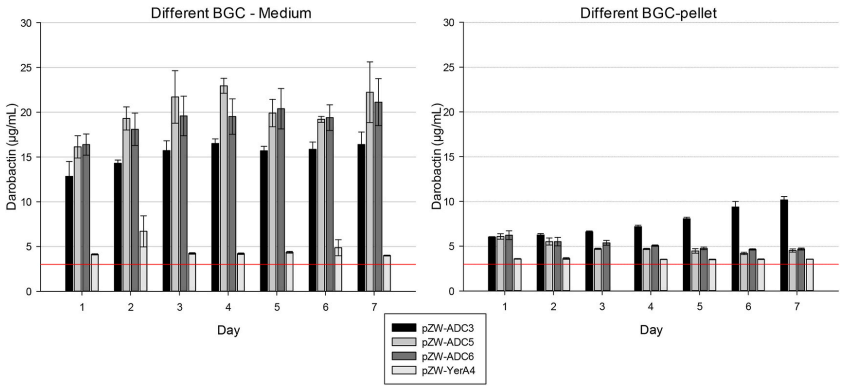


Fig. 7. Bar diagram showing DAR production of different BGCs in *E. coli* Rosetta™(DE3). Heterologous expression of the streamlined DAR BGC from *P. khanii* HGB1456 (pZW-ADC3), the native BGC from *P. khanii* HGB1456 (pZW-ADC5), the native BGC from *P. khanii* DSM3369 (pZW-ADC6), and the native BGC from *Y. frederiksenii* ATCC33641 (pZW-YerA4) in *E. coli* Rosetta™(DE3). Data were collected from three biological replicates at day 1 to 7; error bars show standard deviation. The red line is the linearity limit of the DAR standard curve. Values below this line are not given. Therefore, the data point of the third day of pZW-YerA4 was not converted to concentration. Expression of the native BGCs from *P. khanii* HGB1456 and of *P. khanii* DSM3369 were in the same range. Heterologous expression of the *Yersinia*-derived BGC resulted also in DAR production, although in much lower amount.

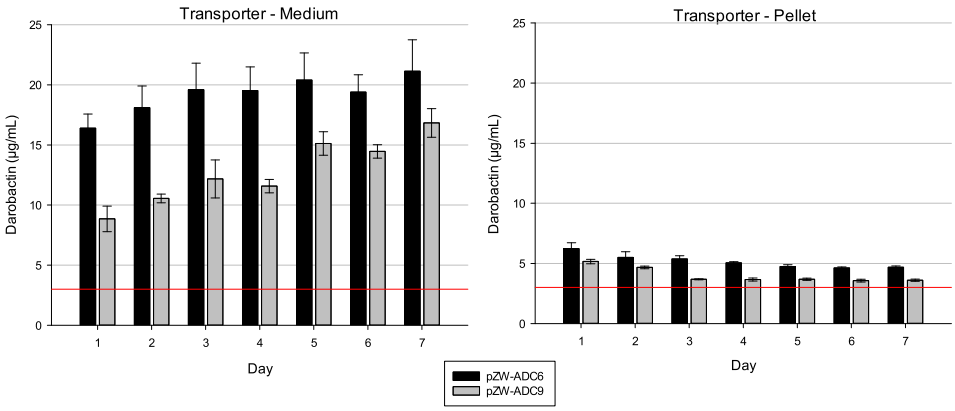
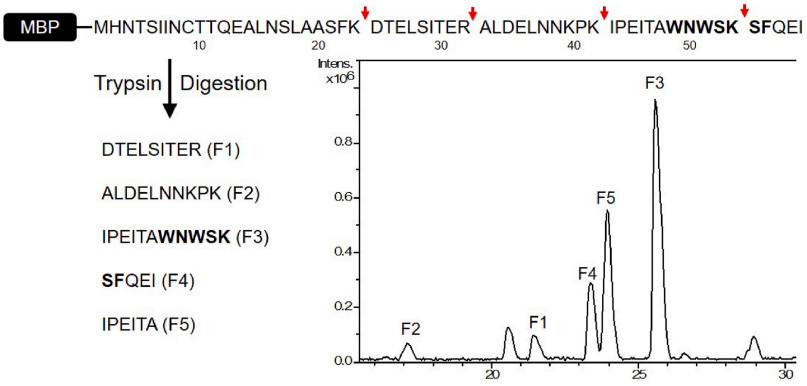


Fig. 8. Comparison of DAR production from constructs that carry transporter genes and without transporter genes. pZW-ADC6 is a construct with the native DAR BGC from *P. khanii* DSM3369, and pZW-ADC9 is constructed from pZW-ADC6 omitting *darBCD*. Red line is the linearity limit of DAR standard curve. Data were collected from three biological replicates at day 1 to 7; error bars show standard deviation. DAR was detectable outside of the cells without *darBCD*.

membrane efflux protein TolC. To investigate if TolC is involved in the export of DAR into the medium, a *tolC* knockout strain (*E. coli* Δ*tolC*) was created and the plasmid pZW-ADC5 was transferred into this host. However, there was no significant difference detectable (i) in DAR production and (ii) in the ratio between the intra- and extracellular concentration. (Fig. S6).

3.5. A dedicated peptidase is not part of the BGC

To obtain DAR from the precursor DarA, it has to be modified and cleaved twice, upstream and downstream of the core heptapeptide. DarE introduces the modification, *i.e.* the formation of the bicyclic ring structure, and proteolytic processing is usually achieved by one or more peptidase(s). However, we did not detect any dedicated peptidase within the DAR BGC. Therefore, it was in question whether the peptidic cleavage is also catalyzated by DarE, or if a peptidase from the host strain can catalyze this proteolysis. To get insights into this, DarA alone was heterologously expressed and purified in *E. coli*. An His-tagged version of DarA could not be obtained. This is commonly observed in heterologous expression of RiPP precursor proteins, most probably due to proteolytic instability. Fusion of a MBP (Maltose Binding Protein) to the N-terminal side improved stability and enabled purification using an



Fragment	Peptide Sequence	Ionization	Theoretical Mass (*m/z*)	Observed Mass (*m/z*)	Error (ppm)
F1	DTELSITER	$[M+2H]^{2+}$	532.2669	532.2689	3.7575
F2	ALDELNNKPK	$[M+2H]^{2+}$	571.3142	571.3163	3.6757
F3	IPEITAWNWSK	$[M+2H]^{2+}$	672.8510	672.8545	5.2017
F4	SFQEI	$[M+H]^{+}$	623.3035	623.3022	2.0857
F5	IPEITA	$[M+H]^{+}$	643.3661	643.3675	2.1760

Fig. 9. Scheme and LCMS chromatogram of trypsin-digested MBP-DarA Red arrow represent trypsin-cutting sites; the core peptide of DarA is in bold. The table shows the fragments including calculated and observed *m/z* ratio.

amylose resin. Then, the purified MBP-DarA was digested with trypsin and analyzed by LCMS (Fig. 9).

In addition to the predicted trypsin-digested fragments (F1, F2, F3 and F4), the LCMS analysis showed another abundant fragment (F5). This is in accordance with a part of F3 that was cleaved exactly at the position where the core peptide should be released from the leader peptide. Since only DarA was introduced into the *E. coli* host strain, it can be concluded that this proteolytic activity relies on a peptidase from the host and not by any of DarBCDE. The cleavage of the C-terminal amino acids might be also catalyzed by an *E. coli* peptidase. However, these end-standing amino acids were not observed. The peptidase(s) processing DarA remain elusive and by this experiment, the option that the peptide is catalyzing a self-cleavage cannot be ruled out. Furthermore, to test if the last three C-terminal amino acids are essential for heterologous expression it was envisaged if these can be omitted. Therefore, we analyzed the expression constructs carrying natural DarA, or a variant in which the last three amino acids were removed. There were no significant changes in production observed (Fig. S8).

3.6. Integration of an additional darA copy

Since *darA* encodes for the precursor peptide that will be modified to DAR, it was investigated whether having two copies of *darA* will improve the DAR yield. A second copy of the *darA* gene was introduced to pZW-ADC3 and pZW-ADC5, creating pZW-ADC3.2 and pZW-ADC5.2. If the intergenic region between *darA* and *darB* was not present in the expression construct, indeed more DAR was produced when a second *darA* copy was introduced (pZWADC3.2). The increase of DAR yield due to the addition of the second *darA* copy was evident in the cell pellet as well as in the medium. On the third day of cultivation, twice the amount of DAR could be detected in the medium of *E. coli* Rosetta™(DE3) pZW-ADC3.2 in comparison to pZW-ADC3. On the other hand, if the native BGC as it is present in *Photorhabdus* was used and an additional *darA* gene was added to this construct (pZW-ADC5.2), the maximum DAR yield that could be obtained was 17.5 mg/L. This is about two third of the maximum yield of *E. coli* Rosetta™(DE3) pZW-ADC5 that reached 27.7 mg/L (Fig. 10).

3.7. Integration of native darA upstream regions

It was observed that in the streamlined construct an additional *darA* copy could double the yield; while in the construct based on the native BGC, this effect was not so prominent. To get an additional data point, in a next step the upstream region *darA* was integrated into the expression construct. To study the effect of putative regulatory elements present in this region concerning heterologous DAR production, two additional constructs were created. The *in silico* analysis of the region up- and downstream *darA* with bacterial promoter prediction tools, *e.g.* BPROM (Solovyev and Salamov, 2011) and CNNprom (Umarov and Solovyev, 2017) revealed several putative transcriptional factor binding sites (TFBS) (Fig. S9) and housekeeping sigma factor binding sites (RpoD16 and RpoD17, Gruber and Gross, 2003). Furthermore, the presence of LexA and FNR (transcriptional repressor at non-stress and aerobic condition, respectively; Butala et al., 2008 and Unden and Schirawski, 1997) binding sites between *darA* and *darB* were detected. Beside LexA, a H-NS (global transcriptional silencer of genes with high AT content; Dorman, 2004) binding site was predicted 393 bp upstream of *darA* (Fig. S9).

Hence, 200 bp and 605 bp located upstream of *darA* were added to pZW-ADC5 yielding pZW-ADC7 and pZW-ADC8, respectively. Interestingly, the DAR production was doubled in pZW-ADC7 in comparison to pZW-ADC5. In contrast, including a 605 bp long region upstream of *darA* in the expression construct (pZW-ADC8) resulted in a DAR production yield half as much as pZW-ADC5 (Fig. 11).

3.8. Quantification of darA and darE transcription

From these experiments, including upstream regions or additional gene copies, it became clear that the *darA* transcript level should have an impact on DAR production. Hence, quantification of *darA* and *darE* transcript level was aimed by qPCR experiments. This revealed that pZW-ADC3, which has no intergenic region, had a *darA*:*darE* transcription ratio of 6.89. Meanwhile pZW-ADC5 that has an intergenic region between *darA* and *darB* showed a ratio of 23.56 and produced higher DAR levels than pZW-ADC3.

Next, all constructs that have the intergenic region between *darA* and *darB*, but carry different lengths of the *darA* upstream region (pZW-ADC5, pZW-ADC7, and pZW-ADC8) were compared. Also here, total DAR production indicated a relation to the *darA*:*darE* transcript ratio. Plasmid pZW-ADC7 that produced the highest amount of DAR (29.7 mg/L) had the highest *darA*:*darE* transcript ratio of 37.42. On the other hand, pZW-ADC8, which resulted in the lowest DAR production among the three constructs, had the lowest *darA*:*darE* transcript ratio of 17.05 (Fig. 12). The overall transcript level of *darA* had no influence of the

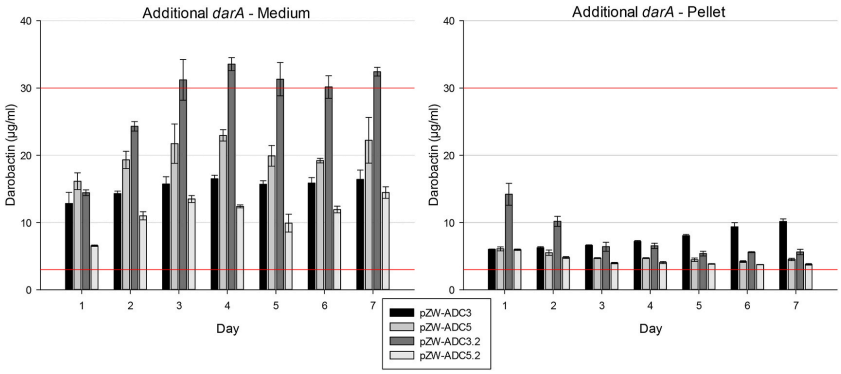


Fig. 10. DAR production in the medium and cell pellet for strains carrying an additional *darA* copy. Bar diagram showing the DAR production from *E. coli* Rosetta™(DE3) that carries streamlined DAR BGC with a single (pZW-ADC3) and double copy of *darA* (pZW-ADC3.2), as well as native DAR BGC from *P. khanii* HGB1456 with a single (pZW-ADC5) and double copy of *darA* (pZW-ADC5.2). Red line is the linearity limit of DAR standard curve. Data were collected from three biological replicates at day 1 to 7; error bars show standard deviation. Double copy of *darA* boosted DAR production in the streamlined DAR BGC, but halved the DAR production from the DAR BGC with intergenic region between *darA* and *darB*.

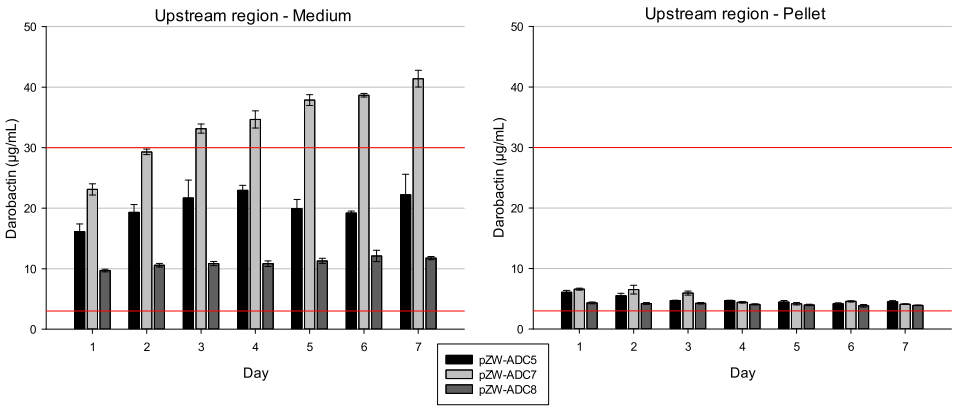


Fig. 11. Comparison of constructs carrying different length of *darA* upstream regions of *Photorhabdus khanii* HGB1456. Bar diagram depicted DAR production from *E. coli* Rosetta™(DE3) that carries DAR BGC without upstream region (pZW-ADC5), with 200 bp upstream region (pZW-ADC7), and with 605 bp upstream region (pZW-ADC8). Data were collected from three biological replicates at day 1 to 7; error bars show standard deviation. The red line is the linearity limit of the DAR standard curve. It can be seen that the 200 bp upstream region improves heterologous expression.

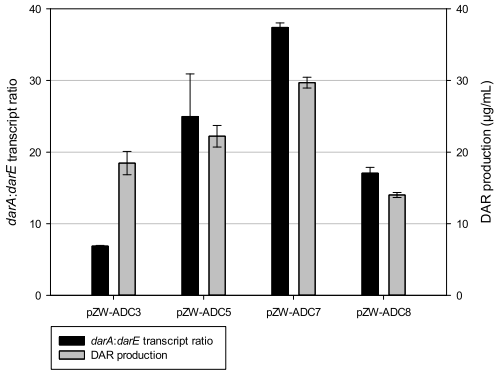


Fig. 12. Ratio of *darA*:*darE* transcripts from expression constructs with different non-coding regions in relation to their DAR production. Construct pZW-ADC7, which gave the highest DAR yield, showed the highest *darA*:*darE* ratio. Transcripts of *darA* and *darE* were quantified by qPCR in triplicates and DAR production was measured from three biological replicates; error bars show standard deviation.

observed DAR yield in these experiments.

4. Discussion

DAR was recently discovered as a natural product most promising for further development as antibiotic lead structure (Imai et al., 2019). It was hard to spot the initial activity of the compound in classic overlay assays, since it seemed that the product of interest was only poorly expressed. For further investigation of this compound, we aimed to increase the production yield to enable biotechnological studies, e.g. derivatization of the compound, in the future. Therefore, a heterologous approach was pursued.

The originally identified producer strains, *i.e. Photorhabdus* species, are Proteobacteria of the order Enterobacterales. Therewith, *Vibrio natriegens* and especially *E. coli* strains emerged as the first option, since these strains are Enterobacterales as well and are extensively studied and established molecular tools exist. *E. coli* as a promising host for heterologous expression of *Photorhabdus*-derived natural products was exemplified before, *e.g.* by the expression of luminmycin (Bian et al., 2012) and anthraquinone (Zhou et al., 2019) from *P. luminescens*, as well as phototemtide A from *P. temperata* (Zhao et al., 2020). Among the host strains tested, the result showed that *Vibrio natriegens* was outperformed by *E. coli* strains, especially Rosetta™(DE3) which produced DAR the best (Fig. 4). An explanation why *E. coli* Rosetta™(DE3) produced higher amounts of DAR in comparison to other related *E. coli* strains might be the presence of the pRARE plasmid, which carries the genes for rare tRNA. The *darE* gene sequence from *P. khanii* HGB1456 has 33 codons that are rarely used in *E. coli.* The result that utilizing codon-optimized *darA* and *darE* resulted in similar production yields of *E. coli* BL21(DE3) and Rosetta™(DE3) supports this hypothesis (Fig. S2). *E. coli* as an established workhorse of modern microbiology and biotechnology represents an excellent choice for future optimization, since reproducibility and process speed can be improved, downstream processing is facilitated by a reduced chemical background and simultaneously regulatory concerns are avoided if the project advances to GMP studies.

4.1. bamA mutation that confers resistance to DAR impaired the growth of heterologous host

A point that has to be considered to select a heterologous host is that the product DAR is efficiently killing *E. coli* cells. Therefore, in preceding studies, DAR-resistant strains were generated and the resistance was traced back to mutations in the *bamA* gene. Hence, the reported three point mutations that confer high-level DAR resistance were introduced into the *bamA* gene of *E. coli* Rosetta™(DE3) to enable higher production titers without killing the producer strain in batch fermentation. BamA is an essential component of the β-barrel assembly machinery (BAM) complex that catalyze the assembly of outer membrane protein (OMP) in Gram-negative bacteria. Modifications in BamA affects the viability of the cell (Konovalova et al., 2017; Wu et al., 2005). The three point mutations in the *bamA* gene were also affecting the *E. coli* virulence, thereby indicating the reduced fitness of the mutant strain (Imai et al., 2019). In the present study, it was observed that the generated mutant strain showed impaired growth compared to the strain with wild type *bamA*. In consequence, the lower viability of the DAR-resistant strain and the lower OD reached during fermentation, resulted in lower DAR

production. This result indicates that having a well-functioning BamA for viability is more beneficial in DAR production than having a mutated BamA through DAR resistance. Therefore, we selected *E. coli* Rosetta™(DE3) with native *bamA* over *E. coli* Rosetta™(DE3) with mutated *bamA*.

4.2. Minimum DAR BGC consists of only darA and darE

In a next step, it was evaluated if the transporter-encoding genes *darBCD* are essential for the heterologous DAR expression or if *darA*, encoding the precursor peptide and *darE*, encoding the radical SAM modification enzyme (Imai et al., 2019), are sufficient. By deletion of *darBCD,* DAR production was not abolished. However, the yield was lower (1.5-fold) than the one reached with the construct including transporter genes (pZW-ADC6). Most interestingly, DAR was also present in the medium, even without the transporter–encoding genes. On the one hand, this clearly defines the minimum DAR BGC, which consists of only *darA* and *darE.* On the other hand, it became clear that in *E. coli,* DAR is present outside the cell even without the specific heterologous transporter genes *darBCD*.

In general, in addition to *darBCD*, which encode for an ABC transporter permease protein, a periplasmic adaptor protein, and an ATP binding domain of the inner membrane ABC transporter, respectively, an outer membrane efflux protein, *e.g.* TolC, would be needed to transport a molecule to the outside. This is for example the case when molecules toxic to the cell are transported out as a form of antibiotic resistant mechanism in *E. coli* (Greene et al., 2018). However, a *ΔtolC E. coli* mutant strain did not exhibit any difference in production level to the parent *E. coli* strain with a functional *tolC* gene, and maintained the ratio between intra- and extracellular DAR. This indicates that DAR is not actively transported to the outside of the cell by a DarBCD-TolC complex in the heterologous expression system. It can be assumed that in a natural system the transporters are needed to guarantee a regulated efflux. However, for the heterologous system, the presence of the transporters is not essential, but the transport system utilized is not known. Beside TolC, three other outer membrane efflux protein (OEP) were identified in *E. coli, i.e.* YjcP, YohG, and YlcB (Sulavik et al., 2001). In addition to that, *E. coli* harbors TolC-independent efflux systems that belong to the major facilitator superfamily (MFS), small multidrug resistance family (SMR), multidrug and toxin extrusion (MATE) family, cation diffusion facilitator family (CDF), and proteobacterial antimicrobial compound efflux family (PACE) (Slipski et al., 2018). Due to the size of DAR, it can be assumed that passive diffusion via porins does not take place.

4.3. The role of unspecific peptidase in the maturation of DarA to DAR

To release DAR from modified DarA by DarE, the modified core peptide in DarA needs to be cleaved. However, we did not find any dedicated peptidase from DAR BGC. Furthermore, we could show that DarBCD were not involved in the DAR biosynthesis. It can be hypothesized that either DarE might also catalyze a peptidolytic cleavage, or a peptidase from the host is involved. By expression of DarA fused to MBP without DarE and purification through an amylose column, we could observe that a fraction of DarA was already cleaved exactly before the core peptide (Fig. 9). This result clearly favors the second explanation that the cleavage of DarA is not catalyzed by DarE. Most likely, a peptidase from the host can catalyze the reaction and DarA can be cleaved at this position even though no modification by DarE took place. This cleavage seems to be specific, since neither ions corresponding to peptides with the sequence IPEIT, nor to IPEITAW were detected. However, the SF (Ser-Phe) fragment, which should be liberated by the cleavage behind the core peptide sequence, was not detected. Either this cleavage is not catalyzed by a peptidase in *E. coli* that can be acquired, or a peptidase cuts first upstream of the core peptide. Then, this fragment will be lost during purification of MBP-fused peptides. Anyway, the fact that the end-standing amino acids can be omitted in expression constructs will facilitate the generation of derivatives, since a cleavage site must not be considered.

4.4. Native DAR BGC from different bacterial strains and the effect of non-coding region in DAR BGC to DAR production

Besides the BGC from *P. khanii* HGB1456 and *P. khani* DSM3369, a *Yersinia*-derived DAR BGC was chosen as an alternative source to be expressed in *E. coli* Rosetta™ (DE3), since the GC content is similar to *Photorhabdus* species. The DAR production result showed that DAR BGC from *Photorhabdus* strains resulted in a more than 3-fold higher DAR production compared to DAR BGC from *Y. frederiksenii* ATCC33641 (Fig. 7). The DAR BGC of the two *Photorhabdus* strains is highly similar; it revealed 99.49% similarity on nucleotide level with 99% query coverage. Despite this high similarity, the intergenic region between *darA* and *darB* of the two *Photorhabdus* strains differs in length. *P. khanii* DSM3369 carries a 136 bp insertion that is lacking in *P. khanii* HGB1456 (Fig. S5). This difference seems to have a positive effect in heterologous production of DAR from *P. khanii* HGB1456 BGC. Meanwhile, the DAR BGC from *Y. frederiksenii* ATCC33641 showed only 63.64% identity with 31% query covered on nucleotide level compared to the DAR BGC from *P. khanii* HGB1456. Hence, the difference in DAR production from these different sources could be attributed to the difference in DAR BGC sequence.

Anyway, for the regulation of a BGC that consists mainly of a precursor and an enzyme necessary for the modification of the first a fine-tuned mechanism to adjust the ration between these two can be expected. In the bottromycin BGC for example, the precursor peptide is enhanced specifically by a regulatory protein that does not affect other genes in the BGC (Vior et al., 2020). In the DAR BGCs, there is an intergenic non-coding region between *darA* and *darB*. It can be speculated that this region carries recognition sequences, which could play a role in regulating the expression level of *darA* and the genes located downstream of it, thus affecting DAR production. The importance of the transcript ratio of *darA* and *darE* was shown by comparison of pZW-ADC3 in which the intergenic region was removed and pZW-ADC5 as a control. Plasmid pZW-ADC5, which encodes production of higher DAR levels than pZW-ADC3, has a higher *darA:darE* ratio than pZW-ADC3 (Fig. 12). This indicates that the intergenic region plays a role in adjusting this transcript ratio and the higher the ratio, the better the DAR production is. A further experiment to increase the *darA:darE* ratio, in which an additional *darA* copy was introduced to pZW-ADC3 (pZW-ADC3.2), increased DAR production as well (Fig. 10). However, introducing an additional *darA* copy in pZW-ADC5 (pZW-ADC5.2 that carries the intergenic region) decreased DAR production (Fig. 10). It seemed like the presence of the intergenic region regulates *darA:darE* transcription level and the system was put out of action by the artificial introduction of an additional *darA* copy.

Interestingly, the 200 bp region upstream of *darA* was quite conserved in the *Photorhabdus* genus (94% identity and 70%–94% query coverage) and was therefore integrated into the expression construct, creating pZW-ADC7. In addition, the 605 bp upstream region of *darA* was also integrated in a construct, despite the fact that this region was less conserved throughout *Photorhabdus* strains carrying the BGC (94% identity and 24%–64% query coverage). Integration of the 200 bp upstream region significantly increased DAR production, while extending the integrated upstream region up to 605 bp diminished it (Fig. 11). Consequently, the effect of the region upstream of *darA* towards the *darA:darE* transcript level was also investigated. Indeed, pZW-ADC7 showed the highest *darA:darE* ratio observed in this study. Hence, it should be concluded that the non-coding region upstream of *darA* has a positive regulatory effect on transcription level and DAR production. Thereby, supporting the hypothesis that a higher *darA:darE* transcript level goes in line with higher DAR production. Within the 200 bp upstream region of *darA*, three open reading frames (ORFs) with the size of

52 aa, 44 aa and 28 aa, respectively, were detected. The first two ORFs were located in the opposite strand, while the last ORF was located on the same strand as the DAR BGC. The first two ORFs were only annotated as hypothetical proteins, while the 28 aa ORF had 92.9% identity to the partial sequence of *relE* (135 aa).

Several putative TFBS, housekeeping sigma factor binding sites (RpoD16 and RpoD17) were predicted up- and downstream *darA*. Regulation of a RiPP BGC by an RNA polymerase sigma factor (RpoD) was before reported for microcin "MccB17" (Duquesne et al., 2007). Furthermore, the presence of LexA and FNR binding sites between *darA* and the following cluster parts enabled transcription repression of downstream genes; thereby, increasing the *darA*:*darE* transcript ratio. In addition, the presence of an H-NS binding site 393 bp upstream of *darA* could indicate a downregulation of *darABCDE* transcription, as it was shown for the construct carrying this region. Future detailed analysis of DAR BGC regulation will support the further fine-tuning of the optimal transcript ratio for heterologous expression.

5. Conclusion

In summary, the heterologous expression system developed in this study forms the basis for future bioengineering approaches, which enable further exploitation of this compound class as promising drug lead structure for antibiotic development. Using this platform, generation of novel derivatives, *e.g.* with improved pharmaceutical properties and altered activity profile becomes feasible. Heterologous expression of DAR was optimized to reach a higher yield in a shorter period compared to the original producer strains. The optimization strategies included the use of different expression hosts, comparison of different DAR BGCs originating from different strains, as well as modifications of the BGC. The minimal BGC, consisting solely of *darA* and *darE* was determined and it was shown that to release DAR from the leader peptide, no designated DAR peptidase has to be co-expressed. A high *darA:darE* transcript ratio is necessary to obtain a high DAR yield. Currently, a heterologous production system using *E. coli* Rosetta™(DE3) is available that enables production of >30 mg/L DAR within 2 days.

Declaration of competing interest

ZGW, NB, KL and TFS are co-inventors on different patent applications related to this research.

Credit author statement

Zerlina G. Wuisan: Conceptualization, Methodology, Formal Analysis, Investigation, Writing – Original Draft, Visualization. **I Dewa M. Kresna:** Methodology, Investigation, Writing. **Nils Böhringer:** Methodology, Validation, Writing – Review and Editing. **Kim Lewis:** Resources, Writing – Review and Editing. **Till F. Schäberle:** Conceptualization, Writing – Review and Editing, Visualization, Supervision, Project administration, Funding acquisition.

Acknowledgements

ZGW and IDMK received financial support from German Academic Exchange Service (DAAD). The authors thank Ute Mettal (Justus-Liebig-University Giessen) for measuring the samples on UPLC-HRMS and Marius Spohn and Regina Zweigert (Branch of Bioresources of the Fraunhofer Institute for Molecular Biology and Applied Ecology) for lyophilization of the samples.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymben.2021.04.007.

References

Bian, X., Plaza, A., Zhang, Y., Müller, R., 2012. Luminmycins A–C, cryptic natural products from Photorhabdus luminescens identified by heterologous expression in Escherichia coli. J. Nat. Prod. 75, 1652–1655. https://doi.org/10.1021/np300444e.

Bouhired, S.M., Crüsemann, M., Almeida, C., Weber, T., Piel, J., Schäberle, T.F., König, G.M., 2014. Biosynthesis of phenylnannolone A, a multidrug resistance reversal agent from the halotolerant myxobacterium nannocystis pusilla B150. Chembiochem 15, 757–765. https://doi.org/10.1002/cbic.201300676.

Butala, M., Žgur-Bertok, D., Busby, S.J.W., 2008. The bacterial LexA transcriptional repressor. Cell. Mol. Life Sci. 66, 82. https://doi.org/10.1007/s00018-008-8378-6.

Crüsemann, M., Reher, R., Schamari, I., Brachmann, A.O., Ohbayashi, T., Kuschak, M., Malfacini, D., Seidinger, A., Pinto-Carbó, M., Richarz, R., Reuter, T., Kehraus, S., Hallab, A., Attwood, M., Schiöth, H.B., Mergaert, P., Kikuchi, Y., Schäberle, T.F., Kostenis, E., Wenzel, D., Müller, C.E., Piel, J., Carlier, A., Eberl, L., König, G.M., 2018. Heterologous expression, biosynthetic studies, and ecological function of the selective Gq-signaling inhibitor FR900359. Angew Chem. Int. Ed. Engl. 57, 836–840. https://doi.org/10.1002/anie.201707996.

Datsenko, K.A., Wanner, B.L., 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. Unit. States Am. 97, 6640–6645. https://doi.org/10.1073/pnas.120163297.

Davis, J.H., Rubin, A.J., Sauer, R.T., 2011. Design, construction and characterization of a set of insulated bacterial promoters. Nucleic Acids Res. 39, 1131–1141. https://doi.org/10.1093/nar/gkq810.

Dorman, C.J., 2004. H-NS: a universal regulator for a dynamic genome. Nat. Rev. Microbiol. 2, 391–400. https://doi.org/10.1038/nrmicro883.

Duquesne, S., Destoumieux-Garzón, D., Peduzzi, J., Rebuffat, S., 2007. Microcins, gene-encoded antibacterial peptides from enterobacteria. Nat. Prod. Rep. 24, 708–734. https://doi.org/10.1039/B516237H.

Flinspach, K., Westrich, L., Kaysser, L., Siebenberg, S., Gomez-Escribano, J.P., Bibb, M., Gust, B., Heide, L., 2010. Heterologous expression of the biosynthetic gene clusters of coumermycin A(1), clorobiocin and caprazamycins in genetically modified Streptomyces coelicolor strains. Biopolymers 93, 823–832. https://doi.org/10.1002/bip.21493.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345. https://doi.org/10.1038/nmeth.1318.

Greene, N.P., Kaplan, E., Crow, A., Koronakis, V., 2018. Antibiotic resistance mediated by the MacB ABC transporter family: a structural and functional perspective. Front. Microbiol. 9 https://doi.org/10.3389/fmicb.2018.00950.

Gruber, T.M., Gross, C.A., 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. Annu. Rev. Microbiol. 57, 441–466. https://doi.org/10.1146/annurev.micro.57.030502.090913.

Guzman, L.M., Belin, D., Carson, M.J., Beckwith, J., 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. 177, 4121–4130.

Hegemann, J.D., Zimmermann, M., Zhu, S., Klug, D., Marahiel, M.A., 2013. Lasso peptides from proteobacteria: genome mining employing heterologous expression and mass spectrometry. Pept. Sci. 100, 527–542. https://doi.org/10.1002/bip.22326.

Imai, Y., Meyer, K.J., Iinishi, A., Favre-Godal, Q., Green, R., Manuse, S., Caboni, M., Mori, M., Niles, S., Ghiglieri, M., Honrao, C., Ma, X., Guo, J.J., Makriyannis, A., Linares-Otoya, L., Böhringer, N., Wuisan, Z.G., Kaur, H., Wu, R., Mateus, A., Typas, A., Savitski, M.M., Espinoza, J.L., O'Rourke, A., Nelson, K.E., Hiller, S., Noinaj, N., Schäberle, T.F., D'Onofrio, A., Lewis, K., 2019. A new antibiotic selectively kills Gram-negative pathogens. Nature 576, 459–464. https://doi.org/10.1038/s41586-019-1791-1.

Konovalova, A., Kahne, D.E., Silhavy, T.J., 2017. Outer membrane biogenesis. Annu. Rev. Microbiol. 71, 539–556. https://doi.org/10.1146/annurev-micro-090816-093754.

Linares-Otoya, L., Linares-Otoya, V., Armas-Mantilla, L., Blanco-Olano, C., Crüsemann, M., Ganoza-Yupanqui, M.L., Campos-Florian, J., König, G.M., Schäberle, T.F., 2017. Identification and heterologous expression of the kocurin biosynthetic gene cluster. Microbiology 163, 1409–1414. https://doi.org/10.1099/mic.0.000538.

Ongley, S.E., Bian, X., Zhang, Y., Chau, R., Gerwick, W.H., Müller, R., Neilan, B.A., 2013. High-titer heterologous production in E. coli of lyngbyatoxin, a protein kinase C activator from an uncultured marine cyanobacterium. ACS Chem. Biol. 8, 1888–1893. https://doi.org/10.1021/cb400189j.

Paulus, C., Rebets, Y., Zapp, J., Rückert, C., Kalinowski, J., Luzhetskyy, A., 2018. New alpiniamides from streptomyces sp. IB2014/011-12 assembled by an unusual hybrid non-ribosomal peptide synthetase trans-AT polyketide synthase enzyme. Front. Microbiol. 9 https://doi.org/10.3389/fmicb.2018.01959.

Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., Khosla, C., 2001. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. Science 291, 1790–1792. https://doi.org/10.1126/science.1058092.

Ruiz, N., Wu, T., Kahne, D., Silhavy, T.J., 2006. Probing the barrier function of the outer membrane with chemical conditionality. ACS Chem. Biol. 1, 385–395. https://doi.org/10.1021/cb600128v.

Shibui, T., Uchida, M., Teranishi, Y., 1988. A new hybrid promoter and its expression vector in Escherichia coli. Agric. Biol. Chem. 52, 983–988. https://doi.org/10.1080/00021369.1988.10868770.

Silhavy, T.J., Kahne, D., Walker, S., 2010. The bacterial cell envelope. Cold Spring Harb. Perspect. Biol. 2 https://doi.org/10.1101/cshperspect.a000414 a000414–a000414.

Slipski, C.J., Zhanel, G.G., Bay, D.C., 2018. Biocide selective TolC-independent efflux pumps in enterobacteriaceae. J. Membr. Biol. 251, 15–33. https://doi.org/10.1007/s00232-017-9992-8.

Solovyev, V., Salamov, A., 2011. Automatic annotation of microbial genomes and metagenomic sequences. In: Li, R.W. (Ed.), Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies. Nova Science Publishers, Hauppauge, NY, pp. 61–78.

Studier, F.W., Moffatt, B.A., 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 189, 113–130. https://doi.org/10.1016/0022-2836(86)90385-2.

Sucipto, H., Pogorevc, D., Luxenburger, E., Wenzel, S.C., Müller, R., 2017. Heterologous production of myxobacterial α-pyrone antibiotics in Myxococcus xanthus. Metab. Eng. 44, 160–170. https://doi.org/10.1016/j.ymben.2017.10.004.

Sulavik, M.C., Houseweart, C., Cramer, C., Jiwani, N., Murgolo, N., Greene, J., DiDomenico, B., Shaw, K.J., Miller, G.H., Hare, R., Shimer, G., 2001. Antibiotic susceptibility profiles of Escherichia coli strains lacking multidrug efflux pump genes. Antimicrob. Agents Chemother. 45, 1126. https://doi.org/10.1128/AAC.45.4.1126-1136.2001.

Umarov, R.K., Solovyev, V.V., 2017. Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks. PLoS One 12, e0171410. https://doi.org/10.1371/journal.pone.0171410.

Unden, G., Schirawski, J., 1997. The oxygen-responsive transcriptional regulator FNR of Escherichia coli: the search for signals and reactions. Mol. Microbiol. 25, 205–210. https://doi.org/10.1046/j.1365-2958.1997.4731841.x.

Vior, N.M., Cea-Torrescassana, E., Eyles, T.H., Chandra, G., Truman, A.W., 2020. Regulation of bottromycin biosynthesis involves an internal transcriptional start site and a cluster-situated modulator. Front. Microbiol. 11 https://doi.org/10.3389/fmicb.2020.00495.

Weinstock, M.T., Hesek, E.D., Wilson, C.M., Gibson, D.G., 2016. Vibrio natriegens as a fast-growing host for molecular biology. Nat. Methods 13, 849–851. https://doi.org/10.1038/nmeth.3970.

WHO publishes list of bacteria for which new antibiotics are urgently needed [WWW Document], n.d. https://www.who.int/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed–. (Accessed 3 May 2020).

Wu, T., Malinverni, J., Ruiz, N., Kim, S., Silhavy, T.J., Kahne, D., 2005. Identification of a multicomponent complex required for outer membrane biogenesis in Escherichia coli. Cell 121, 235–245. https://doi.org/10.1016/j.cell.2005.02.015.

Yamanaka, K., Reynolds, K.A., Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Nizet, V., Dorrestein, P.C., Moore, B.S., 2014. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc. Natl. Acad. Sci. U.S.A. 111 https://doi.org/10.1073/pnas.1319584111, 1957–1962.

Zhao, L., Vo, T.D., Kaiser, M., Bode, H.B., 2020. Phototemtide A, a cyclic lipopeptide heterologously expressed from Photorhabdus temperata Meg1, shows selective antiprotozoal activity. Chembiochem 21, 1288–1292. https://doi.org/10.1002/cbic.201900665.

Zhou, K., Zhou, L., Lim, Q. 'En, Zou, R., Stephanopoulos, G., Too, H.-P., 2011. Novel reference genes for quantifying transcriptional responses of Escherichia coli to protein overexpression by quantitative PCR. BMC Mol. Biol. 12, 18. https://doi.org/10.1186/1471-2199-12-18.

Zhou, Q., Bräuer, A., Adihou, H., Schmalhofer, M., Saura, P., Grammbitter, G.L.C., Kaila, V.R.I., Groll, M., Bode, H.B., 2019. Molecular mechanism of polyketide shortening in anthraquinone biosynthesis of Photorhabdus luminescens. Chem. Sci. 10, 6341–6349. https://doi.org/10.1039/C9SC00749K.

Supporting Information

Optimization of Heterologous Darobactin A Expression and Identification of the Minimal Biosynthetic Gene Cluster

Zerlina G. Wuisan[1], I. Dewa Kresna[1], Nils Böhringer[1,4], Kim Lewis[2], Till F. Schäberle[1,3,4,*]

[1]Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, 35392 Giessen, Germany.
[2]Antimicrobial Discovery Center, Northeastern University, Department of Biology, Boston, Massachusetts, USA 02115.
[3]Branch for Bioresources of the Fraunhofer Institute for Molecular Biology and Applied Ecology, 35394 Giessen, Germany.
[4]German Center for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Giessen, Germany.
*Corresponding author. *E-mail address*: till.f.schaeberle@agrar.uni-giessen.de

Table S1. Bacterial strains used in this study

Strain	Source
E. coli BAP1	(Pfeifer et al., 2001)
E. coli BAP1 ΔtolC:: aac(3)-IV (Apr^R)	This study
E. coli BL21(DE3)	Merck KGaA (Darmstadt, Germany)
E. coli Rosetta™(DE3)	Merck KGaA (Darmstadt, Germany)
E. coli Rosetta™(DE3)(Dar^R)	This study
E. coli TOP10	Invitrogen (California, USA)
P. khanii DSM3369	DSMZ (Braunschweig, Germany)
P. khanii HGB1456	(Imai et al., 2019)
V. natriegens Vmax™	Synthetic Genomics (California, USA)
Y. frederiksenii ATCC 33641	LGC Standards GmbH (Wesel, Germany)

Table S2. Primers used in this study

Name	Sequence (5' → 3')
ADC-F7	GTATAAGAAGGAGATATACAATGCATAATACCTTAAATGA
ADC-R1	AATAGCATTCATTTATGGCTCTCCTTTTAAATTTCCTGGAAGCTTT
ADC-F2	AAAGCTTCCAGGAAATTTAAAAGGAGAGCCATAAATGAATGCTATT
ADC-R7	TGCTCAGCGGTGGCAGCAGCTTACGCCGCGATGGTTTGTT
ADC-F9	GTATAAGAAGGAGATATACACCGATGATATACTTTTATTA
ADC-F10	GTATAAGAAGGAGATATACAAAACTCACTACATTTTAGCG
ADC9-R	ATGGGGATTATTGTGTCCATTTATGGCTACCGTTCCTTAC
ADC9-F	GTAAGGAACGGTAGCCATAAATGGACACAATAATCCCCAT
pro.CO.F	GTATAAGAAGGAGATATACAATGCACAACACCTCTATCAT
pro.CO.R	ATCGGGATGATGGTGTCCATTTAGATTTCCTGGAAAGATTTAGAC
RS.CO.F	GTCTAAATCTTTCCAGGAAATCTAAATGGACACCATCATCCCGAT
RS.CO.R	TGCTCAGCGGTGGCAGCAGCAGCAGCGATGGTCTGTTTGA
YerA-F4	GTATAAGAAGGAGATATACAAGGAGGTTATTTAAATGGAGAATTAT
YerA-R4	TGCTCAGCGGTGGCAGCAGCTCAGTAGATACTGGCGATAT
GFP-F	AACAAACCATCGCGGCGTAAGAGGAGAAATTAAGCATGCG
GFP-R2	TAGTTATTGCTCAGCGGTGGTTATTTGTATAGTTCATCCATGCC
ADC5-GFP-F2	TGGATGAACTATACAAATAACCACCGCTGAGCAATAACTA
ADC5-GFP-R	CGCATGCTTAATTTCTCCTCTTACGCCGCGATGGTTTGTT
T5-F	TTGCTTTCAGGAAAATTTTTCTGTATAATAGATTCCCCATCTTAGTATATTAG TT
T5-R	GAATCTATTATACAGAAAAATTTTCCTGAAAGCAACAACGCAATTAATGTAA GTT
J23101-F	TTTACAGCTAGCTCAGTCCTAGGTATTATGCTAGCCCCATCTTAGTATATTA GTT
J23101-R	GCTAGCATAATACCTAGGACTGAGCTAGCTGTAAACAACGCAATTAATGTAA GTT
J23101-F2	GATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTG
J23101-R2	CATTGTATATCTCCTTCTTATACTTAACTAATATACTAAGATGGG
J23101-ADC5-GFP-R	ACTCATTAGGCACCGGGATCTCGACCGATGCCCTTGAGAG
pPro-F	AACTTACATTAATTGCGTTGCCGGATAAAGCGTTCGCG
pPro-R	AACTAATATACTAAGATGGGAGCCCATCCTTTGTTATCAA
bamA-recF	ACTATCTGGATCGCGGTTATGC
bamA-recR	TTCACAGCAGTCTGGATACGAG
A-F3	TCAAGAAGCACTCAATTCTC
A-R3	GTGATCTCAGGGATCTTAGG
E-F3	GGCCAACATCCCATAAAGTC
E-R3	ACTTCCTCCAGGATCATCAC
rrsA-F3	CTCTTGCCATCGGATGTGCCCA
rrsA-R3	CCAGTGTGGCTGGTCATCCTCTCA
TolCKO-F	ATGAAGAAATTGCTCCCCATTCTTATCGGCCTGAGCCTTTATTCCGGGGAT CCGTCGACC
TolCKO-R	TCAGTTACGGAAAGGGTTATGACCGTTACTGGTGGTAGTGTGTAGGCTGGA GCTGCTTC
Daro.MBP.Xa.F	TTTAAGAAGGAGATATACATATGAAAATAGAAGAAGGTAAACTGGTAATCTG G
Daro.MBP.Xa.R	CCTTCCCTCGATCCCGAGGTTGTTGTTATTGTTATTGT
Daro.Xa.pro.F	ACCTCGGGATCGAGGGAAGGATGCACAACACCTCTATCATCAAC
Daro.Xa.pro.R	TCCACCAGTCATGCTAGCCATTAGATTTCCTGGAAAGATTTAGACCAGT

Table S3. qPCR plate scheme design

Template		A-F3 and A-R3 (A)	E-F3 and E-R3 (E)	rrsA-F3 and rrsAR3 (16S)	No-RT control		
					A	E	16S
cDNA dilution	1						
	1:5						
	1:25						
	1:125	96-well plate					
	1: 625						
	1:3125						
	1:15625						
NTC	H_2O						

Table S4. Normalized amount of *rrsA*, *darA*, *and darE* transcript

	Slope			Cycle threshold (C_q)*			darA:darE	DAR (µg/mL)
	rrsA	darA	darE	rrsA	darA	darE		
pZW-ADC3	-3.436	-3.455	-3.420	6.173	9.083	11.840	6,88389	18,46441
pZW-ADC5	-3.402	-3.3364	-3.398	6.857	7.750	12.603	23,56114	22,21879
pZW-ADC7	-3.436	-3.428	-3.362	6.733	8.507	13.63	37,42173	29,69882
pZW-ADC8	-3.344	-3.330	-3.336	8.630	8.300	12.527	17,05235	14,00512

*Cycle threshold is the average of technical triplicates

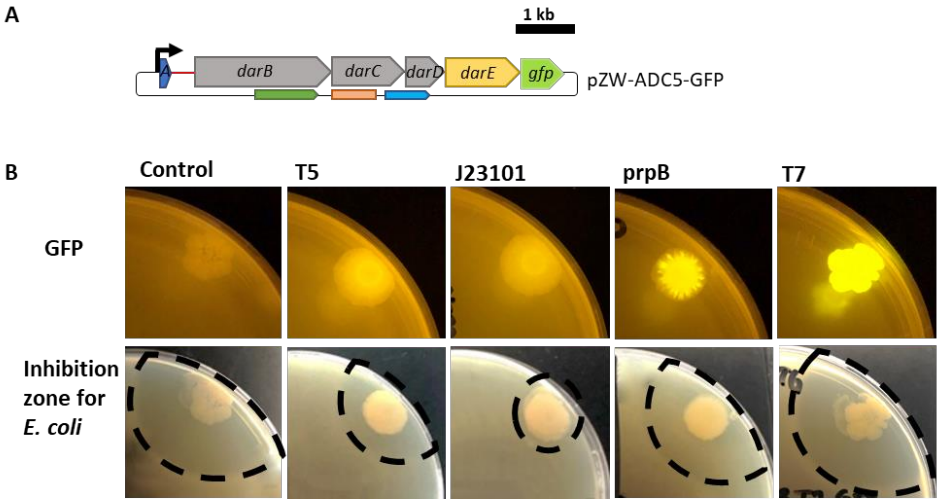


Fig. S1: Qualitative analysis of different promoters used. A, schematic vector map; the DNA sequence encoding for GFP was inserted downstream of the DAR BGC to serve as reporter. Black arrow shows the position where the respective constitutive promoter (T5 and J23101) and inducible promoter (T7 and prpB) was integrated. **B**, Photographs of agar plates with the respective clone. GFP was observed using a blue-light transilluminator at 470 nm to compare the strength of each promoter. T7 promoter showed the brightest fluorescence, followed by prpB promoter. The inhibition zones (highligted by the dashed lines) against *E. coli* MG1655 BamA6 also confirmed that T7 was the strongest promoter, followed by prpB, T5, and J23101. As a control without *gfp*, pZW-ADC5 was used.

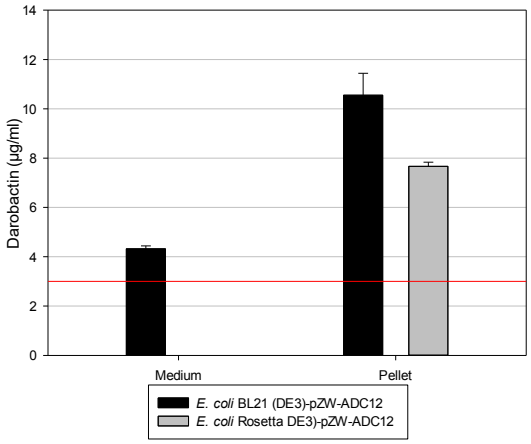


Fig. S2: DAR production by codon-optimized minimum DAR BGC in *E. coli* BL21(DE3) and *E. coli* Rosetta™(DE3). *E. coli* BL21(DE3)-pZW-ADC12 produced more DAR than *E. coli* Rosetta™(DE3)-pZW-ADC12. The red line marks the lower linearity border of the calibration curve. Therefore, values below the border are not shown. Data were collected from three biological replicates from one-day culture; error bars show standard deviation.

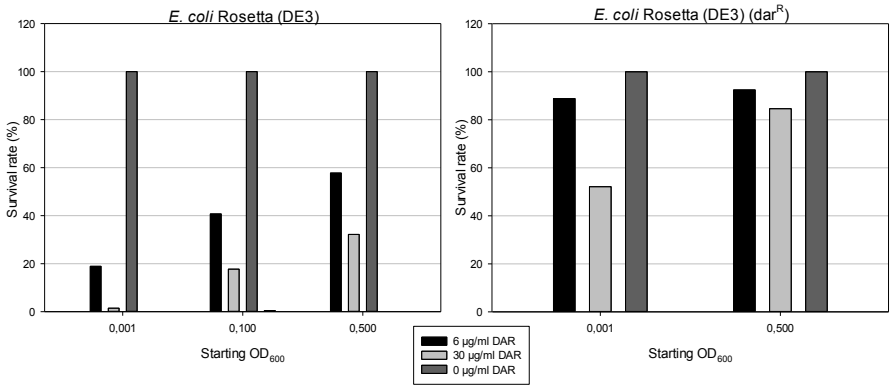


Fig. S3: *E. coli* Rosetta™(DE3) and *E. coli* Rosetta™(DE3) (dar^R) survival rate in the presence of different concentration of DAR. The growth of *E. coli* strains without DAR treatment was considered as 100% survival rate. Survival rate of both strains increased with the increase of the starting OD_{600}. *E. coli* Rosetta™(DE3) (dar^R) had higher survival rate compared to its parent strain.

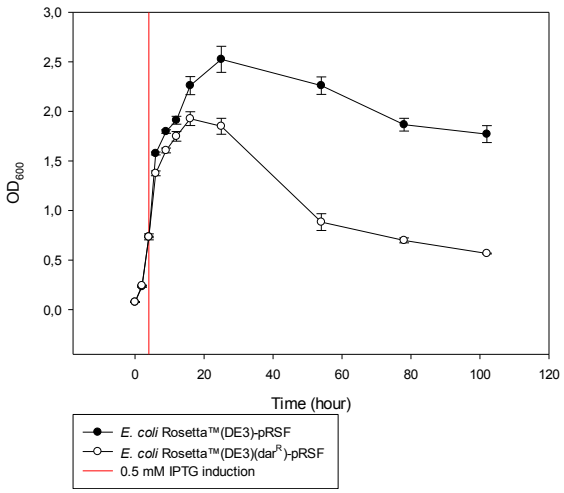


Fig. S4: Growth curve of *E. coli* Rosetta™(DE3) and *E. coli* Rosetta™(DE3)(Dar^R) that carries empty expression vector, pRSFDuet™-1. Red line refers to the time

point when the cultures were induced with 0.5 mM IPTG induction. Data were collected from three biological replicates; error bars show standard deviation.

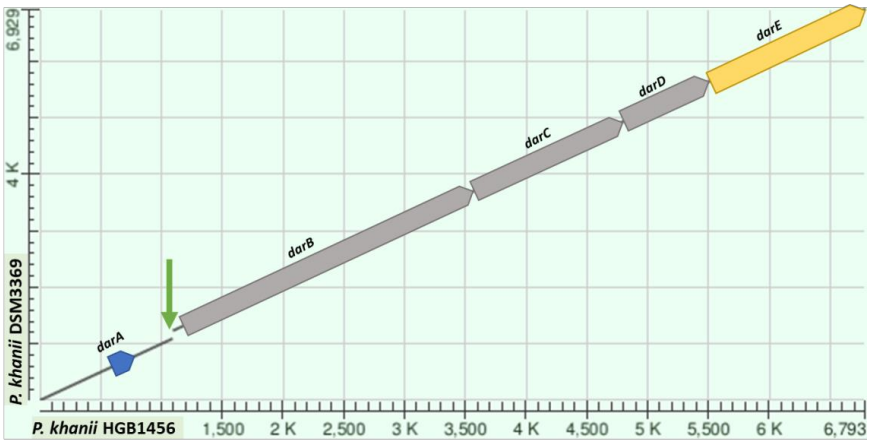


Fig. S5: Dot plot depicting alignment of DAR BGC from *P. khanii* HGB1456 and *P. khanii* DSM3369. DAR BGC including 605 bp upstream region and all intergenic regions from *P. khanii* HGB1456 (6793 bp) was aligned to *P. khanii* DSM3369 (6929 bp) using blastn suite-2 sequences program (National Library of Medicine, Bethesda, MD, USA; https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Zhang et al., 2000). On DNA level, they have 99% identity. DAR BGC from *P. khanii* DSM3369 has a 136 bp longer intergenic region between *darA* and *darB*, marked by green arrow.

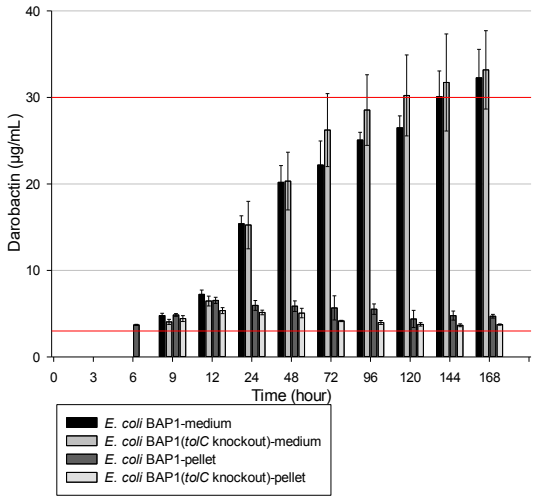


Fig. S6: Comparison of DAR production in *E. coli* strains with or without TolC. *E. coli* BAP1 and *E. coli* BAP1 Δ*tolC* were transformed by pZW-ADC5, and DAR production was compared. DAR is produced in the same range by both hosts, with the same intra-/extracellular ratio. The red line indicates the linearity limit of the DAR standard curve. Data were collected from three biological replicates at mentioned time point; error bars show standard deviation.

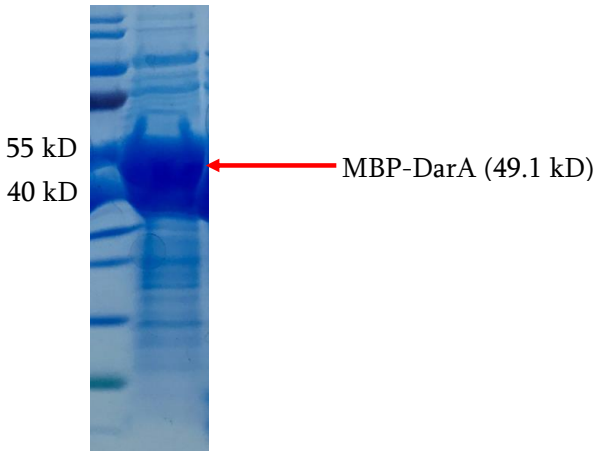


Fig. S7: SDS-page analysis of MBP-DarA. The red arrow shows the target protein (MBP-DarA), which fits to the calculated molecular weight of 49.1 kDa.

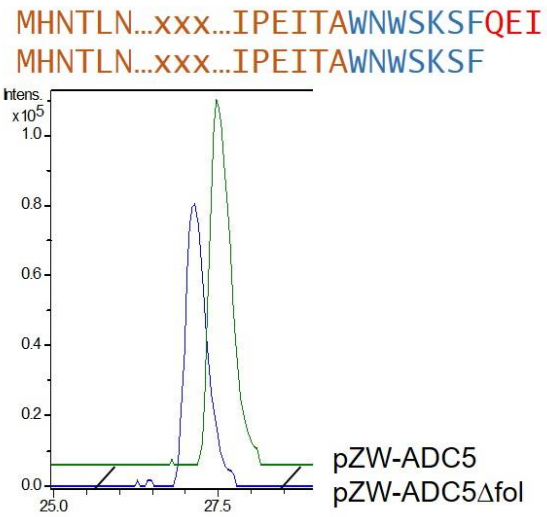


Fig. S8: DAR production from DarA precursor with or without follower peptide. In the shown amino acid sequence the core peptide is indicated in blue and the follower (if present) in red. The chromatogram shows similar production for both expression constructs.

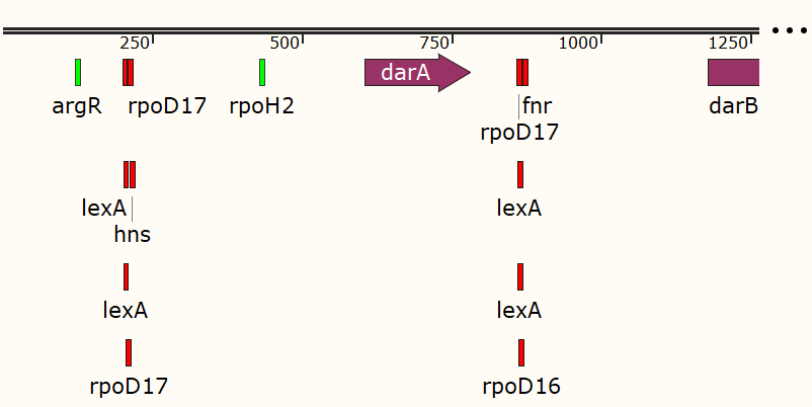


Fig. S9: Transcriptional factor binding site prediction. *In silico* analysis of non-coding region upstream and downstream of *darA* using BPROM (Solovyev and Salamov, 2011) and CNNprom (Umarov and Solovyev, 2017). Red indicates transcription factor binding site by both tools, and green indicates prediction by BPROM only.

Sequence S1. Codon optimized DNA sequence of *darA* from *Photorhabdus namnaonensis*

ATGCACAACACCTCTATCATCAACTGCACCACCCAGGAAGCTCTGAACTCTCTGGCTGCTTCTTT
CAAAGACACCGAACTGTCTATCACCGAACGTGCTCTGGACGAACTGAACAACAAACCGAAAATC
CCGGAAATCACCGCTTGGAACTGGTCTAAATCTTTCCAGGAAATCTAA

Sequence S2. Codon optimized DNA sequence of *darE* from *Photorhabdus namnaonensis*

ATGGACACCATCATCCCGATCAAATACCTGAACGCTGACGAATCTTCTATCCTGAAAAAATCTCC
GAAAATCAACTACCGTCAGCTGGCTTGCCGTATCATCGGTGAAATCCCGGCTGAAAAAATCCTG
GACGACGACGAACTGGCTCTGTACAACGAAGAAATCGGTATCCACTTCTCTCCGGAAATCATCAA
CGCTAACAAACTGGTTGTTGTTGTTAAAGCTACCCGTCTGTGCAACCTGCGTTGCACCTACTGCC
ACTCTTGGGCTGAAGGTAAAGGTAACACCCTGACCTTCTTCAACCTGATGCGTTCTATCCACCGT
TTCCTGTCTATCCCGAACATCAAACGTTTCGAATTCGTTTGGCACGGTGGTGAAGTTACCCTGCT
GTCTGTTAACTACTTCAAAAAACTGATCTGGCTGCAGGAACAGTTCAAAAAACCGGACCAGGTTA
TCACCAACTCTGTTCAGACCAACGCTGTTAACATCCCGGAAGACTGGCTGGTTTTCCTGAAAGGT
ATCGGTATGGGTGTTGGTATCTCTGTTGACGGTATCCCGGAAATCCACGACTCTCGTCGTCTGG
ACTACCGTGGTCGTCCGACCTCTCACAAAGTTGCTGCTGGTATGAAAAAACTGCGTTCTTACGGT
ATCCCGTACGGTGCTCTGGTTGTTGTTGACCGTGACGTTTACGAATCTAACATCGAAAAAATGCT
GTCTTACTTCTACGAAATCGGTCTGACCGACATCGAATTCCTGAACATCGTTCCGGACAACCGTT
GCCAGCCGGGTGACGACCCGGGTGGTTCTTACATCACCTACCACAACTACATCAACTTCCTGTC
TAACGTTTTCCGTGTTTGGTGGAACGACTACCAGGACAAAATCAACATCCGTCTGTTCCACGGTT
TCATCGACTCTATCAAATCTTCTCAGAAAAAAATCTCTGACTGCTACTGGGCTGGTAACTGCTCTC
AGGAAATCATCACCCTGGAACCGAACGGTACCGTTTCTGCTTGCGACAAATACGTTGGTGCTGA
AGGTAACAACTACGGTTCTATCATCGACAACGACCTGGGTCACCTGCTGATCAAATCTAACACCA
ACAAAAACCACCTGAAAGAAGAAATCGAATCTTACGAAAAAATGCACCAGTGCAAATGGTTCCAC
CTGTGCAACGGTGGTTGCCCGCACGACCGTGTTACCAACCGTAAACACAACCCGAACTACAACG
ACTTCTGCTGCGGTACCGGTGGTCTGCTGGAAATCATCAAACAGACCATCGCTGCTTAA

References

Imai, Y., Meyer, K.J., Iinishi, A., Favre-Godal, Q., Green, R., Manuse, S., Caboni, M., Mori, M., Niles, S., Ghiglieri, M., Honrao, C., Ma, X., Guo, J.J., Makriyannis, A., Linares-Otoya, L., Böhringer, N., Wuisan, Z.G., Kaur, H., Wu, R., Mateus, A., Typas, A., Savitski, M.M., Espinoza, J.L., O'Rourke, A., Nelson, K.E., Hiller, S., Noinaj, N., Schäberle, T.F., D'Onofrio, A., Lewis, K., 2019. A new antibiotic selectively kills Gram-negative pathogens. Nature 576, 459–464. https://doi.org/10.1038/s41586-019-1791-1

Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., Khosla, C., 2001. Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of E. coli. Science 291, 1790–1792. https://doi.org/10.1126/science.1058092

Solovyev, V., Salamov, A., 2011. Automatic Annotation of Microbial Genomes and Metagenomic Sequences, in: Metagenomics and Its Applications in Agriculture, Biomedicine and Environmental Studies (Ed. R.W. Li). Nova Science Publishers, Hauppauge, NY, pp. 61–78.

Umarov, R.K., Solovyev, V.V., 2017. Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks. PLOS ONE 12, e0171410. https://doi.org/10.1371/journal.pone.0171410

Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A Greedy Algorithm for Aligning DNA Sequences. J. Comput. Biol. 7, 203–214. https://doi.org/10.1089/10665270050081478

6. Chapter V

The radical-SAM enzyme DarE catalyzing the formation of the intramolecular *C-C* and *C-O-C* rings in darobactin biosynthesis

This project is done in collaboration with the working groups of Kim Lewis, Kenichi Yokoyama, Yvain Nicolet and Till F. Schäberle.
A first manuscript from this project is currently under review and attached at the end of the chapter.

Summary: Here, the characterization of the radical SAM (RaS) enzyme DarE is presented. Insights how DarE is catalyzing the formation of C-C and C-O-C (ether) rings during darobactin biosynthesis are described. While C-C bond formation by RaS enzymes has been previously reported, to the best of our knowledge an ether bond formation was not reported until now. Furthermore, it is intriguing that both different rings are catalyzed by a single RaS enzyme. To analyze this *in vivo,* the substrate DarA linked to a maltose binding (MBP) protein was heterologously expressed with and without DarE. Subsequently, it was analyzed if DarA-MBP was modified. This was achieved by trypsin cleavage of the precursor substrate and subsequent analysis by LCMS.

Contribution: I Dewa Made Kresna designed and planed the experimental set up. He did literature research, planed and performed the experiments described in the first part of the chapter. He analyzed these results and put them into a discussion. He summarized the *in vivo* results in the style of a draft manuscript. These data were the precondition for the *in vitro* results mainly presented in the submitted manuscript.

Radical SAM enzyme catalyzing the formation of C-C and C-O-C rings in Darobactin

I Dewa M. Kresna[a,b], Nils Böhringer[a,b,c], Zerlina G. Wuisan[a,b,c], Till F. Schäberle[a,b,c]

[a] Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, Ohlebergsweg 12, 35392 Giessen, Germany

[b] Department of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Ohlebergsweg 12, 35392 Giessen, Germany

[c] German Center for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Giessen, Germany

Abstract

Radical *S*-adenosylmethionine (RaS) enzymes initiate a diverse set of radical reactions. The minimal biosynthetic gene cluster (BGC) of darobactin consists solely of the precursor-encoding *darA* and the RaS enzyme-encoding *darE*. Darobactin A (DAR) is a ribosomally synthesized and post-translationally modified peptide (RiPP) consisting of a 7 amino acids core peptide (W^{1}-N^{2}-W^{3}-S^{4}-K^{5}-S^{6}-F^{7}) with two intramolecular bonds – an ether bond (*C-O-C*) between W^{1} and W^{3} and a *C-C* bond between W^{3} and K^{5}. To the best of our knowledge, RaS enzymes catalysing two different biochemical reactions were not described before. Since DarE is the only modification enzyme in the BGC, it was hypothesized that DarE is responsible for the two different bonds formation. Within this work, we confirm the hypothesis by *in vivo* experiments. DarA was heterologously expressed in *E. coli* with or without co-expression of DarE. DarA was purified and analysed by HR UPLC-MS/MS. DAR was only observed when DarE was co-expressed. Modifications occurred before the core peptide was released from the DarA-leader.

Introduction

The bicyclic heptapeptide darobactin A (DAR) is a ribosomally synthesized, post-translationally modified peptide (RiPP). As a promising antibiotic, DAR is a further example for the exquisite biological activities and manifold chemical scaffolds of this growing natural product class. The underlying biosynthetic gene cluster (BGC) consists of the precursor peptide-encoding *darA*, the transport encoding genes *darBCD*, and *darE*. The latter is encoding a member of the radical S-adenosylmethionine (RaS) enzyme super-family, which catalyse installation of post-translational modifications on the precursor[1]. Most intriguingly, two

completely different modifications, i.e. formation of a *C-C* and of a *C-O-C* ether bond, were predicted to be catalyzed by DarE. The *C-C* bond between lysine and tryptophan (K-W crosslink) is precedent by the natural product streptide[2]. An example of a *C-O-C* bond formation was recently described; however, in this mechanism the oxygen in the formed ether bond is derived from the hydroxyl group of threonine, which is attached to the peptide backbone[3]. In contrast, the ether bond formed in DAR biosynthesis represents a novel W-W ether crosslink. Once the two connected rings are installed on the heptapeptidic core peptide (W^1-N^2-W^3-S^4-K^5-S^6-F^7), the structure is rendered a rigid β-strand conformation. Thereby, it mimics the recognition signal of native substrates and seals the open lateral gate of BamA[4]. To verify the involvement of DarE in modification of the precursor DarA, *in vivo* experiments were performed. DarA alone, or in combination with DarE, was heterologously expressed, purified, and processing of the precursor was analyzed by LCMS.

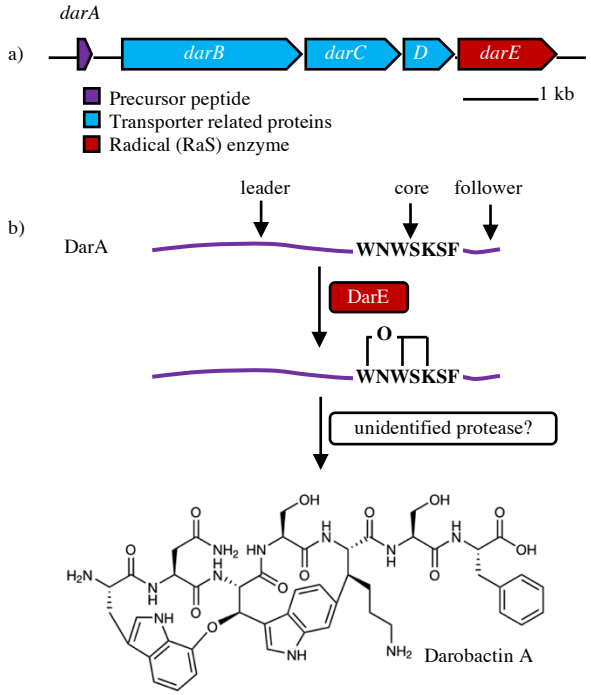


Figure 1. a) The BGC architecture of darobactin A and b) the proposed biosynthetic pathway of darobactin A

Result and Discussion

Under *in vivo* production conditions, only the completely modified DarA precursor, i.e. both intramolecular rings formed and trimmed to darobactin A, was observed. To get further insights

into the DarA modification, a maltose-binding protein (MBP) tag was fused to DarA and heterologous expression was done in *E. coli,* with or without co-expression of DarE. Thereafter, MBP-DarA was purified, digested with trypsin and analyzed by LCMS. It has to be noted that darobactin A was not cleaved by trypsin, even though a lysine, which is a trypsin recognition site, is present at position 5 of the core peptide. However, also after 24 h incubation with trypsin, darobactin A degradation was not detected by LCMS analysis. This could be attributed to the fact that the intramolecular bonds between the amino acid side chains result in a rigid bicyclic heptapeptide, as shown in the darobactin A crystal structure[4]. The rigid compact structure might be the reason that darobactin A does not represent a substrate for trypsin.

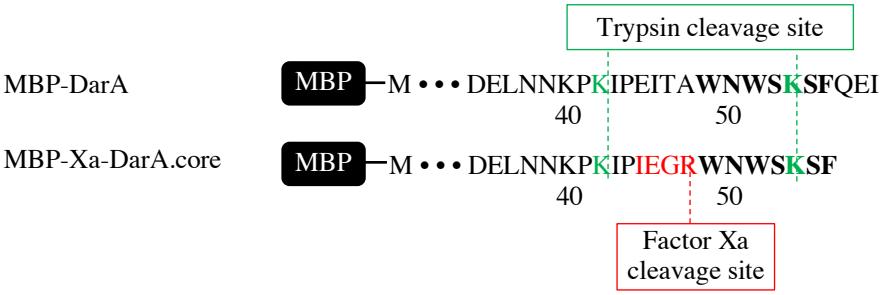
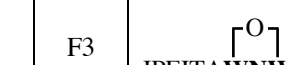
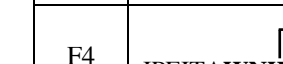
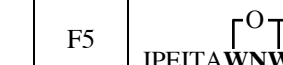
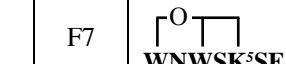


Figure 2. Amino acid sequence of DarA from MBP-DarA differs from MBP-Xa-DarA.core. Bold letters indicate the core peptide. Red and green letters indicate the factor Xa and trypsin recognition site, respectively. Red and green dotted lines represent the cleavage site of factor Xa and trypsin, respectively.

In the here used system, MBP-tagged DarA (MBP-DarA, Figure 2) variants that can be modified intracellularly were affinity purified from cell lysate in a first step, and secondly trypsin digestion was performed. Comparing the resulting peptide fragments detected from strains with or without DarE (F1, Table 1), in both the expected linear peptide between the trypsin cutting sites was present. Furthermore, the fragment that represents exactly the N-terminal rest of the leader peptide upstream of the core peptide (F2) was detected, indicating that this proteolytic step is not DarE catalyzed. Fragments of linear trypsin-digested peptides carrying either the *C-O-C* ring (F3), or the *C-C* ring (F4) were detected; thereby not indicating a clear consecutive order of the ring closures, and proving that trypsin digestion takes place if only one ring is formed. The *C-C* ring between the β-carbon of the K^5 residue and the W^3 indole ring does not prevent cleavage, which might be attributed to the fact that the covalent bond of the K^5 side chain to W^3 is not hindering proteolytic access to the backbone peptide bond.

Scanning the trypsin-digested MBP-DarAs for the presence of fragments with both rings attached (F5), did not yield any result (Table 1 and S1). This indicates that these fragments were already released from the leader peptide due to the proteolytic activity of a yet unidentified protease that is cutting before W^{1}. Hence, during amylose purification, these fragments are lost.

Table 1. The list of investigated fragments of DarA after treatment with trypsin (F1-5) or factor Xa (F6-7). Core peptide is marked with bold letters. The corresponding LCMS chromatograms are shown in Figure S1-6.

Fragment code	Investigated fragments	Description	Ionization	Co-expressed with DarE	
				Yes	No (control)
F1	IPEITA**WNWSK**^{5}	Non modified version	$[M+2H]^{2+}$	672.8554 *m/z* error: 5.77 ppm	672.8551 *m/z* error: 5.32 ppm
F2	IPEITA	Successful unidentified protease activity in the first site	$[M+H]^{+}$	643.3692 *m/z* error: 3.93 ppm	643.3675 *m/z* error: 1.29 ppm
F3		C-O-C ring modification and successful trypsin digestion after K^{5}	$[M+2H]^{2+}$	679.8434 *m/z* error: 3.31 ppm	-
F4		C-C ring modification and successful trypsin digestion after K^{5}	$[M+2H]^{2+}$	671.8470 *m/z* error: 4.92 ppm	-
F5		Full modification with unsuccessful trypsin digestion after K^{5} and successful unidentified protease activity in the second site	$[M+H]^{+}$ $[M+2H]^{2+}$ $[M+3H]^{3+}$	-	-
F6	**WNWSK^{5}SF**	Non modified version of a fragment from MBP-Xa-DarA.core after factor Xa digestion	$[M+2H]^{2+}$	477.7326 *m/z* error: 10.5 ppm	477.7335 *m/z* error: 12.35 ppm
F7		Full 2 rings attached fragment (DAR) from MBP-Xa-DarA.core after factor Xa digestion	$[M+2H]^{2+}$	483.7072 *m/z* error: 4.57 ppm	-

In consequence, if both rings are formed and completely modified DarA heptapeptide was liberated from the precursor, DAR should be observable in the cell or even the cultivation medium. Hence, cell and medium extracts were analyzed and indeed, the mass of DAR was detected (Figure 3). In contrast, a bicyclic peptide consisting of the DAR core and the follower

peptide was not observed, neither in cell, nor in medium crude extracts. This showed that even though no dedicated protease is located in the DAR BGC, the trimming to a heptapeptide is highly efficient, also in the heterologous *E. coli* system. Therefore, it was projected to alter the putative recognition site of the unknown protease to test if in that way a bicyclic DAR precursor attached to the leader can be identified. The four amino acids E-I-T-A (AA45-48) upstream of the core were replaced to the factor Xa cleaving site I-E-G-R, yielding a mutated DarA (MBP-Xa-DarA.core, Figure 2). The follower peptide of MBP-Xa-DarA.core was removed, since it was reported that the follower peptide is not necessary for DAR biosynthesis[5].

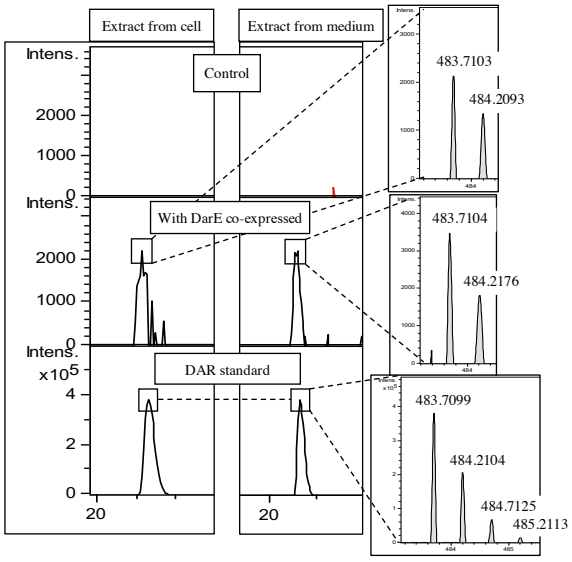


Figure 3. Extracted ion chromatograms (EICs) of the crude extract from the cell and culture media of *E. coli* BAP1 expressing MBP-DarA with and without DarE co-expression. The EICs correspond to the $[M+2H]^{2+}$ of DAR. The mass of DAR standard is 483.7099 *m/z* (error: 1.0140 ppm). The mass of DAR from cell and culture media crude extract are 483.7103 *m/z* (error: 1.8410 ppm) and 483.7104 *m/z* (error: 2.0477 ppm), respectively. The mass of DAR cannot be found in any samples of the control (without DarE co-expression).

As before, the MBP-Xa-DarA.core was heterologously expressed with and without co-expression of DarE. Following amylose purification, the samples were analyzed by LCMS before and after Factor Xa treatment. This time, the linear core peptide (F6) was not detected before Factor Xa treatment, but was released by factor Xa, independent of the DarE presence. Furthermore, DAR was present if DarE was co-expressed and absent if not (F7, Table 1 and Figure 4), indicating that the bicyclic core peptide was attached to its leader before factor Xa

treatment. Noteworthy, using the artificial cleavage site, the heptapeptide core with only one, either *C-O-C* or *C-C*, ring closure was not detected anymore, and the mass of DAR was not found neither in the cell nor in medium crude extracts. The *in vivo* results showed that DarE catalyzes both ring formations and should act on a leader-bound core peptide. Trimming of the precursor to yield DAR is prevented by alteration of the recognition/cutting site.

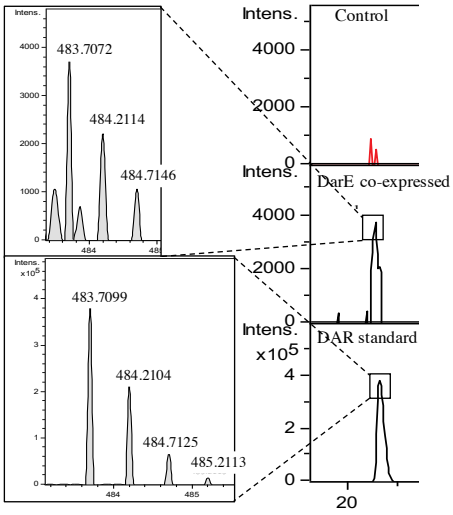


Figure 4. The extracted LCMS chromatogram of factor Xa digestion treatment of purified MBP-Xa-DarA.core with and without co-expression of DarE. The chromatograms were extracted to $[M+2H]^{2+}$ of DAR. The mass of DAR observed from samples with DarE co-expression was 483.7072 *m/z* (error = 4.5678 ppm), and from the standard 483.7099 *m/z* (error = 1.0140 ppm). Ions corresponding to DAR were not found in the control (without DarE co-expression). More LCMS chromatogram details are shown in Figure S6.

Conclusion

Using MPB-fused DarA with co-expression of DarE to observe the full modification of DarA before the cleavage was not successful. However, both partially modified DarA precursors (only *C-C* or *C-O-C* bond formation) could be observed indicating an independent formation. The likely reason for the unobserved full modified DarA is because of the high activity rate of the unidentified protease activity that releases the full modified core peptide, thus loss during amylose purification. In consequence, the release of completely modified core peptide, which essentially is DAR, can be observed in the cell and culture medium crude extracts. By replacing 4AA upstream the core peptide to the synthetic Factor Xa cleaving site prevents the proteolytic

release. In consequence, DAR cannot be detected in the cell and culture medium crude extract anymore. Using the latter approach, the full modification of DarA was observed; thereby, confirming the involvement of DarE in the formation of *C-O-C* and *C-C* rings in DAR biosynthesis.

Material and Method

Construction of expression vectors. Plasmid pET24c.MBP.Xa.DarA was used to heterologously express MPB-DarA, as described previously[5]. For a co-expression with DarE, plasmid pET24c.MBP.Xa.darA was linearized by overnight re-striction with enzyme SalI (NEB), dephosphorylated (1 h, 37 °C) using alkaline phosphatase (Thermo Fisher) and purified using the Zymo-research KIT, yielding pET24c.MBP.Xa.darA-SalI. The *darE* gene was amplified by PCR with the primer pair SAM.F and SAM.R, using plasmid pET24c.RSE.His as template (Sequence S1), in which a synthetic *E. coli*-codon-optimized *darE* version from *Photorhabdus namnaonensis* presents. The resulting 1360 bp DNA fragment, consisting of a ribosomal binding site and *darE* was purified by agarose gel and in a second step by the Zymo-research large fragment KIT. The linearized vector and the *darE* fragment were assembled via Gibson Assembly[6], yielding pET24c.MBP.Xa.darA.darE. The mix was transferred to *E. coli* cells and confirmed by PCR and DNA restriction analysis.

Construct pET24c.MBP.Xa.1.Xa.2.darE was used to co-express MBP-Xa-DarA.core and DarE, while pET24c.MBP.Xa.1.Xa.2 was used to express MBP-Xa-DarA.core. pET24c.MBP.Xa.1.Xa.2.darE was made by amplifying pET24c.MBP.Xa.darA.darE with primers Xa.core.darA.F and Xa.core.darA.R, yielding a 7972 bp fragment. The latter was purified, assembled by Gibson Assembly[6] and transferred to *E. coli* cells. The construct was corroborated by PCR, DNA restriction analysis and Sanger sequencing. pET24c.MBP.Xa.1.Xa.2 was made similarly, except pET24c.MBP.Xa.darA was used as template.

Protein expression and purification. The expression of (MBP-fused) proteins/peptide was done following standard procedures. In brief, expression was induced using isopropyl β-D-1-thiogalactopyranoside (IPTG) (final concentration: 0.1 mM) at an OD_{600} of ~0.5. Following induction, the culture was incubated at 18 °C, shaking at 180 rpm, overnight. Cells were concentrated by centrifugation and opened by sonification on ice. Purification was achieved

using amylose resin (NEB). The resulting elution fraction containing the desired protein was concentrated using Amicon ultra-15 spin columns. Protein concentration was determined by Bradford assay. For 500 mL cultivation, 0.918, 1.024, 0.84 and 0.966 mg of MBP-DarA with and without DarE co-expressed and MBP-Xa-DarA.core with and without DarE co-expressed respectively could be isolated. Prior to SDS-Page analysis (Figure S7), disulphide bonds in the protein samples were reduced. Hence, 50 μL of 100 mM ammonium bicarbonate was added to 50 μL sample protein and dithiothreitol (DTT) was added to a final concentration of 5 mM, incubated 56 °C, 30 min. Thereafter, 15 mM final concentration of iodoacetamide (IAA) was added and incubated for 30 min in dark room.

Protein digestion. Trypsin digestion was done according to Promega manufacturer´s protocol. Factor Xa digestion was done by adding 1 μL of factor Xa protease (NEB) into 100 μL sample solution. The solution (protein/peptide/darobactin + trypsin/factor Xa) was incubated overnight at 37 °C for trypsin and room temperature for factor Xa. To quench the reaction, 0.5% (v/v) trifluoroacetic acid was added and 1 μL was used to verify on pH paper that the pH is below 2. Thereafter, the solution was centrifuged at max speed for 2 min. The clear supernatant was transferred to a new 1.5 mL tube. A stage tip was prepared by adding 2 layers of C18 material to a 200 μL pipette tip. Then, solvents (each time 20 μL) were added to the stage tip prior to centrifugation (1000x g, 2 min). In consecutive order: MeOH, solution B (0.5% formic acid in 80/20% MeCN/water), and twice solution A (0.5% formic acid in 5/95% MeCN/water). Next, the trypsin digested sample was applied to the stage tip and centrifuged (1000x g, 4 min). Washing was done with 20 μL solution A (1000x g, 2 min), before the stage tip was moved to a new 1.5 mL tube and 100 μL of solution B were added (1000x g, 2 min) for elution. The eluted solution was subjected to LCMS.

Extraction of E. coli cells and culture medium for LCMS analysis. 0.2 mL of an overnight pre-culture of the respective strain was used to inoculate 20 mL lysogeny broth liquid medium containing 50 µg/mL kanamycin and incubated at 37 °C, 180 rpm until reaching an OD_{600} ~0.5 and induced with 0.5 mM final concentration of IPTG. The culture was then incubated at 30 °C shacked for 4 days. Thereafter, the culture was centrifuged at max speed for 2 min to separate cells and medium. For analysis of the culture medium, 200 μL of the clear medium were treated with C18 material prior to LCMS measurement as explained before (trypsin digestion). For analysis of the cell extract, the cell pellet was dissolved with 200 μL of 1:1

MeCN:water (v/v) solution, resuspended and lysed using an ultrasonic bath for 2 x 15 min. Then, the suspension was centrifuged at max speed for 2 min to separate the debris. Thereafter, 100 μL of clear supernatant was subjected to LCMS.

LCMS. The LCMS was performed on a Dionex Ulti-mate3000 (Thermo scientific) using an EC10/2 Nucleoshell C18 2.7 μm column (Macherey-Nagel) HPLC coupled to a microTOFq II (Bruker) ESI-qTOF-HRMS. The LC program was running in the following gradient (A: H_2O, 0.1% FA; B: MeOH, 0.1% FA): 0-5 min 10% B, 5-35 min 10-100% B, 35-50 min 100% B, 50-60 min 10% B. The flow was 200 μL/min and the column temperature was 45 °C.

References

1. Imai, Y. *et al*. A new antibiotic selectively kills gram-negative pathogens. *Nature* **576**, 459-464 (2019). Doi: s41586-019-1791-1
2. Schramma, K. R. *et al*. Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink. *Nature Chemistry* **7**, 431-437 (2015). Doi: 10.1038/NCHEM.2237
3. Clark, K. A., *et al*. Aliphatic ether bond formation expands the scope of radical SAM enzymes in natural product biosynthesis. *J. Am. Chem. Soc*. **141**, 10610-10615 (2019). Doi: 10.1021/jacs.9b05151
4. Kaur, H., *et al*. The antibiotic darobactin mimics a β-strand to inhibit outer membrane insertase. *Nature* 593, 125-129 (2021). Doi: 10.1038/s41586-021-03455-w
5. Wuisan, Z. G. *et al*. Optimization of heterologous Darobactin A expression and identification of the minimal biosynthetic gene cluster. *Metabolic Engineering* **66**, 123-136 (2021). Doi: 10.1016/j.ymben.2021.04.007
6. Gibson, D. G. *et al*. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343-345 (2019). Doi: 10.1038/nmeth.1318

6. Chapter V: Supporting Information

Radical SAM enzyme catalyzing the formation of *C-C* and *C-O-C* rings in Darobactin

I Dewa M. Kresna[a,b], Nils Böhringer[a,b,c], Zerlina G. Wuisan[a,b,c], Till F. Schäberle[a,b,c]

[a] Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, Ohlebergsweg 12, 35392 Giessen, Germany

[b] Department of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Ohlebergsweg 12, 35392 Giessen, Germany

[c] German Center for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Giessen, Germany

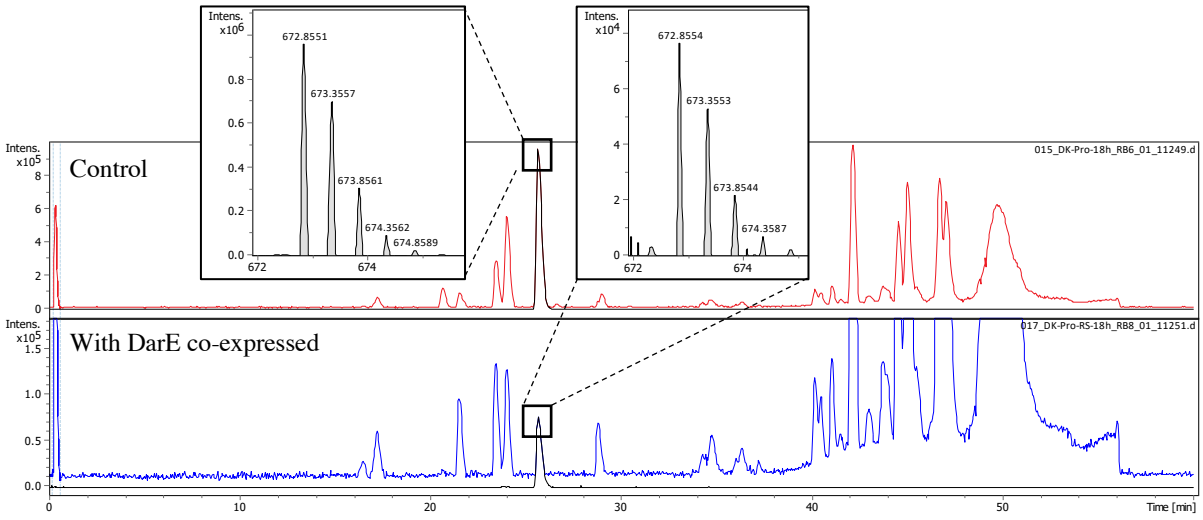


Figure S1. The base peak chromatogram of purified MBP-DarA after trypsin digestion. The blue and red line represent the MBP-DarA co-expressed with and without DarE (control), respectively. The black line represent the extracted ion chromatogram to $[M+2H]^{2+}$ of fragment F1 (see Table 1 or S1). The observed F1 mass from the sample with DarE co-expressed is 672.8554 *m/z* (error: 5.77 ppm), while from the control sample is 672.8551 *m/z* (error: 5.32 ppm).

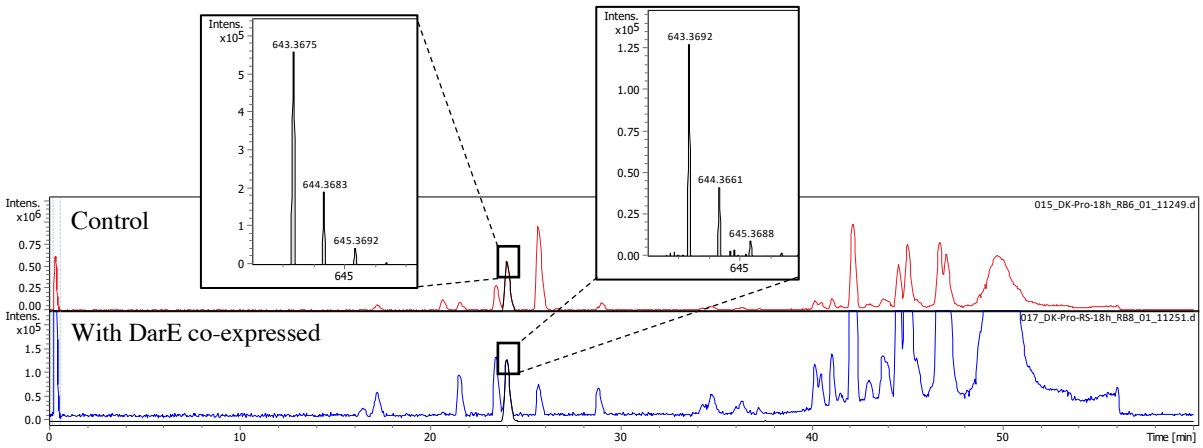


Figure S2. The base peak chromatogram of purified MBP-DarA after trypsin digestion. The blue and red line represent the MBP-DarA was co-expressed with and without DarE (control) respectively. The black line represent the extracted ion chromatogram to $[M+H]^+$ of fragment F2 (see Table 1 or S1). The observed F2 mass from the sample with DarE co-expressed is 643.3692 *m/z* (error: 3.93 ppm), while from the control sample is 643.3675 *m/z* (error: 1.29 ppm).

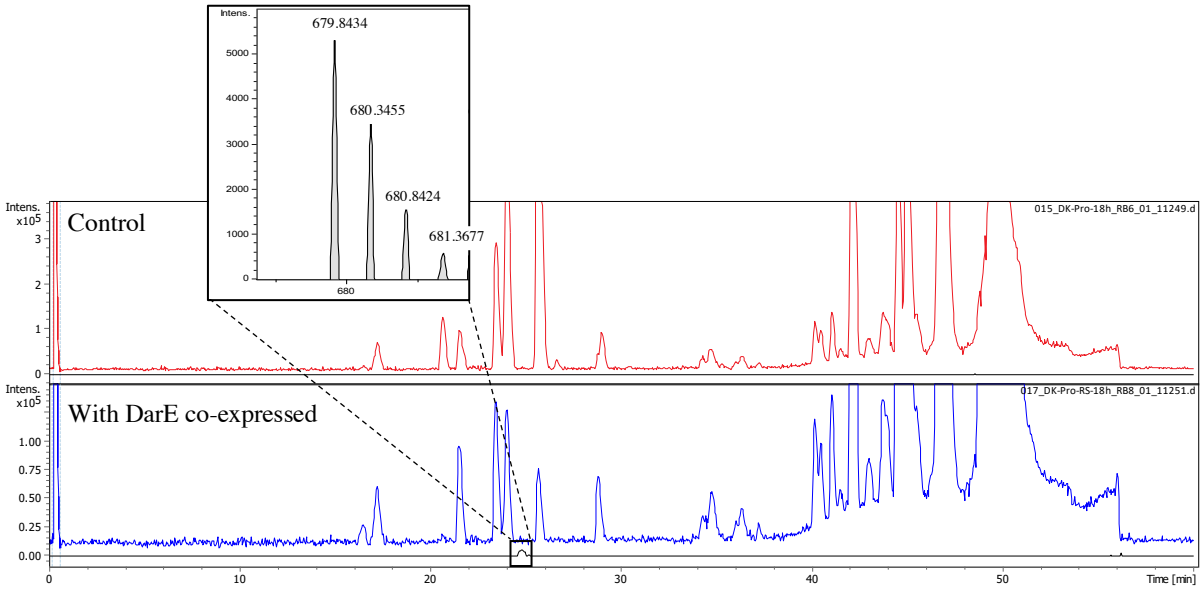


Figure S3. The base peak chromatogram of purified MBP-DarA after trypsin digestion. The blue and red line represent the MBP-DarA was co-expressed with and without DarE (control) respectively. The black line represent the extracted ion chromatogram to $[M+2H]^{+2}$ of fragment F3 (see Table 1 or S1). The observed F3 mass from the sample with DarE co-expressed is 679.8434 *m/z* (error: 3.31 ppm), while from the control sample, this mass did not appear.

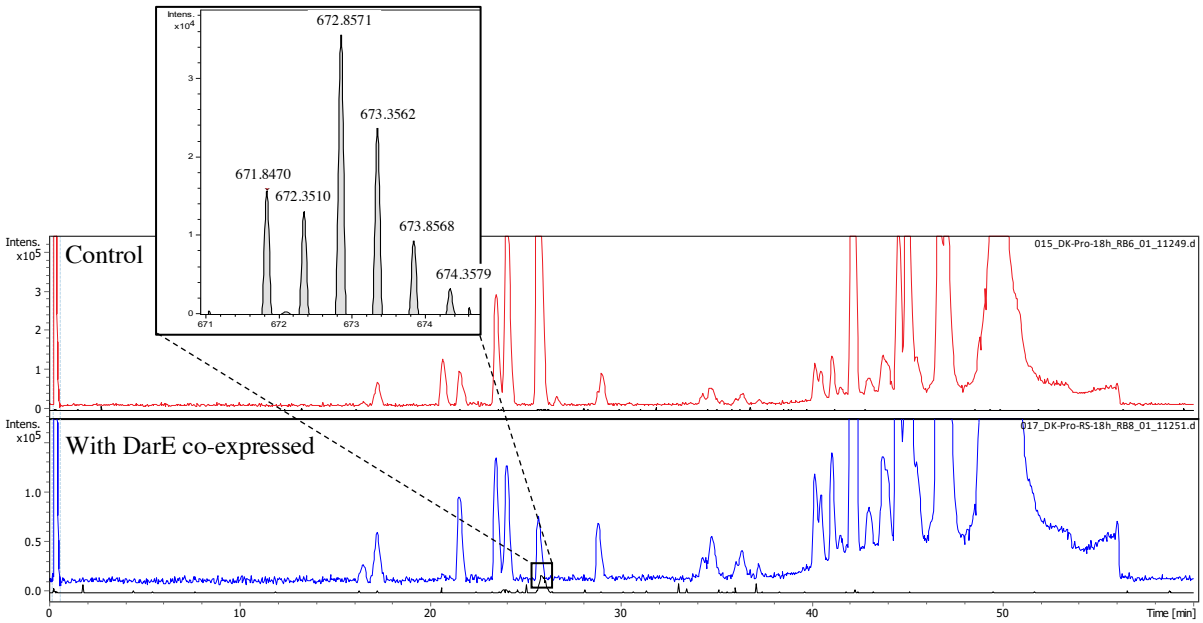


Figure S4. The base peak chromatogram of purified MBP-DarA after trypsin digestion. The blue and red line represent the MBP-DarA was co-expressed with and without DarE (control) respectively. The black line represent the extracted ion chromatogram to $[M+2H]^{+2}$ of fragment F4 (see Table 1 or S1). The observed F4 mass from the sample with DarE co-expressed is 671.8470 *m/z* (error: 4.92 ppm), while from the control sample, this mass did not appear. The third isotopic pattern of F4 (672.8571 *m/z*) cannot be distinguished with the first isotopic mass of $[M+2H]^{+2}$ of F1 since they were coming out (overlapped) in the close retention time (see Figure S1).

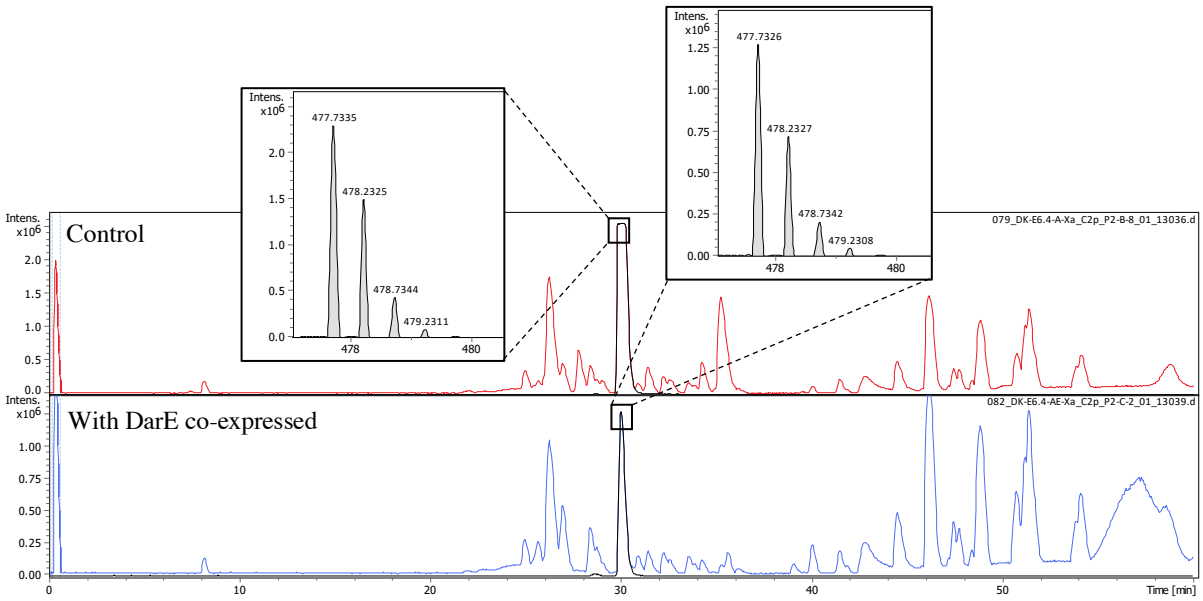


Figure S5. The base peak chromatogram of purified MBP-Xa-DarA.core after factor Xa digestion. The blue and red line represent the MBP-DarA was co-expressed with and without DarE (control) respectively. The black line represent the extracted ion chromatogram to $[M+2H]^{2+}$ of fragment F6 (see Table 1 or S1). The observed F6 mass from the sample with DarE co-expressed is 477.7326 *m/z* (error: 10.46 ppm), while from the control sample is 477.7335 *m/z* (error: 12.35 ppm).

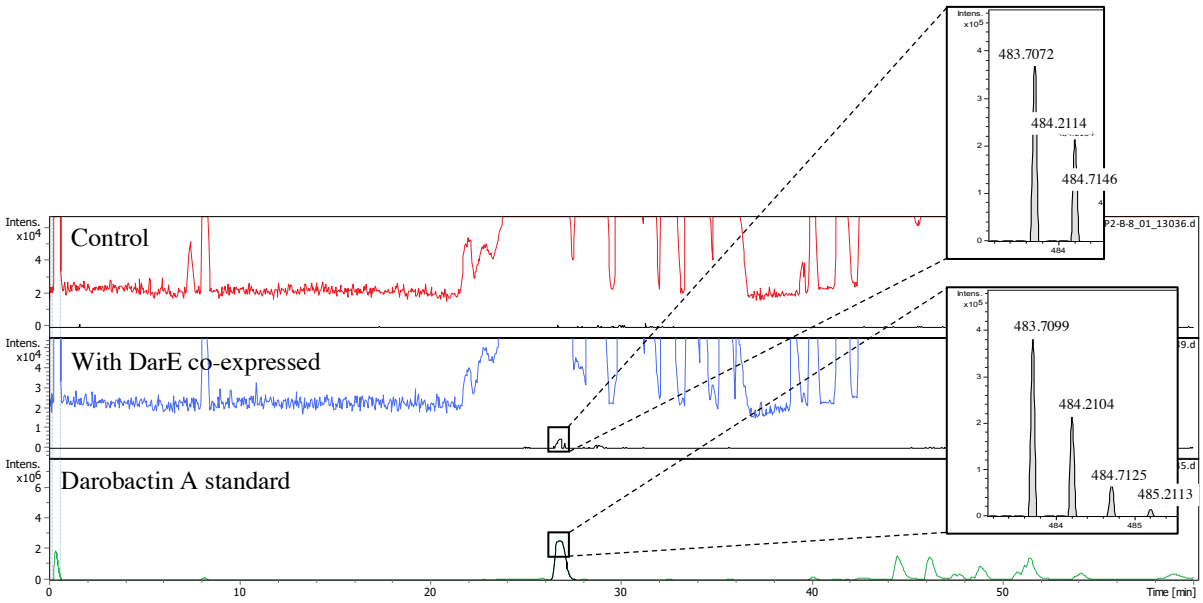


Figure S6. The base peak chromatogram of purified MBP-Xa-DarA.core after factor Xa digestion and darobactin A standard. The blue and red line represent the sample of MBP-DarA was co-expressed with and without DarE (control) respectively. The green line represent the sample of darobactin A standard. The black line represent the extracted ion chromatogram to $[M+2H]^{2+}$ of fragment F7 (darobactin A). The observed F7 mass from the sample with DarE co-expressed is 483.7072 *m/z* (error: 4.5678 ppm), while from the darobactin A standard sample is 483.7099 *m/z* (error: 1.0140 ppm). Darobactin A mass could not be found in the control sample.

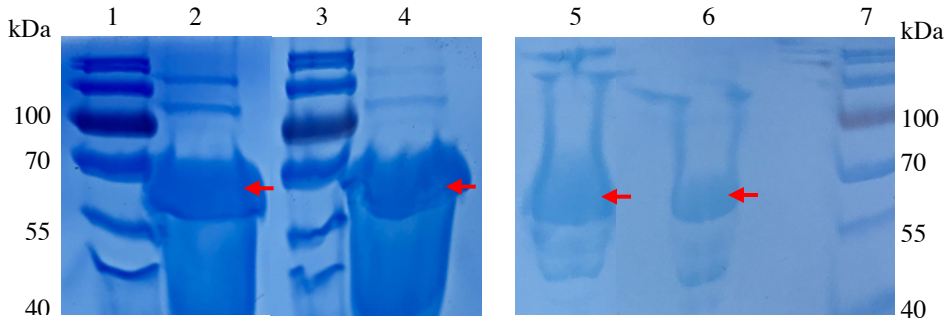


Figure S7. SDS-Page analysis of purified MBP-DarA and MBP-Xa-DarA.core. Line 1, 3 and 7: protein ladder; line 2: MBP-DarA expressed without DarE co-expressed; line 4: MBP-DarA with DarE co-expressed; line 5: MBP-Xa-DarA.core expressed without DarE co-expressed; line 6: MBP-Xa-DarA.core with DarE co-expressed. Theoretical molecular weight of MBP-DarA is 49.1 kDa and for MBP-Xa-DarA.core is 48.7 kDa.

Table S1. DarA fragments containing the core peptide after digestion with trypsin. The core peptide is marked with bold letters. The first site cut off located before the core peptide and the second one located after the core peptide. More LCMS details of the detected fragments mass are shown in Figure S1-6.

Fragment code	Investigated fragments	Description	Ionization	Co-expressed with DarE	
				Yes	No
F1	IPEITA**WNWSK**^{5}	Non modified version	$[M+2H]^{2+}$	672.8554 *m/z* error: 5.77 ppm	672.8551 *m/z* error: 5.32 ppm
F2	IPEITA	Successful peptidase activity in the 1st site	$[M+H]^{+}$	643.3692 *m/z* error: 3.93 ppm	643.3675 *m/z* error: 1.29 ppm
F3	IPEITA**WNWSK**^{5} (O)	*C-O-C* ring modification and successful Trypsin digestion after K^{5}	$[M+2H]^{2+}$	679.8434 *m/z* error: 3.31 ppm	-
F3.1	IPEITA**WNWSK**^{5}**SF** (O)	*C-O-C* ring modification with unsuccessful Trypsin digestion after K^{5} and successful peptidase activity in the 2nd site	$[M+H]^{+}$ $[M+2H]^{2+}$ $[M+3H]^{3+}$	-	-
F3.2	IPEITA**WNWSK**^{5}**SF**QEI (O)	*C-O-C* ring modification with unsuccessful Trypsin digestion after K^{5} and intact 2nd site of peptidase activity	$[M+H]^{+}$ $[M+2H]^{2+}$ $[M+3H]^{3+}$	-	-
F4	IPEITA**WNWSK**^{5}	*C-C* ring modification and successful Trypsin digestion after K^{5}	$[M+2H]^{2+}$	671.8470 *m/z* error: 4.92 ppm	-
F4.1	IPEITA**WNWSK**^{5}**SF**	*C-C* ring modification with unsuccessful Trypsin digestion after K^{5} and successful peptidase activity in the 2nd site	$[M+H]^{+}$ $[M+2H]^{2+}$ $[M+3H]^{3+}$	-	-
F4.2	IPEITA**WNWSK**^{5}**SF**QEI	*C-C* ring modification with unsuccessful Trypsin digestion after K^{5} and intact 2nd position of peptidase activity	$[M+H]^{+}$ $[M+2H]^{2+}$ $[M+3H]^{3+}$	-	-

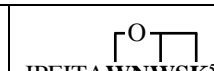
F5	 IPEITA**WNWSK^{5}SF**	Full modification with unsuccessful Trypsin digestion after K^{5} and successful peptidase activity in the 2nd site	$[M+H]^{+}$ $[M+2H]^{2+}$ $[M+3H]^{3+}$	-	-
F5.1	 IPEITA**WNWSK^{5}SF**QEI	Full modification with unsuccessful Trypsin digestion after K^{5} and intact 2nd position of peptidase activity	$[M+H]^{+}$ $[M+2H]^{2+}$ $[M+3H]^{3+}$	-	-
F5.2	 IPEITA**WNWSK^{5}**	Full modification with successful Trypsin digestion after K^{5}	$[M+H]^{+}$ $[M+2H]^{2+}$ $[M+3H]^{3+}$	-	-
F6	**WNWSK^{5}SF**	Non modified version of core peptide	$[M+2H]^{2+}$	477.7326 *m/z* error: 10.5 ppm	477.7335 *m/z* error: 12.35 ppm
F7	 **WNWSK^{5}SF**	Core peptide with 2 rings attached (darobactin A)	$[M+2H]^{2+}$	483.7072 *m/z* (error: 4.5678 ppm)	-
F7.1	 **WNWSK^{5}SF**	Core peptide with *C-O-C* ring attached	$[M+H]^{+}$ $[M+2H]^{2+}$ $[M+3H]^{3+}$	-	-
F7.2	 **WNWSK^{5}SF**	Core peptide with *C-C* ring attached	$[M+H]^{+}$ $[M+2H]^{2+}$ $[M+3H]^{3+}$	-	-

Table S2. Primer list used in this study

No.	Primer Name	Primer Sequence
1	SAM.F	GGATCCGAATTCGAGCTCCGTTTGTTTAACTTTAAGAAGGAGATATACATATG GAC
2	SAM.R	CGAGTGCGGCCGCAAGCTTGTCAAGCAGCGATGGTCTGTTTGAT
3	Xa.core.darA.F	TGGAACTGGTCTAAATCTTTCTAATGGCTAGCATGACTGGTGGAC
4	Xa.core.darA.R	AAAGATTTAGACCAGTTCCACCTTCCCTCGATCGGGATTTTCGGTTTGTTGTTC AG TTCG

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGA
CCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCC
GGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCG
ACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCC
CTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTAT
CTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTT
AACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAA
ATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTC
TTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTT
TGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTG
GTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGG
TTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCT
TTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTAT

TCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAA
TCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTT
CTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGA
TAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAAC
ATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGA
TAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGT
TGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACT
GTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGC
GTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTG
CAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCG
AAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCAC
CACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCA
GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGG
GCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTA
CAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGG
CAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTG
TCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAA
AAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTG
CGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCG
AACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTA
CGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAG
TTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACAC
CCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGG
GAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATC
AGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGA
AGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATG
CCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGA
TACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGA
TGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTC
CACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCG
TTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCA
GCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAG
CCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATG
GCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATT
CCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGAC
CCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGAT
AGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCC
CGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAAC
CTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAG
GGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGT
TGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGG
ATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCG

GACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACG
ATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCG
CTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAG
AACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAGTC
GCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACG
CCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCA
GCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTAC
CATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGG
CGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGC
CACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACG
TGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTAT
AACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAA
AGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGC
AGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCG
CCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAA
GTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCC
GGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTC
ACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAG
ATATACAT**ATG**GACACCATCATCCCGATCAAATACCTGAACGCTGACGAATCTTCTATCCTGAAAAAAT
CTCCGAAAATCAACTACCGTCAGCTGGCTTGCCGTATCATCGGTGAAATCCCGGCTGAAAAAATCCTGG
ACGACGACGAACTGGCTCTGTACAACGAAGAAATCGGTATCCACTTCTCTCCGGAAATCATCAACGCTA
ACAAACTGGTTGTTGTTGTTAAAGCTACCCGTCTGTGCAACCTGCGTTGCACCTACTGCCACTCTTGGGC
TGAAGGTAAAGGTAACACCCTGACCTTCTTCAACCTGATGCGTTCTATCCACCGTTTCCTGTCTATCCCG
AACATCAAACGTTTCGAATTCGTTTGGCACGGTGGTGAAGTTACCCTGCTGTCTGTTAACTACTTCAAAA
AACTGATCTGGCTGCAGGAACAGTTCAAAAAACCGGACCAGGTTATCACCAACTCTGTTCAGACCAACG
CTGTTAACATCCCGGAAGACTGGCTGGTTTTCCTGAAAGGTATCGGTATGGGTGTTGGTATCTCTGTTGA
CGGTATCCCGGAAATCCACGACTCTCGTCGTCTGGACTACCGTGGTCGTCCGACCTCTCACAAAGTTGCT
GCTGGTATGAAAAAACTGCGTTCTTACGGTATCCCGTACGGTGCTCTGGTTGTTGTTGACCGTGACGTTT
ACGAATCTAACATCGAAAAAATGCTGTCTTACTTCTACGAAATCGGTCTGACCGACATCGAATTCCTGA
ACATCGTTCCGGACAACCGTTGCCAGCCGGGTGACGACCCGGGTGGTTCTTACATCACCTACCACAACT
ACATCAACTTCCTGTCTAACGTTTTCCGTGTTTGGTGGAACGACTACCAGGACAAAATCAACATCCGTCT
GTTCCACGGTTTCATCGACTCTATCAAATCTTCTCAGAAAAAAATCTCTGACTGCTACTGGGCTGGTAAC
TGCTCTCAGGAAATCATCACCCTGGAACCGAACGGTACCGTTTCTGCTTGCGACAAATACGTTGGTGCTG
AAGGTAACAACTACGGTTCTATCATCGACAACGACCTGGGTCACCTGCTGATCAAATCTAACACCAACA
AAAACCACCTGAAAGAAGAAATCGAATCTTACGAAAAAATGCACCAGTGCAAATGGTTCCACCTGTGC
AACGGTGGTTGCCCGCACGACCGTGTTACCAACCGTAAACACAACCCGAACTACAACGACTTCTGCTGC
GGTACCGGTGGTCTGCTGGAAATCATCAAACAGACCATCGCTGCTGCACTCGAGCACCACCACCACCAC
CACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATA
ACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATC
CGGAT

Sequence S1. The DNA sequence of pET24.RSE.His (circular). The gene *darE* represents a codon-optimized version of the gene derived from *Photorhabdus namnaonensis* strain PB45.5 Phpb_contig000089 (WP_065391756.1). Red letters represent the DNA sequence amplified by primers SAM.F and SAM.R. Bold letters represents the start codon of *darE*. The stop codon was integrated by the 3' end of the primer SAM.R.

Journal of the American Chemical Society

Characterization of a Radical SAM Oxygenase for the Ether Crosslinking in Darobactin Biosynthesis

Journal:	*Journal of the American Chemical Society*
Manuscript ID	ja-2022-05565v
Manuscript Type:	Article
Date Submitted by the Author:	26-May-2022
Complete List of Authors:	Nguyen, Hai; Duke University School of Medicine Kresna, I Dewa Made; Justus Liebig Universitat Giessen Fachbereich fur Agrarwissenschaften Okotrophologie und Umweltmanagement, Institute for Insect Biotechnology Böhringer, Nils; Justus Liebig Universitat Giessen Fachbereich fur Agrarwissenschaften Okotrophologie und Umweltmanagement, Institute for Insect Biotechnology Ruel, Jeremie ; Institut de Biologie Structurale de la Mora, Eugenio; Institut de Biologie Structurale, Metalloproteins Unit Kramer, Jil-Christine; Justus Liebig Universitat Giessen Fachbereich fur Agrarwissenschaften Okotrophologie und Umweltmanagement, Institute for Insect Biotechnology Lewis, Kim ; Northeastern University, Biology Nicolet, Yvain; Institut de Biologie Structurale, Metalloproteins Unit Schäberle, Till; Justus Liebig Universitat Giessen Fachbereich fur Agrarwissenschaften Okotrophologie und Umweltmanagement, Institute for Insect Biotechnology Yokoyama, Kenichi; Duke University School of Medicine, Biochemistry; Duke University

Characterization of a Radical SAM Oxygenase for the Ether Crosslinking in Darobactin Biosynthesis

Hai Nguyen[a], I Dewa M. Kresna[b], Nils Böhringer[b,c], Jeremie Ruel[d], Eugenio de la Mora[d], Jil-Christine Kramer[b], Kim Lewis[f], Yvain Nicolet[d,*], Till F. Schäberle[b,c,e,*], Kenichi Yokoyama[a,g,*]

[a] Department of Biochemistry, Duke University School of Medicine, Durham, NC 27710

[b] Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, Ohlebergsweg 12, 35392 Giessen, Germany

[c] German Center for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Giessen, Germany

[d] University of Grenoble Alpes, CEA, CNRS, IBS, Metalloproteins Unit, F-38000 Grenoble, France

[e] Department of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Ohlebergsweg 12, 35392 Giessen, Germany

[f] Antimicrobial Discovery Center, Department of Biology, Northeastern University, Boston, MA 02115

[g] Department of Chemistry, Duke University, Durham, NC 27710

KEYWORDS: radical SAM enzymes, oxygenases, antibiotic, natural product, biosynthesis, RiPPs

ABSTRACT: Darobactin A is a ribosomally synthesized, post-translationally modified peptide (RiPP) with potent and broad-spectrum anti-Gram-negative antibiotic activity. The structure of darobactin A is characterized by an ether and a C-C crosslinking. However, the specific mechanism of the crosslink formation, especially the ether crosslink, remain elusive. Here, using *in vitro* enzyme assays, we demonstrate that both crosslinks are formed by the DarE radical SAM enzyme in an O_2-dependent manner. The relevance of the observed activity to darobactin A biosynthesis was demonstrated by proteolytic transformation of the DarE product into darobactin A. Furthermore, DarE assays in the presence of $^{18}O_2$ or [^{18}O]water demonstrated that the oxygen of the ether crosslink originates from O_2 and not from water. These results demonstrate that DarE is a radical SAM enzyme that uses oxygen as a co-substrate in its physiologically relevant function. Since radical SAM enzymes are generally considered to function under anaerobic environment, the discovery of a radical SAM oxygenase represents a significant change in the paradigm and suggests that these radical SAM enzymes function in aerobic cells.

Introduction

The bicyclic heptapeptide darobactin A is a ribosomally synthesized, post-translationally modified peptide (RiPP) with antibiotic activities against diverse sets of Gram-negative bacteria including many clinically important pathogens[1]. darobactin A exhibits its antibiotic activity by inhibiting BamA, a protein in the Bam complex responsible for the assembly of proteins in the outer membrane of Gram-negative bacteria[1, 2]. Darobactin A has a rigid *β*-strand conformation through the ether and C-C crosslinks on the heptapeptide core structure of (W^1-N^2-W^3-S^4-K^5-S^6-F^7, see Figure 1 for the structure), which mimics the recognition signal of native substrates and blocks the open lateral gate of BamA[2]. However, the mechanism by which this unique and biologically important bicyclic structure is formed remains unknown.

The darobactin biosynthetic gene cluster (BGC) consists of the precursor peptide-encoding *darA*, the transporter genes *darBCD*, and a radical *S*-adenosyl-L-methionine (SAM) enzyme gene *darE* (Figure 1a). Earlier studies have established that co-expression of DarA and DarE is sufficient for the production of darobactin A in *E. coli*[1, 3, 4] (Figure 1b). Radical SAM enzymes form a large enzyme superfamily[5] and catalyze various radical-mediated reactions[6]. These enzymes harbor an oxygen-sensitive 4Fe-4S cluster to catalyze the reductive cleavage of *S*-adenosyl-L-methionine (SAM) to transiently generate 5′-deoxyadenosyl radical (5′-dA•) that then in most cases abstracts a H-atom from the substrate or, in other cases, adds to the substrate to carry out various radical reactions. Since the 4Fe-4S clusters of radical SAM enzymes are oxygen-sensitive, characterization of these enzymes has been performed under strictly anaerobic conditions and it has been largely assumed that these enzymes must function in the absence of oxygen. So far, no radical SAM enzymes has been reported to catalyze oxygenase reactions that require molecular oxygen as the co-substrate.

In the past decade, many radical SAM enzymes have been identified to be responsible for the maturation of RiPPs. In particular, radical SAM enzymes in the SPASM subfamily[7], characterized by the presence of auxiliary 4Fe-4S clusters (AUX) in addition to the canonical radical SAM 4Fe-4S cluster, were reported to catalyze the formation of various crosslinks during RiPPs biosynthesis. Such crosslinks include C-C bonds between aromatic and aliphatic carbons[8,9], thioethers via a Cys residue[10], and an ether via Thr alcohol[11]. However, all of these

reported crosslinks utilize functional groups already available in the precursor peptides. In contrast, the ether ring in darobactin A is unique in that the oxygen atom of the ether must be post-translationally installed. The origin of this oxygen atom and the mechanism of its installation remain ambiguous.

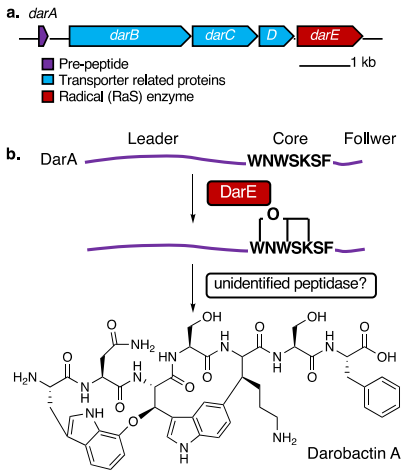


Figure 1. Darobactin A biosynthesis. **a.** The BGC of darobactin A. **b.** The proposed biosynthetic pathway of darobactin A.

Recently, when this manuscript was in preparation, Guo et al reported *in vitro* characterization of *Photorhabdus khanii* DarE with N-terminal His-tag[12]. The authors reported formation of DarA with a ether crosslink without the C-C crosslinking as the major product and proposed water as the source of the ether oxygen. However, no evidence was provided whether such compound may be converted into darobactin A, and hence the biological relevance of this observation remained unclear.

Here, we report successful *in vitro* reconstitution of darobactin A biosynthesis. Importantly, our ^{18}O-labeling study provides strong evidence that the oxygen atom of the ether crosslink is derived from molecular oxygen and not from water, making DarE as the first radical SAM oxygenase. These results extend the repertoire of the reactions catalyzed by radical SAM enzymes and provide the evidence for their function under aerobic environment.

Materials and Methods.

General. Sodium dithionite (SDT) was purchased from Sigma-Aldrich. β-Mercaptoethanol (βME) was from Calbiochem. Dithiothreitol (DTT) was from Amresco. G-25 Sephadex resin was from GE Healthcare. Ni-NTA agarose resin was from Qiagen. Strep-Tactin XT 4Flow high-capacity resin was from IBA LifeSciences. *Escherichia coli* DH5α and BL21(DE3) competent cells were from Invitrogen. All anaerobic experiments were carried out in an MBRAUN glovebox maintained at 10 ± 2 °C with an O_2 concentration < 0.1 ppm. All anaerobic buffers were degassed on a Schlenk line and equilibrated in the glovebox overnight. All plastic devices were evacuated in the antechamber of the glovebox overnight before use. All HPLC experiments were performed on a Hitachi L-2130 pump equipped with an L-2455 diode array detector, an L-2485 fluorescence detector, an L-2200 autosampler, and an L-2300 column oven maintained at 40 °C. UV−vis absorption spectra were determined using a U-3900 UV−vis ratio recording double-beam spectrometer (Hitachi). All mass spectra were recorded on a 6224 accurate-mass time-of-flight mass spectrometer (Agilent) equipped with a Dual ESI source, and accurate mass data were obtained by internal calibration using a secondary nebulizer to continuously deliver the reference solution. LCMS analysis was conducted on a 6224 TOF LC/MS system equipped with a Phenomenex Kinetex C18 EVO column (3 x 100 mm, 2.6 µm particle), monitored by a diode array detector (254 nm) under the following gradient: hold 100% A (100:3:0.3 H2O:MeOH:formic acid) for 0 ~ 0.5 min, then ramp 0% ~ 50% B (100:3:0.3 MeCN:H_2O:formic acid) from 0.5 ~ 8 min, flow rate: 0.5 mL/min.

Heterologous production of darobactin A: The *Pseudoalteromonas luteoviolacea darA* and *darE* genes were codon optimised for *E. coli* using the Java Codon Adjustment Tool (jCAT)[13] and the resulting sequences were synthesized by Eurofins Genomics (Eberswalde, Germany). The resulting synthetic *darA* and *darE* genes were PCR amplified using the primer pairs darA-f/darA-r and darE-f/darE-r (Table S1) using Q5 polymerase (NEB Biolabs, New Brunswick, USA) according to the manufacturer's manual. All PCR products were gel purified using 1% or 2% TAE agarose gels and DNA was recovered using the Zymo Research Large Fragment DNA Recovery Kit (Zymo Research, USA) according to the manufacturer's manual. An empty pRSFduet-1 plasmid (Novagen) was digested using *Nde*I/*Avr*II. All fragments were assembled using homemade isothermal assembly master mix[14] and the assembled plasmids were transferred to *E. coli* Top10 using standard electroporation methodology and selected on LB_{Kan}. The identity of the plasmids was corroborated by test restriction. The correctly assembled plasmid was transferred to *E. coli* Bap1 and *E. coli* Bap1 harboring pGro7 (Takara) by electroporation. *E. coli* Bap1 + pRSFduet-1 (empty vector) was used as a negative control.

For heterologous Darobactin production, 20 mL LB_{Kan} was inoculated with an overnight preculture of *E. coli* Bap1 + pRSFduet-1, *E. coli* Bap1 + pRSF-darAE$_{Pse}$ of *E. coli* Bap1 + pRSF-darAE$_{Pse}$ + pGro7 (Takara) cells and the cultures were incubated at 30° C to an OD_{600} of ~0.5. Transcription was induced by addition of IPTG to 1 mM final concentration. After 3 days of cultivation, 0.5 mL of culture was separated into supernatant and pellet by centrifugation and the cleared supernatant was applied to self-packed C18 stage tips. The stage tips were washed using ddH_2O, the material was eluted in 50 µL 80:20 MeCN/H_2O and 5 µL of the concentrated supernatants were analysed by LCMS using a 20 µg/mL darobactin A solution as standard.

Expression and purification of *Pl*DarA: The *Pseudoalteromonas luteoviolacea darA* gene with the His_6- and SUMO-tags fused at its 5'-end was synthesized by Genscript with the codon usage optimized for *E. coli* and cloned into the *Nco*I-*Bam*HI site of pET16b (pET16b-SUMO-*Pl*DarA). *Pl*DarA was then expressed in *E. coli* BL21(DE3) cells harboring pET16b-SUMO-*Pl*DarA and pGro7. The *E. coli* cells was grown in LB medium (50 mL) with 100 mg/L ampicillin and 30 mg/L chloramphenicol, and incubated at 37 °C, 220 rpm overnight until saturation. An aliquot (30 mL) of the overnight culture then was used to inoculate 1.5 L of LB medium with the same antibiotics, which was grown at 37 °C, 220 rpm until

OD_{600} = 0.6 – 0.8. Protein expression was induced with 60 mg/L IPTG and the culture was incubated at 18 °C, 220 rpm for 20 h. The cells were harvested by centrifugation, washed with Buffer A (50 mM Tris-HCl pH 7.6, 150 mM NaCl, 10% v/v glycerol), frozen with liquid nitrogen, and stored at -20 °C. Approximately 6.1 g of wet cell paste was obtained per liter of culture.

SUMO-*Pl*DarA purification: In a typical purification, 20 g of cell pellet was suspended and homogenized in 80 mL of buffer A supplemented with 5 mM β-mercaptoethanol (β-ME) and 0.8 ml protease inhibitor cocktail (100X proteasearrest, G-Biosciences). The cell suspension was lysed by 2 passages through a French pressure cell operating at 14,000 psi. The resulting lysate was cleared by centrifugation (20,000 x *g*, 20 min, 4 °C. The supernatant was incubated with Ni-NTA agarose resin (Qiagen, 10 mL equilibrated in buffer A containing 20 mM imidazole and 1.0 µl benzonase / 30 mL supernatant) for 1 hour. The resin was washed with 10 volume of buffer A with 20 mM imidazole and 3 mM β-ME, and the protein eluted with buffer A with 400 mM imidazole and 3 mM β-ME. Fractions containing SUMO-*Pl*DarA were exchanged into buffer A using a Sephadex G-25 column. The concentration of SUMO-*Pl*DarA was determined by Bradford assay using BSA as a standard. The resultant protein solution was frozen in liquid N_2 and stored at -80 °C. Typically, 14 mg of SUMO-*Pl*DarA was prepared from each gram of wet cell paste. The SUMO tag of SUMO-*Pl*DarA (100 mg) was cleaved by SUMO-protease (0.3 mg) at 0 °C for one hour. The resulting solution was passed through Ni-NTA agarose resin (10 ml equilibrated in distilled water). The protein was eluted with distilled water. Fractions containg *Pl*DarA were brought into an mBraun anaerobic glove box ([O_2] < 0.1 ppm) and exchanged into anaerobic water using a Sephadex G-25 column. Concentration of *Pl*DarA was determined by HPLC.

Expression of His-*Pl*DarE: pRSF-His-*Pl*DarE was created by amplifying $darE_{Pse}$ using the primers his-darE-f/his-darE-r (Table S1) using Q5 polymerase (NEB) according to the manufacturer's manual. The amplified gene was gel purified using 1% TAE agarose gels and the ZymoResearch Large Fragment DNA Recovery Kit according to the manufacturer's manual. The purified fragment was then digested with *Bam*HI/*Sal*I and cloned into the corresponding site of pRSFduet-1. Both fragments were fused using homemade isothermal assembly master mix[14], transferred to *E. coli* Top10 using standard electroporation methodology and selected on LB_{Kan}. Correct assembly was verified by sequencing and the plasmid was designated as pRSF-His-*Pl*DarE. To express His-*Pl*DarE, *E. coli* Bap1 cells were transformed with pRSF-His-*Pl*DarE and pGro7. The resulting transformant was grown in LB medium (50 mL) with 50 mg/L kanamycin and 30 mg/L chloramphenicol, and incubated at 37 °C, 220 rpm overnight until saturation. An aliquot (30 mL) of the overnight culture then was used to inoculate 1.5 L of terrific broth containing 50 mg/L cysteine, 30 mg/L iron (III), and 0.5 g/L L-arabinose with the same antibiotics, which was grown at 37 °C, 220 rpm until OD_{600} = 0.6 – 0.8. Protein expression was induced with 60 mg/L IPTG, and the culture was incubated at 18 °C, 220 rpm for 20 h. The cells were harvested by centrifugation, washed with Buffer A (50 mM Tris pH 7.6, 150 mM NaCl, 10% glycerol), frozen with liquid nitrogen, and stored at -20 °C. Approximately 22 g of wet cell paste was obtained per liter of culture.

Purification of His-*Pl*DarE: His-*Pl*DarE was anaerobically purified typically from a 10-30 g cell paste of *E. coli* BAP1 harboring pRSF-His-*Pl*DarE. The cell pellet was brought into an mBraun anaerobic glove box ([O_2] < 0.1 ppm). Then, each gram of cell pellet was suspended and homogenized in 4 mL of anaerobic buffer A supplemented with 5 mM β-ME. The cell suspension was brought out of the glovebox and lysed by 2 passages through a French pressure cell operating at 14,000 psi under constant Ar flow. The resulting lysate was cleared by centrifugation (20,000 x *g*, 20 min, 4 °C) using centrifugal tubes filled with Ar gas and brought back to the glovebox. All subsequent purification steps were carried out under strict anaerobic conditions in the glove box ([O_2] < 0.1 ppm) maintained at 10 °C. The dark brown-colored supernatant was incubated with Ni-NTA agarose resin (Qiagen, 20 mL equilibrated in buffer A containing 40 mM imidazole and 1.0 µL benzonase / 30 mL supernatant) for 1 hour. The resin was washed with 10 volumes of buffer A with 20 mM imidazole and 3 mM β-ME, and the protein eluted with buffer A with 400 mM imidazole and 3 mM β-ME. Fractions containing dark brown colored His-*Pl*DarE were exchanged into buffer A using a Sephadex G-25 column. The concentration of His-*Pl*DarE was determined by Bradford assay using BSA as a standard. The amounts of $Fe^{2+/3+}$ were quantified by following published protocols[15]. Typically, the preparation yielded His-*Pl*DarE with 4.0 ± 0.8 Fe per monomer. The resulting purified His-*Pl*DarEwas immediately used for anaerobic reconstitution carried out at 10 °C by a slow addition of 8-12 eq. $Fe^{II}(NH_4)_2(SO_4)_2$ and Na_2S per His-*Pl*DarE monomer over the course of 10 min. The amounts of $Fe^{II}(NH_4)_2(SO_4)_2$ and Na_2S were adjusted based on the amounts of $Fe^{2+/3+}$ associated with the as-isolated His-*Pl*DarE. The resulting mixture was incubated for 60 min at 10 °C. The protein was then desalted using a Sephadex G-25 column equilibrated with buffer A. The amounts of Fe and sulfide were determined by the ferrozine assay.[15] The resultant protein solution was frozen in liquid N_2 and stored at -80 °C. Typically, 0.8 mg of His-*Pl*DarE was prepared from each gram of wet cell paste.

Expression and Purification of untagged *Pl*DarE: The Twin-strep II tag and the Factor Xa cleavage site (MASAWSHPQFEKGGGSGGGSGGSAWSHPQFEKSGIEG R) were introduced to the 5'-end of the codon optimized *P. luteoviolacea darE* gene by PCR. The pRSF-His-*Pl*DarE plasmid was amplified using a primer pair pRSF-strep-DarE-f/r (Table S1) and ligated with a strep-f/r oligo pair (Table S1) using the InFusion kit (Takara). The resulting plasmid pRSF-Strep-*Pl*DarE was then used as a template to amplify the Strep-*Pl*DarE gene using a primer pair pET-strep-*Pl*DarE-f/r (Table S1). The resulting PCR products were digested with *Nde*I and *Hin*dIII and introduced into the corresponding site of pET30b to yield pET30-Strep-*Pl*DarE. Strep-*Pl*DarE was expressed in an identical way to His-*Pl*DarE using *E. coli* Bap1 with pET30-Strep-*Pl*DarE and pGro7. Approximately 6.9 g of wet cell paste was obtained per liter of culture. To purify Strep-*Pl*DarE, the *E. coli* cells were lysed in the same manner as that for His-*Pl*DarE. The cleared lysate was loaded onto the Strep-Tactin XT 4Flow high-capacity resin and the column was washed with 10 column volumes of buffer W (100 mM Tris-HCl pH 8.0, 150 mM NaCl), and the protein eluted with buffer BXT (100 mM Tris-HCl pH 8.0, 150 mM NaCl, and 50 mM biotin). Fractions containing dark brown colored *Pl*DarE were exchanged into buffer W using a Sephadex G-25 column. The concentration of

*Pl*DarE was determined by Bradford assay using BSA as a standard. The strep-tag was removed by an incubation with Factor Xa protease (1 μg per 1 mg protein, New England Biolabs) in buffer W supplemented with 2 mM $CaCl_2$ at 10 °C for 4h. Then, the Factor Xa protease was removed by Xarrest resin, and the cleaved strep tag was removed using Strep-tactin resin. The 4Fe-4S cluster was reconstituted as described above for His-*Pl*DarE, and the resulting reconstituted *Pl*DarE contained 12.2 ± 0.5 Fe per monomer. Typically, 2.0 mg of *Pl*DarE was prepared from each gram of wet cell paste.

DarE activity assay: His-*Pl*DarE or *Pl*DarE (50 μM) was anaerobically incubated with DarA (50 μM) in the presence of SAM (1.0 mM), flavodoxin (10 μM), flavodoxin reductase (2.0 μM), NAD(P)H (1.0 mM), and oxygen (100 μM) in an air tight glass vial with a total volume of 50 μL buffer H supplemented with 5 mM DTT for 3h at 25 °C. The reaction was initiated by adding DarA and oxygen saturated buffer H (20 mM HEPES pH 8.0). After 3 h of incubation, the reaction was diluted twice and boil-quenched at 100 °C for 5 min. After removal of precipitation by centrifugation, an aliquot (40 μL) of the supernatant was injected to HPLC equipped with an XSelect Peptide CSH C18 column (Waters) equilibrated in 0.1% TFA in water (Solvent A). The elution was made with a flow rate at 1.0 mL/min using solvent A and solvent B (0.1% TFA in MeCN): 3% B for 5 min, 3-27% B for 5-15 min, and 27-30% B for 15-50 min. Chromatography was monitored by the L-2455 diode array detector.

Purification of DarE products: A DarE reaction (1.0 ml) was performed in the conditions described above and quenched with 1 ml buffer Q (0.2% TFA and 6% MeCN in water). After removal of precipitation by centrifugation, an aliquot (500 μL) of the supernatant was injected to and chromatographed on HPLC under the above-mentioned conditions. P1 or P2 were monitored by the L-2455 diode array detector and eluted at 44 or 30 min, respectively. After combining all the fractions containing P1 or P2, TFA and MeCN were removed by a centrifugal evaporator. The resulting solution was lyophilized and redissolved in water. Concentration of P1 or P2 was determined based on light absorption using $\varepsilon_{280nm} = 11\ mM^{-1}cm^{-1}$ assuming their absorptivity similar to unmodified tryptophan. HRMS (ESI-TOF) calcd. for P1 ($C_{281}H_{448}N_{74}O_{89}$, $[M+7H]^{+7}$, ^{13}C = 3) 899.0493, found 899.0417; P2 ($C_{281}H_{446}N_{74}O_{90}$, $[M+7H]^{+7}$, ^{13}C = 4) 901.1897, found 901.1809.

Proteinase K treatment of DarE products: P1 or P2 (2.5 μg) was treated with proteinase K (100 ng) in 20 μL buffer K (50 mM Tris-HCl pH 8.0 and 5.0 mM $CaCl_2$) for one hour at 37 °C. The reaction mixture was then quenched with an equal volume of buffer Q. Precipitation was removed, and an aliquot (20 μL) of supernatant was analyzed by HPLC. Another aliquot (5 μL) was analyzed by LC-HRMS. HRMS (ESI-TOF) calcd. for K-digested P1 ($C_{35}H_{43}N9O_8$, z = +2) 359.6696, found 359.6689; K-digested P2 ($C47H55N_{11}O_{12}$, z = +2) 483.7094, found 483.7097.

$^{18}O_2$ incorporation experiments: $^{18}O_2$-saturated buffer H was prepared by degassing buffer H on a Schlenck line followed by a refill with $^{18}O_2$ gas (97% enrichment, Cambridge Isotope Laboratory). *Pl*DarE assays with $^{18}O_2$ was performed in a way identical to the assays described above, except that an aliquot (10 μL) of each reaction mixture was added proteinase K (100 ng) and buffer K (40 μL). The resulting solution was incubated at 37 °C for one hour, and subsequently centrifugated to collect supernatant for LCMS analysis.

$H_2{}^{18}O$ incorporation experiments: *Pl*DarE assay (0.5 mL) with $H_2{}^{18}O$ was performed as described above, except that HEPES buffer was prepared in 95% ^{18}O enriched water. The final ^{18}O enrichment in the assay mixture was ~80%. P2 was purified by HPLC and characterized as described above.

Results and Discussion

In search for soluble expression in *E. coli*, we screened several DarEs from different organisms and identified that DarE from *Pseudoalteromonas luteoviolacea* (*Pl*DarE) can be expressed as a soluble protein. When *Pl*DarE was co-expressed with *P. luteoviolacea* DarA (*Pl*DarA) in *E. coli*, darobactin A production was observed (Figure S1a). The identity of darobactin A was confirmed based on the LC-MS comparison with the structurally characterized darobactin A standard (Figure S1).

Based on the observation in *E. coli*, we expressed and purified *Pl*DarE in *E. coli*. Initially, we used N-terminally His_6-tagged *Pl*DarE (His-*Pl*DarE) that was purified to >90% purity (Figure S2) with the broad light absorption feature characteristic for 4Fe-4S cluster proteins (Figure S3). As-isolated His-*Pl*DarE harbored 3.6 ± 0.9 Fe per monomer, which was increased to 8.8 ± 0.3 Fe per monomer after chemical reconstitution of the cluster. Since DarE is a member of the SPASM subfamily with conserved Cys ligands for the three 4Fe-4S clusters, supported by sequence alignment (Figure S4) and an AlphaFold structural prediction (Figure S5), the theoretical Fe content of the fully loaded DarE is 12 per monomer. Thus, the observed amount of Fe likely represents only partial loading of the three clusters.

The catalytic function of His-*Pl*DarE was assessed using recombinant *Pl*DarA. To this end, we expressed *Pl*DarA in *E. coli* as an *N*-terminal fusion with His_6-tag and a SUMO protein. After purification using a Ni-NTA column, the His-SUMO tag was removed by treating with the SUMO protease to prepare the full length and untagged DarA. The identity of the purified DarA was confirmed by LCMS (Figure S6). When the purified DarA was incubated with His-*Pl*DarE under anaerobic conditions in the presence of SAM and the flavodoxin system as a reductant, we observed formation of a product (P0, Figure 2a). This compound exhibited a UV-vis absorption spectrum similar to unmodified DarA and distinct from Darobactin A (Figure 2b), suggesting the absence of modifications on the tryptophan indole rings. LC-HRMS analysis of P0 showed the molecular weight 2 Da smaller than DarA (Figure 2c). Therefore, we tentatively assigned P0 to be DarA with a single C-C crosslink, likely between W3 and K5.

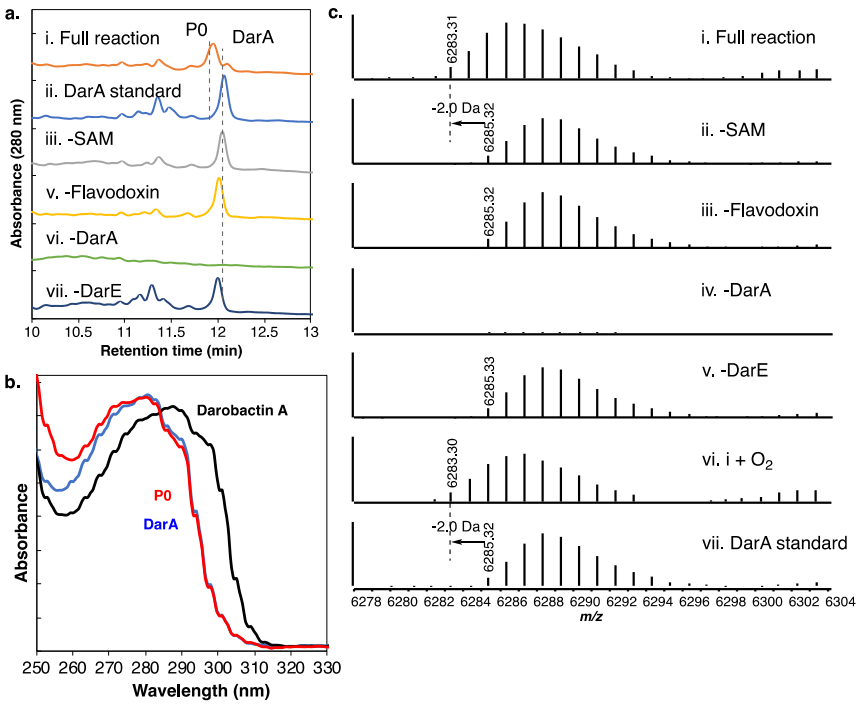


Figure 2. Characterization of His-*Pl*DarE. **a.** HPLC characterization of His-DarE assays. **b.** UV-vis spectra of P0, DarA, and darobactin A. **c.** Extracted and deconvoluted mass spectra of His-*Pl*DarE assays. Mass spectra were extracted from the retention time that covers unmodified *Pl*DarA, and P0.

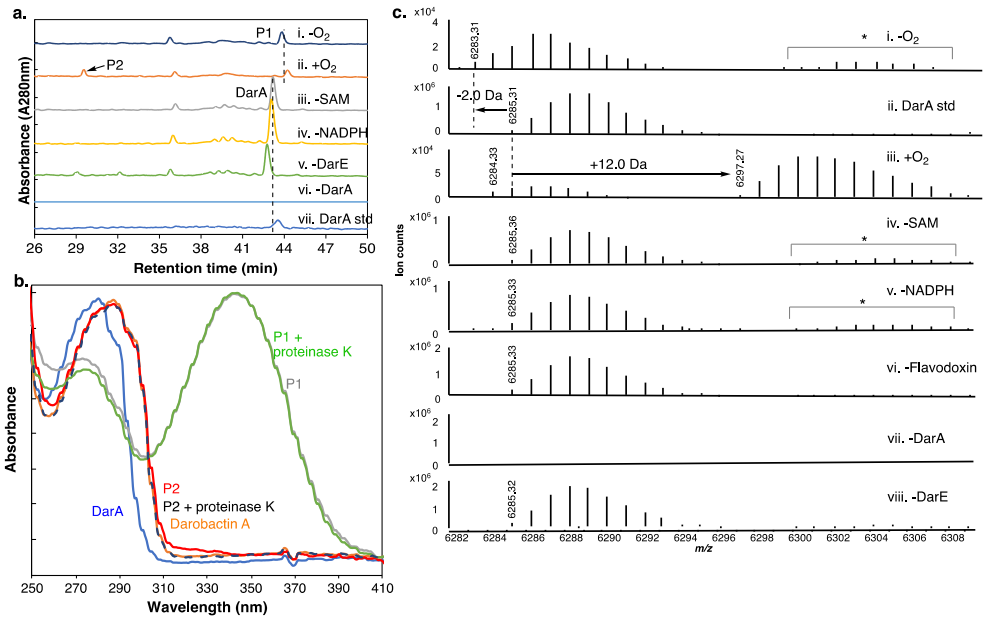


Figure 3. Activity assays of *Pl*DarE. **a.** HPLC analyses of *Pl*DarE assay in the absence of O_2 (i), in the presence of O_2 (ii), ii without SAM (iii), ii without NADPH (iv), ii without *Pl*DarE (v), ii without *Pl*DarA (vi), and *Pl*DarA standard (vii). **b.** UV-vis absorption spectra of P1 (gray), P1 + proteinase K (green), P2 (red), P2 + proteinase K (dashed black), darobactin A (orange), and *Pl*DarA (blue). **c.** Extracted and deconvoluted mass spectra of DarE assays. Mass spectra were extracted from the retention time that covers unmodified *Pl*DarA, P1, and P2. The mass peaks indicated with asterisks (*) have molecular weight close to *Pl*DarA with a 14 Da modification, but they were also present in the controls without SAM or NADPH, and hence unlikely the product of the radical SAM enzyme activity.

Since no ether crosslink formation was reproducibly observed with His-*Pl*DarE even after extensive screening of the reaction conditions, we hypothesized that the *N*-terminal His-tag may be interfering with the DarE function. Therefore, we created an *N*-terminally Strep-tagged DarE with a Factor Xa cleavage site immediately preceding the first codon of DarE (Strep-*Pl*DarE) for a traceless tag removal. This protein was expressed in *E. coli* in a manner similar to His-*Pl*DarE and purified using a Strep-tactin resin. Subsequently, the Strep-tag was cleaved by treatment with Factor Xa protease. The resulting enzyme contained 7.5 ± 0.7 eq. of Fe. Subsequent chemical reconstitution increased the Fe content to 12.2 ± 0.5 eq. per monomer, consistent with the presence of three 4Fe-4S clusters per monomer. This result is in a sharp contrast to His-*Pl*DarE, where we observed only ~9 Fe per monomer.

The resulting untagged *Pl*DarE was tested for its catalytic function using *Pl*DarA. When we investigated its activity under strictly anaerobic conditions (< 0.1 ppm O_2), we observed a single product (P1, Figure 3a, trace i) with a unique UV-absorption band at ~350 nm (Figure 3b). P1 was produced only in the presence of SAM, NADPH (for the flavodoxin and flavodoxin reductase to serve as a reductant), DarE, and DarA, supporting that it is produced by the radical SAM activity of DarE. Furthermore LC-HRMS analysis revealed its molecular weight 2 Da smaller than DarA (Figure 3c). Importantly, P1 was distinct from P0 observed in the His-*Pl*DarE reaction and P0 was not detectable in the assays with *Pl*DarE fully loaded with three 4Fe-4S clusters.

Since we did not observe any oxygen insertion, we tested the activity of DarE in the presence of oxygen. Upon testing several assay conditions, we observed the formation of two products (Figure 3a). One of them co-migrated with P1 formed in the absence of O_2, but the other product (P2) migrated at a distinct retention time. The formation of P1 and P2 accompanied the production of 5'-dA (Figure S7), supporting that their production requires the reductive SAM cleavage. The UV-vis absorption spectrum of P2 was distinct from that of P1 or DarA and agreed with that of darobactin A (Figure 3b). P2 was produced only in the presence of O_2, SAM, and *Pl*DarE (Figure 3a). In addition P2 production was observed only when flavodoxin/flavodoxin reductase was used as a reductant and not when using either sodium dithionite or Ti(III) citrate as a reductant. LCMS characterization of this species revealed its molecular weight 12 Da larger than DarA (Figure 3c), which is consistent with the presence of two crosslinks and an addition of an oxygen atom.

To better characterize P1 and P2, we purified them by HPLC (Figure S8) and digested with proteinase K. The proteinase K treatment of each of these compounds yielded only one major

product with UV absorption features identical to the parent molecules (Figure 4a and 4b). Based on the quantitation of the HPLC peak before and after the proteinase K treatment, the transformation was quantitative (>90% yield). Importantly, the proteinase K digest of P2 co-migrated with darobactin A on HPLC and LCMS (Figures 4b, and S9) and showed identical molecular weight to darobactin A (Figure 4c). Moreover, the MS/MS fragmentation pattern agreed with darobactin A (Figure S10) with the observation of the key fragment ions characteristic to the ether crosslinking, including those with hydroxyindole (Figures S10 and S11). These observations provided strong evidence that P2 has the ether and C-C crosslinks identical to darobactin A and the physiological function of DarE is the C-C and ether crosslink formation using molecular oxygen as the source of ether oxygen.

The molecular weight of the P1 proteinase K digest (Figure 4c, panel i) was consistent with a pentapeptide (WNWSK) from the core sequence of DarA with a loss of 2 Da. The MS/MS fragmentation pattern was consistent with the loss of two hydrogen atoms from the W3 residue (Figure S10 and S11). No fragment ions specific to a C-C crosslink were observed, and the unique UV-vis absorption band at 350 nm suggests the presence of an extended aromatic system. Therefore, although we cannot fully eliminate a possibility of W1-W3 C-C crosslink, we characterize P1 as DarA with Cα-Cβ desaturation of W3. To test if P1 is an intermediate of the P2 formation, we incubated purified P1 with DarE. However, even after prolonged incubation (>10 h), P1 was not consumed, and no products, including P2, were observed. Based on this observation, we propose P1 as an off-pathway shunt product.

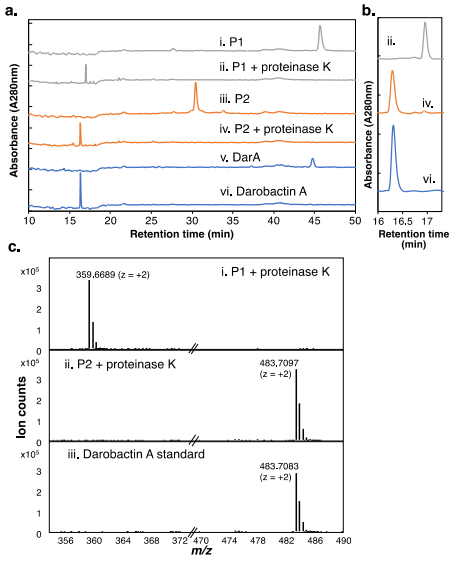


Figure 4. Characterization of *Pl*DarE products. Shown are HPLC chromatograms (**a** and **b**), and mass spectra (**c**) of P1, P2, P1 + proteinase K, P2 + proteinase K, and darobactin A standard. **b** is a magnified view of ii, iv, and vi in **a**. See Figure S9 for the EIC and full range MS spectra.

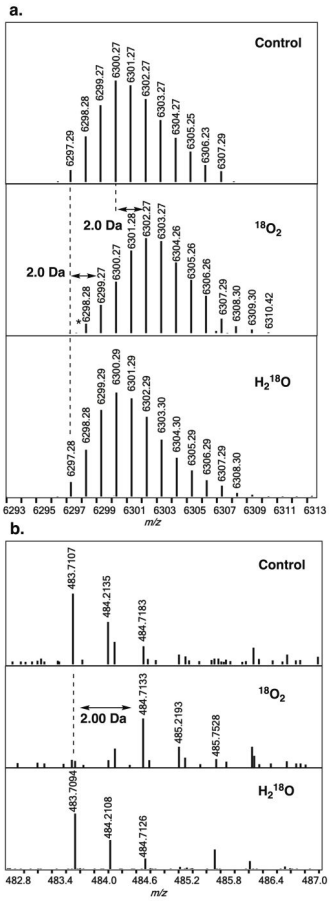


Figure 5. Characterization of P2 formed in the presence of $^{18}O_2$ or [^{18}O]water. **a.** Mass spectra of P2 formed in the presence of natural isotope O_2, $^{18}O_2$ or [^{18}O]water. **b.** Mass spectra of proteinase K digests of P2 formed in the presence of natural isotope O_2, $^{18}O_2$ or [^{18}O]water. Shown are the z = +2 ions. Calculated m/z for [^{18}O]darobactin A = 484.7110 $[M+2H]^{2+}$; Observed m/z 484.7133 (4.8 ppm). Calculated m/z for darobactin A = 483.7089 $[M+2H]^{2+}$; Observed m/z 483.7107 (3.7 ppm) and 483.7094 (1.0 ppm). * indicates the presence of natural abundance isotope P2 due to the imperfect $[^{18}O_2]O_2$ gas exchange. (see Figure S12 for calculation of the ^{18}O enrichment).

To obtain further evidence for the use of O_2 as the source of ether oxygen, we performed *Pl*DarE assay in the presence of $^{18}O_2$. The MS characterization of P2 formed in the presence of $^{18}O_2$ suggested an increase of the molecular weight by 2 Da compared to P2 from the assay with natural isotope O_2

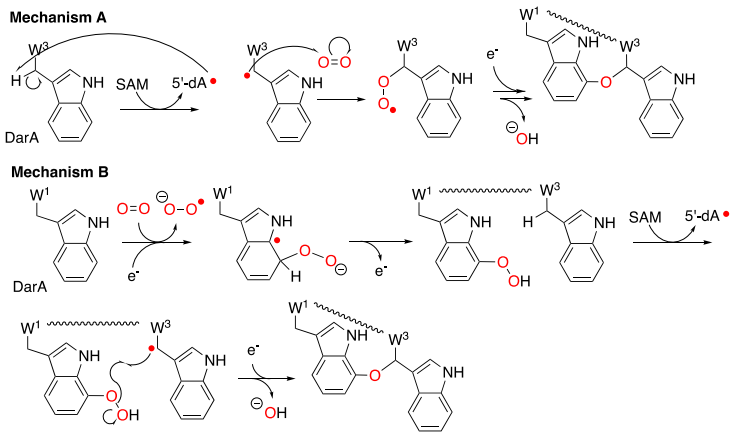
(Figure. 5a). Based on the isotope signal pattern, the ^{18}O enrichment of P2 generated in the presence of $^{18}O_2$ was calculated as 81 ± 2% (Figure S12), which likely represents the imperfect $^{18}O_2$ replacement. Furthermore, when we converted P2 produced in $^{18}O_2$ into darobactin A, we observed a 2 Da increase of the molecular weight of darobactin A (Figure 5b). The ^{18}O enrichment in darobactin A was estimated to be 85 ± 3% based on the ratio of the MS signal intensities at *m/z* 483.71 and 484.71 $[M+2H]^{2+}$. To test if ^{18}O in water is incorporated into P2, we performed the DarE assay in [^{18}O]water (80% enrichment) with natural isotope O_2. The resulting P2 and its proteinase K digest exhibited mass spectra indistinguishable from P2 formed in natural isotope water (Figure 5a and 5b). The comparison of the *m/z* 483.71 and 484.71 signal intensities in the proteinase K digest revealed no detectable level of ^{18}O-incorporation (<5%, Figure 5b). These observations together suggest that DarE incorporates one of the oxygen atoms of O_2 into the ether crosslink of P2.

Our observations of DarE's ability to incorporate ^{18}O from O_2 and not from water is distinct from the recent proposal that *N*-terminally His-tagged *P. khanii* DarE (His-*Pk*DarE) incorporates ^{18}O from water into its product (DarA with a 14 Da modification). Although we cannot eliminate the possibility that *Pl*DarE and *Pk*DarE (61% amino acid sequence identity) catalyze the ether crosslinking in two distinct mechanisms, the reported characterization of the His-*Pk*DarE product remains significant ambiguity. The structure was proposed solely based on MS without showing specific fragment ions for the ether crosslink. The authors did not describe if the DarA+14Da species can be maturated to darobactin A. No results were reported for His-*Pk*DarE assays in the presence of O_2. In addition, the *Pk*DarE sample was partially reconstituted with only 8 Fe/DarE. Therefore, further characterizations are required to make conclusions about the mechanism of ether crosslink formation by *Pk*DarE.

Still, the comparison of the results between His-*Pk*DarE and His-*Pl*DarE assays revealed an important commonality. In both cases, the major product is missing either or both of the ether and/or C-C crosslinks. His-*Pk*DarE produced *Pk*DarA +14 Da species as the major product and His-*Pl*DarE produced *Pl*DarA -2 Da (P0). Even when the His-*Pl*DarE assays were performed in the presence of O_2, P0 remained the major product (Figure 2c). Importantly, both His-*Pk*DarE and His-*Pl*DarE were loaded with only up to 8-9 Fe per monomer, whereas the untagged *Pl*DarE can be fully loaded with 12 Fe per monomer, suggesting the successful assembly of all three 4Fe-4S clusters. Therefore, the observed functional difference between His-tagged and untagged DarE is likely caused by the amount of cluster loading.

The ability of DarE to use O_2 as a co-substrate is unprecedented among radical SAM enzymes and is of significant mechanistic interest. Here, we propose two possible mechanisms of O_2-dependent ether crosslinking by DarE (Scheme 1). In mechanism A, DarE abstracts the Hβ-atom of W3 and the resulting radical adds to O_2. Subsequent ether formation requires a transfer of at least one electron and would proceed through a reductive cleavage of O-O bond or nucleophilic attack by the W1 indole ring. Whether a SAM cleavage is required for this process is unknown. Alternatively, in mechanism B, DarE reduces O_2 to superoxide that can then add to the W1 indole ring to form a peroxide intermediate. Subsequent radical-mediated C-O bond formation would yield the ether crosslinking.

Scheme 1. Proposed mechanism of DarE-catalyzed ether crosslink formation.



The proposed mechanisms of ether crosslink formation require an injection of at least one electron (Scheme 1). While the donor of this electron is currently unknown, one of the candidates is the AUX cluster. A recent report on SuiB[16], a radical SAM enzyme that catalyzes C-C crosslinking during the biosynthesis of RiPP streptide, suggested that the AUX cluster serves as an electron acceptor for the C-C bond crosslinking. Therefore, if the AUX clusters of DarE serve as both the electron donor and acceptor, their redox state must be appropriately tuned for each step of catalysis without being over-reduced or oxidized. Such redox adjustment may only be achievable by a more physiologically relevant reductant, flavodoxin, and not by chemical reductants, which may

explain the absence of ether bond formation in the presence of a chemical reductant.

While the use of O_2 as a co-substrate is unprecedented for radical SAM enzymes and was unexpected due to the oxygen-sensitive nature of radical SAM enzymes, accumulating evidence suggests that radical SAM enzymes could function in aerobic cells. A recent study of Dph1-Dph2 suggested that this non-canonical radical SAM enzyme in diphthamide biosynthesis is associated with an Fe-protein Dph3 that can donate Fe and repair an oxidatively damaged 4Fe-4S cluster of Dph1-Dph2[17]. While the mechanism of cluster maintenance and the physiological reductase are unknown for most other radical SAM enzymes, the presence of such mechanisms in cells would allow radical SAM enzymes to function under aerobic conditions. Our discovery of DarE as a radical SAM oxygenase emphasizes that at least some radical SAM enzymes function under an aerobic cellular environment and can tolerate certain levels of O_2 in the presence of the physiological reductant or 4Fe-4S cluster maintenance machinery.

ASSOCIATED CONTENT

Supporting Information

Supporting figures and a supporting table include oligo sequence, SDS-PAGE and UV-vis spectra of DarA and DarE; MS and MS/MS spectra of DarA, modified DarA, proteolyzed DarE products, darobactin A, and 5'-dA; AlphaFold model of DarE. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

* Kenichi Yokoyama (ken.yoko@duke.edu)
Till F. Schäberle (Till.F.Schaeberle@agrar.uni-giessen.de)
Yvain Nicolet (yvain.nicolet@ibs.fr)

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

This work was supported by Duke University School of Medicine and in part National Institute of General Medical Sciences R01 GM112838 (to K.Y.), and by National Institute of Allergy and Infectious Diseases R01 AI158388 (to K.L.). I Dewa M. Kresna thanks to the Deutscher Akademischer Austauschdienst (DAAD) for his PhD scholarship. The German Federal Ministry of Education and Research (BMBF) and the Hesse State Ministry of Higher Education Research and Arts (HMWK) supported research in the Schäberle lab. J. R., E. de la M. and Y.N thank the French National Research agency for financial support through the contract ANR-20-CE44-0005. They used the platforms of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), which is supported by FRISBI (ANR-10-INBS-05-02) and GRAL and financed within the Université Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003).

ACKNOWLEDGMENT

We acknowledge Leo Padva for cloning initial expression constructs and thank Yu Imai for providing darobactin A.

ABBREVIATIONS

BGC, biosynthetic gene cluster; RiPP, ribosomally synthesized and post-translationally modified peptide; SAM, *S*-adenosylmethionine; AA, amino acid.

REFERENCES

(1) Imai, Y.; Meyer, K. J.; Iinishi, A.; Favre-Godal, Q.; Green, R.; Manuse, S.; Caboni, M.; Mori, M.; Niles, S.; Ghiglieri, M.; et al. A new antibiotic selectively kills Gram-negative pathogens (vol 576, pg 459, 2019). *Nature* **2020**, *580* (7802), E3-E3. DOI: 10.1038/s41586-020-2063-9.

(2) Kaur, H.; Jakob, R. P.; Marzinek, J. K.; Green, R.; Imai, Y.; Bolla, J. R.; Agustoni, E.; Robinson, C. V.; Bond, P. J.; Lewis, K.; et al. The antibiotic darobactin mimics a beta-strand to inhibit outer membrane insertase. *Nature* **2021**, *593* (7857), 125-+. DOI: 10.1038/s41586-021-03455-w.

(3) Wuisan, Z. G.; Kresna, I. D. M.; Bohringer, N.; Lewis, K.; Schaberle, T. F. Optimization of heterologous Darobactin A expression and identification of the minimal biosynthetic gene cluster. *Metab Eng* **2021**, *66*, 123-136. DOI: 10.1016/j.ymben.2021.04.007.

(4) Gross, S.; Panter, F.; Pogorevc, D.; Seyfert, C. E.; Deckarm, S.; Bader, C. D.; Herrmann, J.; Muller, R. Improved broad-spectrum antibiotics against Gram-negative pathogens via darobactin biosynthetic pathway engineering. *Chemical Science* **2021**, *12* (35), 11882-11893. DOI: 10.1039/d1sc02725e.

(5) Sofia, H. J.; Chen, G.; Hetzler, B. G.; Reyes-Spindola, J. F.; Miller, N. E. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res* **2001**, *29* (5), 1097-1106. DOI: 10.1093/nar/29.5.1097.

(6) Broderick, J. B.; Duffus, B. R.; Duschene, K. S.; Shepard, E. M. Radical S-adenosylmethionine enzymes. *Chem Rev* **2014**, *114* (8), 4229-4317, Research Support, N.I.H., Extramural Research Support, U.S. Gov't, Non-P.H.S. DOI: 10.1021/cr4004709.

(7) Grell, T. A.; Goldman, P. J.; Drennan, C. L. SPASM and twitch domains in S-adenosylmethionine (SAM) radical enzymes. *J Biol Chem* **2015**, *290* (7), 3964-3971. DOI: 10.1074/jbc.R114.581249.

(8) Schramma, K. R.; Bushin, L. B.; Seyedsayamdost, M. R. Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink. *Nat Chem* **2015**, *7* (5), 431-437. DOI: 10.1038/nchem.2237.

(9) Nguyen, T. Q. N.; Tooh, Y. W.; Sugiyama, R.; Nguyen, T. P. D.; Purushothaman, M.; Leow, L.; Hanif, K.; Yong, R. H. S.; Agatha, I.; Winnerdy, F. R.; et al. Post-translational formation of strained cyclophanes in bacteria. *Nature Chemistry* **2020**, *12* (11), 1042-+. DOI: 10.1038/s41557-020-0519-z.

(10) Fluhe, L.; Knappe, T. A.; Gattner, M. J.; Schafer, A.; Burghaus, O.; Linne, U.; Marahiel, M. A. The radical SAM enzyme AlbA catalyzes thioether bond formation in subtilosin A (vol 8, pg 350, 2012). *Nature Chemical Biology* **2012**, *8* (8), 737-737. DOI: 10.1038/nchembio0812-737a.

(11) Clark, K. A.; Bushin, L. B.; Seyedsayamdost, M. R. Aliphatic Ether Bond Formation Expands the Scope of Radical SAM Enzymes in Natural Product Biosynthesis. *J Am Chem Soc* **2019**, *141* (27), 10610-10615. DOI: 10.1021/jacs.9b05151.

(12) Guo, S.; Wang, S.; Ma, S.; Deng, Z.; Ding, W.; Zhang, Q. Radical SAM-dependent ether crosslink in daropeptide

biosynthesis. *Nat Commun* **2022**, *13* (1), 2361. DOI: 10.1038/s41467-022-30084-2 From NLM Medline.

(13) Grote, A.; Hiller, K.; Scheer, M.; Münch, R.; Nörtemann, B.; Hempel, D. C.; Jahn, D. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* **2005**, *33* (suppl_2), W526-W531. DOI: 10.1093/nar/gki376 (acccessed 5/20/2022).

(14) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **2009**, *6* (5), 343-345. DOI: 10.1038/nmeth.1318.

(15) Fish, W. W. Rapid Colorimetric Micromethod for the Quantitation of Complexed Iron in Biological Samples. *Methods in enzymology* **1988**, *158*, 357-364. DOI: Doi 10.1016/0076-6879(88)58067-9.

(16) Balo, A. R.; Caruso, A.; Tao, L.; Tantillo, D. J.; Seyedsayamdost, M. R.; Britt, R. D. Trapping a cross-linked lysine-tryptophan radical in the catalytic cycle of the radical SAM enzyme SuiB. *Proceedings of the National Academy of Sciences of the United States of America* **2021**, *118* (21), e2101571118. DOI: 10.1073/pnas.2101571118.

(17) Zhang, Y. G.; Su, D.; Dzikovski, B.; Majer, S. H.; Coleman, R.; Chandrasekaran, S.; Fenwick, M. K.; Crane, B. R.; Lancaster, K. M.; Freed, J. H.; et al. Dph3 Enables Aerobic Diphthamide Biosynthesis by Donating One Iron Atom to Transform a [3Fe-4S] to a [4Fe-4S] Cluster in Dph1-Dph2. *J Am Chem Soc* **2021**, *143* (25), 9314-9319. DOI: 10.1021/jacs.1c03956.

Supporting information for

Characterization of DarE as a radical SAM oxygenase for the ether crosslinking in darobactin biosynthesis

Hai Nguyen[a], I Dewa M. Kresna[b], Nils Böhringer[b,c], Jeremie Ruel[d], Eugenio de la Mora[d], Jil-Christine Kramer[b], Kim Lewis[f], Yvain Nicolet[d,*], Till F. Schäberle[b,c,e,*], Kenichi Yokoyama[a,g,*]

[a] Department of Biochemistry, Duke University School of Medicine, Durham, NC 27710

[b] Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, Ohlebergsweg 12, 35392 Giessen, Germany

[c] German Center for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Giessen, Germany

[d] University of Grenoble Alpes, CEA, CNRS, IBS, Metalloproteins Unit, F-38000 Grenoble, France

[e] Department of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Ohlebergsweg 12, 35392 Giessen, Germany

[f] Antimicrobial Discovery Center, Department of Biology, Northeastern University, Boston, MA 02115

[g] Department of Chemistry, Duke University, Durham, NC 27710

Table of Contents

Table S1. PCR primers

Name	Sequence
darA-f	5'-GTATAAGAAGGAGATATACAATGATCGTTGAAGCTCCGAA-3'
darA-r	5'-TGCAGCTACCTCCTTAGAAAGATTTAGACCAGTTCCAAGCG-3'
darE-f	5'-CTTTCTAAGGAGGTAGCTGCAATGCTGGGTGACATCTCTGT-3'
darE-r	5'-TGCTCAGCGGTGGCAGCAGCTTAGATAGATTCTTCGATAACTTTCAGCAG-3'
his-darE-f	5'-ACCATCATCACCACAGCCAGATGCTGGGTGACATCTCTGT-3'
his-darE-r	5'-ATTATGCGGCCGCAAGCTTGTTAGATAGATTCTTCGATAACTTTCAGCAG-3'
pRSF-strep-DarE-f	CCGCAGTTCGAGAAAAGCGGCATCGAAGGTCGTATGCTGGGTGACATCTCTGTTAAAG
pRSF-strep-DarE-r	GTGTGACCAAGCACTTGCCATGGTATATCTCCTTATTAAAGTTAAAC
strep-f	AGTGCTTGGTCACACCCCCAATTTGAGAAGGGTGGTGGCTCCGGCGGTGGCTCTGGCGG-TAGCGCGTGGAGCCACCCGCAGTTCGAGAAA
strep-r	TTTCTCGAACTGCGGGTGGCTCCACGCGCTACCGCCAGAGCCACCGCCGGAGCCAC-CACCCTTCTCAAATTGGGGGTGTGACCAAGCACT
pET-strep-DarE-f	AAAAACATATGGCAAGTGCTTGGTCACACC
pET-strep-DarE-r	AAAAAAAGCTTTTAGATAGATTCTTCGATAACTTTCAGC

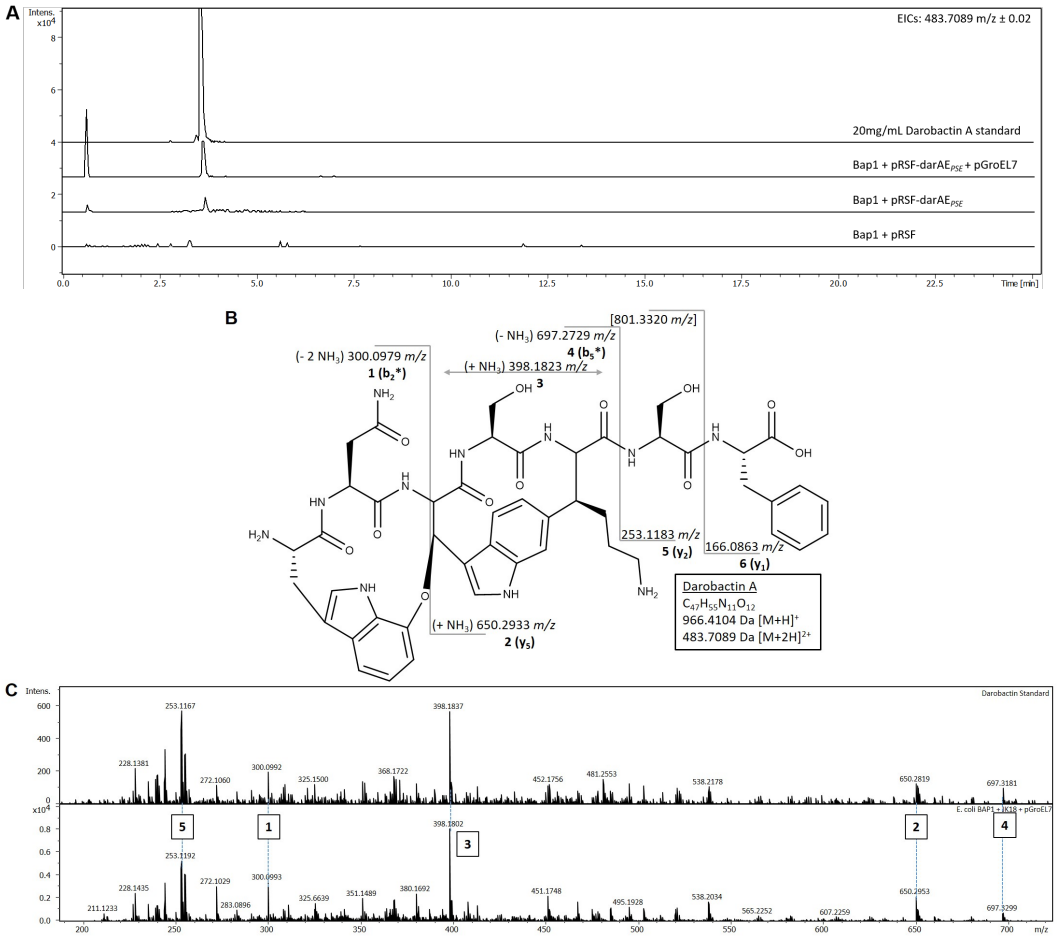


Figure S1. Production of darobactin A in *E. coli*.

A) Extracted ion chromatograms (EICs) of selected *E. coli* culture extracts and a darobactin A standard. B) Darobactin A structure with signature fragment ions indicated[1]. C) MS/MS spectra comparison of a darobactin A standard and the corresponding peak of the heterologous expression using construct pRSF-darAE$_{PSE}$. The reported signature ions are indicated.

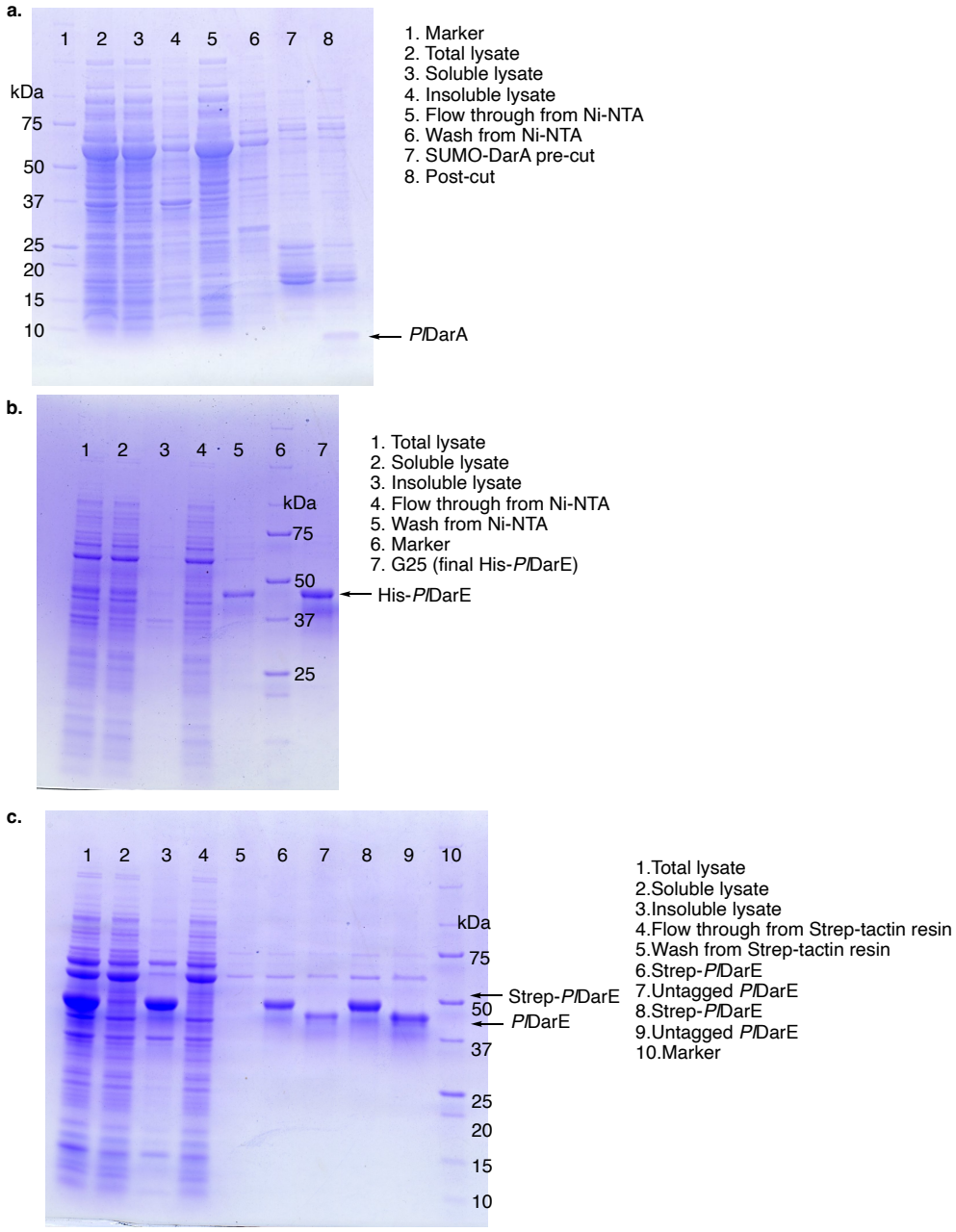


Figure S2. SDS-PAGE of *Pl*DarA and *Pl*DarE purification.

Shown are SDS-PAGE for *Pl*DarA (**a**), His-*Pl*DarE (**b**), and Strep-*Pl*DarE (**c**).

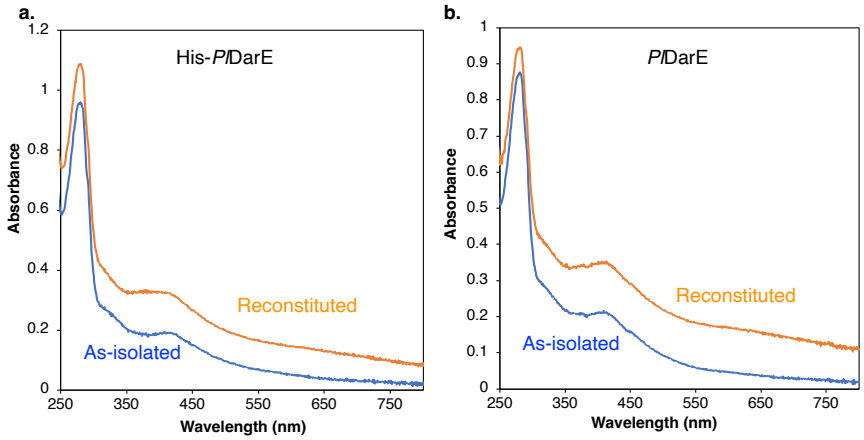


Figure S3. UV-vis spectra of DarE.

Shown are UV-vis spectra of His-*Pl*DarE (**a**) and untagged *Pl*DarE (**b**) before and after the 4Fe-4S cluster re-constitution. Concentrations of all the proteins were normalized to 10 μM.

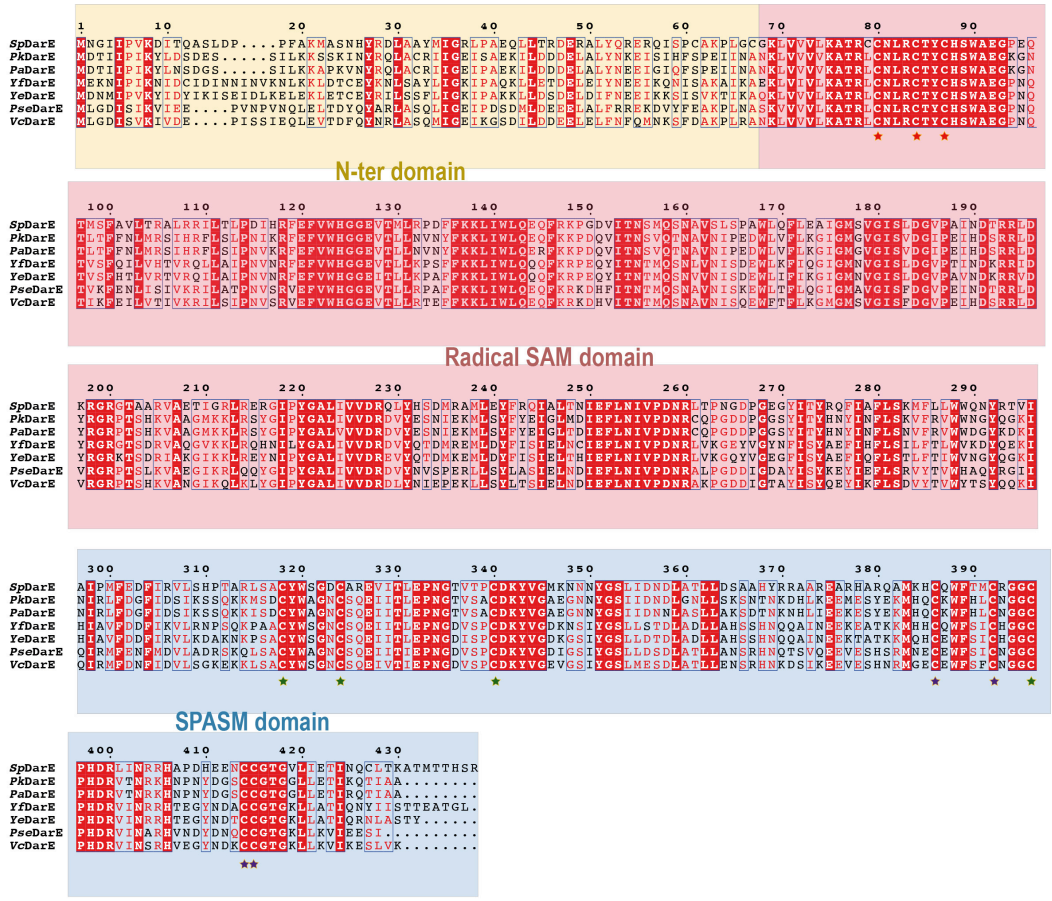


Figure S4. Sequence alignment of DarE.

Multiple sequence alignment of different DarE proteins from *Sodalis praecaptivus* (*Sp*), *Photorhabdus khanii* NC19 (*Pk*), *Photorhabdus asymbiotica subsp. Asymbiotica* (*Pa*), *Yersinia frederiksenii* (*Yf*), *Yersinia enterocolitica subsp. Enterocolitica* (*Ye*), *Pseudoalteromonas luteoviolacea* S4054 (*Pse*) and *Vibrio crassostreae* (*Vc*). Residues strictly and semi-conserved are indicated in red boxes and in red, respectively. The three domains are indicated by rectangles in gold (N-ter), pink (radical SAM) and blue (SPASM) rectangles. The conserved cysteine residues are highlighted by small stars (radical SAM, auxiliary I and II clusters in red, green and purple, respectively).

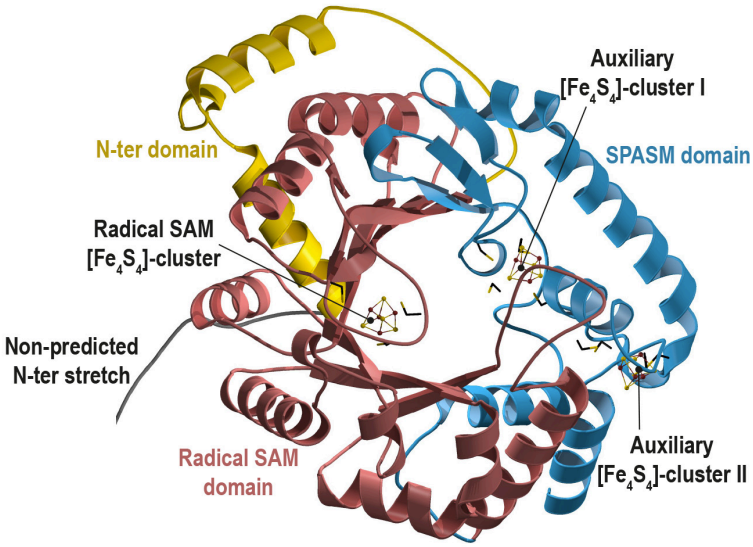
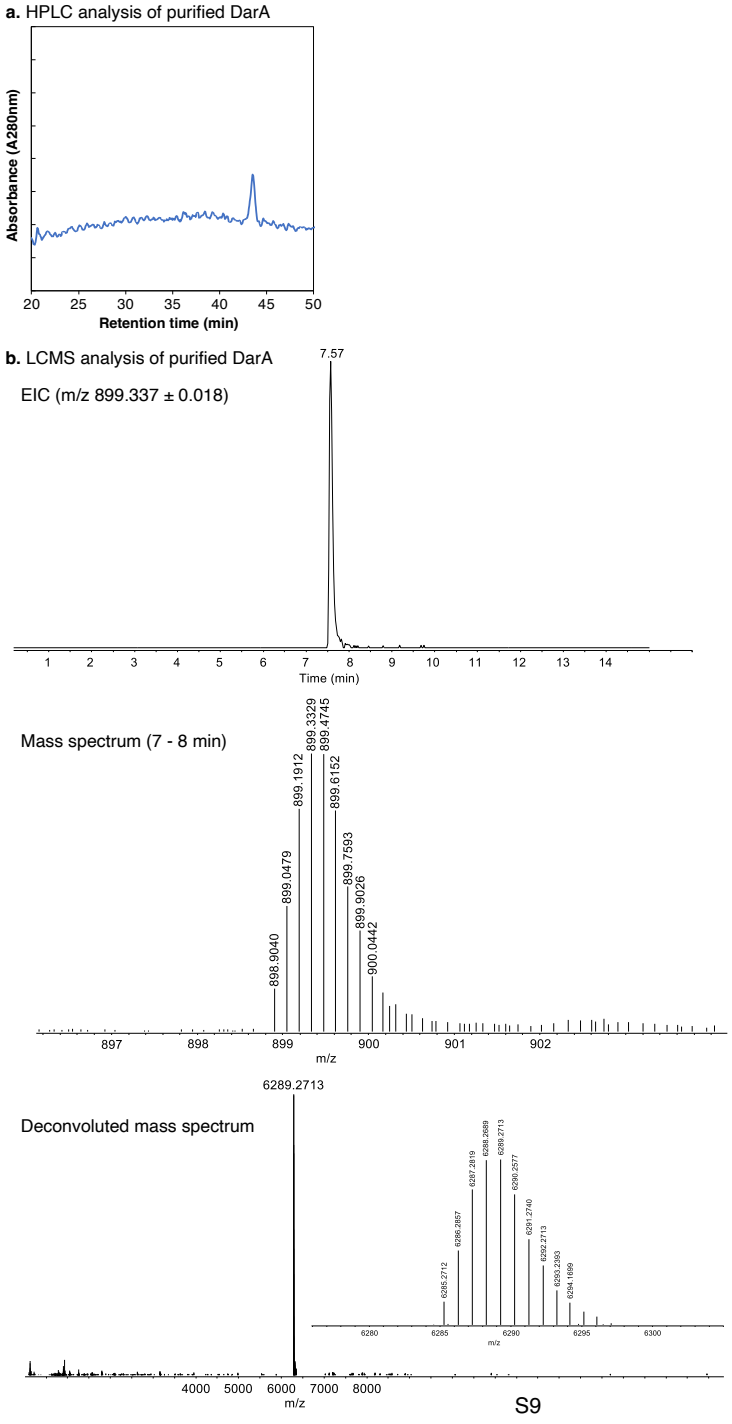


Figure S5. AlphaFold model of the *Pl*DarE.

Pseudoaltermononas luteoviolacea DarE structure prediction using the AlphaFold algorithm[2]. Besides the fifteen first residues (depicted in gray), the overall model displays good prediction scores. According to this prediction, DarE should best resemble the anSME crystal structure (PDB ID 4K37; 10.1073/pnas.1302417110), sharing an RMSD of 2.88Å for 292 aligned Cα atoms despite only 28% sequence identity. The model can be split into three distinct structural domains: a specific N-terminal stretch (in gold), the radical SAM domain (in pink) and a C-terminal SPASM domain (in blue), often found in radical SAM enzymes[3]. SPASM domains contain conserved cysteine residues responsible for the binding of two auxiliary [Fe_4S_4]-clusters. Hence, the predicted model displays three clusters of strictly conserved cysteine residues. C80, C84 and C87 from the Cx_3Cx_2C motif are located at the top of the radical SAM domain and are responsible for the binding of the radical SAM [Fe_4S_4]-cluster. C318, C324, C340 and C396 on one hand and C386, C392, C414 and presumably C415 on the other hand are presumably responsible for the binding of the auxiliary clusters I and II, respectively, hence supporting the existence of three [Fe_4S_4] clusters in DarE. The cysteine residues are depicted as sticks. [Fe_4S_4]-clusters deduced from superposition of anSME structure (PDB ID 4K37) are also presented as ball-and-sticks, further supporting the existence and location of these clusters in DarE.


a. HPLC analysis of purified DarA
Absorbance (A280nm)
20 25 30 35 40 45 50
Retention time (min)
b. LCMS analysis of purified DarA
EIC (m/z 899.337 ± 0.018)
7.57
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Time (min)
Mass spectrum (7 - 8 min)
898.9040
899.0479
899.1912
899.3329
899.4745
899.6152
899.7593
899.9026
900.0442
897 898 899 900 901 902
m/z
Deconvoluted mass spectrum
6289.2713
6285.2712
6286.2857
6287.2819
6288.3088
6289.2713
6290.2577
6291.2740
6292.2713
6293.2363
6294.1699
6280 6285 6290 6295 6300
m/z
4000 5000 6000 7000 8000
m/z

S9

Figure S6. Characterization of purified DarA.

a. HPLC chromatogram of DarA monitoring the light absorption at 280 nm. The chromatography was performed on a XSelect CSH Phenyl-Hexyl column (Waters) **b.** LC-HRMS analysis of DarA. Calculated monoisotropic *m/z* = 898.9070 $[M+7H]^{+7}$. Observed m/z = 898.9040 (3.3 ppm). Calculated monoisotropic molecular weight is 6285.296. Observed molecular weight 6285.271 (4.0 ppm). The chromatography was performed on a Phenomenex Kinetix EVO C18 (3x100mm, 2.6 um).

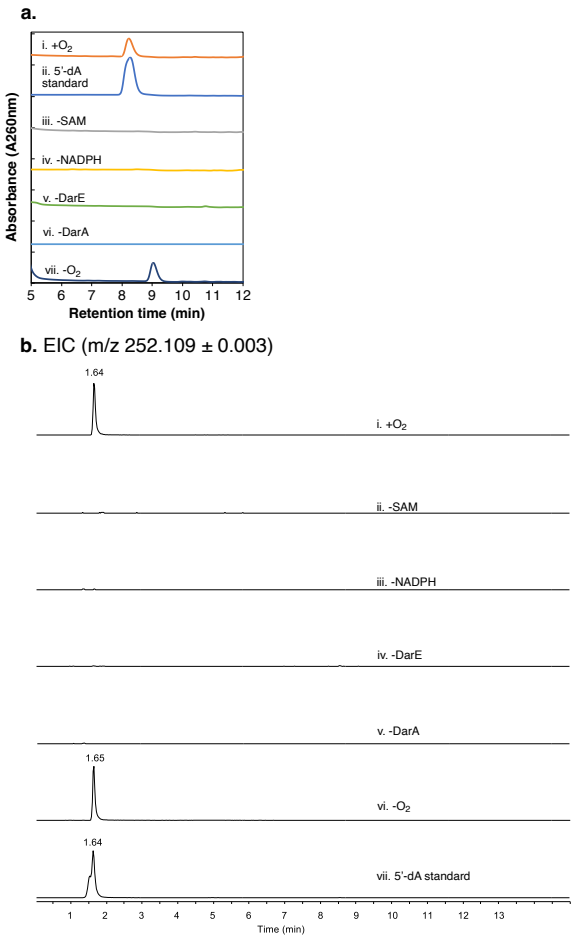


Figure S7. Characterization of 5'-dA in DarE assays.

HPLC (**a**) and LCMS (**b**) analyses of DarE assays. Calculated m/z = 252.1091 $[M+H]^+$. Observed m/z = 252.1099 (3.2 ppm).

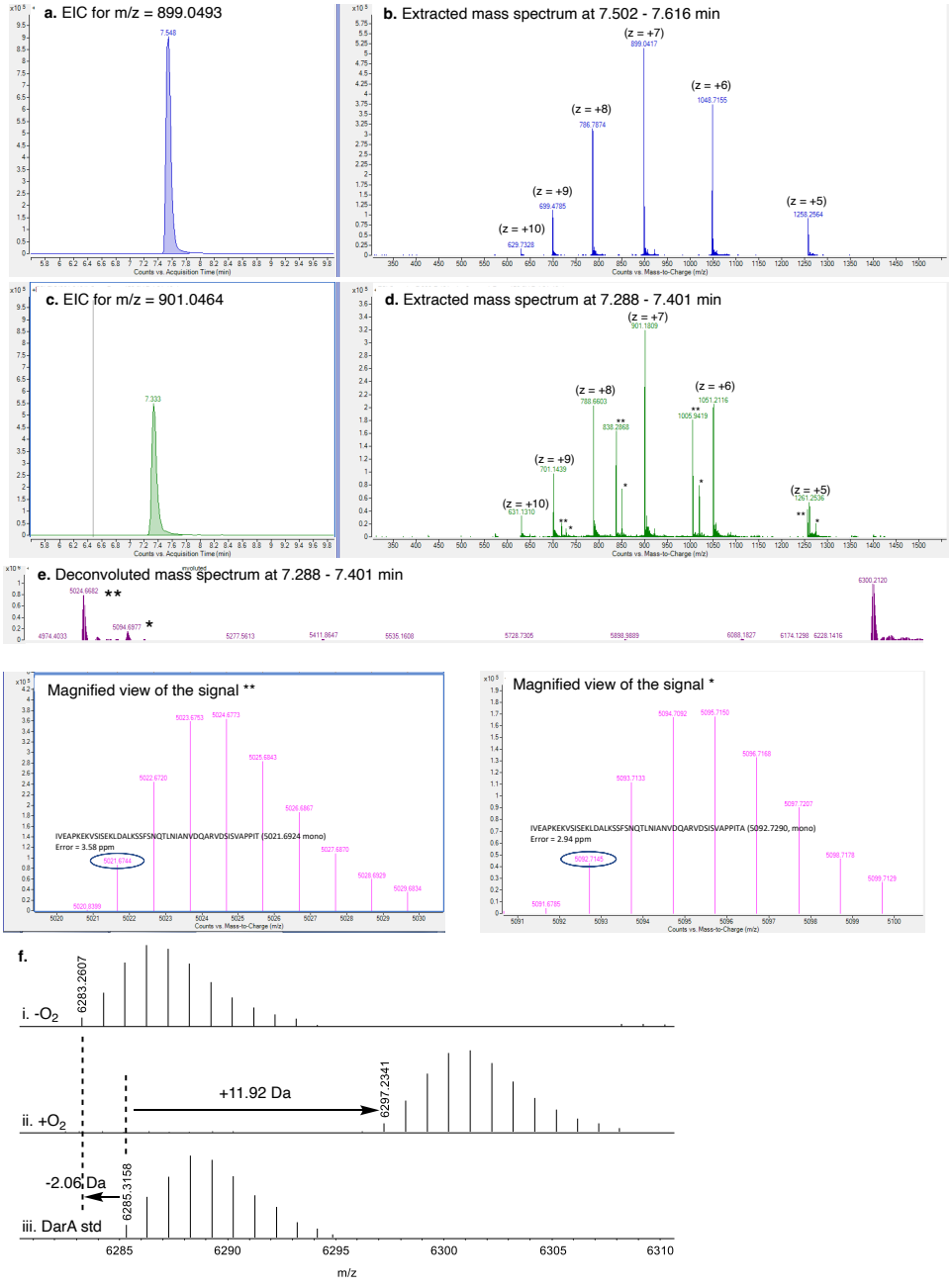

a. EIC for m/z = 899.0493
b. Extracted mass spectrum at 7.502 - 7.616 min
(z = +7)
(z = +6)
(z = +8)
(z = +9)
(z = +5)
(z = +10)
c. EIC for m/z = 901.0464
d. Extracted mass spectrum at 7.288 - 7.401 min
e. Deconvoluted mass spectrum at 7.288 - 7.401 min
Magnified view of the signal **
IVEAPKEKVSISEKLDALKSSFSNQTLNIANVDQARVDSISVAPPIT (5021.6924 mono)
Error = 3.58 ppm
Magnified view of the signal *
IVEAPKEKVSISEKLDALKSSFSNQTLNIANVDQARVDSISVAPPITA (5092.7290, mono)
Error = 2.94 ppm
f.
i. -O2
6283.2607
ii. +O2
+11.92 Da
6297.2341
-2.06 Da
6285.3158
iii. DarA std
6285
6290
6295
6300
6305
6310
m/z

Figure S8. LCMS analysis of purified P1 and P2.

Extracted ion chromatograms (EIC) and mass spectra of P1 (**a** and **b**) and P2 (**c** and **d**). The EIC window was set to 20 ppm. Deconvoluted mass spectrum of P2 (**e**). Shown in asterisks are the leader peptide (*) and the leader peptide without the C-terminus Ala (**) copurified with P2. The signals shown in * and ** correspond to that of the leader peptides or the leader peptide without the C-terminal Ala residue, which migrated very close to but with slightly different retention time compared to P2. **f.** Comparison of the deconvoluted spectra of P1, P2, and unmodified DarA.

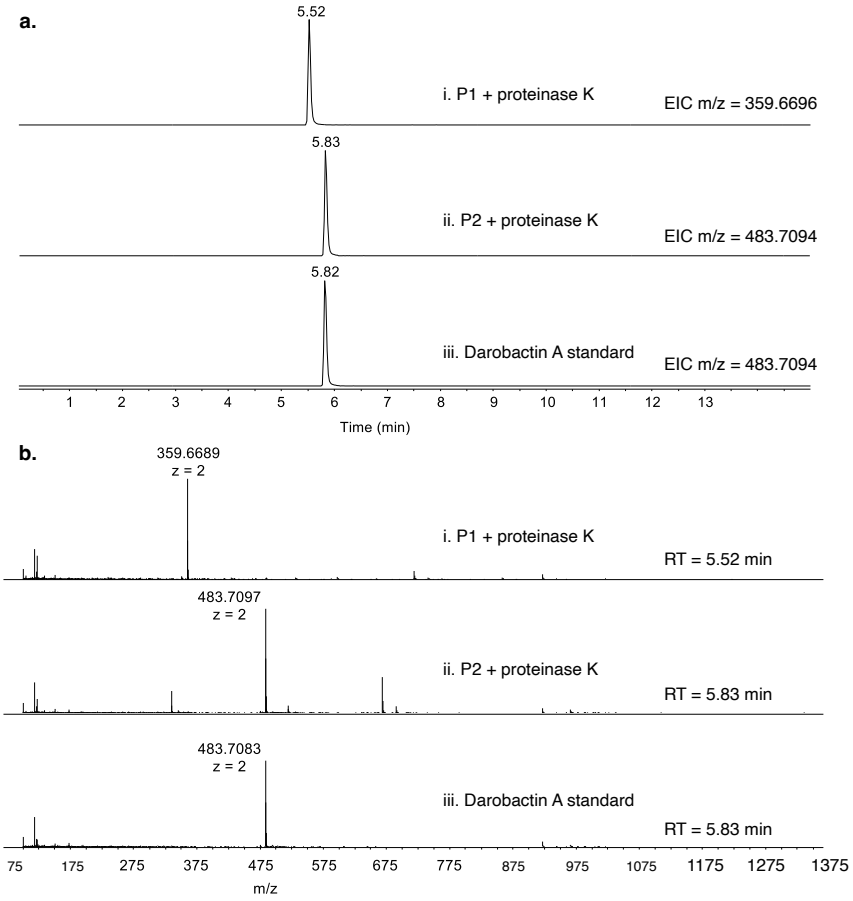


Figure S9. LCMS analysis of proteinase K-digested P1 and P2.

Shown are the EIC (**a**) and extracted mass spectra (**b**) of the indicated samples. EIC window was 20 ppm

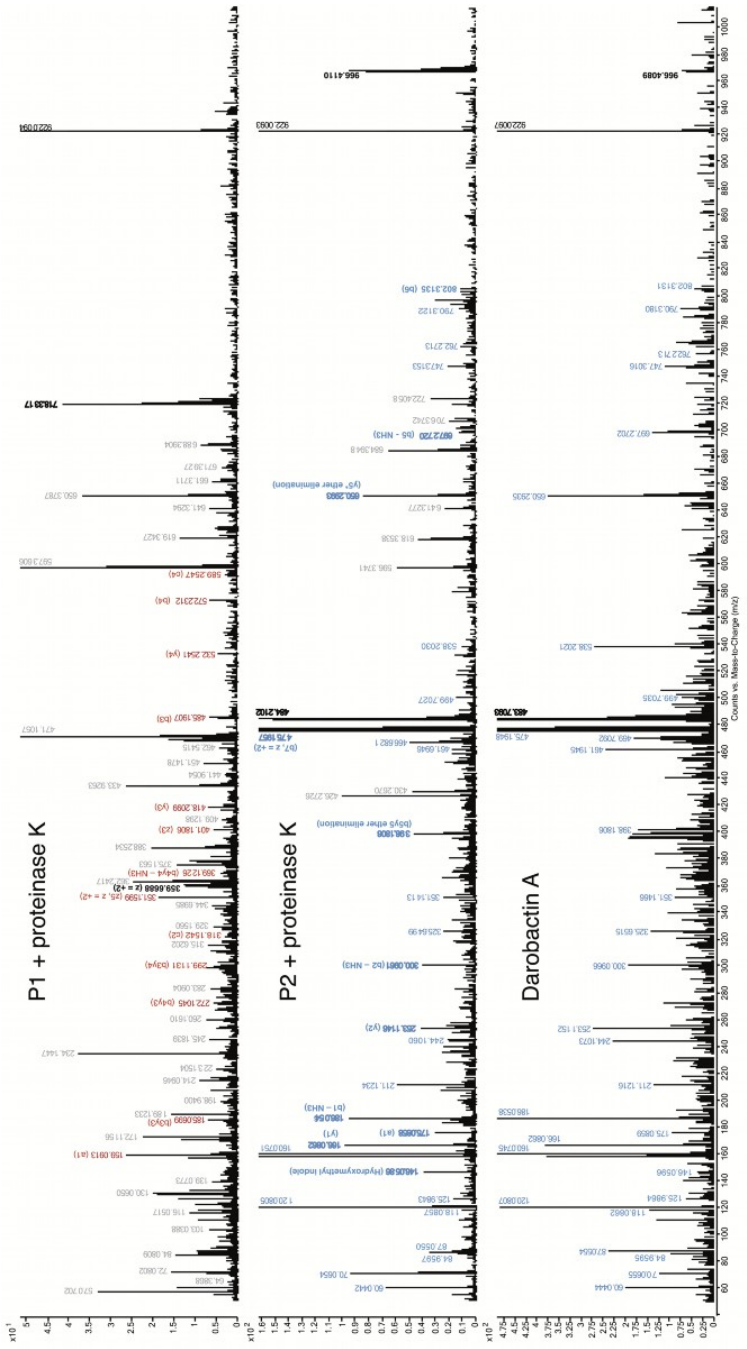
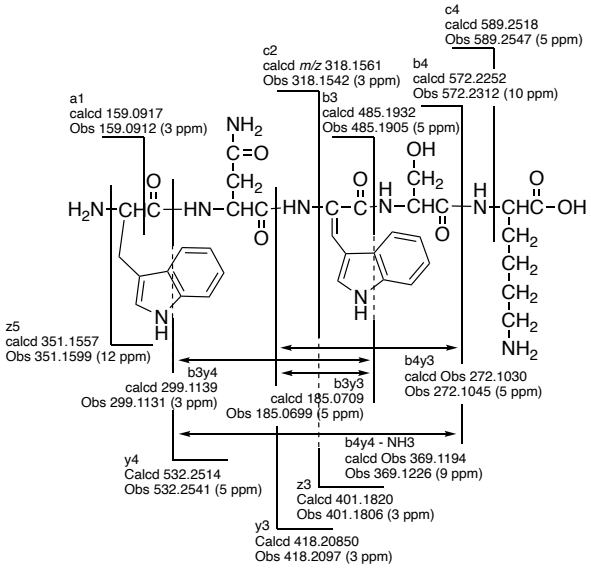

P1 + proteinase K
P2 + proteinase K
Darobactin A
Counts vs. Mass-to-Charge (m/z)

Figure S10. MS/MS analysis of darobactin A and proteinase K-digested P1 and P2.
Signal assignments are shown in brackets. In the P1 + proteinase K spectrum, assigned signals are shown in magenta. Signals observed in both darobactin A and the P2 + proteinase K were shown in blue and those signals that have been assigned were shown in bold.

a. Fragment ions observed in P1 digested with proteinase K



b. Fragment ions observed in P2 digested with proteinase K

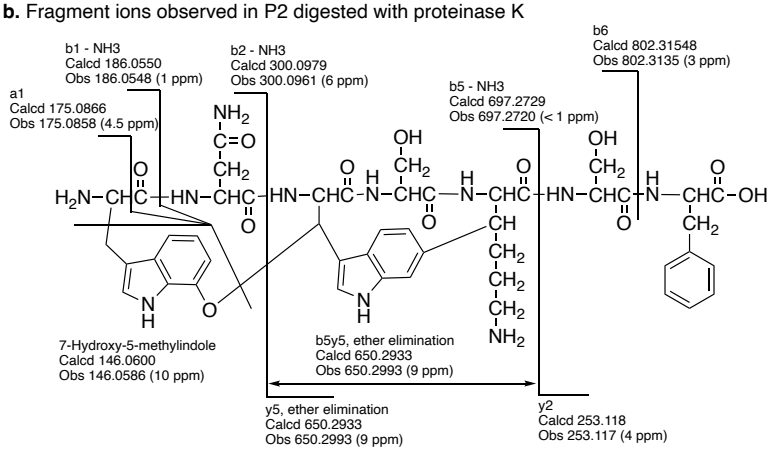


Figure S11. Structural assignment of the fragment ions observed in the MS/MS analysis of proteinase K-digested P1 (a) and P2 (b).

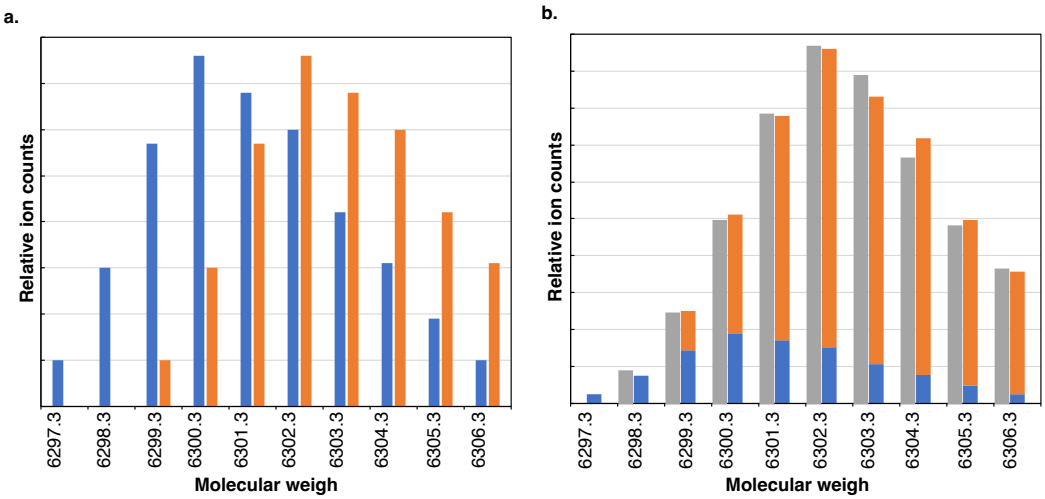


Figure S12. Quantitation of the ^{18}O enrichment in P2 formed in the presence of $[^{18}O_2]O_2$.

^{18}O enrichment of P2 formed in the presence of $[^{18}O_2]O_2$ was determined by comparing the relative abundance of isotope mass signals. The relative abundance of the MS signals for natural isotope P2 (0% ^{18}O enrichment) were first experimentally determined (blue bars in **a**) and then used to calculate the isotope distribution of P2 with 100% ^{18}O enrichment (orange bars in **a**). Then, the experimentally determined mass spectrum of P2 formed in the presence of $^{18}O_2$ (gray bars in **b**) was simulated as a sum of mass spectra of natural isotope P2 (blue bars in **b**) and 100% ^{18}O-labeled P2 (orange bars in **b**). The ^{18}O enrichment was determined based on the ratio of the two reference spectra by minimizing the difference between the simulated and experimental mass spectra of [^{18}O]P2. The panel **b** shows the result of simulation with 81% ^{18}O enrichment.

Reference

(1) Böhringer, N.; Green, R.; Liu, Y.; Mettal, U.; Marner, M.; Modaresi, S. M.; Jakob, R. P.; Wuisan, Z. G.; Maier, T.; Iinishi, A.; et al. Mutasynthetic Production and Antimicrobial Characterization of Darobactin Analogs. *Microbiol Spectr* **2021**, *9* (3), e01535-01521. DOI: doi:10.1128/spectrum.01535-21.
(2) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589. DOI: 10.1038/s41586-021-03819-2.
(3) Grell, T. A.; Goldman, P. J.; Drennan, C. L. SPASM and twitch domains in S-adenosylmethionine (SAM) radical enzymes. *J Biol Chem* **2015**, *290* (7), 3964-3971. DOI: 10.1074/jbc.R114.581249.

7. Chapter VI

Systematic investigation of DarE flexibility generating a new darobactin derivative with broader gram-negative bioactivity

Kresna, I D. M., Wuisan, Z. G., Joshi, S., Mettal, U., Patras, M., Brinkmann, S., Böhringer, N., Marner, M., Brinkrolf, K., Reiter, S., Schmitt, S., Schäberle, T. F.

Status: Gathering data

Summary: The substrate flexibility of the radical SAM enzyme DarE was investigated systematically. Therefore, the part of the *darA* gene that is encoding for the heptapeptide core was mutated. Every proteinogenic amino acid was thereby introduced to each position, yielding possible derivatives library. Each construct was expressed heterologously in *E. coli* and the extracts were subsequently analyzed by UPLC-MS. In addition, a qualitative antimicrobial overlay assay was performed. In that way, the substrate flexibility of DarE was revealed and the antibiotic activity of the derivatives enables first insights into structure-activity-relationships. However, minimum inhibitory concentration (MIC) determination of the derivatives directly from extracts was unsuccessful. Furthermore, a bigger library containing all possible combinations of amino acids was also generated and the high throughput methods to screen the library for activity are still under optimization.

Contribution: I Dewa Made Kresna designed and planed the strategies and experimental set up. He performed the experiments, analyzed the results and put them into a discussion.

Systematic investigation of DarE flexibility generating a new darobactin derivative with broader gram-negative bioactivity

I Dewa M. Kresna[a,b], Zerlina G. Wuisan[a,b,c], Sanjata Joshi,[a] Ute Mettal,[a,b] Maria Patras,[b] Stephan Brinkmann,[b] Nils Böhringer,[a,b,c] Michael Marner,[b] Karina Brinkrolf,[b] Silke Reiter,[d] Steven Schmitt,[d] Till F. Schäberle[a,b,c]

[a] Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, Ohlebergsweg 12, 35392 Giessen, Germany.

[b] Department of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Ohlebergsweg 12, 35392 Giessen, Germany.

[c] German Center for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Giessen, Germany.

[d] Bioprocess Laboratory, Department of Biosystem Science and Engineering, ETH Zurich, Basel, Switzerland.

Abstract

Darobactin has shown impressive gram-negative antimicrobial and low toxicity properties making it a potent candidate in the drug development pipeline. Thus, it is an interesting compound to be derivatized to further increase its potential. The biosynthetic gene cluster (BGC) has been characterized and heterologous expression was achieved. Using this established system, the corresponding genes were manipulated to produce derivatives. The codons encoding for an amino acid of the core darobactin peptide were systematically mutated to encode for other proteogenic residues. The resulting derivatives library enabled the determination which substrates are accepted by DarE – the only modification enzyme involved in darobactin maturation. By LCMS analysis it was observed that 68 out of 133 heptapeptides were maturated by intramolecular ring closures. Furthermore, an antimicrobial overlay assay using *E. coli* as test pathogen was performed. This revealed that only 36 bicyclic heptapeptides resulted in a detectable inhibition zone. Due to the fact that a direct approximation of the minimum inhibitory concentration (MIC) directly from the extracts was unsuccessful, selected ones were expressed and isolated for further analysis. In a next step, a rational attempt was done by combining amino acid exchanges in one construct (B9), which indeed showed an improved and broader gram-negative activity profile. Furthermore, a big randomized library was constructed containing all possible amino acid combinations at position 2, 4, 6 and 7 in the heptapeptide core of DarA was generated. The high throughput activity screening for active products from this big library is still underway.

Introduction

Multi resistance gram negative pathogens such as *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacteriaceae* have been categorized as a threat for human health prioritizing them as pathogens against which drugs must be developed.[1] Sustainable discovery and development of new antibiotics are essential, since pathogens evolve resistance to traditional antibiotics. In 2019, darobactin A that selectively kills gram-negative pathogens was reported.[2] Darobactin is a ribosomally synthetized and post translationally modified peptide (RiPP) antibiotic, which has seven amino acids (AA) in its core peptide – W^{1}-N^{2}-W^{3}-S^{4}-K^{5}-S^{6}-F^{7}. Its biosynthetic gene cluster (BGC) consists of the structural gene *darA* encoding a precursor peptide, *darBCD* encoding transport-related proteins, and *darE* encoding a radical SAM enzyme.[2, 3] By expressing *darA* and *darE* only, darobactin production can be achieved, concluding that DarE is the only modification enzyme necessary for darobactin maturation.[3] DarE catalyzes the formation of a *C-O-C* bond between W^{1} and W^{3} and a *C-C* bond between W^{3} and K^{5} on the core peptide of DarA. Thereafter, the DarA leader and follower peptide are cleaved off, releasing darobactin prior to secretion by the ABC transporter DarBCD.[2] Like many RiPP precursor peptides, DarA consists of a leader, core and follower peptide. The protease activity to release the core peptide seems to be independent of DarE. The release of the modified heptapeptide from the precursor was suspected to be catalyzed by still unidentified protease(s) from the host or by possible self-cleavage of DarA.[3] Recently, Groß and coworkers characterized the protease DarF from the darobactin producer *Pseudoalteromonas luteoviolacea* S4054.[4] However, DarF activity towards the leader or the follower peptide was not described. Instead, DarF might have a role as a feedback-regulated self-detoxification or self-resistance mechanism, in which high darobactin concentrations lead to *darF* gene expression and consequently the degradation of darobactin A. Thereby, setting a limit to darobactin concentrations not harmful for the native producer.[4] Furthermore, the follower peptide has been reported to be not essential, since its removal did not significantly interfere with darobactin production yields.[3]

Derivatization of existing hit compounds plays an important role for developing them into a lead compound[5]. So far, from database-guided discovery, six other darobactin derivatives have been identified and initially characterized – darobactin B-F.[6] Further derivatization by engineering the biosynthetic pathway has led to many un-natural darobactin derivatives – darobactin 1-21.[4] This has been possible by employing a heterologous expression system.

Heterologous expression systems use related host species to express the foreign BGCs. It is beneficial if the host is well investigated and cultivation conditions and DNA manipulation techniques are established,[7] as it was exemplified by reaching higher production rates of darobactin in *E. coli* and *Vibrio natriegens* cells.[3,4]

Besides the aforementioned six natural darobactin derivatives, there might be other derivatives in nature, which have not been investigated so far and thus are not present in the database yet. The observed limitations in respect to darobactin derivatives might be ascribed to the flexibility of DarE towards its substrate (the core peptide) or to the fact that other "not conserved" AAs might abolish the activity properties, which would lead to deselection of these during evolution. In the present study, the flexibility of the core peptide in sequence data was investigated to answer the question which AAs are naturally observed at which position and which combination could result in better darobactin properties.

Result and Discussion

From a BLASTp investigation using the DarA core peptide sequence as the query against the NCBI database, *dar* BGCs could be identified from *Photorhabdus*, *Yersinia*, *Pseudoalteromonas*, and *Vibrio* strains (Figure S1). In this study, the basis was the BGC originating from *Photorhabdus khanii* HGB1456. The BGC was cloned into the pRSFduet vector (pZW-ADC5Δfol)[3]. However, the codons encoding the follower peptide of DarA were removed. PCR amplification using degenerated or non-degenerated primers was performed and Gibson assembly was employed to clone the modified core peptides (Figure 1). The resulting vectors were transferred into the expression host *E. coli* BAP1 and were used to heterologously express darobactin derivatives. The latter host strain has shown a better production of darobactin A compared to *E. coli* BL21(DE3).[3] After fermentation, the cell pellet was separated from the culture medium, lysed and centrifuged. The resulting supernatant was subjected to LCMS analysis to detect the mass corresponding to darobactin derivatives. The expected mass of a corresponding derivative can be calculated from the mass of the core peptide, taking into account the two characteristic ring closures – a *C-O-C* and a *C-C* bond. In the control experiment, the host producer carrying an empty pRSFduet vector was also treated likewise.

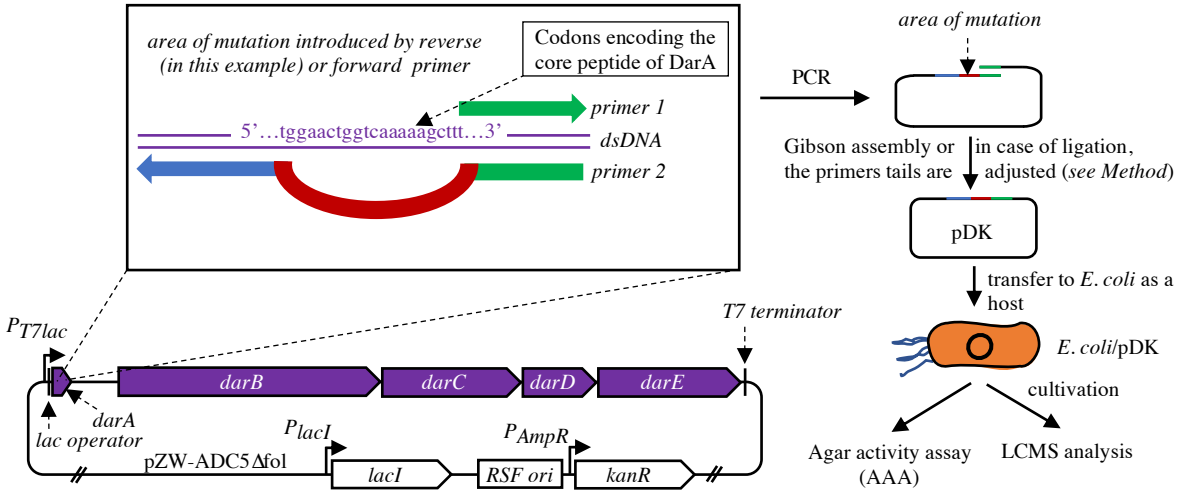


Figure 1. General schematic overview of the workflow for the generation of modified darobactin BGCs based on plasmid pZW-ADC5Δfol. The BGC originates from *P. khanii* HGB1456 and the nucleotides encoding the follower peptide in DarA were removed.

Analyzing the LCMS data, it can be concluded that DarE accepts the modified heptapeptide sequence as substrate and efficiently processes it, if a mass corresponding to the darobactin derivative can be detected in the respective sample and not in the control. On the other hand, if the mass of the expected derivative cannot be observed in the respective sample, it is considered that DarE is not able to recognize and/or process this particular core peptide sequence. However, it is also possible, that the expression of a particular derivative is below the detection limit. There are also other possibilities, for example (1) the number of ring closures formed might deviate from two (e.g. only one ring closure or even three ring closures), or (2) upon changing the AAs in the core peptide, other unexpected atoms might be incorporated into the newly formed bonds. Yet, those possibilities are not covered and discussed in this study.

A qualitative overview of the potential antibiotic activity from the analogues produced in this study was derived from agar plate based activity assay (AAA) against the representative gram-negative pathogen *E. coli* MG1655 BamA6. This strain shows an increased sensitivity against the BamA inhibitor darobactin due to additional two AAs between codon 218 and 219 in its BamA.[8] The MIC of this strain against darobactin decreases to 0.25 µg/mL (16-fold) compared to the WT (4 µg/mL). As this assay is a qualitative activity test, it should be noted, that the clear zone (halo effect) in the assay is the combined effect of the active properties of a substance and how well it is produced. Thus, it is possible that a product, which is only expressed on a low level but is active, might not be observable.

Increase and decrease of the core peptide size

Within the heptapeptide of Darobactin, two rings – between W^1 and W^3 and between W^3 and K^5 are installed by DarE. The possibility to change the size of these rings and the core peptide was conducted. To decrease the size of the rings, a residue within the rings was removed. Meanwhile, to increase the size, alanine was introduced. However, none of these constructs resulted in the expression of the expected product (darobactin A5-8 and A11-12, Table 1). Decreasing the length of the core peptide from the C-terminus is possible, though only by one AA (darobactin A9). Shortening the sequence by two AAs abolishes the expression of the expected product (darobactin 11). In parallel, the study by Groß and coworker showed a similar result. No product was observed when two AA from the C-terminal of the core peptide were removed.[4] However, although it is possible to obtain a product, which is shortened by one AA from the C-terminus, the halo effect in the AAA was abolished, thus showing the vital involvement of F^7 in the antibiotic activity of darobactin A. Darobactin A resembles a typical *E. coli* β-signal sequence, both by altering hydrophobicity and by the presence of the F^7. It is in agreement with the fact that upon binding, F^7 occupies the place of the C-terminal aromatic residue (consensus signal sequence) of an *E. coli* BamA substrate.[9]

Moreover, introducing additional alanine (A) upstream of the core peptide only yields darobactin A (construct pDK-2). Originally, A48 is the residue upstream W^1 and one can expect that A48 might be essential for the recognition/activity of the unidentified protease to release darobactin. As a consequence, an additional A might not prevent the unspecific protease activity. When R was introduced instead of A, amide bond cleavage still took place (construct pDK-3). Hence, it can be speculated that the W^1 might be the recognition site, or that an exonuclease cleaves towards the C-terminus of DarA, which will then directly stop before the modified W^1 in the case where the *C-O-C* bond is installed. Hence, the *C-O-C* ring closure might prevent the unidentified protease to digest further. Similarly, the addition of A to the C-terminus of the core peptide also leads only to the formation of darobactin A (construct pDK-4). It can be speculated that F^7 in original DarA represents the protease recognition site to cleave off the follower peptide. When the F^7 was replaced with W, a mixture (±50:50) of products with and without additional A at the C-terminus was observed (construct pDK-05, Table 1 and Figure S11), showing that the unidentified protease is still partially able to cleave after W^7. Replacement of F^7 to W might negatively influence the unidentified protease activity.

By comparison to the original DarA with its 3 AA follower peptide, darobactin derivatives with additional one, two, or three follower AAs cannot be detected after expression (construct pZW-ADC5), indicating a highly efficient protease activity when position 7 in the core peptide is occupied by F. Conclusively, even though additional AAs in the C-terminal region seem to be possible, the desired end product cannot be fully generated (due to the involvement of the unidentified protease). Overall, it suggests that one additional AA at the N- or C-terminus of the core peptide can be tolerated by DarE (since both ring closures can be detected) and results in darobactin A. To obtain further insight into DarE flexibility, in respect to end products that maintain activity, in the further experiments the size of the core peptide was preserved at 7 AAs.

Table 1. Overview of experiments with increasing and decreasing size of the core peptide. Italic letters represent residues between which the *C-O-C* bond is formed. Underlined letters represent residues between which the *C-C* bond is formed. In the AAA column, "+" indicates an observable halo inhibition zone, while "-" indicates absence. More LCMS analysis details are shown in Figure S11. AAA documentations are shown in Figure S2-3.

Construct	Core and follower peptide	Investigated end product	Darobactin	Chemical formula	Detected $[M+2H]^{2+}$ ion (*m/z*)	Error (ppm)	Agar activity assay (AAA)
pZW-ADC5Δfol	WNWSKSF	*WN*<u>*W*</u>SK<u>S</u>F	A	$C_{47}H_{55}N_{11}O_{12}$	483.7114	5.2	+
pDK-2	AWNWSKSF	A*WN*<u>*W*</u>SK<u>S</u>F	A1	$C_{50}H_{60}N_{12}O_{13}$	-	-	+
		WN<u>*W*</u>SK<u>S</u>F	A	$C_{47}H_{55}N_{11}O_{12}$	483.7072	3.5	
pDK-3	RWNWSKSF	R*WN*<u>*W*</u>SK<u>S</u>F	A2	$C_{53}H_{67}N_{15}O_{13}$	-	-	+
		WN<u>*W*</u>SK<u>S</u>F	A	$C_{47}H_{55}N_{11}O_{12}$	483.7103	2.9	
pDK-4	WNWSKSFA	*WN*<u>*W*</u>SK<u>S</u>FA	A3	$C_{50}H_{60}N_{12}O_{13}$	-	-	+
		WN<u>*W*</u>SK<u>S</u>F	A	$C_{47}H_{55}N_{11}O_{12}$	483.7073	3.3	
pDK-5	WNWSKSWA	*WN*<u>*W*</u>SK<u>S</u>WA	A4	$C_{52}H_{61}N_{13}O_{13}$	538.7342	2.4	+
		WN<u>*W*</u>SK<u>S</u>W	9	$C_{49}H_{56}N_{12}O_{12}$	503.2168	5.0	
pDK-6	WANWSKSF	*WAN*<u>*W*</u>SK<u>S</u>F	A5	$C_{50}H_{60}N_{12}O_{13}$	-	-	-
pDK-7	WNAWSKSF	*WNA*<u>*W*</u>SK<u>S</u>F	A6	$C_{50}H_{60}N_{12}O_{13}$	-	-	-
pDK-8	WNWASKSF	*WN*<u>*W*</u>*A*SK<u>S</u>F	A7	$C_{50}H_{60}N_{12}O_{13}$	-	-	-
pDK-9	WNWSAKSF	*WN*<u>*W*</u>S*A*K<u>S</u>F	A8	$C_{50}H_{60}N_{12}O_{13}$	-	-	-
pDK-10	WNWSKS-	*WN*<u>*W*</u>SK<u>S</u>	A9	$C_{38}H_{46}N_{10}O_{11}$	410.1732	3.7	-
pDK-11	WNWSK--	*WN*<u>*W*</u>SK	19	$C_{35}H_{41}N_9O_9$	-	-	-
pDK-12	W - WSKSF	*W*<u>*W*</u>S<u>K</u>SF	A11	$C_{43}H_{49}N_9O_{10}$	-	-	-
pDK-13	WNW- KSF	*WN*<u>*W*</u>*K*SF	A12	$C_{44}H_{50}N_{10}O_{10}$	-	-	-
pZW-ADC5	WNWSKSF QEI	*WN*<u>*W*</u>SK<u>S</u>F	A	$C_{47}H_{55}N_{11}O_{12}$	483.7121	6.6	+
		WN<u>*W*</u>SK<u>S</u>FQ	A13	$C_{52}H_{63}N_{13}O_{14}$	-	-	
		WN<u>*W*</u>SK<u>S</u>FQE	A14	$C_{57}H_{70}N_{14}O_{17}$	-	-	
		WN<u>*W*</u>SK<u>S</u>FQEI	A15	$C_{63}H_{81}N_{15}O_{18}$	-	-	

AA exchange in the core peptide

This experiment was performed by replacing one by one of the proteogenic AAs (only 1 AA exchange at a time) at each position of the core peptide, excluding the original residue in darobactin A. In total, 133 derivative constructs were made, which equals 19 possible AAs multiplied with the 7 AA positions in the core peptide.

Table 2. Overview of the AA exchange experiment of the core peptide. Italic letters represent residues between which the *C-O-C* bond is formed. Underlined letters represent residues between which the *C-C* bond is formed. In the AAA column, "+" indicates an observable halo inhibition zone, while "-" indicates absence. More LCMS analysis details are shown in Figure S12-17. AAA documentations are shown in Figure S4-10.

Construct	Core peptide	Investigated end product after expression	Darobactin	Chemical Formula	Observed $[M+2]^{2+}$ ion (*m/z*)	Error (ppm)	Agar activity assay (AAA)
pDK-W1A	ANWSKSF	*A*N*W*S*K*SF	W1A	$C_{39}H_{50}N_{10}O_{12}$	undetected	-	-
pDK-W1C	CNWSKSF	*C*N*W*S*K*SF	W1C	$C_{39}H_{50}N_{10}O_{12}S$	undetected	-	-
pDK-W1D	DNWSKSF	*D*N*W*S*K*SF	W1D	$C_{40}H_{50}N_{10}O_{14}$	undetected	-	-
pDK-W1E	ENWSKSF	*E*N*W*S*K*SF	W1E	$C_{41}H_{52}N_{10}O_{14}$	undetected	-	-
pDK-W1F	FNWSKSF	*F*N*W*S*K*SF	W1F	$C_{45}H_{54}N_{10}O_{12}$	undetected	-	-
pDK-W1G	GNWSKSF	*G*N*W*S*K*SF	W1G	$C_{38}H_{48}N_{10}O_{12}$	undetected	-	-
pDK-W1H	HNWSKSF	*H*N*W*S*K*SF	W1H	$C_{42}H_{52}N_{12}O_{12}$	undetected	-	-
pDK-W1I	INWSKSF	*I*N*W*S*K*SF	W1I	$C_{42}H_{56}N_{10}O_{12}$	undetected	-	-
pDK-W1K	KNWSKSF	*K*N*W*S*K*SF	W1K	$C_{42}H_{57}N_{11}O_{12}$	undetected	-	-
pDK-W1L	LNWSKSF	*L*N*W*S*K*SF	W1L	$C_{42}H_{56}N_{10}O_{12}$	undetected	-	-
pDK-W1M	MNWSKSF	*M*N*W*S*K*SF	W1M	$C_{41}H_{54}N_{10}O_{12}S$	undetected	-	-
pDK-W1N	NNWSKSF	*N*N*W*S*K*SF	W1N	$C_{40}H_{51}N_{11}O_{13}$	undetected	-	-
pDK-W1P	PNWSKSF	*P*N*W*S*K*SF	W1P	$C_{41}H_{52}N_{10}O_{12}$	undetected	-	-
pDK-W1Q	QNWSKSF	*Q*N*W*S*K*SF	W1Q	$C_{41}H_{53}N_{11}O_{13}$	undetected	-	-
pDK-W1R	RNWSKSF	*R*N*W*S*K*SF	W1R	$C_{42}H_{57}N_{13}O_{12}$	undetected	-	-
pDK-W1S	SNWSKSF	*S*N*W*S*K*SF	W1S	$C_{39}H_{51}N_{10}O_{13}$	undetected	-	-
pDK-W1T	TNWSKSF	*T*N*W*S*K*SF	W1T	$C_{40}H_{52}N_{10}O_{13}$	undetected	-	-
pDK-W1V	VNWSKSF	*V*N*W*S*K*SF	W1V	$C_{41}H_{54}N_{10}O_{12}$	undetected	-	-
pDK-W1Y	YNWSKSF	*Y*N*W*S*K*SF	W1Y	$C_{45}H_{54}N_{10}O_{13}$	undetected	-	-
pDK-N2A	WAWSKSF	*WA*W*S*K*SF	13	$C_{46}H_{54}N_{10}O_{11}$	462.2146	18.6	-
pDK-N2C	WCWSKSF	*WC*W*S*K*SF	5	$C_{46}H_{54}N_{10}O_{11}S$	undetected	-	-
pDK-N2D	WDWSKSF	*WD*W*S*K*SF	N2D	$C_{47}H_{54}N_{10}O_{13}$	undetected	-	-
pDK-N2E	WEWSKSF	*WE*W*S*K*SF	N2E	$C_{48}H_{56}N_{10}O_{13}$	491.2133	9.4	-
pDK-N2F	WFWSKSF	*WF*W*S*K*SF	N2F	$C_{52}H_{58}N_{10}O_{11}$	undetected	-	-
pDK-N2G	WGWSKSF	*WG*W*S*K*SF	N2G	$C_{45}H_{52}N_{10}O_{11}$	undetected	-	-
pDK-N2H	WHWSKSF	*WH*W*S*K*SF	N2H	$C_{49}H_{56}N_{12}O_{11}$	495.2238	13.9	+
pDK-N2I	WIWSKSF	*WI*W*S*K*SF	N2I	$C_{49}H_{60}N_{10}O_{11}$	483.2272	4.6	-

pDK-N2K	WKWSKSF	*WKWSKSF*	N2K	$C_{49}H_{61}N_{11}O_{11}$	490.7385	7.3	+
pDK-N2L	WLWSKSF	*WLWSKSF*	N2L	$C_{49}H_{60}N_{10}O_{11}$	483.2329	7.2	-
pDK-N2M	WMWSKSF	*WMWSKSF*	N2M	$C_{48}H_{58}N_{10}O_{11}S$	492.2132	11.4	+
pDK-N2P	WPWSKSF	*WPWSKSF*	N2P	$C_{48}H_{56}N_{10}O_{11}$	undetected	-	-
pDK-N2Q	WQWSKSF	*WQWSKSF*	11	$C_{48}H_{57}N_{11}O_{12}$	490.7212	9.2	+
pDK-N2R	WRWSKSF	*WRWSKSF*	N2R	$C_{49}H_{61}N_{13}O_{11}$	504.7427	9.3	+
pDK-N2S	WSWSKSF	*WSWSKSF*	E	$C_{46}H_{54}N_{10}O_{12}$	470.2057	4.9	-
pDK-N2T	WTWSKSF	*WTWSKSF*	12	$C_{47}H_{56}N_{10}O_{12}$	477.2134	4.6	-
pDK-N2V	WVWSKSF	*WVWSKSF*	N2V	$C_{48}H_{58}N_{10}O_{11}$	476.2233	3.6	-
pDK-N2W	WWWSKSF	*WWWSKSF*	N2W	$C_{54}H_{59}N_{11}O_{11}$	519.7304	6.3	+
pDK-N2Y	WYWSKSF	*WYWSKSF*	N2Y	$C_{52}H_{58}N_{10}O_{12}$	508.2223	6.3	-
pDK-W3A	WNASKSF	*WNASKSF*	W3A	$C_{39}H_{50}N_{10}O_{12}$	undetected	-	-
pDK-W3C	WNCSKSF	*WNCSKSF*	W3C	$C_{39}H_{50}N_{10}O_{12}S$	undetected	-	-
pDK-W3D	WNDSKSF	*WNDSKSF*	W3D	$C_{40}H_{50}N_{10}O_{14}$	undetected	-	-
pDK-W3E	WNESKSF	*WNESKSF*	W3E	$C_{41}H_{52}N_{10}O_{14}$	undetected	-	-
pDK-W3F	WNFSKSF	*WNFSKSF*	W3F	$C_{45}H_{54}N_{10}O_{12}$	undetected	-	-
pDK-W3G	WNGSKSF	*WNGSKSF*	W3G	$C_{38}H_{48}N_{10}O_{12}$	undetected	-	-
pDK-W3H	WNHSKSF	*WNHSKSF*	W3H	$C_{42}H_{52}N_{12}O_{12}$	undetected	-	-
pDK-W3I	WNISKSF	*WNISKSF*	W3I	$C_{42}H_{56}N_{10}O_{12}$	undetected	-	-
pDK-W3K	WNKSKSF	*WNKSKSF*	W3K	$C_{42}H_{57}N_{11}O_{12}$	undetected	-	-
pDK-W3L	WNLSKSF	*WNLSKSF*	W3L	$C_{42}H_{56}N_{10}O_{12}$	undetected	-	-
pDK-W3M	WNMSKSF	*WNMSKSF*	W3M	$C_{41}H_{54}N_{10}O_{12}S$	undetected	-	-
pDK-W3N	WNNSKSF	*WNNSKSF*	W3N	$C_{40}H_{51}N_{11}O_{13}$	undetected	-	-
pDK-W3P	WNPSKSF	*WNPSKSF*	W3P	$C_{41}H_{52}N_{10}O_{12}$	undetected	-	-
pDK-W3Q	WNQSKSF	*WNQSKSF*	W3Q	$C_{41}H_{53}N_{11}O_{13}$	undetected	-	-
pDK-W3R	WNRSKSF	*WNRSKSF*	W3R	$C_{42}H_{57}N_{13}O_{12}$	undetected	-	-
pDK-W3S	WNSSKSF	*WNSSKSF*	W3S	$C_{39}H_{50}N_{10}O_{13}$	undetected	-	-
pDK-W3T	WNTSKSF	*WNTSKSF*	W3T	$C_{40}H_{52}N_{10}O_{13}$	undetected	-	-
pDK-W3V	WNVSKSF	*WNVSKSF*	W3V	$C_{41}H_{54}N_{10}O_{12}$	undetected	-	-
pDK-W3Y	WNYSKSF	*WNYSKSF*	W3Y	$C_{45}H_{54}N_{10}O_{13}$	472.2044	7.4	-
pDK-S4A	WNWAKSF	*WNWAKSF*	14	$C_{47}H_{55}N_{11}O_{11}$	475.7119	1.1	+
pDK-S4C	WNWCKSF	*WNWCKSF*	6	$C_{47}H_{55}N_{11}O_{11}S$	491.6983	1.8	+
pDK-S4D	WNWDKSF	*WNWDKSF*	S4D	$C_{48}H_{55}N_{11}O_{13}$	497.7080	3.4	-
pDK-S4E	WNWEKSF	*WNWEKSF*	S4E	$C_{49}H_{57}N_{11}O_{13}$	504.7143	0.8	-
pDK-S4F	WNWFKSF	*WNWFKSF*	S4F	$C_{53}H_{59}N_{11}O_{11}$	513.7271	0.0	-
pDK-S4G	WNWGKSF	*WNWGKSF*	S4G	$C_{46}H_{53}N_{11}O_{11}$	468.7034	0.4	-
pDK-S4H	WNWHKSF	*WNWHKSF*	S4H	$C_{50}H_{57}N_{13}O_{11}$	508.7239	3.1	+
pDK-S4I	WNWIKSF	*WNWIKSF*	S4I	$C_{50}H_{61}N_{11}O_{11}$	496.7342	1.4	+
pDK-S4K	WNWKKSF	*WNWKKSF*	S4K	$C_{50}H_{62}N_{12}O_{11}$	504.2444	8.1	+
pDK-S4L	WNWLKSF	*WNWLKSF*	S4L	$C_{50}H_{61}N_{11}O_{11}$	497.7360	2.2	+
pDK-S4M	WNWMKSF	*WNWMKSF*	S4M	$C_{49}H_{59}N_{11}O_{11}S$	505.7127	0.8	+
pDK-S4N	WNWNKSF	*WNWNKSF*	S4N	$C_{48}H_{56}N_{12}O_{12}$	497.2172	5.8	+
pDK-S4P	WNWPKSF	*WNWPKSF*	S4P	$C_{49}H_{57}N_{11}O_{11}$	undetected	-	-
pDK-S4Q	WNWQKSF	*WNWQKSF*	S4Q	$C_{49}H_{58}N_{12}O_{12}$	504.2231	2.0	+
pDK-S4R	WNWRKSF	*WNWRKSF*	S4R	$C_{50}H_{62}N_{14}O_{11}$	518.2466	6.2	+
pDK-S4T	WNWTKSF	*WNWTKSF*	4	$C_{48}H_{57}N_{11}O_{12}$	490.7176	1.8	+
pDK-S4V	WNWVKSF	*WNWVKSF*	S4V	$C_{49}H_{59}N_{11}O_{11}$	489.7275	0.8	+
pDK-S4W	WNWWKSF	*WNWWKSF*	S4W	$C_{55}H_{60}N_{12}O_{11}$	undetected	-	-
pDK-S4Y	WNWYKSF	*WNWYKSF*	S4Y	$C_{53}H_{59}N_{11}O_{12}$	521.7288	8.2	-
pDK-K5A	WNWSASF	*WNWSASF*	15	$C_{44}H_{48}N_{10}O_{12}$	undetected	-	-

pDK-K5C	WNWSCSF	*WNWSCSF*	K5C	$C_{44}H_{48}N_{10}O_{12}S$	undetected	-	-
pDK-K5D	WNWSDSF	*WNWSDSF*	K5D	$C_{45}H_{48}N_{10}O_{14}$	undetected	-	-
pDK-K5E	WNWSESF	*WNWSESF*	K5E	$C_{46}H_{50}N_{10}O_{14}$	undetected	-	-
pDK-K5F	WNWSFSF	*WNWSFSF*	K5F	$C_{50}H_{52}N_{10}O_{12}$	undetected	-	-
pDK-K5G	WNWSGSF	*WNWSGSF*	K5G	$C_{43}H_{46}N_{10}O_{12}$	undetected	-	-
pDK-K5H	WNWSHSF	*WNWSHSF*	K5H	$C_{47}H_{50}N_{12}O_{12}$	undetected	-	-
pDK-K5I	WNWSISF	*WNWSISF*	K5I	$C_{47}H_{54}N_{10}O_{12}$	undetected	-	-
pDK-K5L	WNWSLSF	*WNWSLSF*	K5L	$C_{47}H_{54}N_{10}O_{12}$	undetected	-	-
pDK-K5M	WNWSMSF	*WNWSMSF*	K5M	$C_{46}H_{52}N_{10}O_{12}S$	undetected	-	-
pDK-K5N	WNWSNSF	*WNWSNSF*	K5N	$C_{45}H_{49}N_{11}O_{13}$	undetected	-	-
pDK-K5P	WNWSPSF	*WNWSPSF*	K5P	$C_{46}H_{50}N_{10}O_{12}$	undetected	-	-
pDK-K5Q	WNWSQSF	*WNWSQSF*	K5Q	$C_{46}H_{51}N_{11}O_{13}$	undetected	-	-
pDK-K5R	WNWSRSF	*WNWSRSF*	D	$C_{47}H_{55}N_{13}O_{12}$	497.7144	5.0	-
pDK-K5S	WNWSSSF	*WNWSSSF*	K5S	$C_{44}H_{48}N_{10}O_{13}$	undetected	-	-
pDK-K5T	WNWSTSF	*WNWSTSF*	K5T	$C_{45}H_{50}N_{10}O_{13}$	undetected	-	-
pDK-K5V	WNWSVSF	*WNWSVSF*	K5V	$C_{46}H_{52}N_{10}O_{12}$	undetected	-	-
pDK-K5W	WNWSWSF	*WNWSWSF*	K5W	$C_{52}H_{53}N_{11}O_{12}$	undetected	-	-
pDK-K5Y	WNWSYSF	*WNWSYSF*	K5Y	$C_{50}H_{52}N_{10}O_{13}$	undetected	-	-
pDK-S6A	WNWSKAF	*WNWSKAF*	16	$C_{47}H_{55}N_{11}O_{11}$	475.7117	0.6	+
pDK-S6C	WNWSKCF	*WNWSKCF*	8	$C_{47}H_{55}N_{11}O_{11}S$	491.6988	2.8	+
pDK-S6D	WNWSKDF	*WNWSKDF*	S6D	$C_{48}H_{55}N_{11}O_{13}$	497.7073	2.0	-
pDK-S6E	WNWSKEF	*WNWSKEF*	S6E	$C_{49}H_{57}N_{11}O_{13}$	504.7142	0.2	+
pDK-S6F	WNWSKFF	*WNWSKFF*	S6F	$C_{53}H_{59}N_{11}O_{11}$	513.7316	8.8	-
pDK-S6G	WNWSKGF	*WNWSKGF*	S6G	$C_{46}H_{53}N_{11}O_{11}$	468.7029	1.5	+
pDK-S6H	WNWSKHF	*WNWSKHF*	S6H	$C_{50}H_{57}N_{13}O_{11}$	508.7223	0.0	+
pDK-S6I	WNWSKIF	*WNWSKIF*	S6I	$C_{50}H_{61}N_{11}O_{11}$	496.7362	2.6	+
pDK-S6K	WNWSKKF	*WNWSKKF*	S6K	$C_{50}H_{62}N_{12}O_{11}$	504.2433	6.0	+
pDK-S6L	WNWSKLF	*WNWSKLF*	S6L	$C_{50}H_{61}N_{11}O_{11}$	496.7365	3.2	+
pDK-S6M	WNWSKMF	*WNWSKMF*	S6M	$C_{49}H_{59}N_{11}O_{11}S$	505.7153	4.4	+
pDK-S6N	WNWSKNF	*WNWSKNF*	S6N	$C_{48}H_{56}N_{12}O_{12}$	497.2161	3.6	+
pDK-S6P	WNWSKPF	*WNWSKPF*	S6P	$C_{49}H_{57}N_{11}O_{11}$	undetected	-	-
pDK-S6Q	WNWSKQF	*WNWSKQF*	S6Q	$C_{49}H_{58}N_{12}O_{12}$	504.2227	1.2	+
pDK-S6R	WNWSKRF	*WNWSKRF*	S6R	$C_{50}H_{62}N_{14}O_{11}$	518.2487	10.2	+
pDK-S6T	WNWSKTF	*WNWSKTF*	S6T	$C_{48}H_{57}N_{11}O_{12}$	490.7179	2.5	+
pDK-S6V	WNWSKVF	*WNWSKVF*	S6V	$C_{49}H_{59}N_{11}O_{11}$	489.7269	0.4	+
pDK-S6W	WNWSKWF	*WNWSKWF*	S6W	$C_{55}H_{60}N_{12}O_{11}$	undetected	-	-
pDK-S6Y	WNWSKYF	*WNWSKYF*	S6Y	$C_{53}H_{59}N_{11}O_{12}$	521.7292	9.0	-
pDK-F7A	WNWSKSA	*WNWSKSA*	17	$C_{41}H_{51}N_{11}O_{12}$	445.6941	2.0	-
pDK-F7C	WNWSKSC	*WNWSKSC*	F7C	$C_{41}H_{51}N_{11}O_{12}S$	461.6818	5.6	-
pDK-F7D	WNWSKSD	*WNWSKSD*	F7D	$C_{42}H_{51}N_{11}O_{14}$	467.6909	6.0	-
pDK-F7E	WNWSKSE	*WNWSKSE*	F7E	$C_{43}H_{53}N_{11}O_{14}$	474.6987	6.0	-
pDK-F7G	WNWSKSG	*WNWSKSG*	F7G	$C_{40}H_{49}N_{11}O_{12}$	438.6861	1.6	-
pDK-F7H	WNWSKSH	*WNWSKSH*	F7H	$C_{44}H_{53}N_{13}O_{12}$	478.7046	1.0	-
pDK-F7I	WNWSKSI	*WNWSKSI*	F7I	$C_{44}H_{57}N_{11}O_{12}$	466.7173	1.3	+
pDK-F7K	WNWSKSK	*WNWSKSK*	F7K	$C_{44}H_{58}N_{12}O_{12}$	474.2228	1.5	-
pDK-F7L	WNWSKSL	*WNWSKSL*	F7L	$C_{44}H_{57}N_{11}O_{12}$	466.7186	4.1	+
pDK-F7M	WNWSKSM	*WNWSKSM*	F7M	$C_{43}H_{55}N_{11}O_{12}S$	undetected	-	-
pDK-F7N	WNWSKSN	*WNWSKSN*	F7N	$C_{42}H_{52}N_{12}O_{13}$	467.1967	1.3	-
pDK-F7P	WNWSKSP	*WNWSKSP*	F7P	$C_{43}H_{53}N_{11}O_{12}$	458.7015	1.1	-
pDK-F7Q	WNWSKSQ	*WNWSKSQ*	18	$C_{43}H_{54}N_{12}O_{13}$	474.2046	1.5	-

pDK-F7R	WNWSKSR	*WN*W̲SK̲SR	F7R	$C_{44}H_{58}N_{14}O_{12}$	448.2281	6.5	-
pDK-F7S	WNWSKSS	*WN*W̲SK̲SS	F7S	$C_{41}H_{51}N_{11}O_{13}$	453.6906	0.2	-
pDK-F7T	WNWSKST	*WN*W̲SK̲ST	F7T	$C_{42}H_{53}N_{11}O_{13}$	460.7010	5.4	-
pDK-F7V	WNWSKSV	*WN*W̲SK̲SV	F7V	$C_{43}H_{55}N_{11}O_{12}$	459.7098	2.0	-
pDK-F7W	WNWSKSW	*WN*W̲SK̲SW	9	$C_{49}H_{56}N_{12}O_{12}$	503.2157	2.8	+
pDK-F7Y	WNWSKSY	*WN*W̲SK̲SY	10	$C_{47}H_{55}N_{11}O_{13}$	491.7073	2.0	+

As shown in Table 2, position 1 (W^1) is not exchangeable, indicating a highly preserved residue. The reason might be that no other proteogenic AAs possess a carbon atom with the same geometric properties, such as bond spacing etc., in this position (C_6 in W^1, Figure 2). This inflexibility is supported by database-guided analysis from the former publication[1,5] that no putative derivatives can be found carrying another AA than tryptophan at this position.

At position 3 (W^3), even though the flexibility is nearly as limited as at position 1, Y is allowed (darobactin W3Y). The difference to the W^1 is that the newly formed *C-C* bond is located at atom C_5 (Figure 2). However, F in position 3 is not allowed (darobactin W3F). With the only difference between Y and F being the additional para-OH group of Y, it can be hypothesized that the second ring in darobactin W3Y is installed between the para-OH group of Y^3 and C_β of K^5. Otherwise, if the new bond is located on any other atom of the Y^3, it could be expected that the same atom could be accessible if F (darobactin W3F) is present in this position, which is not the case. Further investigation is necessary to confirm this hypothesis. In the activity test no halo was observed for the darobactin W3Y producer. That the bioactivity is abolished by this heptapeptide sequence could be the reason that it was deselected by evolution and a Y in position 3 was not observed in database-guided analysis[2, 6]. The two unique intramolecular cycles pre-organize darobactin A in a rigid β-strand conformation; thereby, creating an antiparallel β-sheet conformation to the β1-strand of BamA via a series of backbone hydrogen bonds.[9] Further, while the peptide backbone of darobactin A is facing BamA upon binding, the side chain residues in positions 1, 3 and 5 face the membrane bilayer,[9] which is a hydrophobic environment. The additional oxygen atom (polar moiety) in darobactin W3Y might not be preferred in the Y^3 side chain that points towards the membrane. This unfavorable residue could attribute to a decreased binding affinity, thus resulting in the loss of the inhibitory effect of this derivative.

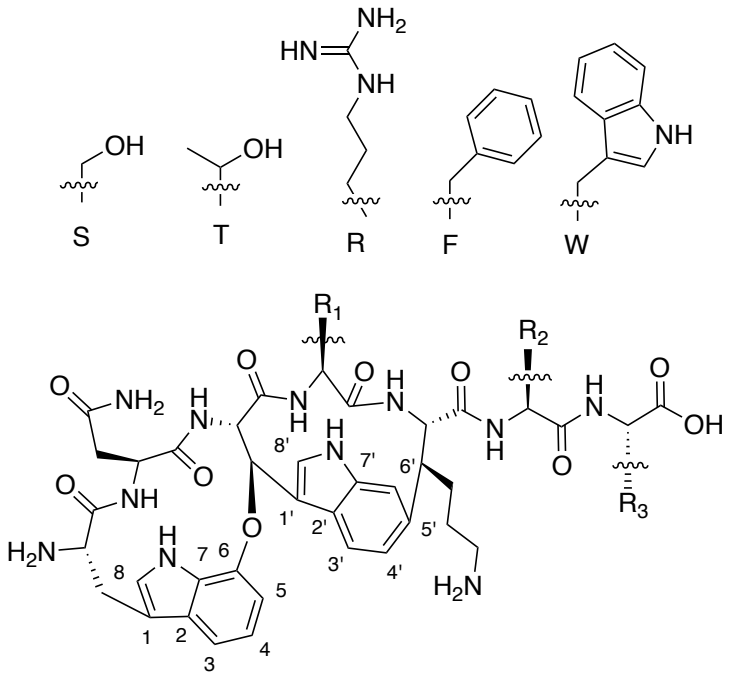


Figure 2. Structure of Darobactin A, B, 9 and B9. Darobactin A: R_1 = S, R_2 = S, R_3 = F; B: R_1 = T, R_2 = R, R_3 = F; 9: R_1 = S, R_2 = S, R_3 = W; and B9: R_1 = T, R_2 = R, R_3 = W.

At position 5 (K^5) the only tolerated AA exchange observed was R (darobactin D). This result is in accordance with the database-guided analysis[2, 6] and is not surprising, as both amino acids belong to the same group of positively charged AAs. Interestingly, H, which also belongs to the same AA class, is not tolerated in this position. Other work has proofed that A was allowed in this position (darobactin 15).[4] The fact that darobactin 15 was unobservable in this study is likely due to its expression yield, which was below the LCMS detection limit. Conclusively, the AA tolerance in this position is limited to DarE flexibility. Upon binding, the side chain of K^5 in darobactin A (which points away from BamA) interacts primarily with the negatively charged phosphate moieties of cardiolipin (CL) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol (PG) lipids.[9] Although R is a positively charged AA like K, the darobactin D producer did not show the halo effect in AAA. In agreement, darobactin D has been characterized to have 8-16 fold higher MIC values than darobactin A.[6]

For the four positions not involved in ring closure, i.e. 2, 4, 6 and 7, the allowed AAs seem to be loosely flexible; thereby, extending the chemical space of so far identified natural and un-natural darobactin derivatives[2, 4, 6]. These positions are exchangeable with almost any other proteogenic AAs. One exception is P, which is special concerning its cyclic non-flexible

structure and which is known to introduce "rigid turn" structures into peptide chains[10, 11]. P is only allowed in position 7 (F^7), the residue which is located the furthest from the macrocyclic rings. Thus, P in this position might not affect (negatively) the catalytic reaction of the ring closures. As supporting evidence, removing F^7 (Table 1, darobactin A9) does not prevent the ring closures. The same logic applies to position 6 (S^6), upon removing (Table 1, darobactin 19) or changing S^6 to P (darobactin S6P) inhibits the catalytic conversion of the pre-peptide by DarE, indicating the importance of position 6 for ring closure.

Moreover, in position 2, only P, C, G, D and F were not allowed in our experiments. In darobactin A, this position is filled with N as an uncharged residue with the polarity coming from the amide moiety. Q, which shares similar properties to N is allowed in this position and this derivative (darobactin N2Q) showed activity in the AAA. Polar AAs containing hydroxyl moieties like S (darobactin E) and T are also allowed, however, those derivatives did not show activity. In agreement, purified darobactin E has been characterized to have 8-16 fold higher MIC values than darobactin A.[6] Furthermore, all positively charged AAs (K, R, and H) are allowed and those derivatives showed activity. In contrast, from the negatively charged AAs only E is allowed, though the derivative did not show activity. Moreover, all the non-polar aliphatic AAs and Y are allowed but did not show activity. Other AAs that are allowed and still showed activity are M and W. In conclusion, even though this position is quite flexible, only derivatives with AAs at position 2 containing an amide moiety, positively charged residues, M and W showed activity. Upon binding, the two hydrogens in the side chain amine group of N^2 of darobactin A form a backbone hydrogen bond with the carboxyl group of K808 and the side-chain hydrogen of N427 in β1 of BamA.[9] Although the mutation of N427A in β1 of BamA of the tested pathogen *E. coli* decreases darobactin A binding by 10-fold, the mutated pathogen is not viable and therefore does not provide a way for bacteria to readily acquire antibiotic resistance.[9] The loss of the halo effect upon replacing the N^2 with other uncharged polar AAs, like S and T might be attributed to the longer bond spacing and the fact that hydroxyl moiety has only one hydrogen which can contribute to the hydrogen bond formation. However, the reason why the derivatives with positively charged AAs, M and W at position 2 showed activity, while with negatively charged AAs and non-polar aliphatic AAs and Y did not, remains elusive.

DarE flexibility towards positions 4 and 6 shows to be similar and highly flexible. Only P and W are not allowed in these positions. In comparison to the natural and unnatural darobactin analogues studied so far, S, T, C, A and S, R, C, A in positions 4 and 6, respectively have been identified[4, 6]. All derivatives with aromatic AAs at position 4 and 6 did not show activity as well as negatively charged AAs, with the only exception being E in position 6. As a "special" residue, the derivative with glycine in position 4 also did not show activity. Upon binding to BamA, S^4 and S^6 in A do not contribute to any bond formation.[9] This is in agreement with the fact that most of the derivatives with mutations in these two positions still showed activity. The loss of inhibition halo by changing S^4 or S^6 to aromatic AAs might be attributed to the steric effect of these bulky AAs. The side chain of S^4 and S^6 is oriented towards the β1 of BamA[9] that changing these residues to bigger AAs could increase the distance of the β-strand anti-parallel binding of others residue in darobactin A and thus weaken the binding affinity. It is unclear though how negatively charged AA in these positions result in abolishment of the inhibition halo.

At position 7, which is the furthest AA to the ring closures, more AA exchanges than at the other positions are allowed and will end up in a bicyclic heptapeptide. Interestingly, only M is not allowed. This is in agreement with an independent study, in which W, Y and A in this position were shown to result in bicyclic heptapeptides.[4] Even though many AAs are tolerated in this position, only a few of the derivatives showed activity. Further, only derivatives with medium-sized hydrophobic (I and L) and aromatic AAs at position 7 showed activity. This is in agreement with the fact that the F^7 is located in a considerable hydrophobic situation near the F426 of BamA[9]. Replacement of F^7 to W was reported to lead to darobactin 9, which has slightly better activity than darobactin A.[4]

Generation of library and screening

As the landscape of darobactin possible derivative has been obtained from the small library (the 133 constructs), in a next step it was aimed to optimize which residue gives the strongest Gram-negative antimicrobial activity for each position in the core peptide. It is not practical to characterize all possible derivatives by purifying each of them one by one. Besides the production capacity, time-consuming and laborious work, the highly soluble water compounds are generally not easy to purify. Therefore, a method to compare the activity directly from the derivative extract to the darobactin A extract as reference was examined. Therefore, the

concentration of a derivative was approximated based on its UV_{280} peak area observed at its specific retention time, which was extrapolated to the standard curve of darobactin A and subsequently normalized by using its calculated molar extinction coefficient (∈) ratio. The representative Gram-negative pathogens applied in the following screening approaches were *E. coli* MG1655 BamA6 or *E. coli* MG1655 BamA6ΔBamB. Both strains are more sensitive than the wild type strain towards BamA inhibitors. While for the actual MIC from the pure compound, *E. coli* ATCC 35218 was used as the pathogen. In the initial preparation, the control extract (from the expression of empty pRSFduet) was examined and showed antimicrobial activity to the growth of the pathogen at some point of volume (critical). Therefore, the dilution range of the experiment was maintained below this critical volume. Several derivatives were selected for method validation and were ranked from the strongest to the weakest. In addition, to validate these results, the same derivatives were produced and purified. Thereafter, their MIC values were determined and compared (Table 3). However, the ranks generated by the two methods are not related. Thus, the method to determine the relative antimicrobial activity from the extract is not valid. The reason could be that the derivative concentration determined by its UV_{280} and ∈ comparison to the darobactin A standard was not valid in the first place.

Table 3. Validation of the relative antimicrobial activity from extract approach. Darobactin A extract was used as comparison and the test strain was *E. coli* MG1655 BamA6 (sensitive pathogen). "+" in the relative MIC means a lower value (stronger activity) and "-" means the opposite. The rank was sorted from the strongest to the weakest. The approach was validated by comparing the generated rank to their actual rank determined by MIC measurement.

Darobactin	Relative antimicrobial approach		Actual MIC	
	Relative MIC (fold)	Rank	μg/mL	Rank
S6T	+2	1	8	2
16	+2	1	64	6
A (control)	1	2	8	2
N2K	1	2	8	2
S6M	1	2	16-8	3
S6L	1	2	>64	7
S6N	1	2	16-8	3
S6I	-2	3	64-32	5
S6Q	<-4	4	8-4	1
S6V	<-8	5	32	4

In parallel, a big darobactin library was generated in a similar fashion as the small library described in the previous experiments. Degenerated primers were used, but in contrast to the small library, a ligation reaction was employed for the assembly (Figure 1). AA randomization was done simultaneously at position 2, 4, 6 and 7 that mutation in the core peptide could occur in more than one position at a time. While the AA in positions 1, 3 and 5 was preserved as W, W and K respectively due to DarE flexibility and loss of activity reasons, which were observed in the proceeding experiments. The randomization codon was "NNK" which covers all the 20 possible amino acids and stop codon. The total possible generated constructs are 21^4 which equals 194,481 possibilities. The randomized codon was "NNK" was incorporated in the reverse primer that as a complement reverses from 5' to 3' becoming "MNN". The template for PCR was pDK-N2P (core peptide: WPWSKSF). pDK-N2P is preferred compared to pZW-ADC5Δfol (core peptide: WNWSKSF) because the product of latter construct is essentially darobactin A (the product that is tried to be excluded), while darobactin N2P cannot be produced due to DarE flexibility reason. The primers of the PCR reaction were modified with phosphate in their 5' end since the assembling method was ligation. The resulting vectors were transferred into the expression host *E. coli* BAP1.darobactin resistance[3] for heterologous expression. To assess the distribution quality within the library, amplicon Illumina sequencing was employed. The result showed a high bias towards specific constructs, only a low number of unique constructs were observed (only 800-2,000 unique sequences from about 50,000 reads). Furthermore, several mutations at undesired positions were detected. Several possible reasons could result in the observed bias, such as (1) some AAs are encoded by more than one codon in the degenerate NKK codon e.g. L has three possible codons while W has only one, (2) miss priming of some unique random nucleotide sequences during PCR, (3) the efficiency of transformation by electroporation that the generated constructs are not entirely transferred and the un-transferred ones will not be present in the library, (4) the different growth behavior of each colony carrying a construct that might produce active or less active (=less toxic) darobactin derivative even though the producer host was darobactin resistant, and (5) in the preparation of amplicon sequencing for quality measurement: assuming that the library has initially well-diversified constructs, during PCR, region of interest for some constructs which were amplified since the first PCR cycle will dominate others which were amplified in the 2nd or latter PCR cycle, thus the amplicon reading might not represent the actual diversity ratio of the library. Not all the potential problems could be easily solved but optimization could be done, such as (1) using lambda phage packaging to transfect the host instead of electroporation,

(2) in amplicon sequencing sample preparation: to calculate the primers amount (mol) to be abundant in comparison to the template so that all the templates could be amplified in the same rate, and (3) to do the amplicon sequencing quality control in triplicate. Although the library diversity was not that rich, it still possesses a considerably large number of unique constructs making it still worthy for further experiments.

Within this thesis, two screening approaches to identify active compounds produced by a particular construct of the big library have been performed. The first trial was more of a classical AAA approach. The library was spread on a production medium (agar) as dense as around 2-5 mm distance between single colonies and grown for several days. Thus, the pathogen contained in the warm melted agar was overlaid on top of the previous production medium and incubated overnight. Colonies giving the halo effect were re-validated by growing them separately and performing regular agar activity assays. However, after two times repetitions (around 480,000 colonies were screened on 480 Petri dishes), all picked colonies were false positive in the re-validated step. One problem was that the halo effect was not that obvious due to the dens between colonies. Unfortunately, if the density would be lowered, more Petri dishes would be needed to achieve the same number of colonies. The experiment is enormously time and labor consuming to manually prepare and check the colonies. Therefore, it was assessed to be not practical for this purpose.

The second trial employed FACS as a high-throughput-screening method. The idea was to create droplets with two water based agarose layers encapsulated in oil (w/w/o)[12]. The heterologous producer carrying the plasmid library was grown sufficiently and then encapsulated in micro droplets (inner layer) using a water-in-oil droplet generator. Thereby, the concentration of cells in medium was selected to obtain one single cell per droplet. Subsequently, the oil encapsulating the water based agarose inner layer (w/o) was broken and the second water based agarose outer layer was introduced (w/w/o) and a fluorescent screening strain was injected into the second (outer) layer of the previous droplets. Thereafter, derivative production from the inner droplet was induced by IPTG and the compound could diffuse from the inner to the outer layer. In theory, active product(s) would kill the screening strain that could be observed by the disappearance of the fluorescence-labeled strain. However, optimization of the conditions is still underway to encounter the identified problems, such as faster growth of the screening strain compared to the producer strain. Furthermore, the Gfp

protein used for detection of fluorescence activity could still be observed even though the cell has died. Therefore, the read-out was not precise.

Other than that, an alternative new method called M*e*[x] (massively parallelized growth assays)[13] is still underway. The idea was that the library would be grown and production is induced. Then, during several time points of cultivation, the library diversity would be investigated by amplicon Illumina sequencing. Based on the fact that darobactin kills *E. coli* cells, it could be expected that a non-resistant producer strain is killed by the compound produced by itself. Thereby, the constructs of interest would be the ones of which number of reads decreases over time in Illumina sequencing. These will be selected for further validation.

Darobactin B9

During the work on this project, two independent reports about optimized darobactin derivatives were published. The analogue darobactin B (core peptide: WNWTKRF) by our group[6] and darobactin 9 (core peptide: WNWSKSW) by Groß, *et al.*[4] have been characterized to have slightly better activity than darobactin A. As both derivatives have unique alterations in the heptapeptide sequence, it would be intriguing to characterize a darobactin derivative derived from the combination of both core peptides (WNWTKRW), which was named darobactin B9 (Figure 2). To get hand on this compound, the DarA-encoding sequence in the construct pZW-ADC5Δfol was modified to encode the core peptide of darobactin B9. After heterologous expression and purification, the MIC values of pure darobactin B9 were examined (Table 4). The initial results showed that darobactin B9 and darobactin B have promising activity towards *P. aeruginosa* and *A. baumanii* and perform better than darobactin A.

Table 4. MIC (μg/mL) of darobactin A, B and B9 against several selected gram negative strains.

Darobactin	A	B	B9
E. coli ATCC35218	8	1	2-1
P. aeruginosa ATCC27853	not determined	16	8
A. baumanii ATCC19606	>64	32	8
K. pneumoniae ATCC30104	4	2-1	4-2

Conclusion

The results of this study extend the vast possibility of darobactin analogues. This study compiled not only a systematic overview which heptapeptides can serve as substrate of DarE, but also provides first insights into a qualitative activity overview of the derivatives processed by DarE. Concerning the end product, the size of the core peptide is heavily dependent on the so far un-identified protease(s) involved in its maturation. Decreasing the length of the core peptide from the C-terminus is possible, though only by one AA and this deletion is accompanied by loss of the halo effect. In the core peptide, W^{1} is highly preserved and thus not exchangeable. W^{3} could only be exchanged to Y but accompanied by loss of the halo effect. While for K^{5}, only exchanging to R could be observed in this study, though Groß and coworkers showed the possibility of A in this position[4]. Yet, these substitutions do not maintain the halo effect. Positions 2, 4, and 6 are considerably flexible and many of the possible substitutions still retain the halo effect. Lastly, position 7 is the most flexible compared to the other positions, but only a few of the tolerated AAs are capable of maintaining the halo effect. In summary, only 68 out of the 133 constructs library expressing derivatives could be maturated and only 36 out of them could maintain the halo zone in the antimicrobial overlay assay. Further investigation is needed to discover the best combination of AA in the core peptide that gives the best activity. A modern, considerable faster, high throughput approach like using FACS and Me^{x} approaches are still under optimization. Once either of them is optimized, small and big darobactin libraries could be screened. The result could potentially provide new better derivatives which have better activity properties. Inspired from literature research, darobactin B9 which has a combination core peptide from darobactin B and 9 possesses a better activity towards *P. aeruginosa* and *A. baumanii* in comparison to darobactin A and B. The activity to those species could be from the W in position 7 like darobactin 9[4].

Material and Method

Constructs generation

The starting construct was the construct of pZW-ADC5Δfol[3] as the template for PCR. It consists of a pRSFduet vector and darobactin BGC from *P. khanii HGB1456*. All mutation to pZW-ADC5Δfol was incorporated using primers or degenerated primers (Figure 1). All primers used in this study and the description of which primers were used for which construct

are listed in Table S1 and S2 respectively. Codon usage for each construct is shown in Table S3-4.

In general, the primers were designed in which consist of several base pairs where they have a melting temperature ™ of 65-66 °C based on Tm calculator by NEB (https://tmcalculator.neb.com/#!/main) set to product group Q5, Polymerase/Kit Q5 high-fidelity DNA Polymerase and primer concentration of 10 nM. Further, each primer pair was used for PCR using pZW-ADC5Δfol as the template (except for making construct pDK-5 and pDK-B9 which were using construct pDK-F7W as the template) with a total volume of 200 µL (50 µL x 4 PCR reactions). Each 50 µL PCR reaction consisted of 34.5 µL ddH_2O, 2.5 µL DMSO, 10 µL Q5 reaction buffer, 1 µL dNTP, 0.5 µL forward primer 100 pmol/µL, 0.5 µL reverse primer 100 pmol/µL, 0.5 µL template DNA and 0.5 µL Q5 DNA polymerase. The PCR program was set as the following: step 1: 98 °C, 10 min; step 2: 98 °C, 10 sec; step 3: 65 °C, 20 sec; step 4: 72 °C, 7 min (30x cycle to step 2); step 5: 72 °C, 10 min; step 6: 4 °C, ∞. The product of PCR was purified through agarose gel purification and assembled using Gibson assembled (GA) method[14]. Thereafter, the GA mix was dialyzed for 20-30 min to remove salts and transferred to *E.coli* BAP1 by electroporation. Survived colonies under 50 µg/mL of kanamycin were tested by PCR with primer E6.4.check.F and E6.4.check.R. Positive PCR colonies were sanger sequenced using primer E6.4.check.F. Sequenced-positive colony was grown at 37 °C overnight LB containing 50 µg/mL of kanamycin and stored in -80 °C with a final concentration of 25% glycerol. This protocol was used to generate constructs in the "increase and decrease the size of the core peptide" and "AA exchange" experiments.

For the AA exchange experiment, in the first round of mutagenesis, only 105 derivative constructs out of 133 can be generated. To obtain the remaining 28 constructs, primers were re-designed according to the remaining missing derivatives. After second round mutagenesis was done. It still left 17 missing constructs. These 17 left constructs were achieved by targeted mutagenesis. Achieved constructs in the first, second, or third round of mutagenesis are detailed shown in Table S2. The complete codon usage of each position is listed in Table S3 and others in Table S4.

Extracts generation

5 µL of a heterologous producer from cryo culture was inoculated to 1 mL of LB containing kanamycin in a 2 mL sterile tube and then incubated at shaker 37 °C overnight with the tube was on "lay" position and not standing to have better aeration. Thus, 0.2 mL of overnight grown pre-culture was inoculated to 20 mL LB containing 50 µg/mL kanamycin in a 50 mL falcon tube. The culture was incubated on laying position at shaker 37 °C until it reached OD_{600} of 0.4-0.6. Further, the culture was induced with a final concentration of 0.5 mM IPTG and continued incubation at shaker 30 °C for 3 days. Thereafter, the culture was centrifuged at max speed, for 3 minutes and the supernatant (medium) was discarded. 400 µL of 50:50 H_2O:MeCN was added to dissolve the pellet/cells and moved to a new 2 mL tube. The cell was lysed by bath ultrasonication for 1 hour. The solution was then centrifuged at max speed for 4 minutes and then 50 µL of the clear solution was subjected to LCMS for analysis. For core peptide containing cysteine, dithiothreitol (DTT) was added to a final concentration of 50 mM and incubated at 37 °C for 40-60 min. The solution was centrifuged again at max speed for 2 min and the clear supernatant was subjected to LCMS.

LCMS

UPLC-HRMS measurements were performed on an Agilent Infinity 1290 UPLC system equipped with an Acquity UPLC BEH C18 1.7 µm (2.1 × 100 mm) column and an Acquity UPLC BEH C18 1.7 µm VanGuard Pre-Column (2.1 × 5 mm; both columns were purchased from Waters) setup coupled to a DAD detector and a micrOTOFQ II mass spectrometer (Bruker). The LC part was run using a gradient (A: H2O, 0.1% FA; B: MeCN, 0.1% FA; Flow: 600 µL min−1): 0 min: 95%A; 0.80 min: 95%A; 18.70 min: 4.75%A; 18.80 min: 0%A; 23.00 min: 0%A; 23.10 min: 95%A; 25.00 min: 95%A and the column oven temperature was set to 45 °C. Calibration of mass spectra was achieved using sodium formate in H_2O/ iPrOH (1:1) as the internal standard. Analysis of mass spectra was accomplished using the software Data Analysis 4.2 (Bruker Daltonics). Identification was done by first extracting the BPC chromatogram to ionization adducts of a derivative. The observed ion was further investigated in the control extract. The observed ion is considered to be the product of a corresponding BGC if the same mass is absent in the control.

Agar activity assay (AAA)

5 µL of heterologous host producer carrying a particular construct from cryo culture was dropped on LB agar containing 50 µg/mL of kanamycin and 1 mM of IPTG (production plate) and was incubated on 37 °C for 3 days. Thereafter, 50 µL of an overnight culture of *E.coli* MG1655 BamA6 was dissolved in 10 mL of warm LB agar containing the same antibiotic and IPTG as previous and then poured on the production plate and continued incubation overnight at 30 °C. The next day, the halo zone on the plate was investigated and documented.

Making the small library

The small library contains all derivative constructs which the AA in the DarA-core peptide was exchanged one at a time. This library excludes the construct of the core peptide sequence for darobactin A. All the 133 constructs from the "AA exchange experiment" were grown separately in a 1 mL LB medium containing 50 µg/mL of kanamycin in a 37 °C shaker overnight. Thereafter, each culture from each construct was normalized with LB containing the same antibiotic as before to OD_{600} of ~2. Thus, 40 µL of each culture were combined (library) and the library plasmids were isolated and transferred to *E. coli* BAP1.darobactin resistance[3]. Directly after transformation and incubation for 45 min, the transformed *E. coli* was centrifuged and redissolved in 1 mL LB containing 50 µg/mL of kanamycin and 25% glycerol for storage in -80 °C. 25 µL of it were plated in LB agar containing 50 µg/mL of kanamycin and incubated at 37 °C overnight. The day after, the plate was analyzed and 42 colonies were grown indicating that around 1680 colonies were successfully transformed.

Making the big library

In general, the big library is the same as the small library with the only difference being AA exchange in DarA-core peptide up to four different positions (positions 2, 4, 6 and 7) at a time. While the AA in positions 1, 3 and 5 was preserved as W, W and K respectively. The randomization codon was "NNK" which covers all the 20 possible amino acids and stop codon. The total possible generated constructs are 21^4 which equals 194,481 possibilities.

For the initial PCR step, the template of pDK-N2P (core peptide: WPWSKSF) was used. pDK-N2P is preferred compared to pZW-ADC5Δfol (core peptide: WNWSKSF) because the

product of later construct is essentially darobactin A (the product that is tried to be excluded), while darobactin N2P cannot be produced due to DarE flexibility reason. The primers of the PCR reaction were modified with phosphate in their 5' end. For the randomization, degenerate primers were used where the core peptide sequence is incorporated in the "reverse primer". The randomized codon was "NNK" and as a complement reverses from 5' to 3' becoming "MNN". The template of N2P was amplified using the primers E11.blunt.P.F and E11.blunt.P.R. The total volume of the PCR reaction was 300 µL (6 reactions x 50 µL each). The ingredient of each 50 µL PCR reaction is explained as in the Material and Method subsection "Construct generation". Thereafter, the PCR product was circulated by ligation using AnzaTM T4 Ligase Master Mix (ThermoFisher) by incubation at 16 °C overnight. Further, the ligation solution was dialyzed for 30 min to remove the salt. Thus, each 2 µL of the dialyzed big library plasmid solution was transferred by electroporation to 50 µL of *E. coli* BAP1.darobactin resistance[3] electro competent cell until all the dialyzed plasmid was consumed. Transformed cells were incubated in a 37 °C shaker for 45 min. An aliquot of it was spread on LB agar containing 50 µg/mL of kanamycin for cell counting and the remaining was kept growing at 37 °C shaker overnight and thereafter, 25% final concentration of glycerol was added for preserving in -80 °C. After calculation of survived colonies on the plate, around 320,000 colonies were successfully transformed. The same technique was used to determine the approximation cells in the overnight culture after transformation which contained about 19.2 million colonies for every 1 mL of the cryo culture.

Preparation for amplicon Illumina sequencing

For small library, plasmid was isolated from overnight grown of 10 µL small library cryo culture to a 5 mL LB containing 50 µg/mL of kanamycin. For the big library, 1 mL of the cryo was directly used for plasmid isolation. The isolated plasmid was used as a template for PCR using primer HTS.F and HTS.R. The PCR product was purified by agarose gel and subjected to amplicon sequencing.

Relative antimicrobial activity from extract

The antimicrobial activity of this approach was calculated relative to the activity of darobactin A extract (crude) as a standard. The concentration of darobactin or its derivative in the extract was calculated by its UV_{280} peak area and extinction coefficient 280 nm ($\in_{280}$) properties

compared to the same parameter of pure darobactin A as the standard compound. The net UV_{280} peak area of a compound (UV_{280}^{net}) was calculated by subtracting the UV_{280} peak area of the compound retention time to the UV_{280} peak area of the same retention time from the control extract. The calculated UV_{280}^{net} was extrapolated to the regression linear curve of standard darobactin A ended up with a calculated concentration in µM. This number would be further multiplied by $\in_{280}^{ratio}$. This ratio number is derived from $\in_{280}^{darobactin\ A}$ (standard) divided by $\in_{280}^{compound}$. The $\in_{280}^{ratio}$ was expected to normalize the use of darobactin A standard regression linear curve to calculate its darobactin derivative concentration. The value of $\in_{280}$ of tryptophane, cysteine and tyrosine are 5500, 125 and 1490 $M^{-1}cm^{-1}$.[15] Using these values and the composition of the AAs in the core peptide, the $\in_{280}^{darobactin\ A}$ was calculated as approximately 11000 $M^{-1}cm^{-1}$ which is derived from 2 Tryptophan (W) (2 x 5500 $M^{-1}cm^{-1}$), plus 0 Cysteine (C) (0 x 125 $M^{-1}cm^{-1}$) and 0 Tyrosine (Y) (0 x 1490 $M^{-1}cm^{-1}$). The value of $\in_{280}^{compound}$ was calculated likewise. The obtained concentration in µM was converted to µg/mL representing the concentration of the compound in the extract. Thereafter, calculated volumes which are needed to a final concentration of 0 (control), 0.01, 0.05, 0.1, 0.25, 0.5, 1, 2, 4, 8, 16 and 32 µg/mL in a total volume of 100 µL was added and dried on a 96-well plate in quadruplet experiments. 100 µL of exponential culture (OD_{600} of 0.1-0.9) that was previously aliquoted and grown from overnight pre-cultures of *E. coli* MG1655.BamA6 was diluted to OD_{600} of 0.001 (approximately 5 × 105 c.f.u. ml^{-1}) in LB containing 50 µg/mL of kanamycin and added to the 96-well plate containing the dried extract. The plate was incubated in a 37 °C shaker overnight. The MIC-crude extract value was determined as the minimum concentration at which no growth of the pathogen could be detected by eyes. The control positive was carbenicillin. The same treatment was likewise performed for darobactin A extract. The relative antimicrobial activity from a derivative was calculated by comparing the MIC-crude extract of a derivative to darobactin A. Note that the antimicrobial assay for the control extract (without any darobactin) was also conducted. The result showed that in a certain volume (critical), the control extract started to show an inhibition effect. Therefore, to prevent this bias, measurement was always maintained below this critical volume.

Production of darobactins

5 µL cryo culture of a heterologous host carrying a particular construct was grown in 200 mL LB containing 50 g/mL kanamycin and incubated at shaker 37 °C overnight. The next day, 1%

pre-culture of the main culture volume was inoculated to each 1 L medium in 2 L Erlenmeyer flask or 2 L medium in 5 L flask. Thereafter, they were incubated at shaker 37 °C until they reached OD_{600} = ~0.5. Thus, culture was induced with IPTG to the final concentration of 0.5 mM and continued incubation at shaker 30 °C for 3 days.

Purification of S6T, 16, N2K, S6M, S6L, S6N, S6I, S6Q and S6

The production volume was 12-16 L. At the time of harvesting, the pellet and the medium were separated by centrifugation. The pellet was moved to several 50 mL falcon tubes and dissolved with 20 mL MeCN for each falcon by shaking the falcon tube with hand until homogenous. The cell was lysed by bath ultra-sonication for 30 minutes. 20 mL of ddH_2O was added to the solution and again bath sonicated for another 30 minutes. Thereafter, the solution was centrifuged at max speed for 30 minutes. The pellet was discarded and the supernatant was moved to a new falcon tube and dried under N_2 gas flow for 3-4 days. The weight of the dried extract of darobactins S6T, 16, N2K, S6M, S6L, S6N, S6I, S6Q and S6V was 1.95, 1.85, 2.3, 1.62, 1.63, 1.79, 2.3, 1.61 and 2.3 g respectively. These dried extracts were washed twice with 30 mL MeOH by ultrasonic bath and centrifugation. 25 mL ddH_2O was added to the washed extract and centrifuged max speed for 10 min. The insoluble part was discarded and water extract was subjected to Flash chromatography (PuriFlash 4125, Interchim). Solvent A and B were water containing 0.1% trifluoracetic acid (TFA) and MeCN containing 0.1% TFA respectively. The column was C18AQ 0120, PuriFlash, Interchim. The flow rate was 46 ml/min. The program was: 0-15 min 10% solvent B, 15-45 min 10-100% solvent B and 45-55 min 100% solvent B. Fraction was collected from minute 15-35 and dried. The weight of the dried extract of darobactins S6T, 16, N2K, S6M, S6L, S6N, S6I, S6Q and S6V after Flash was 244.2, 191.1, 118.7, 246, 233.8, 185.2, 118.7, 253.6, and 150 mg respectively. These dried crude extracts were washed with MeOH and the insoluble part was dissolved with 2 mL of MilliQ water. These water extracts were subjected to HPLC (Agilent Technology 1200 series). For HPLC purification, the column was Nucleodur C18 Gravity-SB, 3 μm, 100x3 mm, Macherey-Nagel. The flow was set to 3 mL/min and the column temperature was set to 40 °C. The solvents were the same as in the flash chromatography setting. The program was set as the following: 0-10 min: 10 % solvent B; 25 min: 50 % solvent B; 25-32.5 min: 100 % solvent B; 32.5-40 min: 10 % solvent B. UV_{280} peaks of darobactins S6T, 16, N2K, S6M, S6L, S6N, S6I, S6Q and S6V at 18.075, 18, 21.5, 18.377, 18.49, 18.022, 20.2, 18.037 and 20.2 min respectively were collected. The weight of the pure darobactins S6T, 16, N2K, S6M, S6L, S6N,

S6I, S6Q and S6V were 2.28, 1.155, 3.2, 0.235, 0.095, 2.825, 0.6, 3.345, 1.9 mg respectively. Pure darobactins were subjected to NMR for structure elucidation and MIC value determination.

NMR

Pure darobactin was dissolved in D_2O for NMR measurements. ^{1}H NMR spectra (600 MHz) and ^{13}C NMR spectra with proton noise decoupling ($^{13}C\{^{1}H\}$-NMR, 150 MHz) were recorded on a Bruker "Avance III HD 600 MHz" spectrometer. NMR spectra were analyzed using the software TopSpin 4.1.1 (Bruker BioSpin GmbH) and MestReNova 14.2.1 (Mestrelab Research, S.L.).

MIC

The MICs were determined by broth microdilution assays in round-bottom 96-well plates.

Production and purification of darobactin B9

The production protocol is similar to the described method for production with some minor differences. The production volume was 48 L. The host producer was *E. coli* Rosetta darobactin resistant[3]. The incubation time for production was 6 days. After production, the cell and medium were separated by centrifugation. The pellet was dissolved in 500 mL of 1:1 MeCN:water until homogeny. The cell suspension was lysed by ultrasonic bath for 1 hour. The suspension was separated by centrifugation and the insoluble part was discarded. The soluble part (cell extract) was subjected to ion-exchange chromatography. The medium part after production was subjected to amberlite purification. The column was 2.5 L XAD16N resin (20-60 mesh, Sigma-Aldrich). The program was as the following: 4L Aceton 100 mL/min, 4 L MeOH 100 mL/min, 12 L ddH_2O 100 mL/min, 48 L sample injection 50 mL/min, 4 L ddH_2O 100 mL/min, 8 L 50% MeOH in water with 0.1% FA 50 mL/min, 5 L 80% MeOH in water with 0.1% FA 50 mL/min (elution) and 5 L MeOH 100 mL/min. The elution solution was combined with the cell extract and was evaporated with a rotary evaporator until all the MeOH evaporates. The water extract was further purified by ion-exchange chromatography. The column was 180 mL of SP Sepharose XL (GE Healthcare). The program was as the following: 1.8 L ddH_2O containing 0.1% FA, pH 3, 15 mL/min; water extract sample injection, 15

mL/min; 1.8 L ddH_2O containing 0.1% FA, pH 3, 15 mL/min; 1.8 L 50 mM ammonium acetate, pH 5, 15 mL/min; 1.8 L 50 mM ammonium acetate, pH 7, 15 mL/min; 1.8 L 50 mM ammonium acetate, pH 9, 15 mL/min; 1.8 L 50 mM ammonium acetate, pH 11, 15 mL/min; 0.9 L 1 M NaCl, 10 mL/min (elution); 0.9 L 1 M NaOH, 10 mL/min; 2 L ddH_2O, 20 mL/min; 0.5 L 20% EtOH in water, 15 mL/min. The elution solution was further subjected to amberlite purification to remove the salt. The column was 200 mL of XAD16N resin (20-60 mesh, Sigma-Aldrich). The flow rate was 13 mL/min. The program was as the following: 1 L MeOH, 1 L ddH_2O, the sample, 0.5 L ddH_2O, 1 L 80% MeOH in water (elution), 1 L MeOH, and 1 L Acetone. The elution was dried yielding 162 mg semi-pure extract. This extract was dissolved in water and subjected to HPLC-prep. The column was MIXER HP1100 4.6 x 60 mm with a flowrate of 15 mL/min. Solvent A and B are water containing 0.1 % FA (A) and MeCN containing 0.1 % FA (B) respectively. The program was as the following: 0-1 min: 10% B; 15 min: 50% B; 18-20 min: 95% B; 23-25 min: 10% B. Darobactin B9 was collected at 10-12 min. The collection was dried yielding 4.4 mg. Pure darobactin B9 was subjected to NMR for structure elucidation and MIC value determination.

Acknowledge

I D M K thanks to DAAD for the PhD scholarship.

References

1. Tacconelly, E. *et al*. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The lancet Infectious Deseases* **18**, 318-327 (2018). doi: 10.1016/S1473-3099(17)30753-3
2. Imai, Y., *et al*. A new antibiotic selectively kills gram-negative pathogens. *Nature* **576**, 459-464 (2019).
3. Wuisan, Z. G., *et al*. Optimization of heterologous darobactin A expression and identification of the minimal biosynthetic gene cluster. *Metabolic Engineering* **66**, 123-136 (2021).
4. Groß, S., *et al*. Improved broad-spectrum antibiotics against gram-negative pathogens via darobactin biosynthetic pathway engineering. *Chemical Science* **12**, 11882-11893 (2021).
5. Miethke, M. *et al*. Towards the sustainable discovery and development of new antibiotics. *Nature Reviews Chemistry* **5**, 726-749 (2021).
6. Böhringer, N., *et al*. Mutasynthetic production and antimicrobial characterization of darobactin analogs. *Microbiology Spectrum* **9**, 1-14 (2021).

7. Li, L. *et al*. Refactoring biosynthetic gene clusters for heterologous production of microbial natural products. *Current Opinion in Biotechnology* **69**, 145-152 (2021).
8. Hagan, C. L. and Kahne, D. The reconstituted escherichia coli Bam complex catalyzes multiple rounds of β-Barrel assembly. *Biochemistry* **50**, 7444-7446 (2011).
9. Kaur, H., *et al*. The antibiotic darobactin mimics a β-strand to inhibit outer membrane insertase, *Nature* **593**, 125-129 (2021).
10. Levitt, M. Effect of proline residues on protein folding. *Journal of Molecular Biology* **145**, 251-263 (1981).
11. Morgan, A. A. and Rubenstein, E. Proline: The distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. *Plos One* **8**, e53785 (2013). doi: 10.1371/journal.pone.0053785
12. Brinkmann, S., *et al*. Two-step generation of monodisperse agarose-solidified double emulsions (w/w/o) excluding an inner oil barrier. *MethodsX* **8**, 1-10 (2021). doi: 10.1016/j.mex.2021.101565
13. Koch, P. *et al*. Discovery of Antimicrobials By M*e*^{x} : Massively Parallelized Growth Assays. *Research Square*. Under review (2021). doi: 10.21203/rs.3.rs-1140231/v1
14. Gibson, D. G. *et al*. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343-345 (2009).
15. Protein A280 Thermo Scientific NanoDrop Spectrophotometers. Available and accessed on "https://tools.thermofisher.com/content/sfs/brochures/Thermo-Scientific-NanoDrop-Products-Protein-Technical-Guide-EN.pdf".

7. Chapter VI: Supporting Information

Systematic investigation of DarE flexibility generating a new darobactin derivative with broader gram-negative bioactivity

I Dewa M. Kresna[a,b], Zerlina G. Wuisan[a,b,c], Sanjata Joshi,[a] Ute Mettal,[a,b] Maria Patras,[b] Mona Abdullahi,[b] Stephan Brinkmann,[b] Nils Böhringer,[a,b,c] Steven Schmitt,[d] Silke Reiter,[d] Karina Brinkrolf,[b] Michael Marner,[b] Till F. Schäberle[a,b,c]

[a] Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, Ohlebergsweg 12, 35392 Giessen, Germany.

[b] Department of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Ohlebergsweg 12, 35392 Giessen, Germany.

[c] German Center for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Giessen, Germany.

[d] Bioprocess Laboratory, Department of Biosystem Science and Engineering, ETH Zurich, Basel, Switzerland.

Table S1. Listed primers used in this tudy

No	Primer	Nucleotide sequence
1	E6.4.W49X.F	NNNaactggtcaaaaagcttctaaagcttatcccatcaggttattttattttcctgaaaaaac
2	E6.4.W49X.R	acctgatgggataagctttagaagctttttgaccagttNNNggccgtgatctcagggatcttagg
3	E6.4.N50X.F	tggNNNtggtcaaaaagcttctaaagcttatcccatcaggttattttattttcctgaaaaaac
4	E6.4.N50X.R	acctgatgggataagctttagaagctttttgaccaNNNccaggccgtgatctcaggg
5	E6.4.W51X.F	tggaacNNNtcaaaaagcttctaaagcttatcccatcaggttattttattttcctgaaaaaac
6	E6.4.W51X.R	acctgatgggataagctttagaagctttttgaNNNgttccaggccgtgatctcagg
7	E6.4.S52X.F	tggaactggNNNaaaagcttctaaagcttatcccatcaggttattttattttcctgaaaaaac
8	E6.4.S52X.R	acctgatgggataagctttagaagcttttNNNccagttccaggccgtgatctc
9	E6.4.K53X.F	tggaactggtcaNNNagcttctaaagcttatcccatcaggttattttattttcctgaaaaaac
10	E6.4.K53X.R	acctgatgggataagctttagaagctNNNtgaccagttccaggccgtga

11	E6.4.S54X.F	tggaactggtcaaaaNNNttctaaagcttatcccatcaggttattttattttcctgaaaaaac
12	E6.4.S54X.R	acctgatgggataagctttagaaNNNttttgaccagttccaggccgtg
13	E6.4.F55X.F	tggaactggtcaaaaagcNNNtaaagcttatcccatcaggttattttattttcctgaaaaaac
14	E6.4.F55X.R	acctgatgggataagctttaNNNgctttttgaccagttccaggccg
15	E6.4.check.F	ccataccgcgaaaggttttgcg
16	E6.4.check.R	aacgtttcatggattctgagatgttaatagcattcat
17	E6.4.W49X.2F	NDSaactggtcaaaaagcttctaaagcttatcccatcaggttattttattttcctgaaaaaac
18	E6.4.W49X.2R	acctgatgggataagctttagaagctttttgaccagttSHNggccgtgatctcaggatcttagg
19	E6.4.N50X.2F	tggNDKtggtcaaaaagcttctaaagcttatcccatcaggttattttattttcctgaaaaaac
20	E6.4.N50X.2R	acctgatgggataagctttagaagctttttgaccaMHNccaggccgtgatctcaggg
21	E6.4.W51X.2F	tggaacRDKtcaaaaagcttctaaagcttatcccatcaggttattttattttcctgaaaaaac
22	E6.4.W51X.2R	acctgatgggataagctttagaagctttttgaMHYgttccaggccgtgatctcagg
23	E6.4.S52X.2F	tggaactggWWKaaaagcttctaaagcttatcccatcaggttattttattttcctgaaaaaac
24	E6.4.S52X.2R	acctgatgggataagctttagaagcttttMWWccagttccaggccgtgatctc
25	E6.4.K53X.2F	tggaactggtcaKDKagcttctaaagcttatcccatcaggttattttattttcctgaaaaaac
26	E6.4.K53X.2R	acctgatgggataagctttagaagctMHMtgaccagttccaggccgtga
27	E6.4.S54X.2F	tggaactggtcaaaaRWGttctaaagcttatcccatcaggttattttattttcctgaaaaaac
28	E6.4.S54X.2R	acctgatgggataagctttagaaCWYttttgaccagttccaggccgtg
29	E6.4.F55X.2F	tggaactggtcaaaaagcDDRtaaagcttatcccatcaggttattttattttcctgaaaaaac
30	E6.4.F55X.2R	acctgatgggataagctttaYHHgctttttgaccagttccaggccg
31	W49.F	aactggtcaaaaagcttctaaagcttatccca
32	W49M.R	tagaagctttttgaccagttcatggccgtgatctcagggatcttagg
33	W49N.R	tagaagctttttgaccagttgttggccgtgatctcagggatcttagg
34	W49G.R	tagaagctttttgaccagttgccggccgtgatctcagggatcttagg
35	N50.F	tggtcaaaaagcttctaaagcttatcccatca
36	N50D.R	ctttagaagctttttgaccaatcccaggccgtgatctcaggga
37	N50F.R	ctttagaagctttttgaccaaaaccaggccgtgatctcaggga
38	N50G.R	ctttagaagctttttgaccagccccaggccgtgatctcaggga
39	N50M.R	ctttagaagctttttgaccacatccaggccgtgatctcaggga
40	N50Q.R	ctttagaagctttttgaccactgccaggccgtgatctcaggga

41	N50V.R	ctttagaagctttttgaccacacccaggccgtgatctcaggga
42	N50W.R	ctttagaagctttttgaccaccaccaggccgtgatctcaggga
43	W51.F	tcaaaaagcttctaaagcttatcccatcaggttattttattttc
44	W51D.R	aagctttagaagctttttgaatcgttccaggccgtgatctcaggg
45	W51G.R	aagctttagaagctttttgagccgttccaggccgtgatctcaggg
46	W51M.2.F	agatcacggcctggaacatgtcaaaaagcttctaaagcttatcccatcaggt
47	W51M.2.R	catgttccaggccgtgatctcaggg
48	S52.F	aaaagcttctaaagcttatcccatcaggttattttattttcc
49	S52F.R	gataagctttagaagcttttaaaccagttccaggccgtgatctca
50	K53.F	agcttctaaagcttatcccatcaggttattttattttcc
51	K53E.R	tgggataagctttagaagctttctgaccagttccaggccgtga
52	K53V.R	tgggataagctttagaagctcactgaccagttccaggccgtga
53	S54.F	ttctaaagcttatcccatcaggttatttttattttcctgaaaaaac
54	S54E.R	tgatgggataagctttagaattcttttgaccagttccaggccgtga
55	S54M.R	tgatgggataagctttagaacatttttgaccagttccaggccgtga
56	F55.F	taaagcttatcccatcaggttattttattttcctgaaaaaaca
57	F55M.R	acctgatgggataagctttacatgctttttgaccagttccaggccg
58	F55W.R	acctgatgggataagctttaccagctttttgaccagttccaggccg
59	E6.4.2.AWNWSKSF.R	tagaagctttttgaccagttccacgcggccgtgatctcagggatcttagg
60	E6.4.2.RWNWSKSF-P.F	tggaactggtcaaaaagcttttaaagcttatcc
61	E6.4.2.RWNWSKSF.3.R	aagctttttgaccagttccaacgggccgtgatctcagggatcttagg
62	E6.4.2.WNWSKSFA.R	acctgatgggataagctttacgcgaagctttttgaccagttccaggcc
63	E6.4.2.WNWSKSWA-P.F	taaagcttatcccatcaggttattttattttcctgaaaaaac
64	E6.4.2.WNWSKSWA.3.R	aataacctgatgggataagctttacgcccagctttttgaccagttccaggc
65	E6.4.2.WANWSKSF.R	agaagctttttgaccagttcgcccaggccgtgatctcaggg
66	E6.4.2.WNAWSKSF.R	ctttagaagctttttgaccacgcgttccaggccgtgatctcagg
67	E6.4.2.WNWASKSF.R	aagctttagaagctttttgacgcccagttccaggccgtgatctc
68	E6.4.2.WNWSAKSF.R	gataagctttagaagctttttcgctgaccagttccaggccgtga
69	E6.4.2.N50.3.F	tggtcaaaaagcttttaaagcttatcccatcagg
70	E6.4.2.WWSKSF.3.R	ataagctttaaaagctttttgaccaccaggccgtgatctcaggga
71	E6.4.2.WNWKSF.R	tgggataagctttagaagcttgaccagttccaggccgtga
72	E11.blunt.P.F	agcttatcccatcaggttattttattttcctgaaaaaaca
73	E11.blunt.P.R	ttaMNNMNNtttMNNccaMNNccaggccgtgatctcaggga
74	HTS.F	tcgtcggcagcgtcagatgtgtataagagacaggcatcattcaaagagactgaactctc
75	HTS.R	gtctcgtgggctcggagatgtgtataagagacagtgaacaacttgattgtttatccccaatgg
76	WNWTKRW.F	tggtaaagcttatcccatcaggttattttattttcct
77	WNWTKRW.R	tgatgggataagctttaccaacgtttggtccagttccaggccgtgatctc

Description: A: Adenine, C: Cytosine, G: Guanine, T: Thymine, R: A or G, Y: C or T, S: G or C, W: A or T, K: G or T, M: A or C, B: C or G or T, D: A or G or T, H: A or C or T, V: A or C or G, N: Any base. 5' phosphate modification was introduced to primer E11.blunt.P.F and E11.blunt.P.R.

Table S2. Table of primer pairs to make each constructs. All construct except construct pDK-5 and pDK-B9 use pZW-ADC5Δfol as the template for PCR.

No.	Construct code	Primer forward	Primer reverse	Achieved from generation* / Des.
1	pDK-W1A	E6.4.W49X.F	E6.4.W49X.R	1st
2	pDK-W1C	E6.4.W49X.F	E6.4.W49X.R	1st
3	pDK-W1D	E6.4.W49X.F	E6.4.W49X.R	1st
4	pDK-W1E	E6.4.W49X.F	E6.4.W49X.R	1st
5	pDK-W1F	E6.4.W49X.2F	E6.4.W49X.2R	2nd
6	pDK-W1G	W49.F	W49G.R	3rd
7	pDK-W1H	E6.4.W49X.2F	E6.4.W49X.2R	2nd
8	pDK-W1I	E6.4.W49X.F	E6.4.W49X.R	1st
9	pDK-W1K	E6.4.W49X.2F	E6.4.W49X.2R	2nd
10	pDK-W1L	E6.4.W49X.F	E6.4.W49X.R	1st
11	pDK-W1M	W49.F	W49M.R	3rd
12	pDK-W1N	W49.F	W49N.R	3rd
13	pDK-W1P	E6.4.W49X.F	E6.4.W49X.R	1st
14	pDK-W1Q	E6.4.W49X.F	E6.4.W49X.R	1st
15	pDK-W1R	E6.4.W49X.F	E6.4.W49X.R	1st
16	pDK-W1S	E6.4.W49X.F	E6.4.W49X.R	1st
17	pDK-W1T	E6.4.W49X.F	E6.4.W49X.R	1st
18	pDK-W1V	E6.4.W49X.F	E6.4.W49X.R	1st
19	pDK-W1Y	E6.4.W49X.F	E6.4.W49X.R	1st
20	pDK-N2A	E6.4.N50X.F	E6.4.N50X.R	1st
21	pDK-N2C	E6.4.N50X.F	E6.4.N50X.R	1st
22	pDK-N2D	N50.F	N50D.R	3rd
23	pDK-N2E	E6.4.N50X.F	E6.4.N50X.R	1st
24	pDK-N2F	N50.F	N50F.R	3rd
25	pDK-N2G	N50.F	N50G.R	3rd
26	pDK-N2H	E6.4.N50X.F	E6.4.N50X.R	1st
27	pDK-N2I	E6.4.N50X.F	E6.4.N50X.R	1st
28	pDK-N2K	E6.4.N50X.F	E6.4.N50X.R	1st
29	pDK-N2L	E6.4.N50X.F	E6.4.N50X.R	1st
30	pDK-N2M	N50.F	N50M.R	3rd
31	pDK-N2P	E6.4.N50X.F	E6.4.N50X.R	1st

32	pDK-N2Q	N50.F	N50Q.R	3rd
33	pDK-N2R	E6.4.N50X.F	E6.4.N50X.R	1st
34	pDK-N2S	E6.4.N50X.F	E6.4.N50X.R	1st
35	pDK-N2T	E6.4.N50X.F	E6.4.N50X.R	1st
36	pDK-N2V	N50.F	N50V.R	3rd
37	pDK-N2W	N50.F	N50W.R	3rd
38	pDK-N2Y	E6.4.N50X.F	E6.4.N50X.R	1st
39	pDK-W3A	E6.4.W51X.F	E6.4.W51X.R	1st
40	pDK-W3C	E6.4.W51X.F	E6.4.W51X.R	1st
41	pDK-W3D	W51.F	W51D.R	3rd
42	pDK-W3E	E6.4.W51X.F	E6.4.W51X.R	1st
43	pDK-W3F	E6.4.W51X.F	E6.4.W51X.R	1st
44	pDK-W3G	W51.F	W51G.R	3rd
45	pDK-W3H	E6.4.W51X.F	E6.4.W51X.R	1st
46	pDK-W3I	E6.4.W51X.F	E6.4.W51X.R	1st
47	pDK-W3K	E6.4.W51X.F	E6.4.W51X.R	1st
48	pDK-W3L	E6.4.W51X.F	E6.4.W51X.R	1st
49	pDK-W3M	W51M.2.F	W51M.2.R	3rd
50	pDK-W3N	E6.4.W51X.F	E6.4.W51X.R	1st
51	pDK-W3P	E6.4.W51X.F	E6.4.W51X.R	1st
52	pDK-W3Q	E6.4.W51X.F	E6.4.W51X.R	1st
53	pDK-W3R	E6.4.W51X.F	E6.4.W51X.R	1st
54	pDK-W3S	E6.4.W51X.F	E6.4.W51X.R	1st
55	pDK-W3T	E6.4.W51X.F	E6.4.W51X.R	1st
56	pDK-W3V	E6.4.W51X.F	E6.4.W51X.R	1st
57	pDK-W3Y	E6.4.W51X.F	E6.4.W51X.R	1st
58	pDK-S4A	E6.4.S52X.F	E6.4.S52X.R	1st
59	pDK-S4C	E6.4.S52X.F	E6.4.S52X.R	1st
60	pDK-S4D	E6.4.S52X.F	E6.4.S52X.R	1st
61	pDK-S4E	E6.4.S52X.F	E6.4.S52X.R	1st
62	pDK-S4F	S52.F	S52F.R	3rd
63	pDK-S4G	E6.4.S52X.F	E6.4.S52X.R	1st
64	pDK-S4H	E6.4.S52X.F	E6.4.S52X.R	1st
65	pDK-S4I	E6.4.S52X.F	E6.4.S52X.R	1st
66	pDK-S4K	E6.4.S52X.2F	E6.4.S52X.2R	2nd
67	pDK-S4L	E6.4.S52X.F	E6.4.S52X.R	1st
68	pDK-S4M	E6.4.S52X.2F	E6.4.S52X.2R	2nd
69	pDK-S4N	E6.4.S52X.F	E6.4.S52X.R	1st
70	pDK-S4P	E6.4.S52X.F	E6.4.S52X.R	1st
71	pDK-S4Q	E6.4.S52X.F	E6.4.S52X.R	1st
72	pDK-S4R	E6.4.S52X.F	E6.4.S52X.R	1st
73	pDK-S4T	E6.4.S52X.F	E6.4.S52X.R	1st

74	pDK-S4V	E6.4.S52X.F	E6.4.S52X.R	1st
75	pDK-S4W	E6.4.S52X.F	E6.4.S52X.R	1st
76	pDK-S4Y	E6.4.S52X.F	E6.4.S52X.R	1st
77	pDK-K5A	E6.4.K53X.F	E6.4.K53X.R	1st
78	pDK-K5C	E6.4.K53X.F	E6.4.K53X.R	1st
79	pDK-K5D	E6.4.K53X.F	E6.4.K53X.R	1st
80	pDK-K5E	K53.F	K53E.R	3rd
81	pDK-K5F	E6.4.K53X.2F	E6.4.K53X.2R	2nd
82	pDK-K5G	E6.4.K53X.F	E6.4.K53X.R	1st
83	pDK-K5H	E6.4.K53X.F	E6.4.K53X.R	1st
84	pDK-K5I	E6.4.K53X.F	E6.4.K53X.R	1st
85	pDK-K5L	E6.4.K53X.F	E6.4.K53X.R	1st
86	pDK-K5M	E6.4.K53X.F	E6.4.K53X.R	1st
87	pDK-K5N	E6.4.K53X.F	E6.4.K53X.R	1st
88	pDK-K5P	E6.4.K53X.F	E6.4.K53X.R	1st
89	pDK-K5Q	E6.4.K53X.F	E6.4.K53X.R	1st
90	pDK-K5R	E6.4.K53X.F	E6.4.K53X.R	1st
91	pDK-K5S	E6.4.K53X.F	E6.4.K53X.R	1st
92	pDK-K5T	E6.4.K53X.F	E6.4.K53X.R	1st
93	pDK-K5V	K53.F	K53V.R	3rd
94	pDK-K5W	E6.4.K53X.2F	E6.4.K53X.2R	2nd
95	pDK-K5Y	E6.4.K53X.2F	E6.4.K53X.2R	2nd
96	pDK-S6A	E6.4.S54X.F	E6.4.S54X.R	1st
97	pDK-S6C	E6.4.S54X.F	E6.4.S54X.R	1st
98	pDK-S6D	E6.4.S54X.F	E6.4.S54X.R	1st
99	pDK-S6E	S54.F	S54E.R	3rd
100	pDK-S6F	E6.4.S54X.F	E6.4.S54X.R	1st
101	pDK-S6G	E6.4.S54X.F	E6.4.S54X.R	1st
102	pDK-S6H	E6.4.S54X.F	E6.4.S54X.R	1st
103	pDK-S6I	E6.4.S54X.F	E6.4.S54X.R	1st
104	pDK-S6K	E6.4.S54X.F	E6.4.S54X.R	1st
105	pDK-S6L	E6.4.S54X.F	E6.4.S54X.R	1st
106	pDK-S6M	S54.F	S54M.R	3rd
107	pDK-S6N	E6.4.S54X.F	E6.4.S54X.R	1st
108	pDK-S6P	E6.4.S54X.F	E6.4.S54X.R	1st
109	pDK-S6Q	E6.4.S54X.F	E6.4.S54X.R	1st
110	pDK-S6R	E6.4.S54X.F	E6.4.S54X.R	1st
111	pDK-S6T	E6.4.S54X.F	E6.4.S54X.R	1st
112	pDK-S6V	E6.4.S54X.F	E6.4.S54X.R	1st
113	pDK-S6W	E6.4.S54X.F	E6.4.S54X.R	1st
114	pDK-S6Y	E6.4.S54X.F	E6.4.S54X.R	1st
115	pDK-F7A	E6.4.F55X.F	E6.4.F55X.R	1st

116	pDK-F7C	E6.4.F55X.F	E6.4.F55X.R	1st
117	pDK-F7D	E6.4.F55X.F	E6.4.F55X.R	1st
118	pDK-F7E	E6.4.F55X.F	E6.4.F55X.R	1st
119	pDK-F7G	E6.4.F55X.F	E6.4.F55X.R	1st
120	pDK-F7H	E6.4.F55X.F	E6.4.F55X.R	1st
121	pDK-F7I	E6.4.F55X.F	E6.4.F55X.R	1st
122	pDK-F7K	E6.4.F55X.F	E6.4.F55X.R	1st
123	pDK-F7L	E6.4.F55X.F	E6.4.F55X.R	1st
124	pDK-F7M	F55.F	F55M.R	3rd
125	pDK-F7N	E6.4.F55X.F	E6.4.F55X.R	1st
126	pDK-F7P	E6.4.F55X.F	E6.4.F55X.R	1st
127	pDK-F7Q	E6.4.F55X.F	E6.4.F55X.R	1st
128	pDK-F7R	E6.4.F55X.F	E6.4.F55X.R	1st
129	pDK-F7S	E6.4.F55X.F	E6.4.F55X.R	1st
130	pDK-F7T	E6.4.F55X.F	E6.4.F55X.R	1st
131	pDK-F7V	E6.4.F55X.F	E6.4.F55X.R	1st
132	pDK-F7W	F55.F	F55W.R	3rd
133	pDK-F7Y	E6.4.F55X.F	E6.4.F55X.R	1st
134	pDK-11	E6.4.S54X.F	E6.4.S54X.R	1st
135	pDK-10	E6.4.F55X.F	E6.4.F55X.R	1st
136	pDK-2	W49.F	E6.4.2.AWNWSKSF.R	-
137	pDK-3	E6.4.2.RWNWSKSF-P.F	E6.4.2.RWNWSKSF.3.R	-
138	pDK-4	F55.F	E6.4.2.WNWSKSFA.R	-
139	pDK-5	E6.4.2.WNWSKSWA-P.F	E6.4.2.WNWSKSWA.3.R	Template: pDK-F7W
140	pDK-6	W49.F	E6.4.2.WANWSKSF.R	-
141	pDK-7	N50.F	E6.4.2.WNAWSKSF.R	-
142	pDK-8	W51.F	E6.4.2.WNWASKSF.R	-
143	pDK-9	S52.F	E6.4.2.WNWSAKSF.R	-
144	pDK-12	E6.4.2.N50.3.F	E6.4.2.WWSKSF.3.R	-
145	pDK-13	S52.F	E6.4.2.WNWKSF.R	-
146	pDK-B9	WNWTKRW.F	WNWTKRW.R	Template: pDK-F7W
147	pDK-big library	E11.blunt.P.F	E11.blunt.P.R	Template: pDK-N2P

* in “AA exchange experiment” as well as construct pDK-11 and pDK-10.

Table S3. Codon usage of each construct in "AA exchange experiment".

AA code	Core peptide position						
	W1	N2	W3	S4	K5	S6	F7
A	GCG	GCG	GCG	GCC	GCG	GCG	GCG
C	TGT	TGC	TGC	TGT	TGT	TGC	TGT
D	GAC	GAT	GAT	GAC	GAC	GAT	GAT
E	GAA	GAA	GAA	GAA	GAA	GAG	GAG
F	TTC	TTT	TTT	TTT	TTT	TTC	
G	GGC	GGC	GGC	GGG	GGC	GGT	GGT
H	CAC	CAT	CAC	CAC	CAC	CAT	CAC
I	ATC	ATT	ATT	ATT	ATA	ATC	ATT
K	AAG	AAA	AAA	AAG		AAA	AAG
L	CTG	TTA	CTG	CTG	TTA	TTA	CTG
M	ATG	ATG	ATG	ATG	ATG	ATG	ATG
N	AAC		AAT	AAC	AAC	AAC	AAT
P	CCA	CCG	CCC	CCG	CCA	CCA	CCT
Q	CAA	CAG	CAG	CAA	CAA	CAA	CAG
R	CGT	CGT	CGT	CGT	CGC	CGG	CGG
S	AGC	AGC	TCC		TCT		AGC
T	ACC	ACC	ACC	ACC	ACG	ACC	ACA
V	GTC	GTG	GTC	GTG	GTG	GTG	GTG
W		TGG		TGG	TGG	TGG	TGG
Y	TAT	TAT	TAT	TAT	TAT	TAT	TAT
STOP	-	-	-	-	-	TAA	TAA

Table S4. Codon usage of each construct in "increasing and decreasing the size of the core peptide experiment" and pDK-B9. pDK-10 and pDK-11 were obtained from random mutagenesis, while the other constructs were obtain from targeted insertion or deletion.

Construct Code	AA sequence of the core peptide	Core peptide codon from 5' to 3' to the direction of transcription
pZW-ADC5Δfol	WNWSKSF	TGGAACTGGTCAAAAAGCTTT
pDK-2	AWNWSKSF	GCGTGGAACTGGTCAAAAAGCTTC
pDK-3	RWNWSKSF	CGTTGGAACTGGTCAAAAAGCTTT
pDK-4	WNWSKSFA	TGGAACTGGTCAAAAAGCTTCGCG
pDK-5	WNWSKSWA	TGGAACTGGTCAAAAAGCTGGGCG
pDK-6	WANWSKSF	TGGGCGAACTGGTCAAAAAGCTTC
pDK-7	WNAWSKSF	TGGAACGCGTGGTCAAAAAGCTTC
pDK-8	WNWASKSF	TGGAACTGGGCGTCAAAAAGCTTC
pDK-9	WNWSAKSF	TGGAACTGGTCAGCGAAAAGCTTC
pDK-10	WNWSKS-	TGGAACTGGTCAAAAAGC

pDK-11	WNWSK--	TGGAACTGGTCAAAA
pDK-12	W - WSKSF	TGGTGGTCAAAAAGCTTT
pDK-13	WNW- KSF	TGGAACTGGAAAAGCTTC
pDK-B9	WNWTKRW	TGGAACTGGACCAAACGTTGG

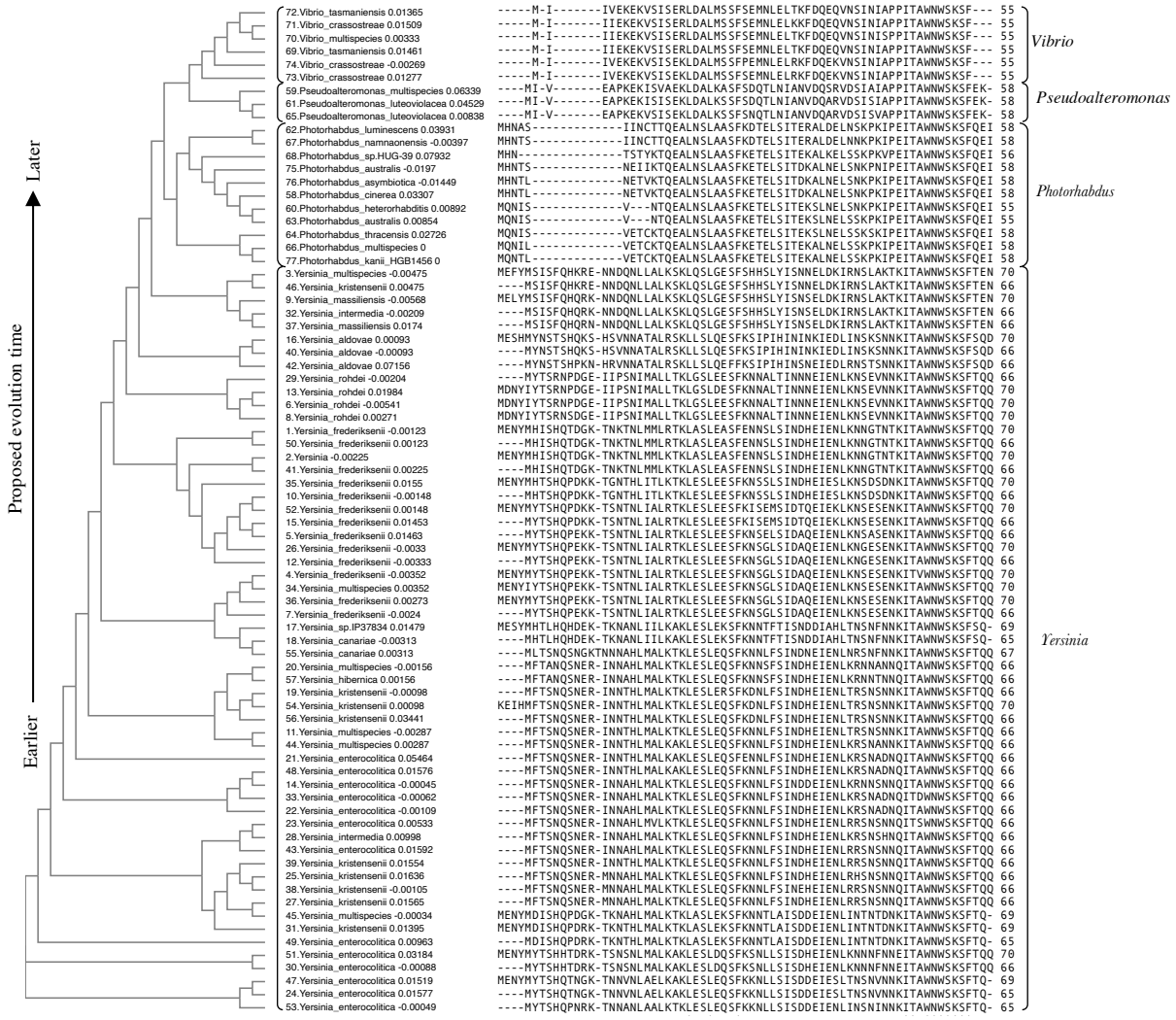
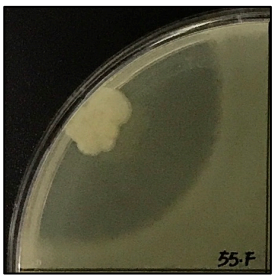
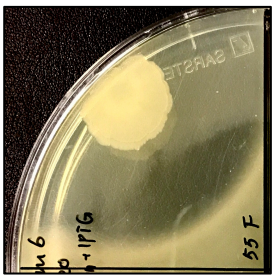


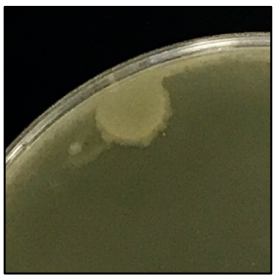
Figure S1. The phylogeny tree of DarA-related pre-peptide from NCBI database. The core peptide of Darobactin A was BLASTp (on 07.08.2020) against database and 24 out of 100 hit have no related to DarA, while 76 of them are related and were aligned using "ClustalW with character counts" algorithm and the alignment result was subjected to phylogeny tree with "distance correction" and "Neighbour-joining" clustering method[1].



pZW-ADC5
WNWSKSF

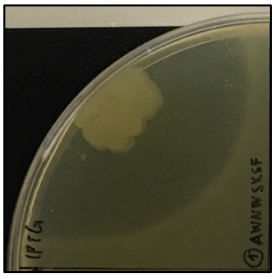


pZW-ADC5Δfol
WNWSKSF

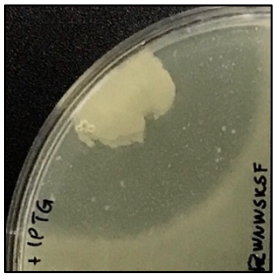


pRSFduet-empty
-

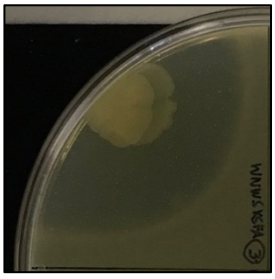
Figure S2. Agar activity assay of the heterologous producer (*E. coli* BAP1) carries a corresponding construct. Following the construct name is the core peptide AA sequence in DarA. The pathogen was *E.coli* MG1655 BamA6.



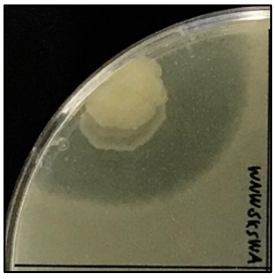
pDK-2
AWNWSKSF



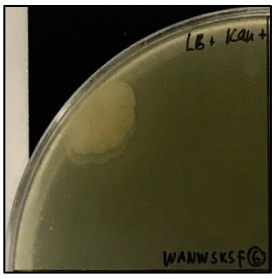
pDK-3
RWNWSKSF



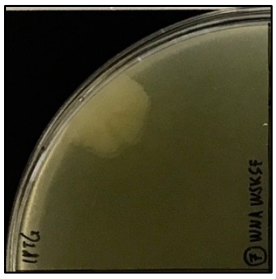
pDK-4
WNWSKSFA



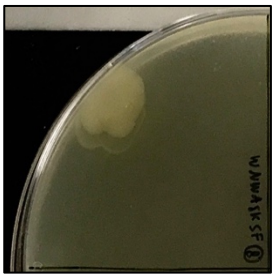
pDK-5
WNWSKSWA



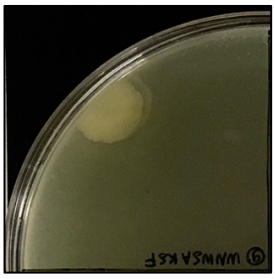
pDK-6
WANWSKSF



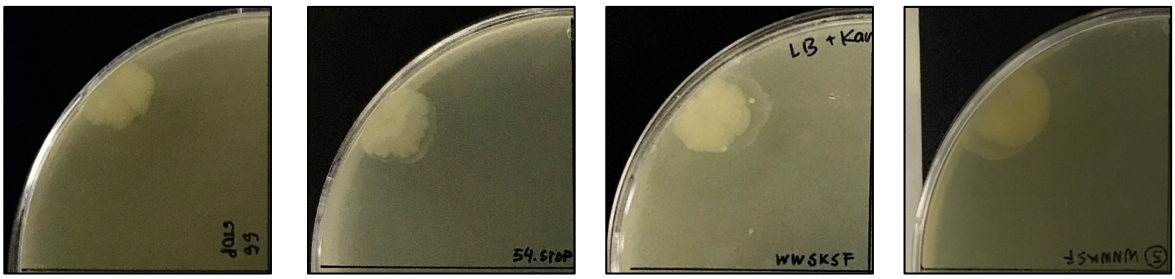
pDK-7
WNAWSKSF



pDK-8
WNWASKSF

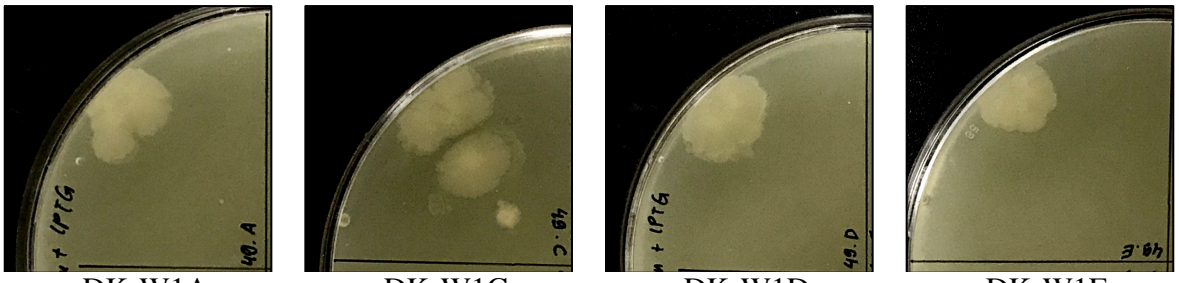


pDK-9
WNWSAKSF

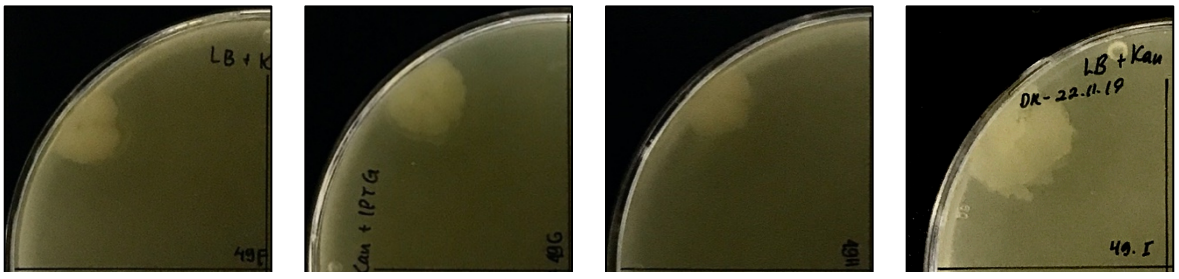


pDK-10 WNWSKS | pDK-11 WNWSK | pDK-12 WWSKSF | pDK-13 WNWKSF

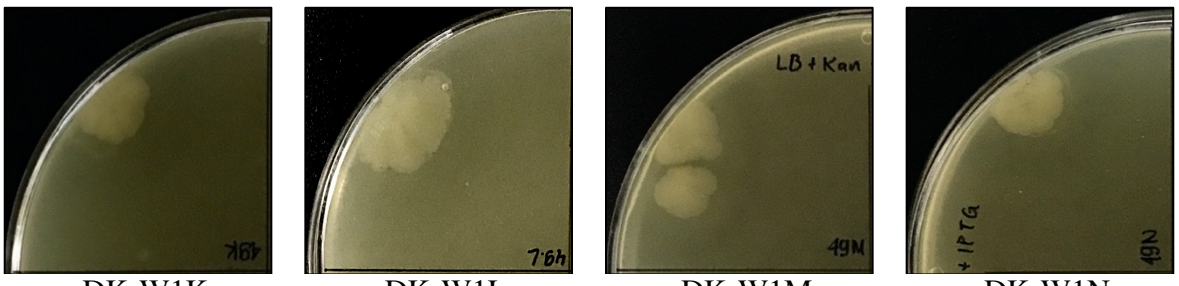
Figure S3. Agar activity assay of the heterologous producer (*E. coli* BAP1) carries a corresponding construct. Following the construct name is the core peptide AA sequence in DarA. The pathogen was *E.coli* MG1655 BamA6.



pDK-W1A ANWSKSF | pDK-W1C CNWSKSF | pDK-W1D DNWSKSF | pDK-W1E ENWSKSF



pDK-W1F FNWSKSF | pDK-W1G GNWSKSF | pDK-W1H HNWSKSF | pDK-W1I INWSKSF



pDK-W1K KNWSKSF | pDK-W1L LNWSKSF | pDK-W1M MNWSKSF | pDK-W1N NNWSKSF

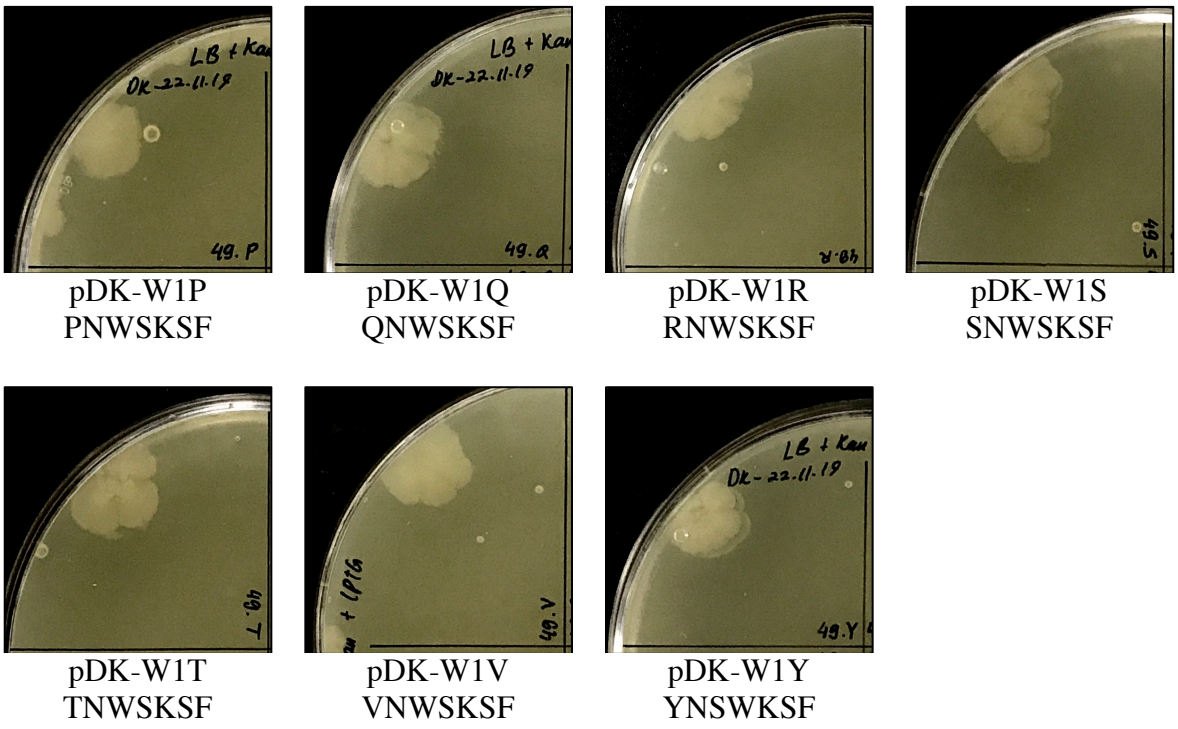
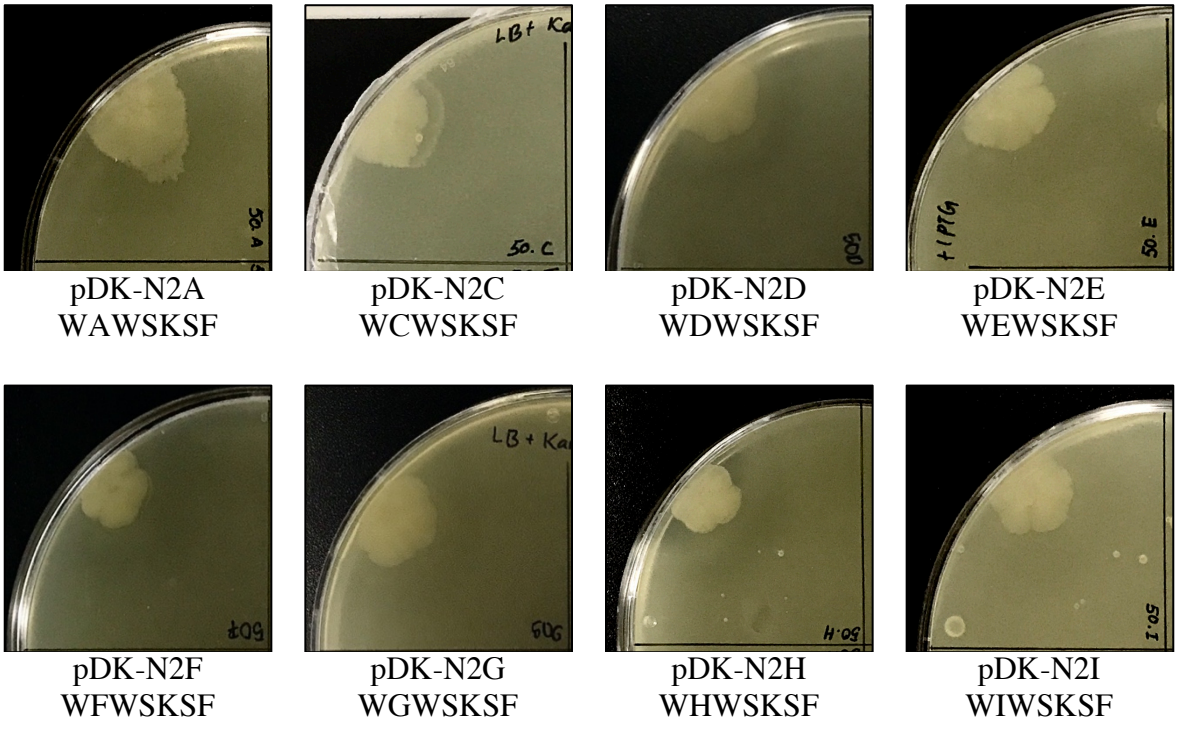
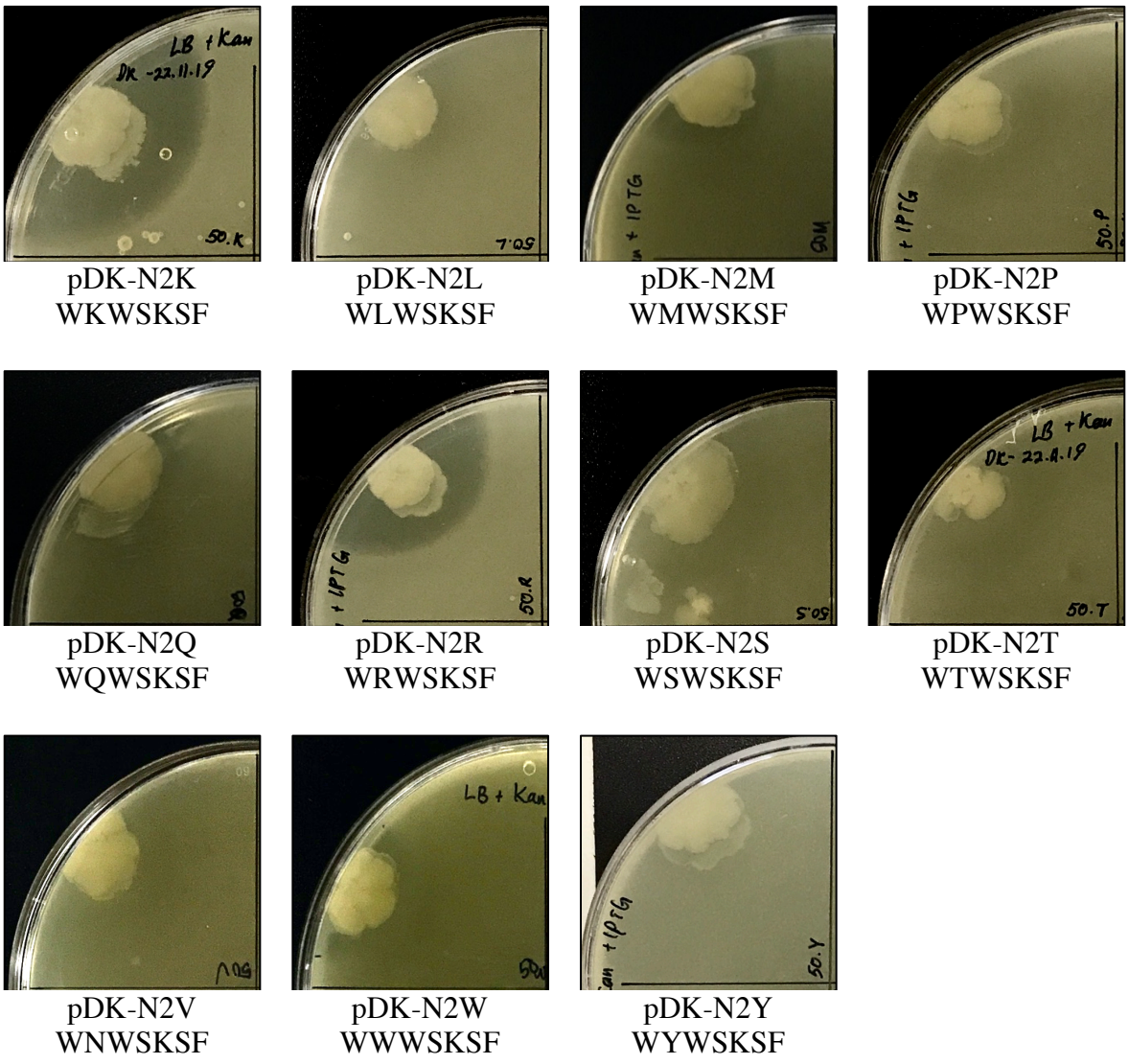


Figure S4. Agar activity assay of the heterologous producer (*E. coli* BAP1) carries a corresponding construct which the AA residue in the position 1 of DarA core peptide was mutated to other residues. Following the construct name is the core peptide AA sequence in DarA. The pathogen was *E.coli* MG1655 BamA6.



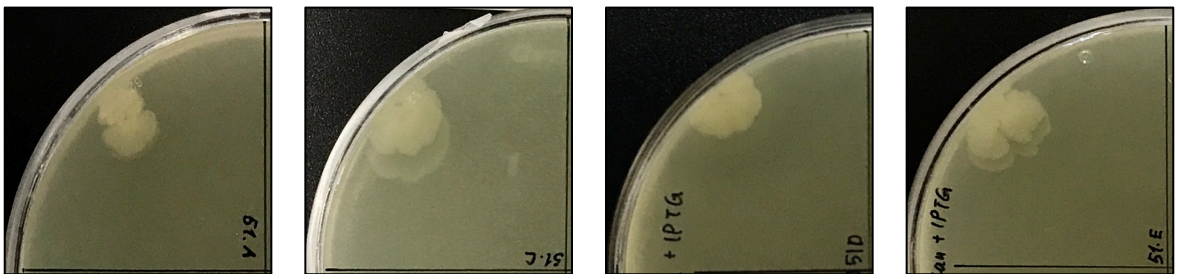


pDK-N2K WKWSKSF | pDK-N2L WLWSKSF | pDK-N2M WMWSKSF | pDK-N2P WPWSKSF

pDK-N2Q WQWSKSF | pDK-N2R WRWSKSF | pDK-N2S WSWSKSF | pDK-N2T WTWSKSF

pDK-N2V WNWSKSF | pDK-N2W WWWSKSF | pDK-N2Y WYWSKSF

Figure S5. Agar activity assay of the heterologous producer (*E. coli* BAP1) carries a corresponding construct which the AA residue in the position 2 of DarA core peptide was mutated to other residues. Following the construct name is the core peptide AA sequence in DarA. The pathogen was *E.coli* MG1655 BamA6.



pDK-W3A WNASKSF | pDKW3C WNCSKSF | pDK-W3D WNDSKSF | pDK-W3E WNESKSF

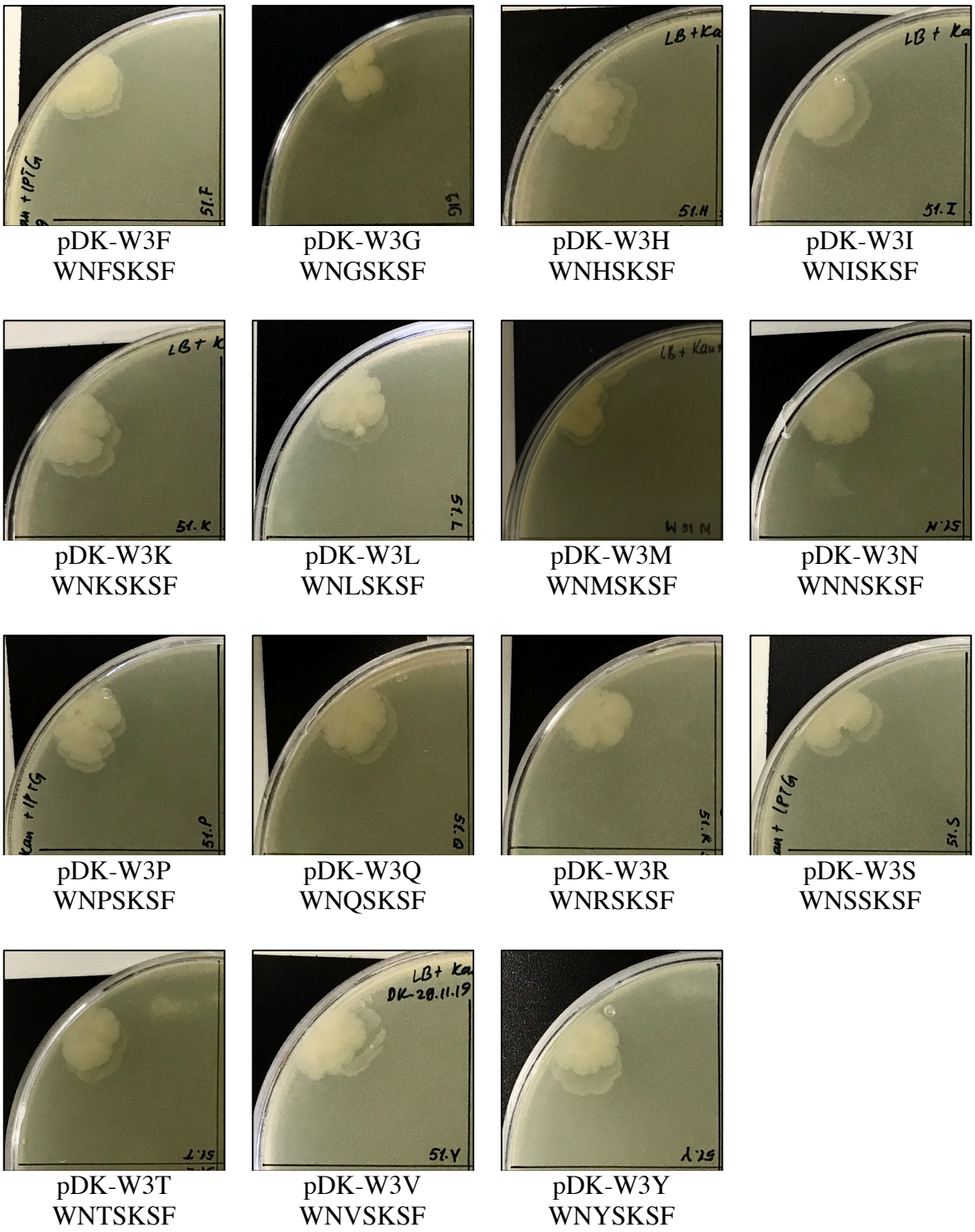
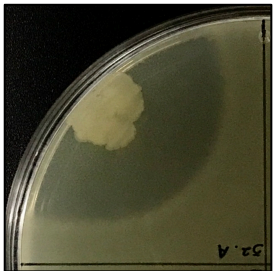
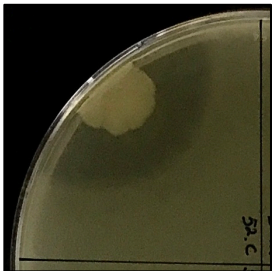


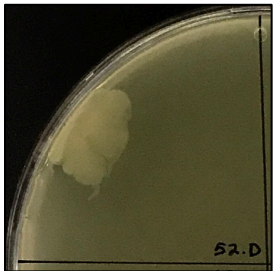
Figure S6. Agar activity assay of the heterologous producer (*E. coli* BAP1) carries a corresponding construct which the AA residue in the position 3 of DarA core peptide was mutated to other residues. Following the construct name is the core peptide AA sequence in DarA. The pathogen was *E.coli* MG1655 BamA6.



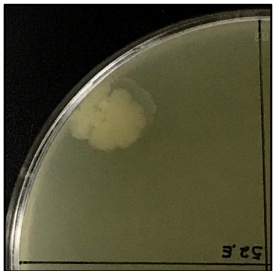
pDK-S4A
WNWAKSF



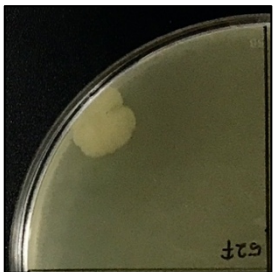
pDK-S4C
WNWCKSF



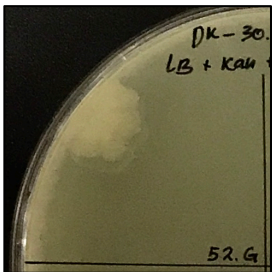
pDK-S4D
WNWDKSF



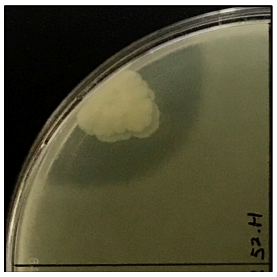
pDK-S4E
WNWEKSF



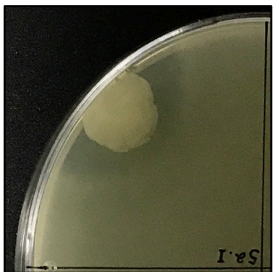
pDK-S4F
WNWFKSF



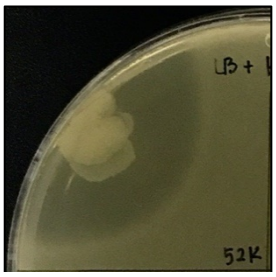
pDK-S4G
WNWGKSF



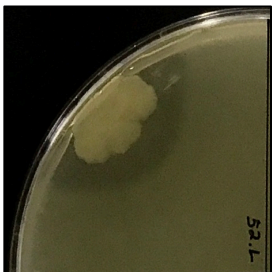
pDK-S4H
WNWHKSF



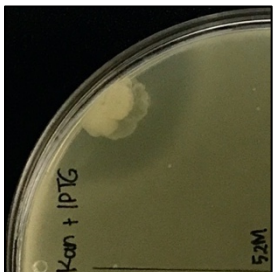
pDK-S4I
WNWIKSF



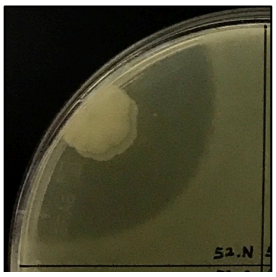
pDK-S4K
WNWKKSF



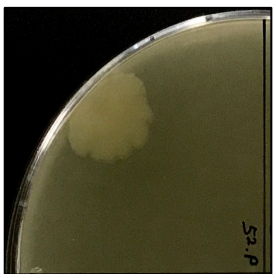
pDK-S4L
WNWLKSF



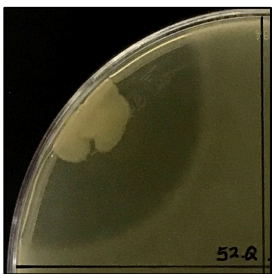
pDK-S4M
WNWMKSF



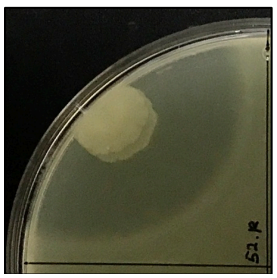
pDK-S4N
WNWNKSF



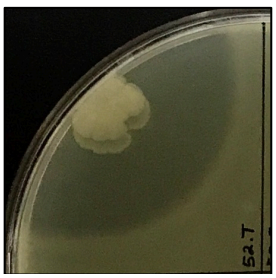
pDK-S4P
WNWPKSF



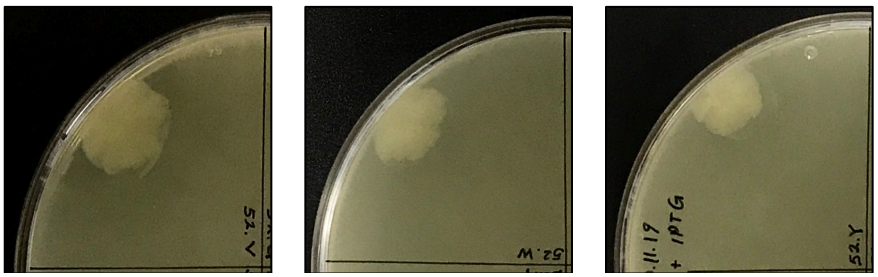
pDK-S4Q
WNWQKSF



pDK-S4R
WNWRKSF

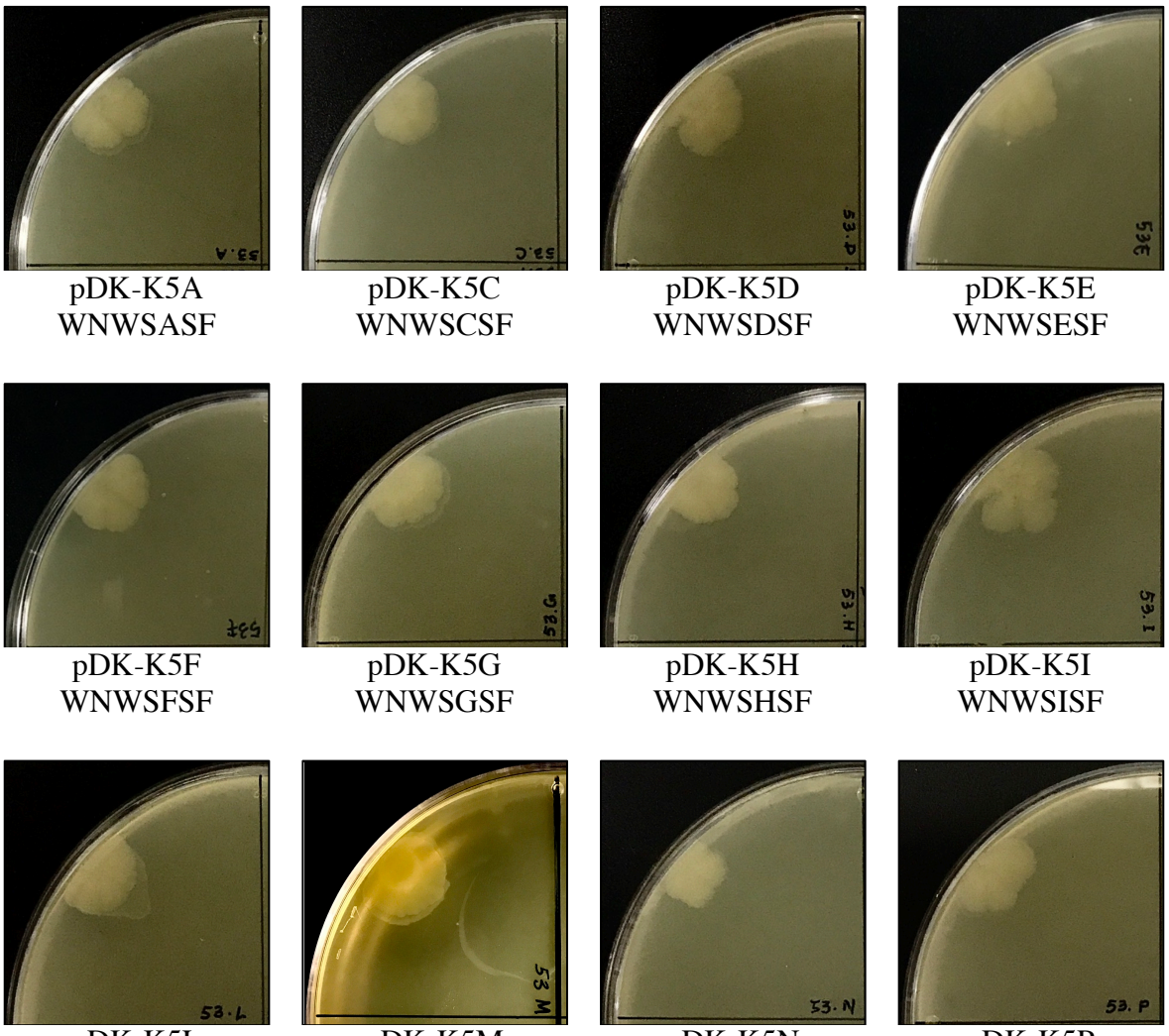


pDK-S4T
WNWTKSF



pDK-S4V WNWVKSF | pDK-S4W WNWWKSF | pDK-S4Y WNWYKSF

Figure S7. Agar activity assay of the heterologous producer (*E. coli* BAP1) carries a corresponding construct which the AA residue in the position 4 of DarA core peptide was mutated to other residues. Following the construct name is the core peptide AA sequence in DarA. The pathogen was *E.coli* MG1655 BamA6.



pDK-K5A WNWSASF | pDK-K5C WNWSCSF | pDK-K5D WNWSDSF | pDK-K5E WNWSESF

pDK-K5F WNWSFSF | pDK-K5G WNWSGSF | pDK-K5H WNWSHSF | pDK-K5I WNWSISF

pDK-K5L WNWSLSF | pDK-K5M WNWSMSF | pDK-K5N WNWSNSF | pDK-K5P WNWSPSF

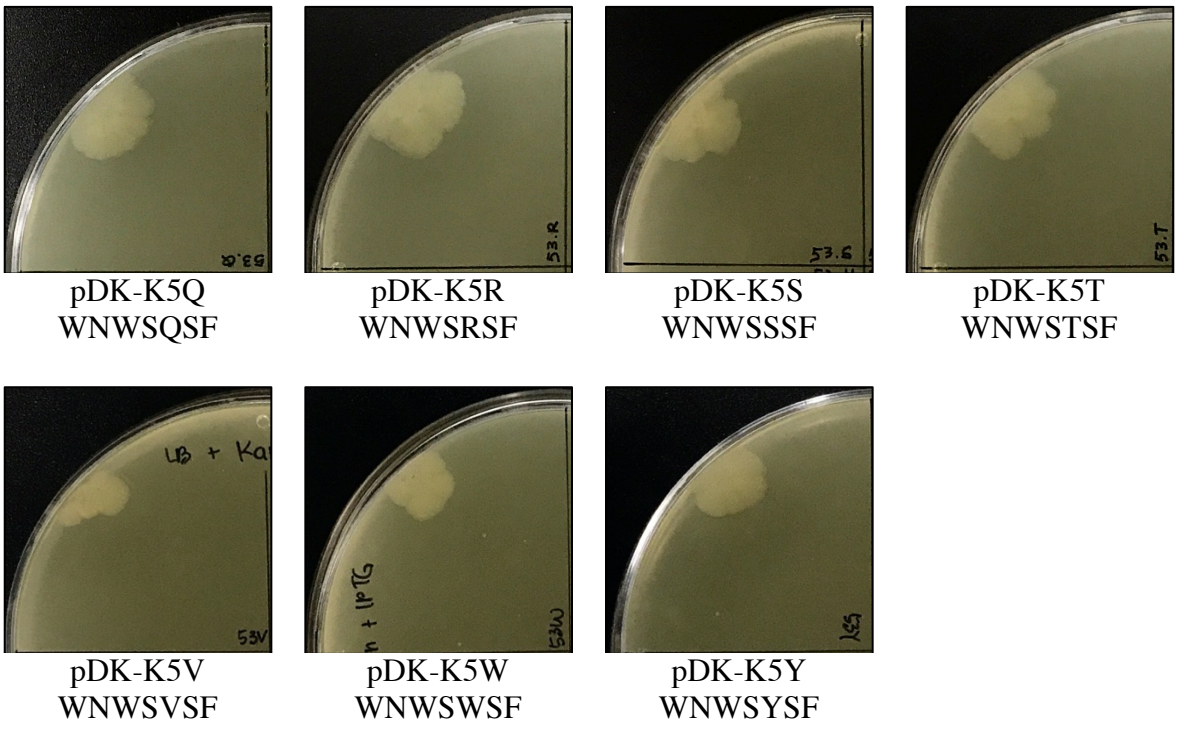
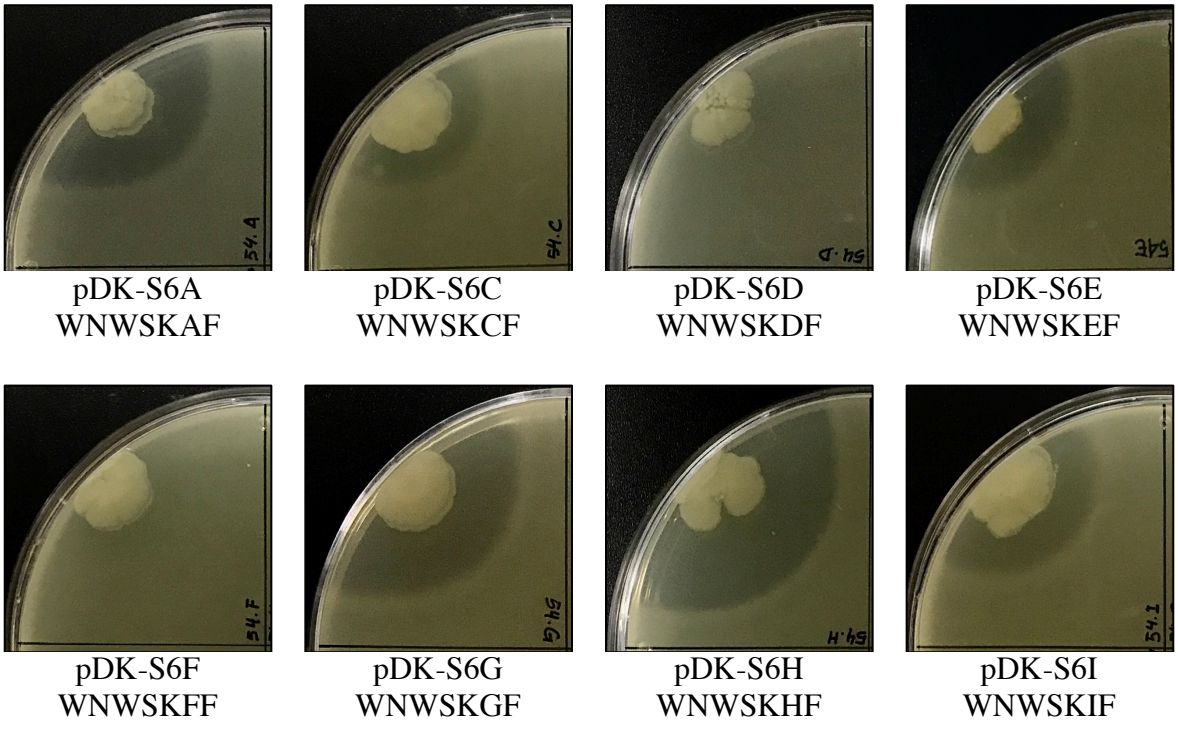


Figure S8. Agar activity assay of the heterologous producer (*E. coli* BAP1) carries a corresponding construct which the AA residue in the position 5 of DarA core peptide was mutated to other residues. Following the construct name is the core peptide AA sequence in DarA. The pathogen was *E.coli* MG1655 BamA6.



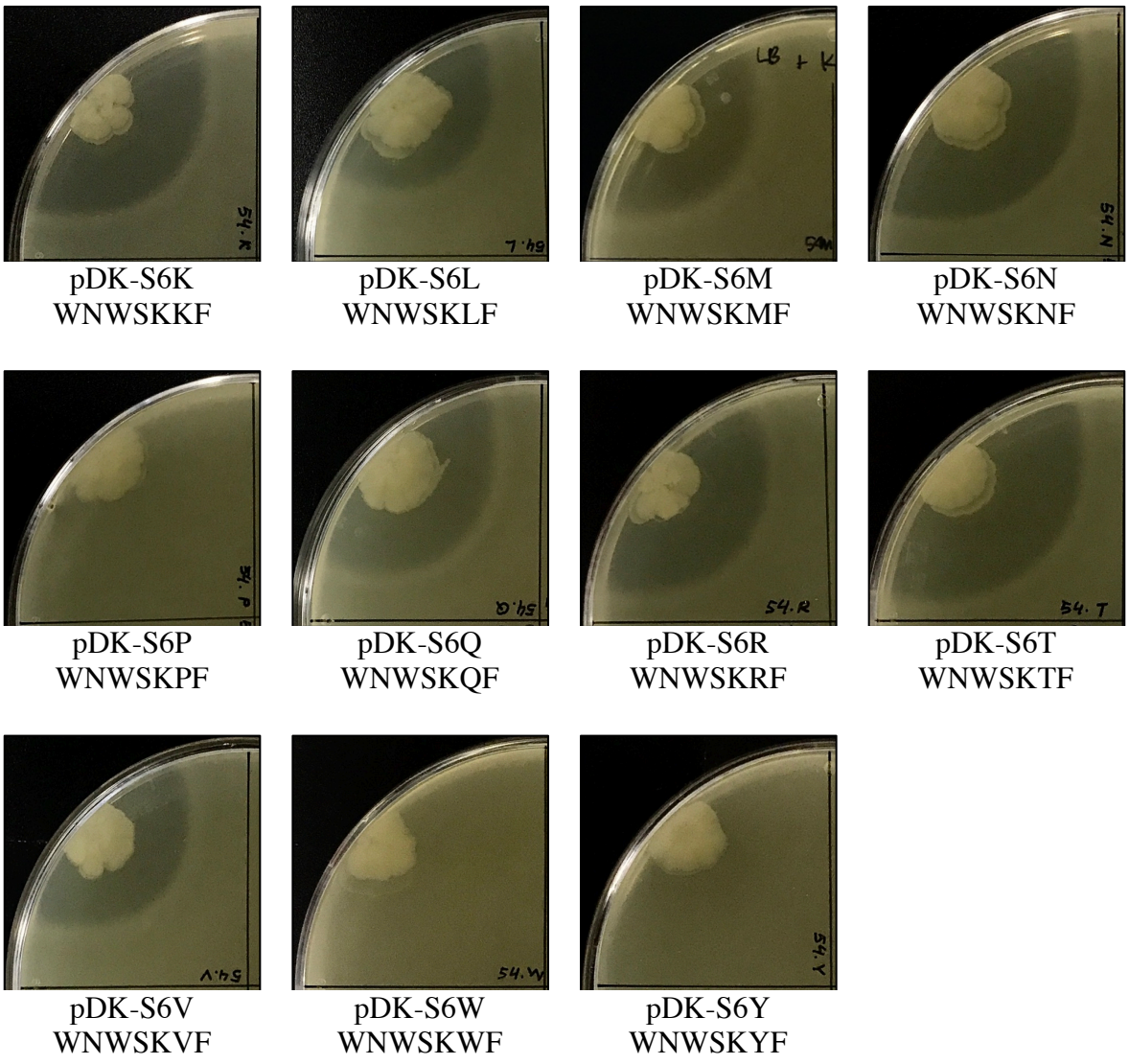
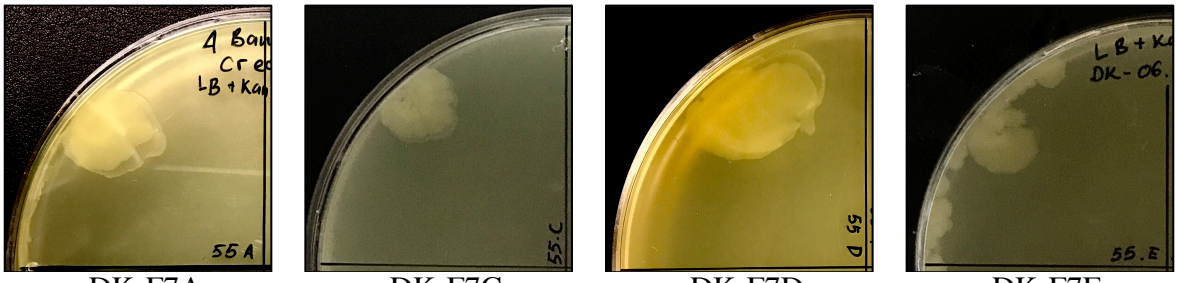


Figure S9. Agar activity assay of the heterologous producer (*E. coli* BAP1) carries a corresponding construct which the AA residue in the position 6 of DarA core peptide was mutated to other residues. Following the construct name is the core peptide AA sequence in DarA. The pathogen was *E.coli* MG1655 BamA6.



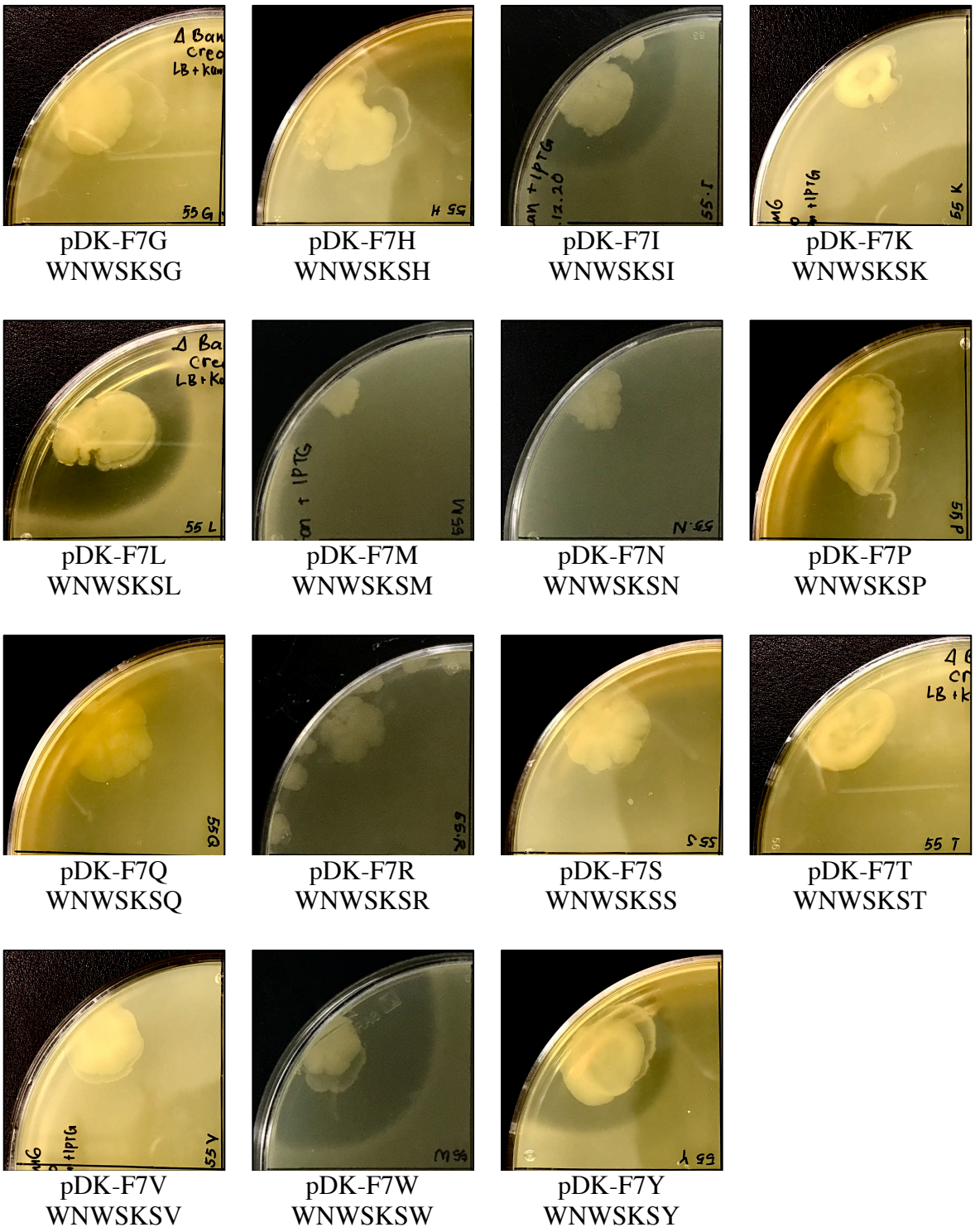
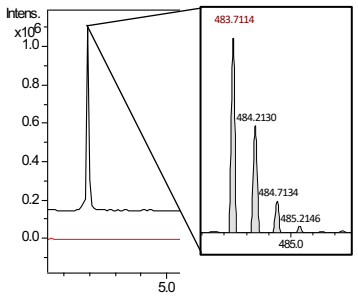
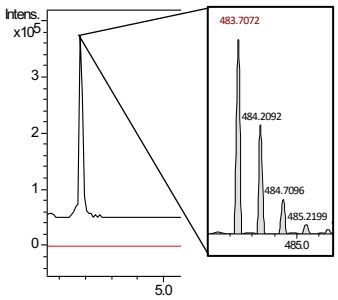


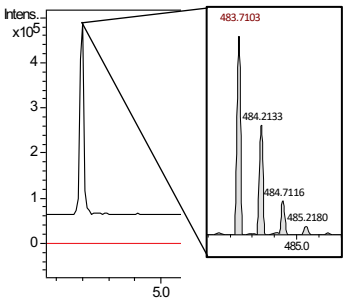
Figure S10. Agar activity assay of the heterologous producer (*E. coli* BAP1) carries a corresponding construct which the AA residue in the position 7 of DarA core peptide was mutated to other residues. Following the construct name is the core peptide AA sequence in DarA. The pathogen was *E.coli* MG1655 BamA6.



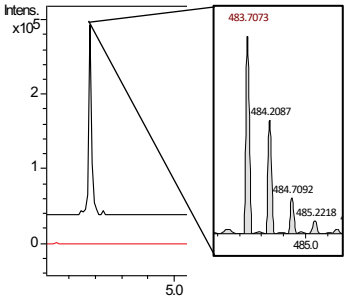
pZW-ADCΔfol
WN<u>W</u>S<u>K</u>SF



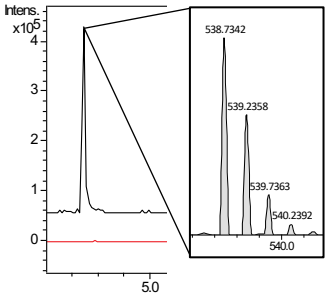
pDK-2
WN<u>W</u>S<u>K</u>SF



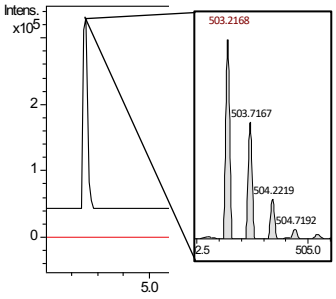
pDK-3
WN<u>W</u>S<u>K</u>SF



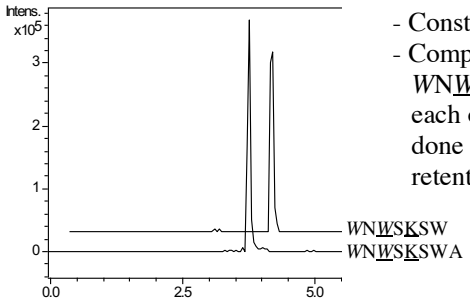
pDK-4
WN<u>W</u>S<u>K</u>SF



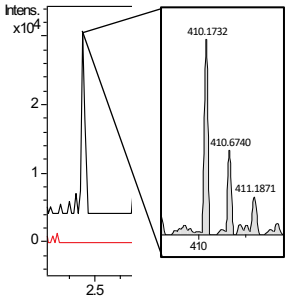
pDK-5
WN<u>W</u>S<u>K</u>SWA



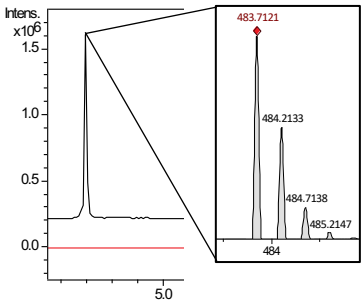
pDK-5
WN<u>W</u>S<u>K</u>SW



- Construct pDK-5
- Comparison EIC between the $[M+2H]^{2+}$ ion of *WN<u>W</u>S<u>K</u>SW* and *WN<u>W</u>S<u>K</u>SWA*. This analysis assumed that the ionization level of each of those two products are similar. The comparison cannot be done by UV_{280} absorption since they were coming out in the same retention time.

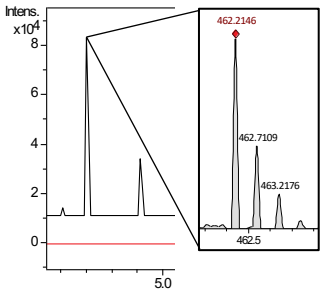


pDK-10
WN<u>W</u>S<u>K</u>S

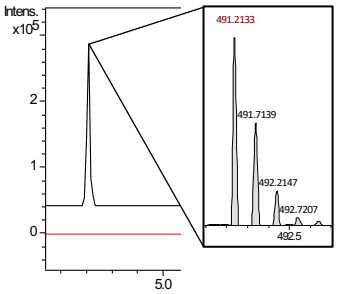


pZW-ADC5
WN<u>W</u>S<u>K</u>SF

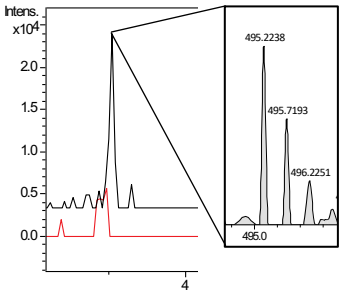
Figure S11. LCMS analysis in the "increasing and decreasing the size of the core peptide experiment". The LCMS chromatogram analysis of investigated product which the mass cannot be detected according to Table 1 is not shown. Following the construct name is the compound which is being analyzed. Italic letters represent in which residues the *C-O-C* bond are formed. Underline letters represent in which residues the *C-C* bond are formed. Black line represent the EIC of $[M+2H]^{2+}$ of the corresponding compound. Red line represents the same EIC parameters from the control.



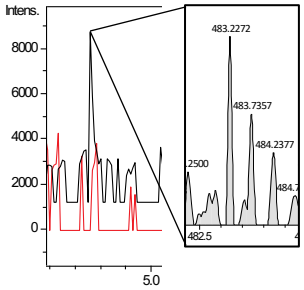
pDK-N2A
*WA*W*S*K*S*F (darobactin 13)



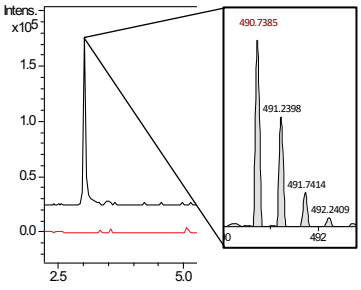
pDK-N2E
*WE*W*S*K*S*F (darobactin N2E)



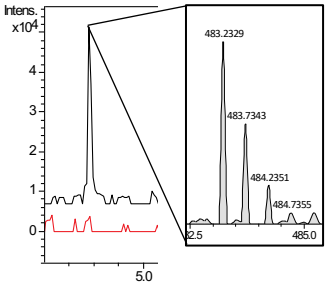
pDK-N2H
*WH*W*S*K*S*F (darobactin N2H)



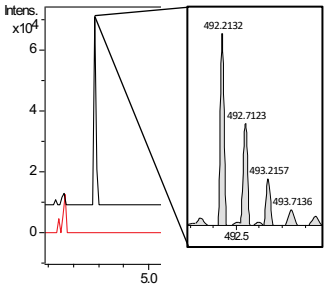
pDK-N2I
*WI*W*S*K*S*F (darobactin N2I)



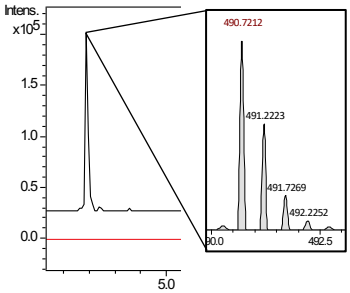
pDK-N2K
*WK*W*S*K*S*F (darobactin N2K)



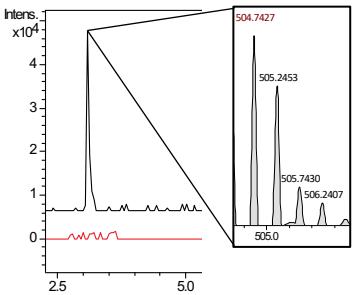
pDK-N2L
*WL*W*S*K*S*F (darobactin N2L)



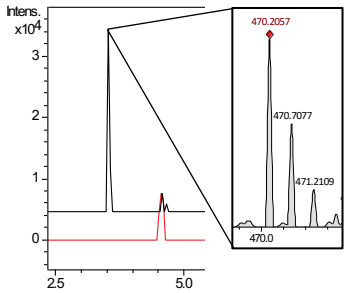
pDK-N2M
*WM*W*S*K*S*F (darobactin N2M)



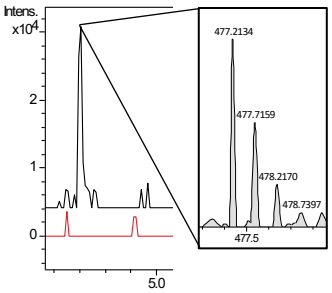
pDK-N2Q
*WQ*W*S*K*S*F (darobactin 11)



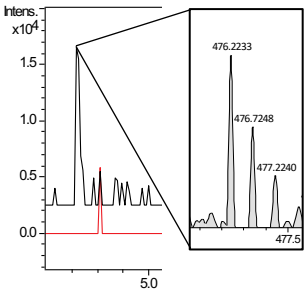
pDK-N2R
*WR*W*S*K*S*F (darobactin N2R)



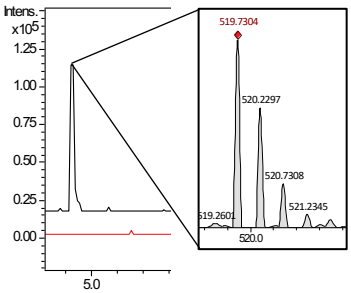
pDK-N2S
WS<u>*W*</u>S<u>K</u>SF (darobactin E)



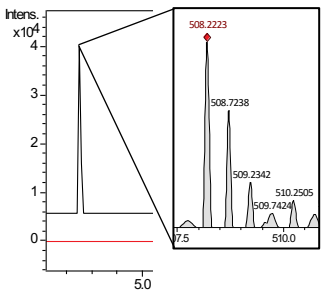
pDK-N2T
WT<u>*W*</u>S<u>K</u>SF (darobactin 12)



pDK-N2V
WV<u>*W*</u>S<u>K</u>SF (darobactin N2V)

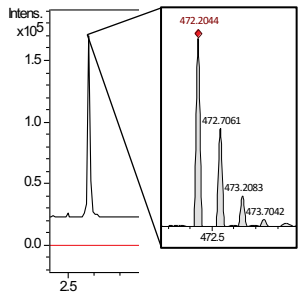


pDK-N2W
WW<u>*W*</u>S<u>K</u>SF (darobactin N2W)



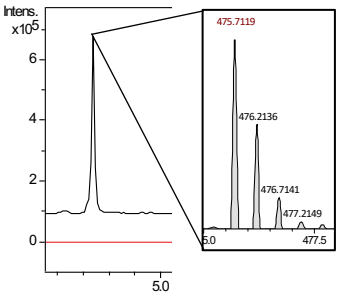
pDK-N2Y
WY<u>*W*</u>S<u>K</u>SF (darobactin N2Y)

Figure S12. LCMS analysis in the "AA exchange experiment" which the AA residue in the position 2 of DarA core peptide was mutated to other residues. The LCMS chromatogram analysis of investigated product which the mass cannot be detected according to Table 2 is not shown. Following the construct name below the LCMS analysis is the compound which is being analyzed. Italic letters represent in which residues the *C-O-C* bond are formed. Underline letters represent in which residues the *C-C* bond are formed. Black line represent the EIC of $[M+2H]^{2+}$ of the corresponding compound. Red line represents the same EIC parameters from the control.

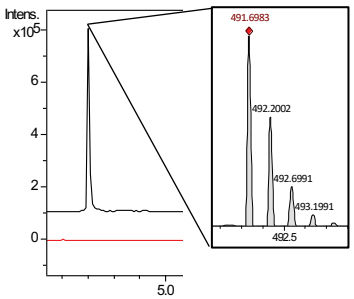


pDK-W3Y
WN<u>*Y*</u>S<u>K</u>SF (darobactin W3Y)

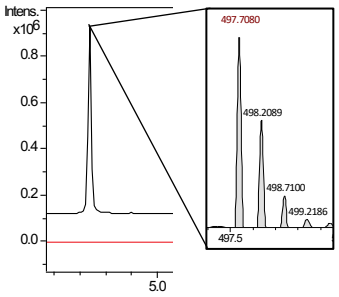
Figure S13. LCMS analysis in the "AA exchange experiment" which the AA residue in the position 3 of DarA core peptide was mutated to other residues. The LCMS chromatogram analysis of investigated product which the mass cannot be detected according to Table 2 is not shown. Following the construct name below the LCMS analysis is the compound which is being analyzed. Italic letters represent in which residues the *C-O-C* bond are formed. Underline letters represent in which residues the *C-C* bond are formed. Black line represent the EIC of $[M+2H]^{2+}$ of the corresponding compound. Red line represents the same EIC parameters from the control.



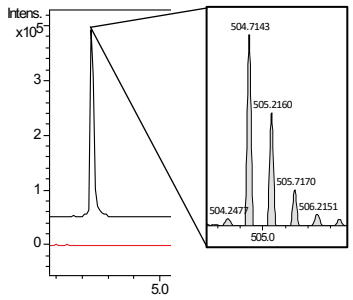
pDK-S4A
WN<u>*W*</u>A<u>K</u>SF (darobactin 14)



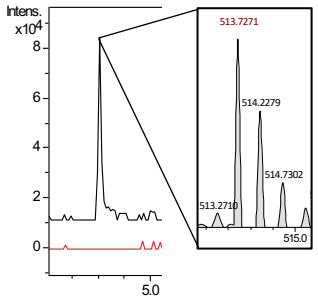
pDK-S4C
WN<u>*W*</u>C<u>K</u>SF (darobactin 6)



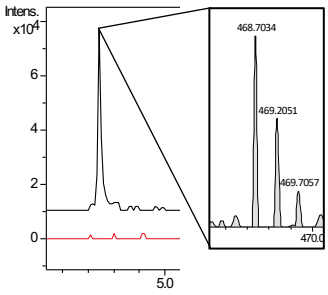
pDK-S4D
WN<u>*W*</u>D<u>K</u>SF (darobactin S4D)



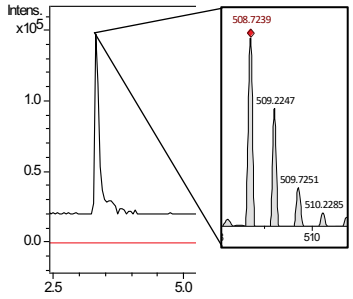
pDK-S4E
*WN*W*E*K*S*F (darobactin S4E)



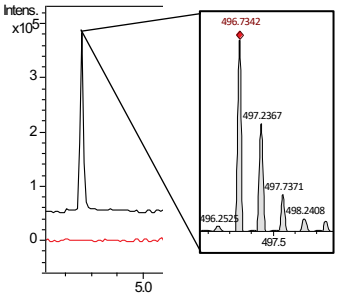
pDK-S4F
*WN*W*F*K*S*F (darobactin S4F)



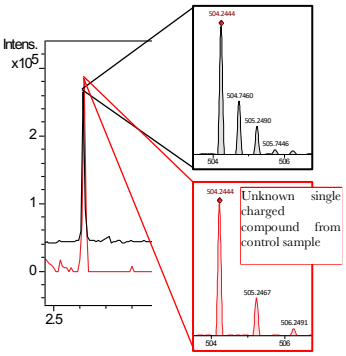
pDK-S4G
*WN*W*G*K*S*F (darobactin S4G)



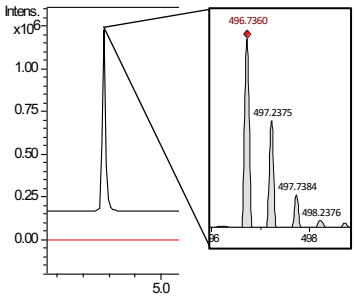
pDK-S4H
*WN*W*H*K*S*F (darobactin S4H)



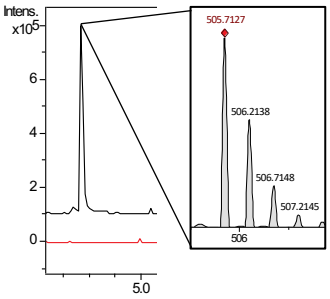
pDK-S4I
*WN*W*I*K*S*F (darobactin S4I)



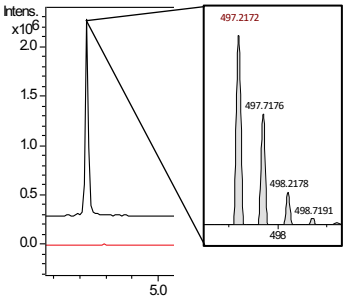
pDK-S4K
*WN*W*K*K*S*F (darobactin S4K)



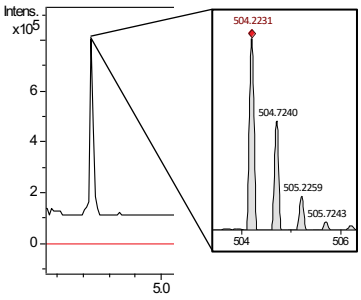
pDK-S4L
*WN*W*L*K*S*F (darobactin S4L)



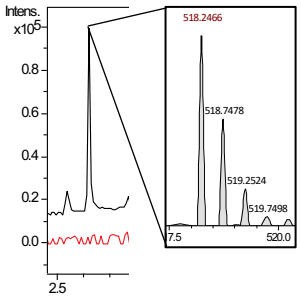
pDK-S4M
*WN*W*M*K*S*F (darobactin S4M)



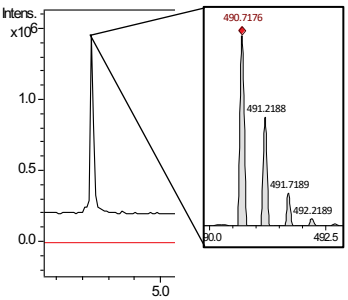
pDK-S4N
*WN*W*N*K*S*F (darobactin S4N)



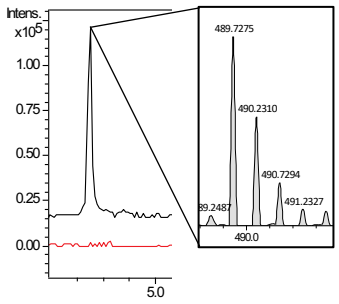
pDK-S4Q
*WN*W*Q*K*S*F (darobactin S4Q)



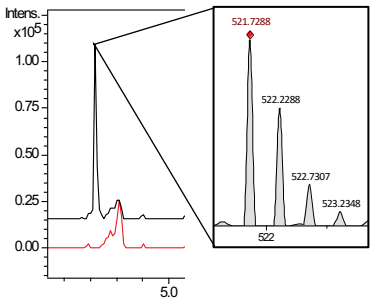
pDK-S4R
*WN*W*R*K*S*F (darobactin S4R)



pDK-S4T
*WN*W*T*K*S*F (darobactin 4)

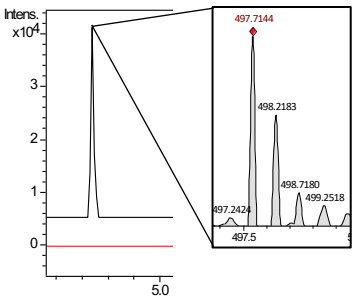


pDK-S4V
*WN*W*V*K*S*F (darobactin S4V)



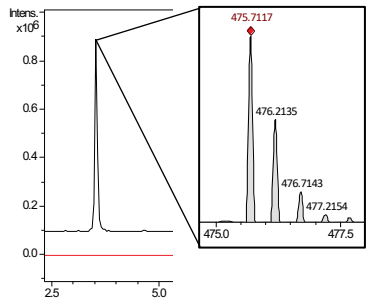
pDK-S4Y
*WN*W*Y*K*S*F (darobactin S4Y)

Figure S14. LCMS analysis in the "AA exchange experiment" which the AA residue in the position 4 of DarA core peptide was mutated to other residues. The LCMS chromatogram analysis of investigated product which the mass cannot be detected according to Table 2 is not shown. Following the construct name below the LCMS analysis is the compound which is being analyzed. Italic letters represent in which residues the *C-O-C* bond are formed. Underline letters represent in which residues the *C-C* bond are formed. Black line represent the EIC of $[M+2H]^{2+}$ of the corresponding compound. Red line represents the same EIC parameters from the control.

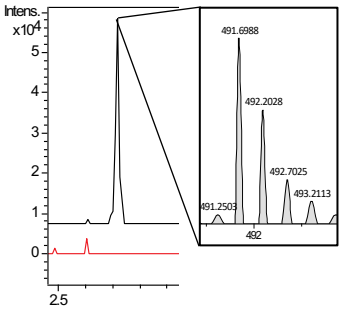


pDK-K5R
*WN*W*S*R*S*F (darobactin D)

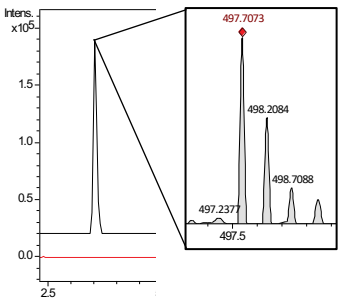
Figure S15. LCMS analysis in the "AA exchange experiment" which the AA residue in the position 5 of DarA core peptide was mutated to other residues. The LCMS chromatogram analysis of investigated product which the mass cannot be detected according to Table 2 is not shown. Following the construct name below the LCMS analysis is the compound which is being analyzed. Italic letters represent in which residues the *C-O-C* bond are formed. Underline letters represent in which residues the *C-C* bond are formed. Black line represent the EIC of $[M+2H]^{2+}$ of the corresponding compound. Red line represents the same EIC parameters from the control.



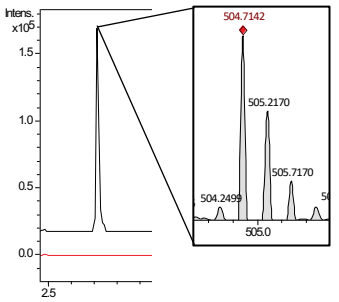
pDK-S6A
*WN*W*S*K*A*F (darobactin 16)



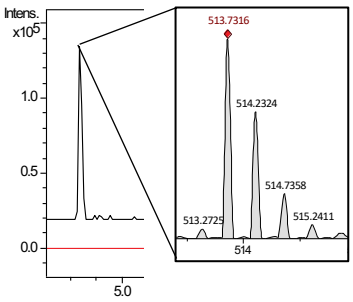
pDK-S6C
*WN*W*S*K*C*F (darobactin 8)



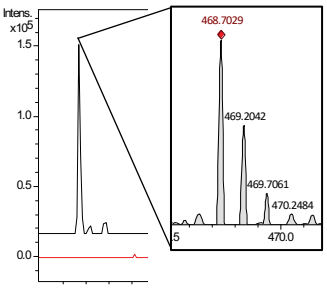
pDK-S6D
*WN*W*S*K*D*F (darobactin S6D)



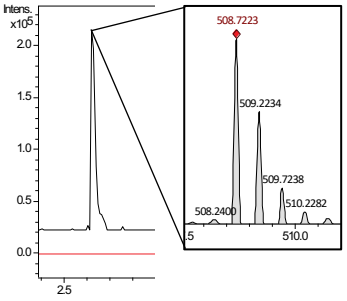
pDK-S6E
*WN*W*S*K*E*F (darobactin S6E)



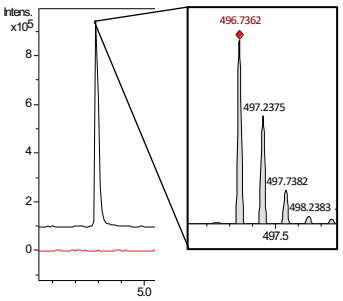
pDK-S6F
*WN*W*S*K*F*F (darobactin S6F)



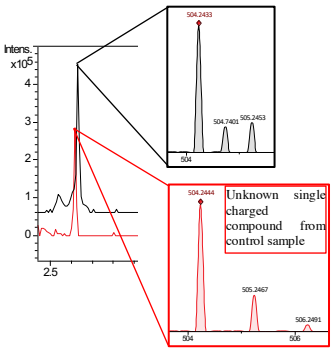
pDK-S6G
*WN*W*S*K*G*F (darobactin S6G)



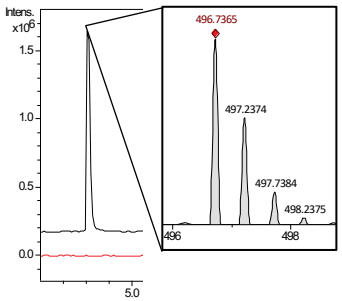
pDK-S6H
*WN*W*S*K*H*F (darobactin S6H)



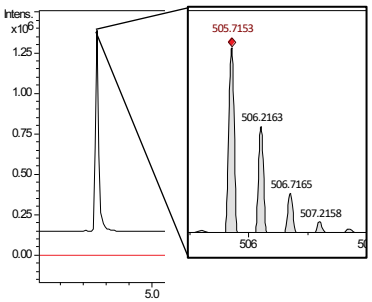
pDK-S6I
*WN*W*S*K*I*F (darobactin S6I)



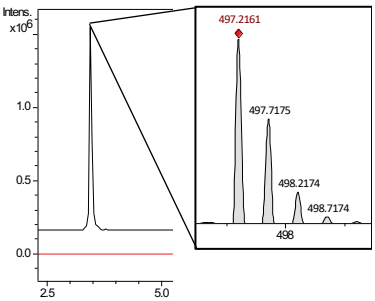
pDK-S6K
*WN*W*S*K*K*F (darobactin S6K)



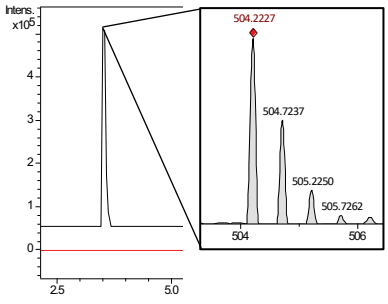
pDK-S6L
*WN*W*S*K*L*F (darobactin S6L)



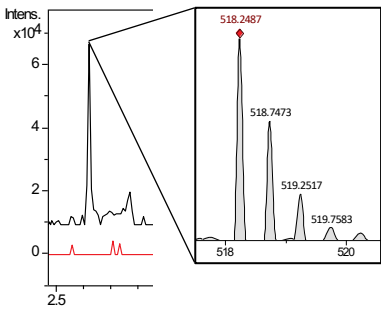
pDK-S6M
*WN*W*S*K*M*F (darobactin S6M)



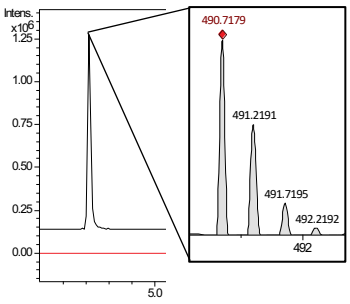
pDK-S6N
*WN*W*S*K*N*F (darobactin S6N)



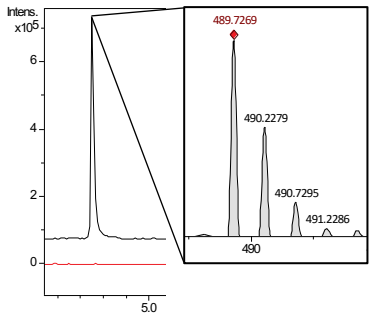
pDK-S6Q
*WN*W*S*K*Q*F (darobactin S6Q)



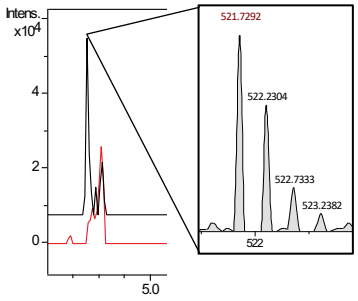
pDK-S6R
*WN*W*S*K*R*F (darobactin S6R)



pDK-S6T
*WN*W*S*K*T*F (darobactin S6T)

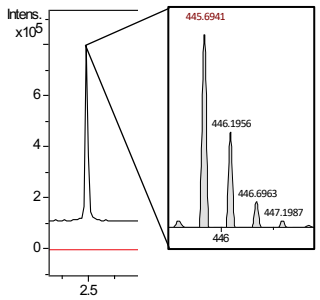


pDK-S6V
*WN*W*S*K*V*F (darobactin S6V)

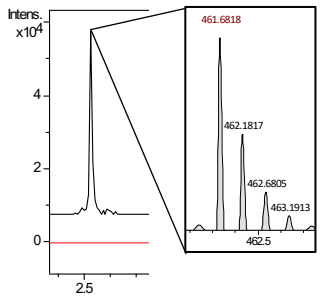


pDK-S6Y
*WN*W*S*K*Y*F (darobactin S6Y)

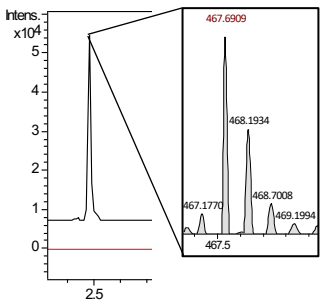
Figure S16. LCMS analysis in the "AA exchange experiment" which the AA residue in the position 6 of DarA core peptide was mutated to other residues. The LCMS chromatogram analysis of investigated product which the mass cannot be detected according to Table 2 is not shown. Following the construct name below the LCMS analysis is the compound which is being analyzed. Italic letters represent in which residues the *C-O-C* bond are formed. Underline letters represent in which residues the *C-C* bond are formed. Black line represent the EIC of $[M+2H]^{2+}$ of the corresponding compound. Red line represents the same EIC parameters from the control.



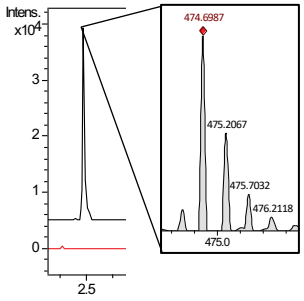
pDK-F7A
*WN*W*S*K*S*A (darobactin 17)



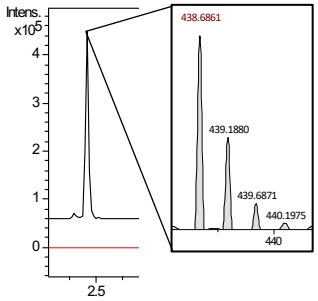
pDK-F7C
*WN*W*S*K*S*C (darobactin F7C)



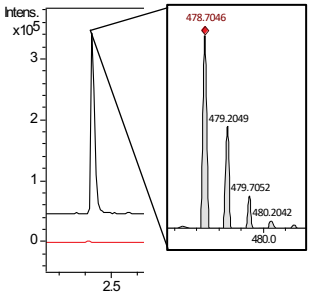
pDK-F7D
*WN*W*S*K*S*D (darobactin F7D)



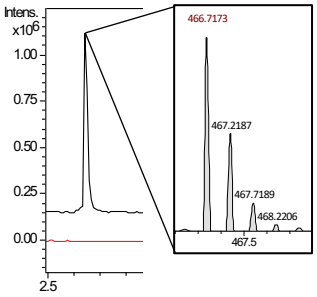
pDK-F7E
*WN*W*S*K*S*E (darobactin F7E)



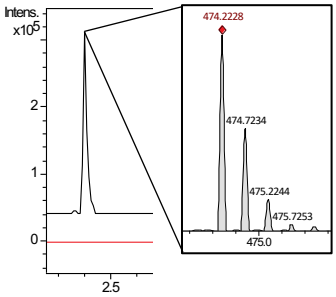
pDK-F7G
*WN*W*S*K*S*G (darobactin F7G)



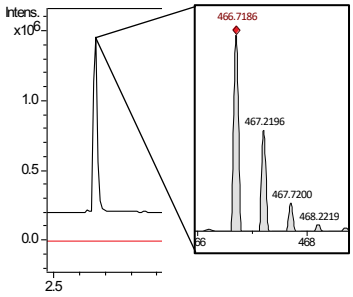
pDK-F7H
*WN*W*S*K*S*H (darobactin F7H)



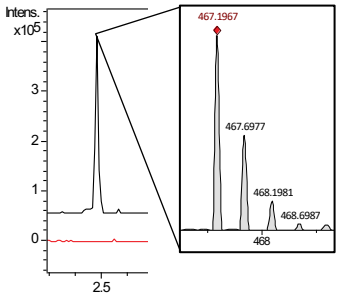
pDK-F7I
*WN*W*S*K*S*I (darobactin F7I)



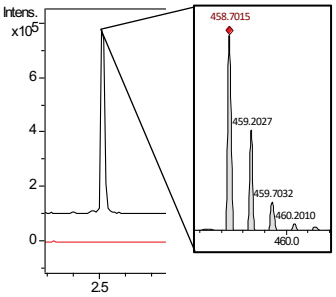
pDK-F7K
*WN*W*S*K*S*K (darobactin F7K)



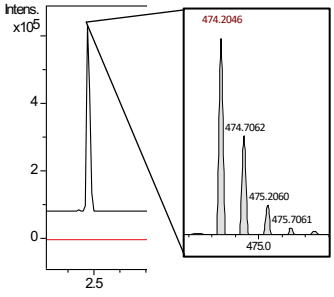
pDK-F7L
*WN*W*S*K*S*L (darobactin F7L)



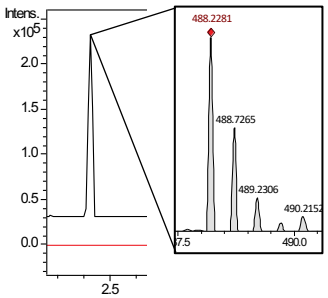
pDK-F7N
*WN*W*S*K*S*N (darobactin F7M)



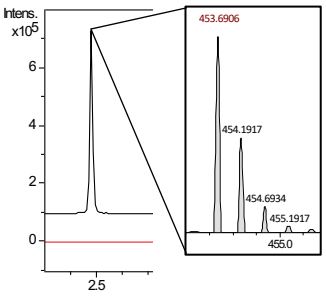
pDK-F7P
*WN*W*S*K*S*P (darobactin F7N)



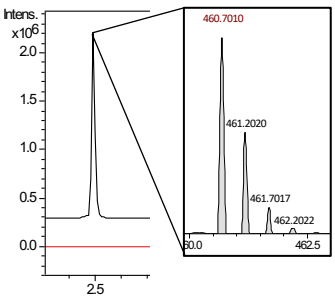
pDK-F7Q
*WN*W*S*K*S*Q (darobactin 18)



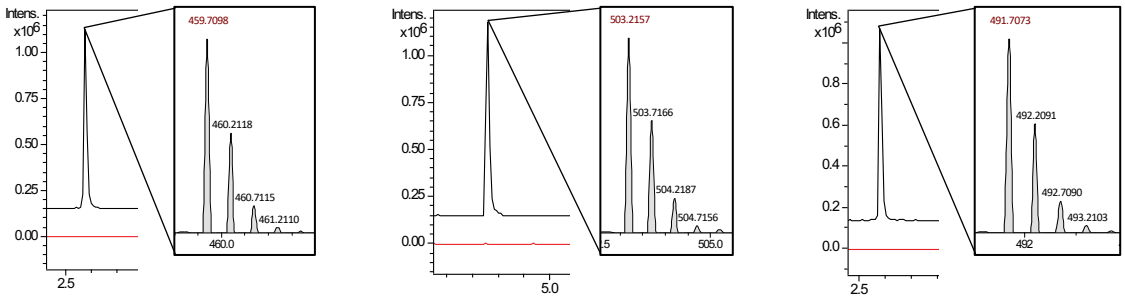
pDK-F7R
*WN*W*S*K*S*R (darobactin F7R)



pDK-F7S
*WN*W*S*K*S*S (darobactin F7S)



pDK-F7T
*WN*W*S*K*S*T (darobactin F7T)



pDK-F7V *W*N*W*S*K*SA (darobactin F7V)

pDK-F7W *W*N*W*S*K*SA (darobactin 9)

pDK-F7Y *W*N*W*S*K*SA (darobactin 10)

Figure S17. LCMS analysis in the "AA exchange experiment" which the AA residue in the position 7 of DarA core peptide was mutated to other residues. The LCMS chromatogram analysis of investigated product which the mass cannot be detected according to Table 2 is not shown. Following the construct name below the LCMS analysis is the compound which is being analyzed. Italic letters represent in which residues the *C-O-C* bond are formed. Underline letters represent in which residues the *C-C* bond are formed. Black line represent the EIC of $[M+2H]^{2+}$ of the corresponding compound. Red line represents the same EIC parameters from the control.

References

1. Madeira, F. *et al*. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* **47**, W636–W641 (2019). doi: 10.1093/nar/gkz268

8. Chapter VII

Genome-mining-guided discovery and characterization of a RiPP BGC originated from *Sulfidibacter corallicola* M133^T

Status: Gathering data

Summary: The goal was to discover new Gram-negative antibiotic compounds, which act as a BamA inhibitor, such as the recently described darobactin. Darobactin is a ribosomally synthesized and post-translationally modified peptide (RiPP). The minimum biosynthetic gene cluster (BGC) consists of *darA* encoding a precursor peptide (pre-peptide) and *darE* encoding the single post-translationally modifying enzyme. DarE is a radical SAM (RaS) enzyme. To identify proteins related to DarE, public databases integrated into the NCBI website using BLASTp were mined *in silico* using a signature of a conserved region of DarE as the query. Out of the hit results, one biosynthetic gene cluster (BGC) named *mcd* consisting of a RaS (*mcdA*) and a pre-peptide (*mcdB*) was discovered. This BGC was synthesized using the natural wild type codons. Thereafter, the BGC was cloned to an expression vector and transferred to *E. coli* as a heterologous host. After heterologous expression and analysis of the resulting extracts by LCMS analysis, an ion corresponding to the core peptide of McdB with additional one oxygen and reduced two hydrogens was identified. Production, purification and characterization of this compound is necessary.

Contribution: I Dewa Made Kresna did the *in silico* research, designed and planed the cloning strategies and the heterologous expression. He did literature research, planned and performed the experiments. He analyzed the results and put them into a discussion.

Genome mining-guided discovery and characterization of a RiPP BGC originated from *Sulfidibacter corallicola* M133^T

I Dewa M. Kresna[a,b], Till F. Schäberle[a,b,c]

[a] Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, Ohlebergsweg 12, 35392 Giessen, Germany.

[b] Department of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Ohlebergsweg 12, 35392 Giessen, Germany.

[c] German Center for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Giessen, Germany.

Abstract

The discovery of new antibiotics is important due to the antimicrobial resistance crisis. Especially multi resistant Gram-negative bacteria are of highest concern in this regard. Recently, the potent gram-negative antibiotic darobactin was discovered. This study reports the identification of a new darobactin-like compound guided by a genome mining approach. The minimum BGC of darobactin consists of *darA* encoding a precursor peptide (pre-peptide) and *darE* encoding a radical SAM (RaS) enzyme. DarE is the single post-translationally modifying enzyme that catalyzes the two intramolecular ring closures in DarA. Genome mining was performed based on the investigation of proteins with similarity to DarE in the NCBI database. A signature sequence representing a conserved motif of DarE proteins was used as the query for a BLASTp search. The hits (putative RaS enzymes) became the basis for further investigation, such as the presence of a pre-peptide in the surrounding area. This led to the identification of a newly named *mcd* BGC consisting of *mcdA* encoding a RaS enzyme and *mcdB* encoding a pre-peptide. Heterologous expression was employed to express the putative BGC. The product was identified by LCMS profiling of the extract and molecular networking. The identified ions match to the putative core peptide sequence of McdB with additional one oxygen and loss of two hydrogens. The modification could be attributed to the formation of a *C-O-C* (ether bond) like it is the case in darobactin. The LCMS-identified product was named mono cyclic (mc-) darobactin.

Introduction

There is a need to discover new antibiotics, since pathogens become resistant to available antibiotics.[1] It is not easy to discover compounds that target gram-negative bacteria.[2,3] In 2019, darobactin, a compound that selectively kills gram-negative pathogens was reported.[4] It is showing no toxicity in mice experiments at the concentrations tested so far, indicating a promising pharmaceutical window and its suitability for development into therapeutics. Due to its potential, many possible derivatives have been studied and characterized to increase its potential[5,6]. Darobactin is a ribosomally synthesized and post-translationally modified peptide (RiPP). The linear heptapeptide core is stabilized by two unusual intramolecular rings. Its minimum biosynthetic gene cluster (BGC) consists of *darA* encoding a precursor peptide and *darE* encoding a radical SAM (RaS) enzyme.[7] Production of darobactin using an optimized heterologous expression *E. coli* system as the host has shown improvement of the yield up to 10 fold and a decrease of cultivation time down to 5 fold in comparison to the original report.[7]

Natural products play an important role as a source for bioactive molecules.[8] The traditional bioactivity-guided isolation approach is the most popular choice to discover new bioactive metabolites. However, because of its popularity, several problems have been identified, such as (1) the rediscovery of known compounds, (2) uncultivable bacteria under laboratory conditions and (3) silent (unexpressed) BGCs.[8] In the last decade, additional data analysis became more and more important to prioritize samples before embarking into laborious isolation and profiling experiments. An increase in publicly available sequence data in combination with great advances in bioinformatics tools for the detection of BGCs, and improvements of molecular biology techniques, today represent an option to apply a genome mining approach.[9] BLAST (Basic Local Alignment Search Tool) is one of many bioinformatic tools that helps to find regions of similarity between biological sequences.[10] The program could compare nucleotide or protein sequences to sequence NCBI databases and calculate the statistical significance.

With all of these developments and the potential of darobactin, it is of interest to discover a new RiPP compound like darobactin. DarE as the single modifying enzyme in its biosynthetic pathway was the basis for genome mining in this study. Initially, the conserved motifs of DarE were analyzed. The motifs were used as a query for BLASTp search throughout public databases integrated into the NCBI website. The hits were further investigated to analyze the

presence of pre-peptide in the surrounding area. A putative BGC named *mcd* BGC was cloned into an expression vector and was expressed heterologously. LCMS chromatogram analysis was performed on the cultivation extract and compared to control (extract from the heterologous expression of empty pRSFduet vector). Two ions corresponding to one compound were identified to be present in the transgenic strain extract and absent in the control extract, indicating a product of the *mcd* BGC. These ions would be the guide for production and purification. The product was named mono cyclic (mc-) darobactin.

Result and Discussion

The general scheme of the workflow is shown in Figure 1. The signature used as query in a BLASTp search was chosen based on the conserved residues of DarE. To identify the conserved residues, a BLASTp search using the protein sequence of DarE from *P. khanii* (acc. number: WP_152962147.1) was performed and the first 100 hits were subsequently aligned using ClustalW[11] (Figure 2).

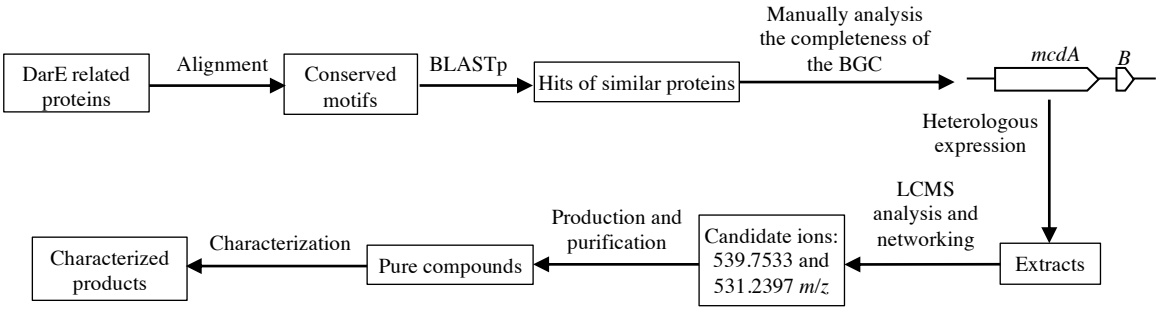


Figure 1. The general scheme of the genome mining-based workflow from the protein of interest, *i.e.* DarE, until characterization of the product derived from the heterologous expression strain, *i.e.* mono cyclic darobactin.

DarE as a RaS enzyme is divided into four domains – RiPP precursor peptide recognition element (RRE), radical SAM, bridging region and SPASM domain as named and described by the crystal structure of SuiB[12]. The function of the RRE domain is still not clear and some RaS enzymes do not have this domain. Therefore, signatures were not taken from this domain. The bridging region is the domain that moves the most (flexible part) upon binding and catalyzation to its substrate as it was described for SuiB.[12] This feature makes an unclear relevance, which residues are important and even though there are two cysteines responsible to ligate the

auxiliary iron-sulphur cluster, these two cysteines are not part of a stable motif, since their position is changing among RaS enzymes[12]. The SPASM domain is a domain that contains the $CX_2CX_5CX_3C$ motifs involved in the ligation/binding of the other two auxiliaries [4Fe-4S] clusters.[13] These two auxiliary clusters are believed to be involved in the electron transfer (redox reaction) during catalysation.[14] A redox role to the auxiliary Fe–S cluster, which is active in the 2+ oxidation state but is converted to the 1+ form during the course of catalysis. The active site cluster, which reductively cleaves SAM, is active in the 1+ oxidation state but is converted to the 2+ form after generation of the 5′-dA•. After one turnover, intramolecular electron transfer from the auxiliary Fe–S cluster to the other would render both in the active oxidation state. Alternatively, each could undergo separate redox reactions with a diffusible redox partner.[14] In DarE, however, the first cysteine in the $CX_2CX_5CX_3C$ motif is mutated to methionine (M382, Figure 3). Instead, there is an additional cysteine at either position 413 or 414, which was suspected to be essential. To investigate whether residues M382, C413, and C414 are essential, constructs with mutation as shown in Table 1 were created from the template of previously constructed darobactin expression plasmid pZW-ADC5[7] containing *dar* BGC from *P. khanii*. Additionally, the essentiality of the cysteine residue at position 260 was also investigated. These constructs with the positive control (pZW-ADC5) were heterologously expressed and the cell extracts were subjected to LCMS for analyzing the ability to produce the final product darobactin (Figure 2). Mutation of either M382 or C260 to A did not abolish darobactin production indicating these residues are not essential. While mutating either C413 or C414 abolished darobactin production indicating their essentiality. Additional experiments to test whether mutation of either C413 or C414 to A can be exchanged in trans by mutation of M382 to C could produce darobactin were performed. However, those constructs could not produce darobactin indicating C413 or C414 cannot be exchanged in trans with C at position 382. Conclusively, the popular motif $CX_2CX_5CX_3C$ in SPASM domain is not entirely fitting to DarE.

Table 1. The variant of darobactin expression constructs containing *dar* BGC from *P. khanii* with and without mutation of codon(s) at gene encoding DarE. In darobactin production column, "+" means darobactin production could be observed, while "-" means otherwise as more detail shown in Figure 2.

Construct name	Mutation in DarE	Darobactin production
pZW-ADC5 (positive control)	-	+
pDK-M382A	M382A	+
pDK-C413A	C413A	-
pDK-C414A	C414A	-
pDK-C260A	C260A	+
pDK-M382C::C413A	M382C and C413A	-
pDK-M382C::C414A	M382C and C414A	-

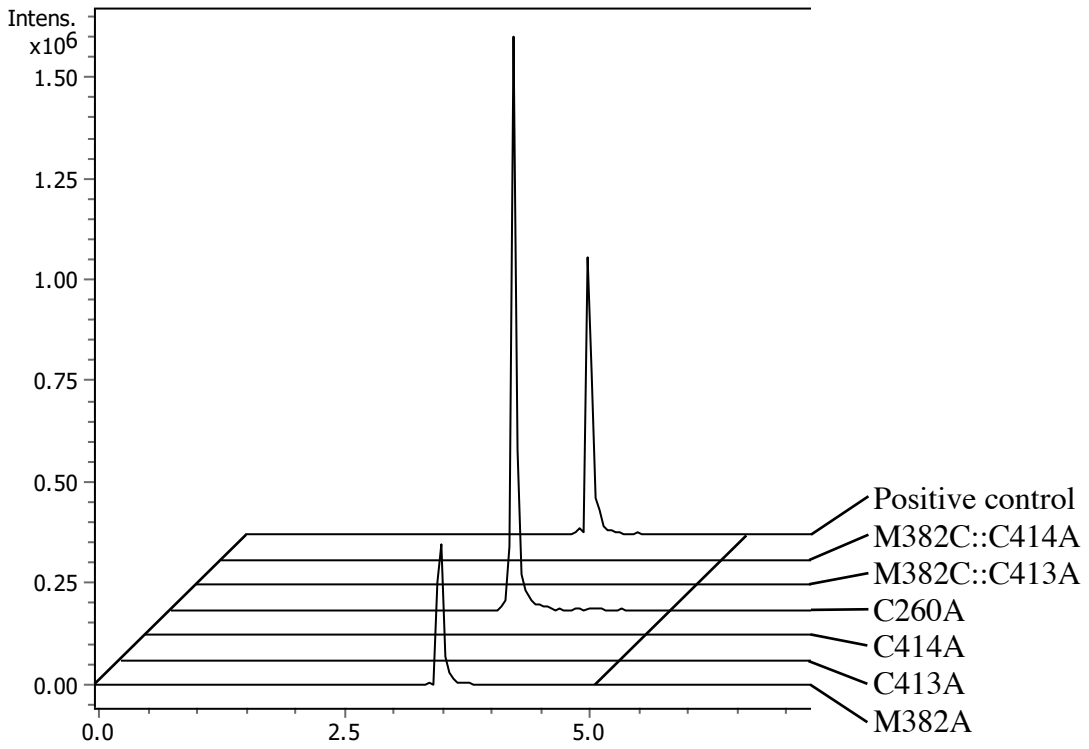


Figure 2. LCMS analysis of the cell extract of heterologous producer expressing variant darobactin constructs (Table 1). The chromatograms were extracted to $[M+2]^{2+}$ ion of darobactin (483.7089 *m/z*).

The radical SAM domain is the domain together with SAM that holds/binds the crucial [4Fe-4S] cluster. Signatures from this domain were considered as the most interesting. However, if the signature was selected too specific, it leads to the known DarE from darobactin BGC. If it is too relaxed, it leads to a broad spectrum of unrelated RaS enzymes. In many cases, no pre-peptide could be identified in the surrounding of the RaS meaning that the putative BGCs either might not produce RiPPs, or the pre-peptide is located somewhere else in the genome. A signature that gave the most relevant hits was EFXWHGGE (Figure 3). After BLASTp search using this signature as query, a BGC originated from *Sulfidibacter corallicola* M133^{T} (isolated from coral *Porites lutea*, China[15]), containing genes encoding a RaS (*mcdA*) and a pre-peptide (*mcdB*), was identified and named *mcd* BGC (Figure 1).

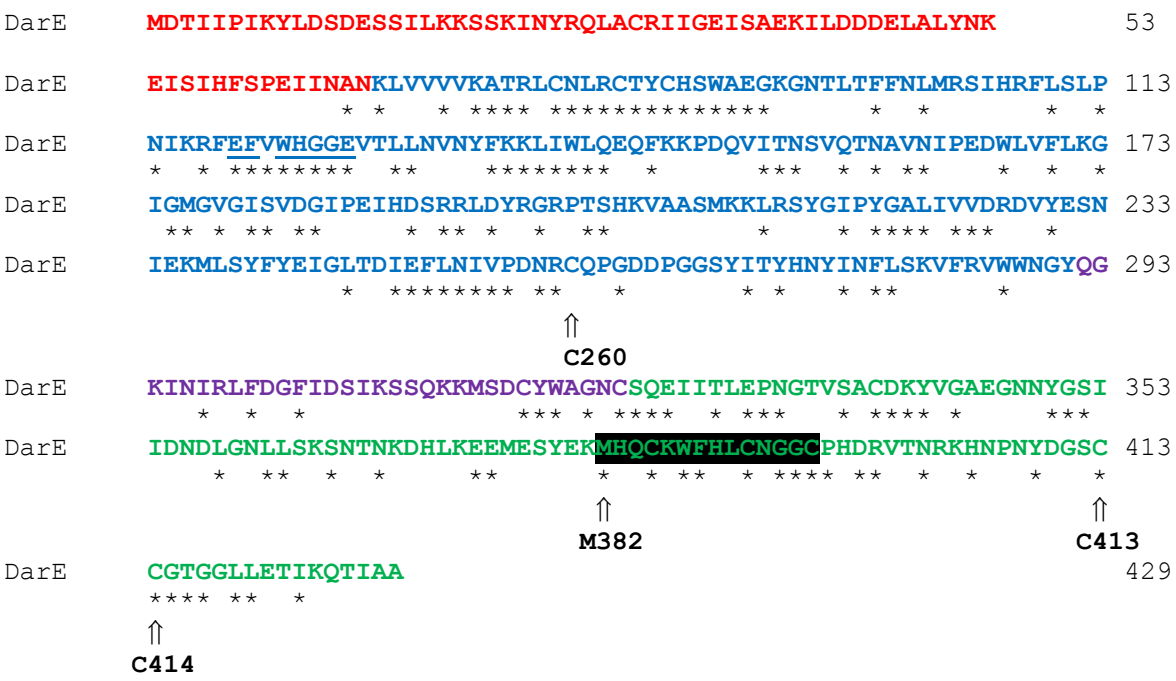


Figure 3. Conserved residues in DarE from *P. khanii*. Asterisks signify the conserved residues. Colored letters represent to which domain the residues are located. Red, blue, purple and green colored residues belong to the domain of RRE, radical SAM, bridging region and SPASM domain, respectively. Underline letters represent the chosen signature. Black highlighted letters represents the $CX_2CX_5CX_3C$ motif.

This *in silico* identified BGC was cloned into a pRSFduet expression vector, yielding pRSF.mcdBGC and subsequently transferred to *E. coli* Rosetta as a heterologous host. The empty pRSFduet vector was used as a control. After heterologous expression, the extracts were subjected to LCMS analysis. By manually analyzing and using molecular networking tools[16]

(Figure S1), two double-charged ions with an *m/z* of 539.7533 and 531.2397 could be identified from the transgenic strain extract and were absent in the control extract, indicating the product of the *mcd* BGC (Figure 4). The difference of 8.5136 *m/z* between both double-charged ions corresponded to the loss of an amino group from the parent mass. The same phenomenon was shown for darobactin ions[7]. The parent $[M+2H]^{2+}$ ion with an *m/z* of 539.7533 matches the chemical formula of $C_{56}H_{63}N_{13}O_{10}$ (error: 8.1 ppm). This chemical formula fits the putative core peptide sequence of McdB (WRWSWPF) with one additional oxygen and loss of two hydrogen atoms. The modification could be attributed to the formation of a *C-O-C* ether bond, like it is the case for darobactins.

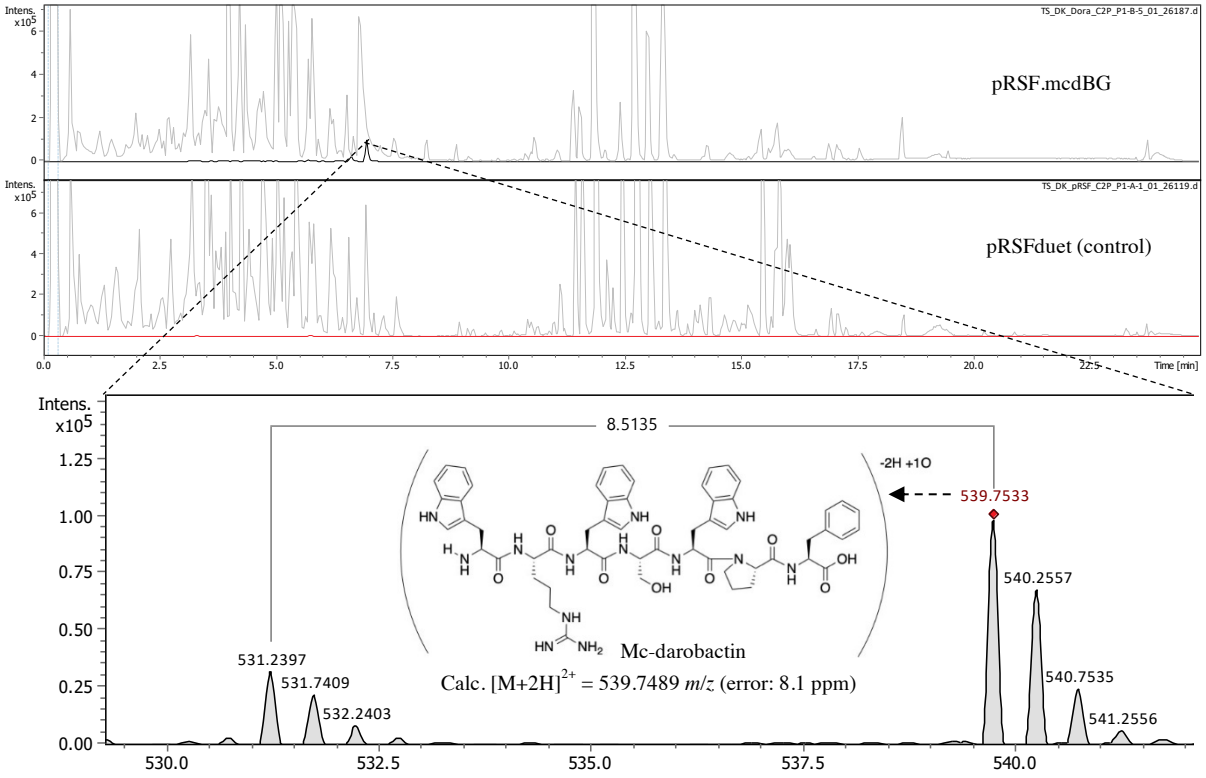


Figure 4. LCMS analysis of extracts derived from fermentation. In the LCMS chromatogram, the grey line represents the BPC, while the black and red lines are EICs for 539.7489 *m/z*, corresponding to the $[M+2H]^{2+}$ of mono cyclic (mc-) darobactin. The ions corresponding to mc-darobactin are absent in the control.

Conclusion

In an effort to discover new darobactin-like compounds, a newly named *mcd* BGC was identified by genome mining approach based on the investigation of similar proteins like DarE

from the NCBI database. The BGC consists of *mcdA*, encoding a RaS enzyme and *mcdB*, encoding a pre-peptide with the core peptide AA sequence of $W^1R^2W^3S^4W^5P^6F^7$. This BGC was synthetized, cloned to the pRSFduet expression vector and heterologously expressed in *E. coli* cell. LCMS analysis of the extract revealed two double-charged ions 539.7533 and 531.2397 *m/z* that were absent in the control extract (empty expression vector) indicating the product of the *mcd* BGC. The latter ion corresponds to the loss of an amino group from the parent ion 539.7533 *m/z*. The parent ion fits the core peptide sequence of McdB with additional one oxygen and loss of two hydrogens. The modification could be expected to be the same as in darobactin - formation of a *C-O-C* (ether) ring (first ring) between W^1 and W^3. A product which possesses an additional second ring (*C-C* ring like in darobactin) was not detected. The second ring (*C-C*) in darobactin is formed between W^3 and K^5. Upon changing of K^5 to W^5, the second ring formation was not occurring as shown in the previous experiment (Chapter VI). This might explain why a second ring is not formed in the core peptide of McdB, in which a W is located at position 5. However, structure elucidation by NMR of mc-darobactin is still pending. It is also of interest to evaluate its bioactivity profile.

Material and Methods

Cloning of mcd BGC

The *mcd* BGC was synthetically generated and purchased. The nucleotide sequence is shown in Sequence S1. The BGC fragment was replicated by PCR using primer dora.F (atattagttaagtataagaaggagatatacatatgaattcttccatgcccctcgc) and dora.R (gcagcagcctaggttaattagaaaggccagctccaacgc) in a total volume of 200 µL (50 µL x 4 PCR reactions). The PCR product (insert) was purified through agarose gel purification. For the expression vector, pRSFduet was used. The plasmid was linearize by PCR using primer pRSF.universal.F (ttaacctaggctgctgccaccg) and pRSF.universal.R (atgtatatctccttcttatacttaactaatatactaagatggggaattgt). The PCR product (linearized vector) was purified by agarose gel. The linearized vector and insert were assembled by Gibson assembly (GA) method[17]. Thereafter, the GA mix was dialyzed for 20-30 min to remove salts and transferred to *E.coli* Rosetta by electroporation. Survived colonies under 50 µg/mL of kanamycin and 25 µg/mL of chloramphenicol were tested by PCR with primer E6.4.check.F (ccataccgcgaaaggttttgcg) and uni.pRSF.check.R (ttgatgttggacgagtcggaatcg). Positive PCR colonies were sanger sequenced using primer E6.4.check.F. Sequenced-positive colony was

grown at 37 °C overnight LB containing the same antibiotics as before and stored in -80 °C with a final concentration of 25% glycerol.

Construction of the variant of darobactin expression constructs with DarE mutations

The initial template construct for PCR was pZW-ADC5. It consists of a pRSFduet vector and modified *dar* BGC from *P. khanii* HGB1456.[7] All mutations were incorporated using primers as shown in Table 1.

Table S1. The primer list used in this study.

Primer	Nucleotide sequence (5' to 3')
M382A	AGGAAATGGAATCATATGAAAAAGCGCATCAATGTAAATGG TTTCATTTGTGTAATGGTGG
M382.R	TTTTTCATATGATTCCATTTCCTCTTTAAGATGATCCTTATTTG
M382C.F	AGGAAATGGAATCATATGAAAAATGCCATCAATGTAAATGGT TTCATTTGTGTAATGGTGG
C413A.F	ACAATCCAAATTATGATGGTTCAGCGTGTGGAACCGGCGGTT TGT
C414A.F	ACAATCCAAATTATGATGGTTCATGTGCGGGAACCGGCGGTT TGTTGGA
C413-14A.R	TGAACCATCATAATTTGGATTGTGCTTCCTG
C260A.F	AATATTGTCCCAGATAACCGAGCGCAGCCGGGTGATGATCCT GG
C260A.R	TCGGTTATCTGGGACAATATTCAGAAATTCAATATCC

The template for PCR to make constructs pDK-M382A, pDK-C413A, pDK-C414A, and pDK-C260A was pZW-ADC5 with primer pair M382A/M382R, C413A.F/C413-14A.R, C414A.F/C413-14A.R, and C260A.F/C260A.R respectively. For the initial amplification for each construct, the total PCR reaction volume was 200 µL (50 µL x 4 PCR reactions). Each 50 µL PCR reaction consisted of 34.5 µL ddH_2O, 2.5 µL DMSO, 10 µL Q5 reaction buffer, 1 µL dNTP, 0.5 µL forward primer 100 pmol/µL, 0.5 µL reverse primer 100 pmol/µL, 0.5 µL template DNA and 0.5 µL Q5 DNA polymerase. The PCR program was set as the following:

step 1: 98 °C, 10 min; step 2: 98 °C, 10 sec; step 3: 65 °C, 20 sec; step 4: 72 °C, 7 min (30x cycle to step 2); step 5: 72 °C, 10 min; step 6: 4 °C, ∞. The product of PCR was purified through agarose gel purification and assembled using Gibson assembled (GA) method[17]. Thereafter, the GA mix was dialyzed for 20-30 min to remove salts and transferred to *E.coli* BAP1 by electroporation. Survived colonies under 50 μg/mL of kanamycin were tested by PCR with primer E6.4.check.F and E6.4.check.R. Positive PCR colonies were verified by sanger sequencing. Sequenced-positive colony was grown at 37 °C overnight LB containing 50 μg/mL of kanamycin and stored at -80 °C with a final concentration of 25% glycerol.

The initial PCR step to make constructs pDK-M382C::C413A and pDK-M382C::C414A used primer pair C413A.F/C413-14A.R and C414A.F/C413-14A.R respectively. The protocol for further steps are the same as explained in the previous paragraph but with a template of pDK-M382C. Construct pDK-M382C was made using the same protocol as explained in the previous paragraph with primer pair M382C.F/M382.R.

Extract generation, LCMS and Networking

5 μL of a heterologous producer carrying a particular construct from cryo culture was inoculated to 1 mL of LB containing 50 μg/mL of kanamycin and 25 μg/mL of chloramphenicol in a 2 mL sterile tube and then incubated at shaker 37 °C overnight. Thus, 0.2 mL of overnight grown pre-culture was inoculated to 20 mL LB containing 50 μg/mL kanamycin and 25 μg/mL of chloramphenicol in a 50 mL falcon tube. The culture was incubated on laying position at shaker 37 °C until it reached OD_{600} of 0.4-0.6. Further, the culture was induced with a final concentration of 0.5 mM IPTG and continued incubation at shaker 30 °C for 3 days. Thereafter, the culture was centrifuged at max speed, for 3 minutes and the supernatant (medium) was discarded. 400 μL of 50:50 H_2O:MeCN was added to dissolve the pellet/cells and moved to a new 2 mL tube. The cell was lysed by bath ultrasonication for 1 hour. The solution was then centrifuged at max speed for 4 minutes and then 50 μL of the clear solution was subjected to LCMS for analysis.

UPLC-HRMS measurements were performed on an Agilent Infinity 1290 UPLC system equipped with an Acquity UPLC BEH C18 1.7 μm (2.1 × 100 mm) column and an Acquity UPLC BEH C18 1.7 μm VanGuard Pre-Column (2.1 × 5 mm; both columns were purchased

from Waters) setup coupled to a DAD detector and a micrOTOFQ II mass spectrometer (Bruker). The LC part was run using a gradient (A: H2O, 0.1% FA; B: MeCN, 0.1% FA; Flow: 600 µL min−1): 0 min: 95%A; 0.80 min: 95%A; 18.70 min: 4.75%A; 18.80 min: 0%A; 23.00 min: 0%A; 23.10 min: 95%A; 25.00 min: 95%A and the column oven temperature was set to 45 °C. Calibration of mass spectra was achieved using sodium formate in H_2O/ iPrOH (1:1) as the internal standard. Analysis of mass spectra was accomplished using the software Data Analysis 4.2 (Bruker Daltonics).

Identification of ions which is present in the expression extract but absent in the control was done using molecular networking GNPS (https://gnps.ucsd.edu/)[15]. LCMS data of the transgenic cultivation extract and the control extract were submitted and the networking result is shown in Figure S1. Hit ions were further investigated by extracting the BPC LCMS chromatogram of the expression extract and control.

References

1. Tacconelly, E. *et al*. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The lancet Infectious Deseases* **18**, 318-327 (2018). doi: 10.1016/S1473-3099(17)30753-3
2. Payne, D. J. *et al*. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nature Reviewers Drug Discovery* **6**, 29-40 (2007).
3. Lewis, K. Platforms for antibiotic discovery. *Nature Reviewers Drug Discovery* **12**, 371-387 (2013).
4. Imai, Y., *et al*. A new antibiotic selectively kills gram-negative pathogens. *Nature* **576**, 459-464 (2019).
5. Groß, S., *et al*. Improved broad-spectrum antibiotics against gram-negative pathogens via darobactin biosynthetic pathway engineering. *Chemical Science* **12**, 11882-11893 (2021).
6. Böhringer, N., *et al*. Mutasynthetic production and antimicrobial characterization of darobactin analogs. *Microbiology Spectrum* **9**, 1-14 (2021).
7. Wuisan, Z. G., *et al*. Optimization of heterologous darobactin A expression and identification of the minimal biosynthetic gene cluster. *Metabolic Engineering* **66**, 123-136 (2021).
8. Atanasov, A. G. *et al*. Natural products in drugs discovery: advances and opportunities. *Nature Reviews Drug Discovery*, **20**, 200-216 (2021).
9. Albarano, L. *et al*. Genome mining as new challenge in natural products discovery. *Marine Drugs* **18**, 1-17 (2020).
10. Altschul, D. F. *et al*. Basic local alignment search tool. *J. Mol. Biol*. **215**, 403-410 (1990).

11. Madeira, F. *et al*. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* **47**, W636–W641 (2019). doi: 10.1093/nar/gkz268
12. Davis, K. M. *et al*. Structures of the peptide-modifying radical SAM enzyme SuiB elucidate the basis of substrate recognition. *PNAS* **114**, 10420-10425 (2017). doi: 10.1073/pnas.1703663114
13. Broderick, J. B. *et al*. Radical *S*-Adenosylmethionine Enzymes. *Chemical Reviews* **114**, 4229-4317 (2014).
14. Schramma, K. R. *et al*. Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink. *Nature Chemistry* **7**, 431-437 (2015).
15. Wang, G. *et al*. Comparative Genomics Reveal the Animal-Associated Features of the *Acanthopleuribacteraceae* Bacteria, and Description of *Sulfidibacter corallicola* gen. nov., sp., nov.
16. Wang, M. *et al*. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology* **34**, 828-837 (2016).
17. Gibson, D. G. *et al*. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343-345 (2009).

8. Chapter VII

Genome mining-guided discovery and characterization of a RiPP BGC originated from *Sulfidibacter corallicola* M133^T

I Dewa M. Kresna[a,b], Till F. Schäberle[a,b,c]

[a] Institute for Insect Biotechnology, Justus-Liebig-University of Giessen, Ohlebergsweg 12, 35392 Giessen, Germany.

[b] Department of Bioresources, Fraunhofer Institute for Molecular Biology and Applied Ecology, Ohlebergsweg 12, 35392 Giessen, Germany.

[c] German Center for Infection Research (DZIF), Partner Site Giessen-Marburg-Langen, Giessen, Germany.

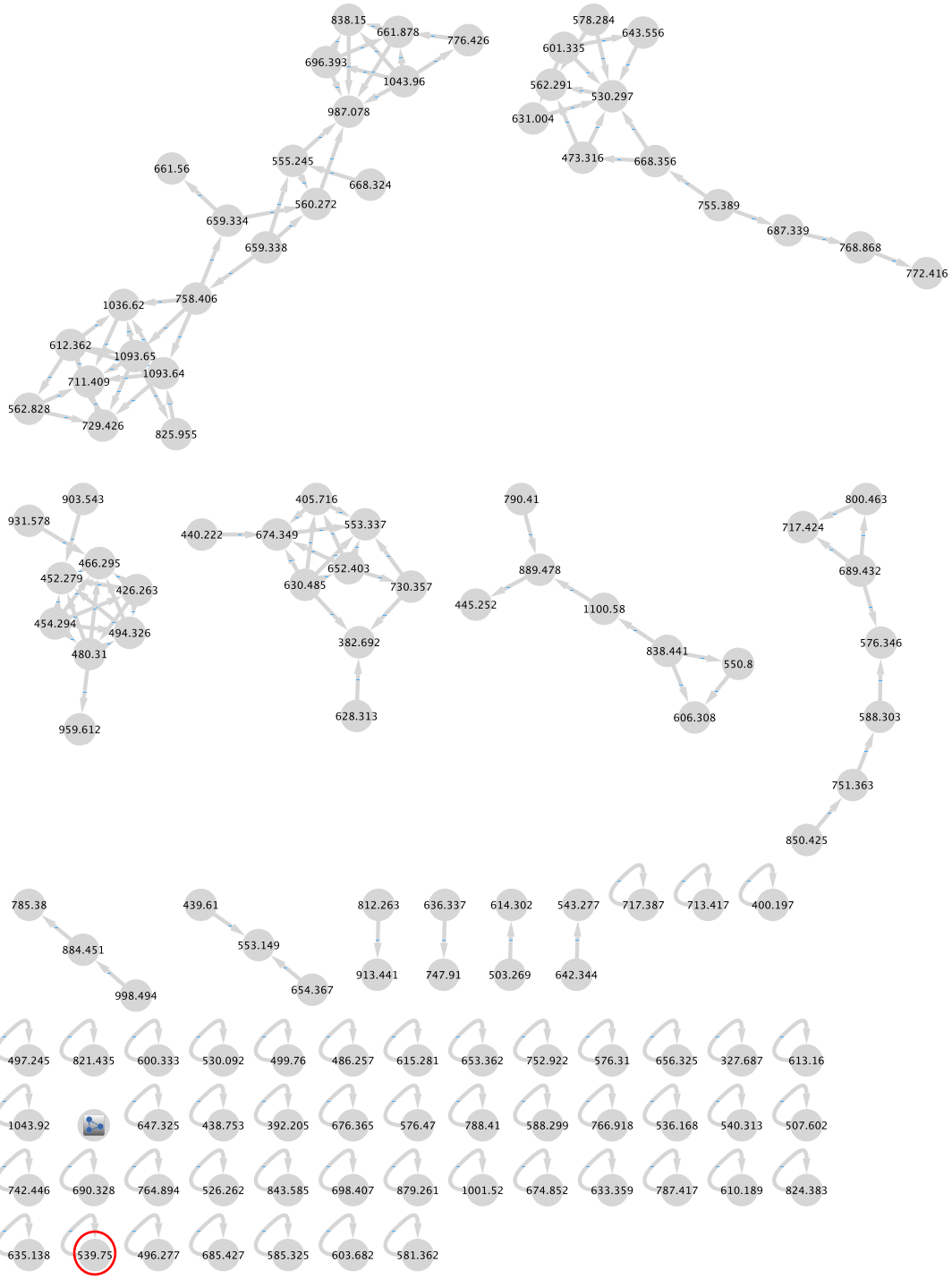


Figure S1. The MS/MS networking result of extracts from the cultivation of transgenic strain expressing *mcd* BGC and control expressing empty pRSFduet expression vector. Red circle represent the parent $[M+2H]^{2+}$ mass ion of mc-darobactin (539.75 *m/z*). This ion only appeared in the transgenic cultivation extract and was absent in the control extract.

atgaattcttccatgcccctcgccggcctggattcacgtaccggtgtgggtacgtttgggatcccgccggaactggccctcgatttcgaa
gccttcaaggcgacgctcagtcctgaagaacaacgggcgctcactgtcctggcggctggtatgaagcaccagtcggccgaggaatt
catcgggaacatccccgaagaagagatcctcaacgcccatgaacggcgggtattcgccgcggaaatggcgcgcgaggtacccgaa
aagagatcgcccagcgcaaaattcgtcatggttctcaaagcgacccgtttgtgtaatttgcgctgtacctactgccgttcctgggcggaa
ggacacaatcaggtcatgtccttccccgttctggttcgcgccattcgcgaggttctctcgatgcgcggccttcagcatgtggaaatcgtct
ggcacggtggcgagatcactttgctgaaaccgaagttcttcaaaaagttgatttgggtgcaacagcagttccggcgccccggccaggt
cattcgcatctccctgcaaacgaacgccacccatttgaccgacgagtggatcgaattcctgtccaccctcaatatcggcgtcgggatca
gtatcgatggaccgccggaggtgcacgaccggcgtcgggtggacaaggacggcaagcccagttctcccttggtggtcgagggcatt
cgccgtttgcggacggccggcattccccacggtgcactggtcgtggtggaccgcttcctccaggacatcggggccgagcgcctgttg
acatactttcacgaaatcggcctgaagaatctcgatttcctcaacatcgtgcccgaaaaccacaagggcgcaacggcttcgagagagg
acggcttcatctgcttcgacgaatacgtcgacttcctccgcgaggcgttcgacgtgtggtggcgcaagtatcgcgacgagatgttcatc
acgacgttcgccgacttgatcccggcgctcagcgacggagccaagccgaacaattgctattgggcgggcaactgcatggtctcgacc
ctgaccatcgaagccaacggcgacatggcgccctgcgacaagttcatcggcgacgagggctcgatcttcggcaacgtcatgcgtca
atcgattgccgacgtcatgcgccagacggcctacatcaagaacgccaatcgcgaacgggcctcgagcctgtcgcgcatgaagacat
gtgacttttccaaatttgccagggcggttgtccccacgaccgcctcgtcaaccggcgccacgtgcccgaccacgaccccgattgttgt
ggcctgaagccgctggttcgacacattcgcgacgccgccatggcctagcgaatccagccccgtttaccctcgcccggatatcgatgc
ggcggagagggaccctgcgacggatctcgtcgccgcgttccgtcgtgtcttcgatcggggcgggggcatttcgaccccactcaaag
aggaggattcccatgaatctgcaaacctacgccactccatctgccgacgtcgtcaatgtagccgacaagctcgccggtttcatgaacg
ccaacggcgtggtctccggaaacgtcaccaagcgcgcccatcaaatggaagccctggccgccgacgacggcagcctgaccctcga
tgccgacgcgttggcacgtatccacgaccgcaacgacaaccacggcaagtggcgttggagctggcctttctaa

Sequence S1. The nucleotide sequence of the synthetically generated and purchased ddDNA fragment containing *mcd* BGC from *Sulfidibacter corallicola* M133^{T}. The red and green colored letter represent the gene of *mcdA* (acc. number: QTD49608.1) and *mcdB* (acc. number: QTD49609.1).

9. Discussion

The nature of evolution driving the development of antimicrobial-resistance is inevitable. The increasing number of antibiotic resistance cases has been a concern since 1960 until present.[1, 2] Natural products are still an important source for antibacterial agents discovery. From 1981 to the third quarter of 2009, ~55% of the approved antibacterial agents are natural products or their derivatives.[3] Many approaches have been developed to discover new antibacterial agents such as culture-based, genome mining and synthetic. The latter approach is based on structure-activity relationship. The synthetic approach is nowadays in trend for example employing a supercomputer for artificial intelligence (AI) and *in silico* simulation[4]. More details about the synthetic approach are not discussed here. The culture-based or classical screening approach is the oldest approach. The principal is based on direct measurement of the ability of a strain to produce antimicrobial compounds in laboratory conditions.[5] Several screening methods have been developed such as the cross-streak method,[6] the spot-on-the-lawn,[7] the well diffusion,[8, 9] disc diffusion method[10] and overlay assay[11]. These are agar-plate-based-experiments, relatively easy and powerful. In general, the tested strain is grown either in liquid medium or on agar plates until it reaches exponential growth phase, in which often the production of secondary metabolites is increased. Thereafter, many variants of the basic methods could be applied: (1) cross-streak the pathogen (indicator) strain to the grown tested strain on the agar plate, (2) drop a grown tested strain on the lawn of an indicator strain, (3) pour whether the grown tested strain or its extract from liquid culture or a round-chopped grown tested strain from agar to a previously chopped well containing a freshly spread lawn of an indicator strain, or (4) drop an extract of tested strain culture into a small paper-based disk. After incubation, the growth behavior i.e. inhibited growth of the indicator strain is observed. Note that these approaches require the growth and the ability to produce the antimicrobial compounds of the tested strain under laboratory conditions. Thereby, after decades of using these screening methods, several problems have been identified: (1) Only approximately 1% of microorganisms are readily cultivable *in vitro* with the majority of all bacteria remaining unculturable using standard method,[12, 13] (2) rediscovery of known compounds[5, 14] and (3) inability to identify compounds due to limited yield as a consequence of silent or BGC expressing only at basal level[15].

Several problems to access unculturable bacteria have been identified i.e. the lack of natural growth factors and slow growing-bacteria outcompete the fast-growing bacteria. Approaches

to cultivate previously unculturable bacteria have been developed e.g. (a) modification of growth media and conditions, (b) high throughput extinction, (c) community culture, (d) coculture, (e) transwell plates with membrane, (f) micromanipulator, (g) optical tweezers, (h) laser microdissection, (i) iChip (diffusion chamber), (j) single cell encapsulation combined with flow cytometry, etc.[16] Point (a) and (b) tries to mimic the natural conditions such as adjustment of media composition including the use of sterilized original water sample,[17 ,18] addition of additives, medium dilution, incubation time, CO_2/O_2 level, temperature, pH, etc.[19, 20, 21] Point (c) and (d) try to cultivate bacteria *in vitro* as a community (together) so that they can share the growth factors.[22, 23] Point (e) employs a plate that has an agar layer containing bacterial community on a membrane filter in the bottom that growth factor and others can diffuse but not bacteria. This plate is placed on the soil slurry in a container allowing natural growth factors to diffuse to the agar layer on the membrane.[24] Point (f), (g), and (h) involve tools e.g. microneedle and laser beam which can selectively pick up, trap, or remove desired or undesired single cell/bacteria from a community so that slow-growing bacteria can be grown separately without having to outcompete to other fast-growing ones.[25, 26] Point (i) and (j) use encapsulation e.g. by membrane semipermeable or agar-based material of single cell bacteria in a community which growth factors can diffuse but not the bacterial cells. These encapsulated bacteria could be cultured whether in an artificially mimicking natural environment (*in vitro*) or on the site of natural environment (*in situ*).[27, 28]

As well for approaches to overcome the second problem have been developed such as antibiotic resistant platform (ARP)[29] and LCMS-based characterization of crude extract or microbial surface profiling[30]. The latter includes "direct detection and selection", "UHPLC/MS profiling and selection", "micro fractionation for active peak identification" and quantification and structure elucidation from small amount compounds"[30]. These approaches use tools, databases and software allowing easier early identification of possible known compounds directly from crude extracts. The last problem, silent BGC, led to the development of genome mining.

Nowadays, the price and time needed to sequence and assemble genomic nucleotides of a bacterial strain are getting cheaper and much faster. Up to April 2020, the number of publicly available genomes are around 400,000![31] 82% of bacterial genomes in RefSeq were produced by the short read of Illumina sequencing as a leading technology followed by PacBio, Roche-454, Ion torrent/proton, Oxford Nanopore and Sanger respectively.[31] Furthermore, most of the

sequencing data are submitted to GenBank (http://www.ncbi.nlm.nih.gov). GenBank is built primarily from the submission of sequence data from authors and from the bulk submission of expressed sequence tag (EST), genome survey sequence (GSS), whole-genome shotgun (WGS) and other high-throughput data from sequencing centres.[32]

9.1. Genome mining

Because of the significant development in sequencing technology and centralized publicly available genome databases, the bioinformatic tools to annotate genes and biosynthetic gene clusters (BGCs) encoding proteins that could produce bioactive compounds are developed in parallel. Many free online bioinformatic tools predicting BGCs are publicly available, such as antiSMASH,[33] BAGEL4,[34] and PRISM[35]. Genome mining approaches could be divided into three methods, which are "ecology-based mining", "functionality-based mining" and "mode-of-action-based mining".[36] One significant benefit of genome mining is the possibility to access potential BGCs independently, regardless of whether the producer strain is culturable or not.

Ecology-based mining

From soil to ocean, from plant roots to animal guts, the ecosystem in which natural products are found is highly diverse.[36] Estimation about the entire bacterial diversity of the sea may be unlikely to exceed 2 x 10^6, while a ton of soil could contain 4 x 10^6 different taxa.[37] A better understanding of microbial community and ecology can be used to chart the metabolic potential and variation and prioritize BGCs that are likely to encode the synthesis of potent antimicrobial molecules.[36] However, due to the fact that only a low number of bacteria can be cultured under laboratory conditions, the metagenomic sequencing approach represents a potential solution. A mix of microorganisms from a sample community can be sequenced and BGCs within the metagenome sequence could be annotated. Bioinformatic tools and literature research are performed to ensure the completeness of the annotated BGCs. Thereafter, potential BGCs are screened to be further investigated.

Functionality-based mining

This approach is primarily based on the predicted function of key protein domains in the biosynthesis.[36] The gene cluster identification algorithms in tools such as antiSMASH,

BAGEL3 and PRISM are based on this principle.[36] In the present study, using this approach, a new Gram-positive antibiotic Polyoxyperuin could be identified (Chapter II). The producer *Streptomyces* sp. s120 isolated from a marine environment (i.e., Peruvian coast) was prioritized due to antimicrobial screening in our lab. Since Streptomycetes are widely known to produce antibiotics, dereplication strategies such as LCMS analysis of the cultivation extract and BGC identification from the genomic sequence were employed. Indeed, by dereplication of its MS/MS fragmentation data, the known compounds naphthyridinomycin and resistomycin were identified.[38] Though, one PKS-NRPS-Hybrid BGC, which has similarity to the known Aurantimycin and Polyoxyperuin BGCs was identified by antiSMASH. Literature and manual investigation suggested that the product of the newly founded BGC belongs to the azinothricin family. Many of these products possess antimicrobial activity. Further LCMS profiling of the cultivation extract revealed the absence of any ions belonging to the known azinothricin family compounds suggesting a new compound. This early stage of information concluded the worthiness of this project to be further investigated yielding successfully the characterization of polyoxyperuin even though the BGC of polyoxyperuins was considered silent. The corresponding products were expressed by the wild type producer strain in many conditions only in very low amounts, if at all. Other examples of using this approach were reported for the systematical genome mining of biosynthetic pathways encoding cyanobactins[39], thiazole/oxazole-modified microcins[40] and enediynes[41]. The goal is to explore the structural diversity of compounds sharing the same classification with a hope to discover new derivatives which possess better features e.g. bioactivity.

Another bioinformatics tool that can be used is BLAST[42], which is integrated into the NCBI database. BLAST is a tool to search or align a query sequence to the whole NCBI database with some statistical scores. Recently, a new potent RiPP antibiotic darobactin that selectively kills gram-negative pathogens was reported.[43] The minimum gene cluster only consists of a precursor peptide (pre-peptide) DarA and a post-translationally modifying radical SAM enzyme (RaS) DarE.[44] Darobactin attacks the new target BamA, an outer membrane chaperon protein.[45] Due to its potential therapeutic features in the *in vivo* mice model,[43] it is of interest to find similar compounds like darobactin. DarE-like proteins were searched by BLASTp using a conserved motif (signature) as the query. The hits led to the discovery of a BGC producing a mono cyclic darobactin. Therefore, the BGC was named *mcd* for mono cyclic darobactin (Chapter VII).

Overall, this result shows that the strain producing polyoxyperuin might have been deprioritized to be further investigated for new bioactive compound discovery if only the classical screening and dereplication approach had been used. In retrospect, functionality-based mining is a great alternative screening approach, significantly contributing to the decision making process. Similarly, the *mcd* BGC was obtained from a metagenomic sequence from a marine sample and the characterization was done without even having the producer strain in hand.

Mode-of-action-based mining

To gather more data of BGCs putatively encoding the production of antimicrobial natural products, two approaches could be employed: resistance-based genome mining and mining for synergistic antibiotics. The resistant-based genome mining employs the identification of a gene responsible for the resistant mechanism, such as transporters, drug-modifying enzymes and paralogous genes encoding 'resistant' copies of housekeeping proteins.[36] Meanwhile, the synergistic antibiotics mining implies the identification of BGCs encoding the production of antimicrobial products that show a synergistic effect supporting the main compound to deactivate/kill the target.[36] A great example is the synergistic effect of cephamycin (β-lactam antibiotic) and clavulanic acid (β-lactamase inhibitor), which are naturally produced by *Streptomyces* clavuligerus.[46, 47] Other examples are the antibiotics lankamycin and lankacidin, which bind complementary sites on the large ribosomal subunit,[48, 49] and the antibiotics griseoviridin and viridogrisein[50].

9.2. Characterization of biosynthetic pathway and engineering synthetic biology

Several factors contribute to the development of the genome mining approach. Prediction tools for genome mining heavily rely on characterized genes (e.g., knowledge about the functionality of the corresponding enzymes). New results can be used to update the algorithm and thus improve the prediction accuracy. Moreover, a better understanding of the biosynthetic pathway opens up the possibility to manipulate the "assembly line" of enzymatic functions involved in the biosynthesis of a given compound to improve its efficiency to obtain higher production yields. For instance, production of a compound might be achieved easier by heterologous expression. Thereby, synthetic promoter can be introduced, co-factors can be added into the

culture medium, etc. (*in vivo*). Another approach would be to express the enzymes involved in the biosynthesis, to purify and assay them *in vitro*. Furthermore, a rational-based modification of the assembly line becomes possible to produce new compounds with improved features for the desired application.

In the present study, the biosynthesis of ambigol (Chapter I)[51] and darobactin (Chapter IV and V) were characterized. The result of characterized ambigol biosynthesis *in vitro* extends our understanding of how polychlorinated aromatic compound is biosynthesized by a terrestrial microorganism. It is noteworthy that to selectively attach a halogen atom to the aromatic substrate, a tiny NRPS-system is applied. Thereby, the substrate is selectively chosen applying this biosynthesis assembly line and efficiently transformed into the product. It can be assumed that due to protein-protein-interactions the enzymes are specifically recruited and the generation of possible side products is decreased. Furthermore, the fact that the characterized halogenase can incorporate bromine to the substrate opens the possibility to produce *in vitro* and *in vivo* polybrominated, or a combination of chlorinated and brominated aromatic compounds.

Investigation of the darobactin biosynthesis has enabled a better understanding how this pathway can be used for heterologous expression (Chapter IV and V). From the initial identification of the darobactin BGC it was revealed that it consists of five genes.[43] However, only two of them – the substrate DarA (pre-peptide) and the modifying (RaS) enzyme DarE – are responsible to mature the product.[44] Although RiPP BGCs are usually small compared to other systems like PKS or NRPS, it is still surprising that such a potent product could be biosynthesized just by expressing two genes. *In vivo* characterization of DarE in the present study suggests that the unusual ether bond formation is catalyzed by DarE and both rings formation could be catalyzed independently of one another. The unusual ether bond formation has not been described so far and ongoing research with collaboration partners suggests the involvement of oxygen from O_2 and not from H_2O (manuscript is in submission), which is surprising since RaS enzymes normally only work in an anaerobic environment due to the ability of O_2 to degrade the [4Fe-4S] cluster(s) in RaS enzymes[52].

Regarding the last step in darobactin maturation, it is still unclear what mechanism and which enzyme is responsible for releasing the core peptide from the pre-peptide after DarE

modification. It has been shown in the present study that by expressing DarA alone, cleavage still takes place, indicating the involvement of (a) yet unidentified protease(s), which is present in the heterologous host and in the original producer (even though not part of the proposed BGC). Furthermore, a self-cleavage mechanism of DarA might be also possible. An indication towards this mechanism is that isolated DarA alone results *in vitro* in the DarA degradation (unpublished data).

In general, the understanding of the biosynthetic pathway leads to a possibility to exploit the biological machinery to produce derivatives with the goal to increase the desired features of the original compounds. Furthermore, improving the production yield of already known molecules is another beneficial strategy.[5] Since the initial report, the darobactin BGC has been modified to produce natural and unnatural derivatives.[53, 54] These derivatives possess slightly stronger antibiotic activity compared to the original darobactin A. In the present study, a new unnatural derivative (i.e. darobactin B9) was characterized. It is active against *P. aeruginosa* and *A. baumanii* strains, showing a bit broader spectrum in comparison to darobactin A (Chapter VI). The flexibility of DarE to modify various DarA core peptides was also systematically investigated. Out of the 133 generated constructs library, 68 could be maturated to the end product and only 36 out of them could maintain the halo zone in the antimicrobial overlay assay. Further investigation of the generated libraries is needed to discover the best combination of AA in the core peptide that gives the desired features. A modern, considerable faster, high throughput approach, like using FACS[55] and M*e*[x56] are envisaged but are still under optimization. Once they are optimized for the constructs and the screening strain, the library could be screened and the potentially new derivatives with better properties can be identified. A simulation-based *in silico* approach could also be an alternative to screen derivative candidates, since a crystal structure of the target in complex with the compound exists.

9.3. Homologous and heterologous expression

The goal of expressing BGCs or genes is primarily the identification and characterization of the corresponding products. As discussed earlier, the main challenges to access the metabolite(s) corresponding to a given BGC are the inability to culture a microbial strain under laboratory conditions and the metabolic regulation that can cause low to silent gene expression. To solve the first problem, many approaches for culturing previously unculturable bacteria have been developed and are reviewed by Pham, V. H. T. *et al*[16]. For the second problem, to

circumvent the natural regulation, one approach is to introduce a synthetic promoter (Chapter II and III). Nowadays, many DNA manipulation techniques such as lambda red,[57] *sacB*-based vector,[58] homologous recombination,[59] CRISPR/Cas systems[60] and more are available for this purpose. However, even with all of these developments, culturing previously unculturable bacteria and DNA modifications of many wild-type strains are still difficult tasks. When it comes to genes or BGCs identified in metagenome sequences, the original strain is mostly not in hand. Hence, heterologous expression is the way forward and it can be tried to use the species that is closest in regard to phylogeny to the original strain and genetically accessible to be employed as a host. Foreign DNA is introduced to the heterologous host and can be either located on a plasmid or integrated into the chromosome. The isolated metagenomic DNA or synthetically generated DNA strands can be used as template for PCR amplifications; thereby, allowing direct access to genes or BGCs without the need for any culturing step of the original strain. However, even though many putative expression strains are well-studied and the approach is performed for decades, other challenges were identified, which do not exist in the homologous expression system. These are: (1) rare codon usage and codon bias, (2) absence of important helper proteins, (3) missing precursors and (4) the toxicity of the products[61]. The scientific community is working on possibilities to overcome these obstacles, e.g. co-expressing the rare tRNAs, helper proteins, biosynthetic pathways for precursors and resistance factors have been implemented and are available.[61] To support homologous and heterologous expression approaches, many DNA cloning and assembling techniques were developed, such as restriction enzyme-based assembly (BioBricks, BglBricks, Pairwazs selection and Golden Gate), overlap-based assembly (InFusion, Isothermal Assembly, SLIC and USER) and *in vivo* recombination (*Bacillus* Domino and Yeast TAR).[62]

Homologous and heterologous expression have been crucial approaches in the present study. The enzymes and/or complete pathways responsible for ambigol and darobactin biosynthesis were constituted by heterologous expression (Chapter I and V). Helper proteins like chaperones and phosphopantetheinyl transferases were co-expressed. Furthermore, using a more controllable environment, heterologous expression of the darobactin BGC increased its production yield up to 10-fold and decreased production time up to 5-fold (Chapter IV). Using the implementation of the heterologous expression system, generation of darobactin derivatives and mc-darobactin became possible (Chapter VI and VII). Lastly, for the production of polyoxyperuin, a homologous expression approach was used (Chapter II). It was considered

more beneficial compared to heterologous expression, due to many reasons. The polyoxyperuin biosynthesis is encoded on a long BGC (~72 kbp), the borders of the BGC were unknown and genes within the BGC show different orientations.

9.4. Future research direction

Due to the development of pathogens becoming resistance to antibiotics and the need for new drugs (antibiotic crisis), discovery and development of new antibiotics are important. To keep pace with resistance development, the technical advances must be used. It can be expected that each step within the discovery and development platform needs innovation and further improvement. However, library generation, high throughput screening and genome mining tools have been the main challenge in the present study. The implementation of creating and screening a library with less effort, time and cost revealed some issues. It was observed that the diversity quality of the generated darobactin library needs to be improved. Randomization of the core peptide by applying degenerated primers results in a significant bias during PCR amplification. Thereby, the sequences are not normally distributed. Furthermore, miss priming was observed from the quality check (which was done by illumina sequencing) of the generated library. Golden gate assembly approaches might be an alternative to get rid of the bias which is introduced by PCR amplification. However, a drawback might be that the number of generated constructs could decrease and meticulous fine tuning to improve the outcome of this approach will be necessary. Furthermore, the transfer of the library by electroporation is still a laborious work- and time-consuming step. Hence, one alternative that should be tested is using MaxPlax lambda phage packaging to transfect the target bacteria[63, 64]. This approach for the DNA transfer is expected to be easier, less laborious, and more efficient.

The next challenge that has to be prioritized is the high throughput library screening. Once the library generated is of good quality, a library screening method must be implemented that can deal with the high number of clones and achieve the desired quality output. As discussed earlier, two approaches – FACS[55] and M*e*[x56] are still under development and/or further adaptation to the system applied here. Another approach that could be considered is *in silico* modeling of the derivatives with the target BamA. A co-crystal structure exists[45] and can be employed to identify *in silico* the derivatives with the best binding affinity to the target.

Nowadays, the genome mining approach is in trend to extract interesting previously inaccessible BGCs from sequence databases. The natural product darobactin is considered to be in its early development stage and further related molecules should be identified and characterized. In the present study, a BLASTp search using a conserved amino acid sequence motif of DarE as the query was employed. Indeed, many DarE-like proteins could be identified. However, the subsequent analysis to detect if the pre-peptide is present adjacent to the DarE-like encoding gene was done manually. It turned out that thousands of hits need to be manually analyzed. Due to this huge number, only around 15% have been analyzed in the present study so far. Hence, bioinformatics tools must be adapted for the search of this new compound class. Then, the available huge sequence data information could be analyzed in an efficient and comprehensive manner.

9.5. References

1. Podolsky, S. H. *Pelgrave Communications* **2018**, *14*, 1-8.
2. Davies, J., Davies, D. *Microbiology and Molecular Biology Reviews* **2010**, *74*, 417-433.
3. Newman, D. J., Cragg, G. M. *J. Nat. Prod.* **2020**, *83*, 770-803.
4. Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cobillos-Ruiz, A., Donghia, N. M., MacNair, C. R., French, S., Carfrae, L. A., Bloom-Ackermann, Z., Tran, V. M., Chiappino-Pepe, A., Badran, A. H., Andrews, I. W., Chory, E. J., Chruch, G. M., Brown, E. D., Jaakkola, T. S., Barzilay, R., Collins, J. J. *Cell* **2020**, *180*, 688-702.
5. Durand, G. A., Raoult, D., Dubourg, G. *International Journal of Antimicrobial Agents* **2019**, *53*, 371-382.
6. Williston, E. H., Zia-Walrath, P., Youmans, G. P. *Journal of Bacteriology* **1947**, *54*, 563-568.
7. Oliveira, L. de C., Silveira, A. M. M., Monteiro, A. de S., Santos, V. L. dos, Nicoli, J. R., Azevedo, V. A. de C., Soares, S. de C., Dias-Souza, M. V., Nardi, R. M. D. *Frontiers in Microbiology* **2017**, *8*, 876.
8. Dubourg, G., Elsawi, Z., Raoult, D. *International Journal of Antimicrobial Agents* **2015**, *46*, 590-593.
9. Adnan, M., Patel, M., Hadi, S. *Peer J* **2017**, *5*, e3085.
10. Brown, D. F. G., Kothari, D. *J. clin. Path.* **1975**, *28*, 779-783.
11. Maricic, N., Dawid, S. *Journal of Visualized Experiments* **2014**, *91*, 1-4.
12. Vartoukian, S. R., Palmer, R. M., Wade, W. G. *FEMS Microbiol. Lett.* **2010**, *309*, 1-7.
13. Xiong, Z.-Q. *Clin. Microbiol.* **2016**, *5*, e132.
14. Silver, L. L. *Clin. Microbiol. Rev.* **2011**, *24*, 71-109.
15. Stasiak, M., Maćkiw, E., Kowalska, J., Kucharek, K., Postupolski, J. *Polish Journal of Microbiology* **2021**, *70*, 421-429.
16. Pham, V. H. T., Kim, J. *Trends in Biotechnology* **2012**, *30*, 475-484.

17. Rappé, M. S., Connon, S. A., Vergin, K. L., Giovannoni, S. J. *Nature* **2002**, *418*, 630-633.
18. Song, J., Oh, H.-M., Cho, J.-C. *FEMS Microbiology Letter* **2009**, *295*, 141-147.
19. Vartoukian, S. R., Palmer, R. M., Wade, W. G. *FEMS Microbiol. Lett.* **2010**, *309*, 1-7.
20. Davis, K. E. R., Joseph, S. J., Janssen, P. H. *Appl. Environ. Microbiol.* **2005**, *71*, 826-834.
21. Janssen, P.H. *New cultivation strategies for terrestrial microorganisms. In Accessing Uncultivated Microorganisms* (Zengler, K., ed.), **2008**, 173–192, ASM Press
22. Nichols, D., Lewis, K., Orjala, J., Mo, S., Ortenberg, R., O'Connor, P., Zhao, C., Vouros, P., Kaeberlein, T., Epstein, S. S. *Applied and Environmental Microbiology* **2008**, *74*, 4889-4897.
23. Brenner, K., You, L., Arnold, F. H. *Trends in Biotechnology* **2008**, *26*, 483-489.
24. Svenning, M. M., Wartiainen, I., Hestness, A. G., Binnerup, S. J. *FEMS Microbiology Ecology* **2003**, *44*, 347-354.
25. Fröhlich, J., König, H. *FEMS Microbiology Reviews* **2000**, *24*, 567-572.
26. Zhang, H., Liu, K.-K. *J. R. Soc. Interface* **2008**, *5*, 671-690.
27. Nichols, D., Cahoon, N., Trakhtenberg, E. M., Pham, L., Mehta, A., Belanger, A., Kanigan, T., Lewis, K., Epstein, S. S. *Appl. Environ. Microbiol.* **2010**, *76*, 2445-2450.
28. Zengler, K., Toledo, G., Rappé, M., Elkins, J., Mathur, E. J., Short, J. M., Keller, M. *PNAS* **2002**, *99*, 15681-15686.
29. Cox, G., Sieron, A., King, A. M., Pascale, G. D., Pawlowski, A. C., Koteva, K., Wright, G. D. *Cell Chemical Biologi* **2017**, *24*, 98-109.
30. Ito, T., Masubuchi, M. *The Journal of Antibiotics* **2014**, *67*, 353-360.
31. Segerman, B. *Frontiers in Cellular and Infection Microbiology* **2020**, *10*, 527102.
32. Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Sayers, E. W. *Nucleic Acid Research* **2013**, *41*, D36-D42.
33. Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., Wezel, G. P. v., Medema, M. H., Weber, T. *Nucleic Acids Research* **2021**, *49*, W29-W35.
34. Heel, A. J. v., Jong, A. d., Song, C., Viel, J. H., Kok, J., Kuipers, O. P. *Nucleic Acids Research* **2018**, *46*, W278-W281.
35. Skinnider, M. A., Johnston, C. W., Edgar, R. E., Dejong, C. A., Merwin, N. J., Rees, P. N., Magarvey, N. A., *PNAS* **2016**, *113*, E6343-E6351.
36. Tracanna, V., Jong, A. d., Medema, M. H., Kuipers, O. P. *FEMS Microbiology Reviews* **2017**, *41*, 417-429.
37. Curtis, T. P., Sloan, W. T., Scannell, J. W. *PNAS* **2002**, *99*, 10494-10499.
38. Linares-Otoya, L., Linares-Otoya, V., Armas-Mantilla, L., Blanco-Olano, C., Crüsemann, M., Ganoza-Yupanqui, M. L., Campos-Florian, J., König, G. M., Schäberle, T. F. *Marine Drugs* **2017**, *15*, 1-14.
39. Leikoski, N., Liu, L., Jokela, J., Wahlsten, M., Gugger, M., Calteau, A., Permi, P., Kerfeld, C. A., Sivonen, K., Fewer, D. P. *Chemistry & Biology* **2013**, *20*, 1033-1043.
40. Cox, C. L., Doroghazi, J. R., Mitchell, D. A. *BMC Genomics* **2015**, *16*, 1-16.
41. Shen, B., Hindra, Yan, X., Huang, T., Ge, H., Yang, D., Teng, Q., Rudolf, J. D., Lohman, J. R. *Bioorganic & Medicinal Chemistry Letters* **2015**, *25*, 9-15.

42. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. *Journal of Molecular Biology* **1990**, *215*, 403-410.
43. Yu, I., *et al*. *Nature* **2019**, *576*, 459-464.
44. Wuisan, Z. G., Kresna, I D. M., Böhringer, N., Lewis, K., Schäberle, T. F. *Metabolic Engineering* **2021**, *66*, 123-136.
45. Kaur, H., Jakob, R. P., Marzinek, J. K., Green, R., Imai, Y., Bolla, J. R., Agustoni, E., Robinson, C. V., Bond, P. J., Lewis, K., Maier, T., Hiller, S. *Nature* **2021**, *593*, 125-129.
46. Ward, J. M., Hodgson, J. E. *FEMS Microbiology Letters* **1993**, *110*, 239-242.
47. Medema, M. H., Trefzer, A., Kovalchuk, A., Berg, M. v. d., Müller, U., Heijne, W., Wu, L., Alam, M. T., Ronning, C. M., Nierman, W. C., Bovenberg, R. A. L., Breitling, R., Takano, E. *Genome Biology and Evolution* **2010**, *2*, 212–224.
48. Suwa, M., Sugino, H., Sasaoka, A., Mori, E., Fujii, S., Shinkawa, H., Nimi, O., Kinashi, H. *Gene* **2000**, *246*, 123-131.
49. Belousoff, M. J., Shapira, T., Bashan, A., Zimmerman, E., Rozenberg, H., Arakawa, K., Kinashi, H., Yonath, A. *PNAS* **2011**, *108*, 2717-2722.
50. Xie, Y., Wang, B., Liu, J., Zhou, J., Ma, J., Huang, H., Ju, J. *ChemBioChem* **2012**, *13*, 2745-2757.
51. Kresna, I D. M., Linares-Otoya, L., Milzarek, T., Duell, E. R., Mohseni, M. M., Mettal, U., König, G. M., Gulder, T. A. M., Schäberle, T. F. *Org. Biomol. Chem.* **2021**, *19*, 2302-2311.
52. Broderick, J. B., Duffus, B. R., Duschene, K. S., Shepard, E. M. *Chem. Rev.* **2014**, *114*, 4229-4317.
53. Böhringer, N., Green, R., Liu, Y., Mettal, U., Marner, M., Modaresi, S. M., Jakob, R. P., Wuisan, Z. G., Maier, T., Iinishi, A., Hiller, S., Lewis, K., Schäberle, T. F. *Microbiology Spectrum* **2021**, *9*, e01535-21.
54. Groß, S., Panter, F., Pogorevc, D., Seyfert, C. E., Deckarm, S., Bader, C. D., Herrmann, J., Müller, R. *Chem. Sci.* **2021**, *12*, 11882-11893.
55. Weng, L., Spoonamore, J. E. *Micromachines* **2019**, *10*, 734.
56. Koch, P., Schmitt, S., Cardner, M., Beerenwinkel, N., Panke, S., Held, M. *Research Square*. Under review (2021).
57. Mosberg, J. A., Lajoie, M. J., Church, G. M. *Genetics* **2010**, *186*, 791-799.
58. Marx, C. J. *BMC Research Notes* **2008**, *1*, 1.
59. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., Court, D. L. *Nature Protocols* **2009**, *4*, 206-223.
60. Cobb, R. E., Wang, Y., Zhao, H. *ACS Synth. Biol.* **2015**, *4*, 723-728.
61. Stevens, D. C., Hari, T. P. A., Boddy, C. N. *Nat. Prod. Rep.* **2013**, *30*, 1391-1411.
62. Ellis, T., Adie, T., Baldwin, G. S. *Integr. Biol.* **2011**, *3*, 109-118.
63. Gurgui, C., Piel, J. (2010) Metagenomic Approaches to Identify and Isolate Bioactive Natural Products from Microbiota of Marine Sponges. In: Streit W., Daniel R. (eds) Metagenomics. Methods in Molecular Biology (Methods and Protocols), vol 668. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-60761-823-2_17
64. EPICENTRE Biotechnologies. Copy-Control™ Fosmid Library Production Protocol (http://www.epibio.com/pdftechlit/ 171Pl079.pdf).

10. List of Scientific Contribution

10.1. Publications:

Kresna, I D. M., Wuisan, Z. G., Schäberle, T. F. Production of antimicrobial compounds by homologous and heterologous expression. Manuscript is under review.

Nguyen, H., Kresna, I D. M., Böhringer, N., Ruel, J., Mora, E. d. l., Kramer, J. C., Lewis, K., Nicolet, Y., Schäberle, T. F., Yokoyama K. (XXXX) Characterization of a Radical SAM Oxygenase for the Ether Crosslinking in Darobactin Biosynthesis. *Journal of the American Chemical Society* **XX**, XXXX-XXXX. Manuscript is under review.

Kresna, I D. M., Wuisan, Z. G., Joshi, S., Mettal, U., Patras, M., Brinkmann, S., Böhringer, N., Marner, M., Brinkrolf, K., Reiter, S., Schmitt, S., Schäberle, T. F. Systematic investigation of DarE flexibility generating a new darobactin derivative with broader gram-negative bioactivity. Manusript in preparation.

Kresna, I D. M., Wuisan, Z. G., Pohl, J. M., Mettal, U., Otoya, V. L., Gand, M., Marner, M., Otoya, L. L., Böhringer, N., Vilcinskas, A., Schäberle, T. F. (2022) Genome-Mining-Guided Discovery and Characterization of the PKS-NRPS-Hybrid Polyoxyperuin Produced by a Marine-Derived Streptomycete. *Journal of Natural Products* **85**, 888-898. doi: 10.1021/acs.jnatprod.1c01018

Wuisan, Z. G., Kresna, I D. M., Böhringer, N., Lewis, K. and Schäberle, T. F. (2021) Optimization of heterologous Darobactin A expression and identification of the minimal biosynthetic gene cluster. *Metabolic Engineering* **66**, 123-136. doi: 10.1016/j.ymben.2021.04.007

Kresna, I D. M., Linares-Otoya, L., Milzarek, T., Duell, E. R., Mohseni, M. M., Mettal, U., König, G. M., Gulder, T. A. M. and Schäberle, T. F. (2021) *In vitro* characterization of 3-chloro-4-hydroxybenzoic acid building block formation in ambigol biosynthesis. *Organic & Biomolecular Chemistry* **19**, 2302-2311. doi: 10.1039/D0OB02372H

10.2. Research project initiation:

Kresna, I D. M., Schäberle, T. F. Genome mining-guided discovery and characterization of a RiPP BGC originated from *Sulfidibacter corallicola* M133^{T}.

10.3. Selected contributions to conferences:

8-9.2018 VAAM-Workshop "The biology of bacteria producing natural products". Frankfurt, Germany.
Poster: Heterologous expression of biosynthetic gene cluster

9.2019 International VAAM Workshop 2019 "Biology of Microorganism Producing Natural Products". Jena, Germany.
Poster: Involvement of a radical SAM in intermolecular cyclisation reaction(s) in Darobactin maturation

9-10.2020 MegaSyn symposium on Megasynthases. Bad Nauheim, Germany.
Poster: *In vitro* analysis of the ambigol BGC – a Tiny Megasynthase

9.2021 The 4th European Conference on Natural Products. Online, Hessen, Germany.

Der Lebenslauf wurde aus der elektronischen Version der Arbeit entfernt.

The curriculum vitae was removed from the electronic version of the paper.