

DISS. ETH NO. 22430

***Unusual Halogenation and Chain Cleavage Reactions in  
Bacterial Polyketide and Peptide Biosynthesis***

A thesis submitted to attain the degree of

**DOCTOR OF SCIENCES of ETH ZURICH**

(Dr. sc. ETH Zurich)

presented by

*Katja Jensen*

*Diplom-Molekularbiomedizinerin, Rheinische Friedrich-  
Wilhelms-Universität Bonn, Germany*

born on 25.05.1986

citizen of Germany

accepted on the recommendation of

Prof. Dr. Jörn Piel (examiner)

Prof. Dr. Julia A. Vorholt (co-examiner)

Prof. Dr. Tobias A.M. Gulder (co-examiner)

2015

This thesis was conducted under supervision of Prof. Jörn Piel from  
2010-2013  
at the Kekulé Institute of Organic Chemistry and Biochemistry at the University of Bonn, Germany  
&  
2013-2014  
at the Institute of Microbiology at the ETH Zürich, Switzerland

*Dedicated to my family and Sebastian*

Et rheinisch Grundgesetz (ein Auszug)  
*Das Rheinische Grundgesetz = Constitution of the Rhineland (an extract)*  
§ 1- Et es wie et es  
*Es ist wie es ist = Things are as they are*  
§ 2- Et kütt wie et kütt  
*Es kommt wie es kommt = What will be will be*  
§ 5- Nix bliev wie et wor  
*Nichts bleibt wie es war = Be open-minded to the new and unknown*  
§ 8- Mach et jot äwer nit ze off  
*Mach es gut aber nicht zu oft = Rather do it well than too often*



# Abstract

---

Many compounds with high bioactivities are specialized (also called secondary) metabolites. They may serve directly as drug compounds or as lead structures towards the highly urgent development for antibiotics, anti-cancer or other drugs. The work presented in this thesis focused on two different biosynthetic machineries that are responsible for the production of two separate types of secondary metabolites. In both cases a basic understanding of biosynthetic peculiarities in chain cleavage and in one case of halogenation was obtained and probable applications as tools were considered.

We first studied the biosynthetic machinery of *trans*-AT polyketide synthases (PKSs), a young class of PKSs that is characterized by its separate acyltransferase (AT). It was explained why several copies of an AT gene are present in many *trans*-AT PKS biosynthetic gene clusters despite the fact that one copy would be sufficient. This was exemplified with the pederin (**5**) gene cluster (*ped*) which harbors two AT gene copies. Our previous studies suggested an involvement of the second AT copy, PedC, in a proofreading mechanism. The presumed hydrolysis of thioesters could be verified in an *in vitro* assay enabling real time tracing of the potential hydrolysis, where different thioester substrates were offered to PedC. PedC was capable of cleaving a wide variety of substrates that differed both in sizes and functionalities. The results are consistent with a highly likely involvement in biosynthetic proofreading. Consequently, the enzyme class was renamed from acyltransferase to acylhydrolase (AH). This is the first described mechanism of *trans*-AT PKS housekeeping and one of very few proofreading mechanism for modular PKSs. AHs might become employed in pathway elucidation and improvement of polyketide yield with a controlled release of intermediates from a pathway of interest.

We assigned genes responsible for the chlorination in the *trans*-AT PKS product oocydin A (**11**), a chlorinated macrolactone. Notably, two genes were involved in the halogenation reaction. Corresponding proteins surprisingly showed the affiliation to a new type of two-component halogenases related to mononuclear non-heme iron,  $\alpha$ -ketoglutarate- and O<sub>2</sub>-dependent halogenases. Firstly, **11** and dechloro-oocydin A (**42**) were isolated and structures were elucidated. A further, previously unknown product of the oocydin gene cluster named oocydin B (**43**) was isolated and its structure could be elucidated. **43** highly resembles haterumalide B (**35**), which was originally isolated from a marine ascidian. This strongly suggests a biosynthetic origin of haterumalides in the **11** gene cluster. A revised biosynthetic model was proposed which explains biosynthesis of **11** and the analog **43** with an extended side chain. Cytotoxicity tests revealed a minute influence of the chlorine atom on the bioactivity. Oocydin B (**43**) with its longer side chain than **11** exhibited lower cytotoxicity than **11** and **42**. An application of the two-component halogenases in chemoenzymatic synthesis is conceivable, since classical chemical synthesis of a vinyl chloride moiety is limited.

During the course of this study the PoyH recognition site was refined. PoyH is an unusual non membrane-bound protease involved in biosynthesis of polytheonamide (**14**), a ribosomally synthesized and posttranslationally modified peptide (RiPP). Single-point mutants of PoyH substrates were generated and applied in an *in vitro* assay to analyze chain cleavage by PoyH and to gain insight into substrate recognition. The refined recognition sequence was introduced to a fully unrelated protein and cleavage could be detected. The protease has thus become accessible as a potential new tool to study other RiPP biosynthetic pathways. It may even potentially find application in more general molecular biology inquiries.

# Zusammenfassung

---

Viele bioaktive Moleküle gehören zur Klasse der Sekundärmetaboliten. Sie können als Leitstrukturen oder als Medikament selber in der dringend nötigen Entwicklung weiterer Antibiotika oder Zytostatika genutzt werden. Die hier vorliegende Arbeit befasst sich mit zwei verschiedenen biosynthetischen Maschinerien, die für die Bildung von Sekundärmetaboliten verantwortlich sind. In beiden Fällen wurden wichtige Einsichten in die grundlegenden biosynthetischen Eigenheiten dieser Enzymkomplexe erhalten und mögliche Anwendungen als Werkzeuge ins Auge gefasst. Der Fokus lag im Besonderen auf verschiedenen Kettenspaltungsreaktionen von biosynthetischen Intermediaten und Halogenierungen.

Die erste untersuchte Biosynthese-Maschinerie gehört zur Klasse der *trans*-AT Polyketidsynthasen (PKS), die sich durch freistehende Acyltransferasen (AT) auszeichnen. Am Beispiel des Pederin (**5**) Genclusters *ped* wurde die Frage aufgeklärt, wieso viele vergleichbare Gencluster mehr als eine Kopie des AT-Gens tragen, obwohl gezeigt wurde, dass eine Kopie ausreichend ist. Es wurde vermutet, dass das kodierte Protein der zweiten AT-Kopie *pedC* an einem hydrolytischen Korrektur-Mechanismus beteiligt sein könnte. Die vermutete Hydrolyse wurde in einem *in vitro* Ansatz übergeprüft. Verschiedene potentielle Thioester-Substrate wurden PedC angeboten und eine eventuelle Hydrolyse in Echtzeit mitverfolgt. Tatsächlich war PedC im Stande Substrate unterschiedlicher Länge und mit verschiedenen Funktionalitäten zu hydrolysieren. Damit liegt eine Beteiligung an einem Korrektur-Mechanismus in der Polyketidbiosynthese nahe und daraufhin wurde die Enzymklasse von Acyltransferase zu Acylhydrolase (AH) umbenannt. Diese nachgewiesene Hydrolyse ist die erste gezeigte Variante für die Durchführung von Fehlerkorrekturen durch *trans*-AT PKS. AHs könnten in der Biosynthese-Aufklärung und Polyketid Titer-Erhöhung durch kontrollierte Ablösung der Intermediate eingesetzt werden.

Weiterhin gelang es, die für die Halogenierung an Oocydin A (**11**), einem halogenierten Makrolakton-Polyketid, verantwortlichen Gene zu bestimmen. Untypischerweise sind zwei Gene an der Halogenierung beteiligt. Die kodierten Proteine sind die ersten Vertreter einer neuen Zwei-Komponenten Halogenase, die Verwandtschaft zu Einzelkernigen nicht-Häm Eisen,  $\alpha$ -Ketoglutarat- und O<sub>2</sub>-abhängigen Halogenasen aufweist. **11** und erstmals Dechloro-Oocydin A (**42**) konnten isoliert und die Struktur mit Hilfe von NMR-Spektroskopie aufgeklärt werden. Ein weiteres unbekanntes Produkt der **11**-Biosynthese wurde isoliert und die Struktur mit Hilfe von NMR-spektroskopischen Untersuchungen aufgeklärt. Die Oocydin B (**43**) genannte Substanz weist eine sehr hohe strukturelle Ähnlichkeit zu Haterumalide B (**35**) auf, das ursprünglich aus einer Seescheide isoliert wurde. Durch diese strukturelle Analogie liegt eine enge biosynthetische Verwandtschaft zwischen **11** und **35** nahe. Ein modifiziertes Biosynthese-Modell wurde vorgeschlagen, das die Biosynthese von **11** und dem langkettigen Analog **43** erklärt. Zytotoxizitäts-Studien zeigten einen minimalen Einfluss des Chlor-Atoms auf die Toxizität im Vergleich zu **11**, wobei die längere Seitenkette von **43** die Zytotoxizität reduzierte im Vergleich zu **11** und **42**. Der Einsatz der Zwei-Komponenten Halogenase in chemoenzymatischen Fragestellungen ist denkbar, da klassisch chemisch-synthetische Ansätze bei Einführungen von Halogenen an Vinylgruppen an ihre Grenzen stoßen.

Im Rahmen dieser Arbeit konnte auch die Erkennungssequenz der Protease PoyH verfeinert werden. Dabei handelt es sich um eine unüblicherweise nicht Membran-gebundene Protease aus der Biosynthese der Polytheonamide (**14**), die zur Klasse der ribosomal synthetisierten und posttranslational modifizierten Peptide (RiPPs) gehören. Es wurden Einzelpunkt-Mutationen eines Substrats generiert und der Protease in einem *in vitro* Assay zur Spaltung angeboten, um Aussagen über die Substraterkennung zu erhalten. Die verfeinerte Erkennungssequenz wurde in ein nicht verwandtes Protein eingebaut und Spaltung des Substrats nachgewiesen. Diese Erkenntnisse machen die Protease zu einem potentiellen neuen Werkzeug in der RiPP-Biosynthese-Forschung oder auch weiteren molekularbiologischen Anwendungen.

# Acknowledgements

---

First of all I would like to thank Prof. Jörn Piel for the opportunity to conduct this interesting research in his lab. He always had an open door, not only giving support, but also giving me freedom in my work. I am glad that I had the opportunity to conduct this research at the University of Bonn but also to get to know Zürich as a researcher and inhabitant.

I thank Prof. Tobias Gulder for always being a contact person and a second (unofficial) supervisor throughout the time in Bonn. Despite our move to Zurich and his move to Munich, he was still willing to become a member of my PhD committee and examination committee.

Thanks to Prof. Julia Vorholt for becoming a member of my PhD and examination committee, although our group was new in the institute and her list of engagements in other committees long.

Thanks to everyone in the Kekulé Institute of Organic Chemistry and Biochemistry at the University of Bonn for the offered facilities and services (HPLC, MS), support and help. A second big thanks goes to the Institute of Microbiology and Immunology at the ETH Zürich. Thanks for the awesome infrastructure and the good times together.

A big thanks goes to Prof. Birgit Piechulla and Nancy Magnus from the University of Rostock for the *Serratia plymuthica* strain and also the generated knock out mutants.

I had several cooperation over the time, thanks to every single one of you, Marco Oldiges, Nicole Brendel and Marco Kai. Without your commitment my research would not have to be possible in this way.

I would like to thank Anna, Alex, Mike W., Brandon, Max, Rebekka and Sebastian for their proofreading. You really helped me writing the thesis in a good way!

Reiko and Brandon introduced me to the "chemical" world, thank you for that, I know it was not always easy.

Thanks to my students Tatjana, Christian and Michael. You all helped a lot at different stages of my research and had great ideas.

I would like to thank Nina Heycke and Cristian Gurgui for their great commitment in science and non-science issues back in Bonn.

All present and past members of the Piel-lab in Bonn and certainly of the Gulder-lab as well. Especially Max C., Christoph, Hülya, Ursel, Fritzi, Max H., Silke, Roy, Guy, Alex, Jana and Minna you all created a pleasant atmosphere with scientific help and support, as well as with much non-scientific support in the form of several beers, laughter, celebrations and so on.

Thanks to everyone from the Piel lab here in Zürich, especially G429. The time of the move was sometimes a bit chaotic and we started with many new people. Together we built up great labs, had the best office and a great group with an expert for every problem.

My final and especial thanks goes to my family, my "Mädels", friends and especially Sebastian. You were always there, supporting me without asking, lifting me up when I had fallen and gave me big hugs when I needed them in hard times. But we shared many good moments and laughter and partied all night long. I am thankful to have you in my life. Thanks!



# Contents

---

|   |             |
|---|-------------|
| <b>ABSTRACT</b> .....   | <b>II</b>   |
| <b>ZUSAMMENFASSUNG</b> .....  | <b>III</b>  |
| <b>ACKNOWLEDGEMENTS</b> .....   | <b>IV</b>   |
| <b>CONTENTS</b> .....   | <b>V</b>    |
| <b>LIST OF FIGURES</b> .....  | <b>VII</b>  |
| <b>LIST OF TABLES</b> .....   | <b>VIII</b> |
| <b>CHAPTER 1. INTRODUCTION</b> .....  | <b>1</b>    |
| 1.1. SPECIALIZED METABOLITES .....  | 1           |
| 1.2. POLYKETIDES .....  | 1           |
| 1.2.1 <i>Type I polyketide synthase</i> .....   | 2           |
| 1.2.1.1 Important domains employed in polyketide biosynthesis .....                         | 6           |
| 1.2.1 <i>Trans-AT PKSs</i> .....  | 13          |
| 1.2.2 <i>Type II PKSs</i> .....   | 16          |
| 1.2.3 <i>Type III PKSs</i> .....  | 17          |
| 1.3 NONRIBOSOMAL PEPTIDE SYNTHETASES .....  | 17          |
| 1.4 RIBOSOMALLY SYNTHESIZED AND POST-TRANSLATIONALLY MODIFIED PEPTIDES .....                | 18          |
| <b>AIM OF THIS THESIS</b> .....   | <b>22</b>   |
| <b>CHAPTER 2. POLYKETIDE PROOFREADING BY AN ACYLTRANSFERASE-LIKE ENZYME</b> .....           | <b>24</b>   |
| 2.1 SUMMARY .....   | 24          |
| 2.2 INTRODUCTION .....  | 24          |
| 2.3 RESULTS .....   | 25          |
| 2.3.1 <i>Study of BaeB, a hydrolase candidate from bacillaene biosynthesis</i> .....        | 25          |
| 2.3.2 <i>Differential roles of the pederin AT-like enzymes in malonyl transfer</i> .....    | 26          |
| 2.3.3 <i>The AT2 PedC hydrolyzes acyl N-acetylcysteamine thioesters</i> .....               | 28          |
| 2.3.4 <i>The AT2 PedC releases ACP-bound acyl units</i> .....                               | 29          |
| 2.4 DISCUSSION .....  | 30          |
| 2.5 SIGNIFICANCE .....  | 31          |
| 2.6 EXPERIMENTAL PROCEDURE.....   | 31          |
| 2.6.1 <i>General</i> .....  | 31          |
| 2.6.2 <i>Bacterial strains and culture conditions</i> .....                                 | 31          |
| 2.6.3 <i>Plasmids and general DNA procedures</i> .....                                      | 32          |
| 2.6.4 <i>Generation of Bacillus amyloliquefaciens mutants and polyketide analysis</i> ..... | 32          |
| 2.6.5 <i>Construction of expression plasmids</i> .....                                      | 33          |
| 2.6.6 <i>Expression and purification of proteins</i> .....                                  | 34          |
| 2.6.7 <i>Acyltransferase assays</i> .....   | 35          |
| 2.6.8 <i>Acyl-SNAC hydrolysis assays</i> .....  | 35          |
| 2.6.9 <i>Acyl-ACP hydrolysis assays</i> .....   | 36          |
| 2.6.10 <i>Phylogenetic analysis</i> .....   | 36          |
| 2.6.11 <i>Synthesis of acyl-SNAC derivatives 18-23</i> .....                                | 36          |
| 2.7 ACKNOWLEDGEMENT .....   | 37          |
| 2.8 SUPPLEMENTAL INFORMATION.....   | 38          |
| <b>CHAPTER 3. DECIPHERING HALOGENATION IN OOCYDIN AND HATERUMALIDE BIOSYNTHESIS</b> .....   | <b>42</b>   |
| 3.1 SUMMARY .....   | 42          |
| 3.2 INTRODUCTION .....  | 42          |

|  |  |            |
|--|--|------------|
| 3.3  | RESULTS .....  | 45         |
| 3.3.1  | <i>Confirmation of <b>11</b> biosynthesis by <i>S. plymuthica</i> 4Rx13</i> .....  | 45         |
| 3.3.2  | <i>Assigning Function to Halogenase Candidates by Gene Inactivation</i> .....  | 46         |
| 3.3.3  | <i>Bioinformatic analysis of <i>OocP</i><sub>4Rx13</sub> and <i>OocQ</i><sub>4Rx13</sub></i> .....                               | 47         |
| 3.3.4  | <i>Clarification of <i>OocP</i><sub>4Rx13</sub> and <i>OocQ</i><sub>4Rx13</sub> influence on the halogenation reaction</i> ..... | 48         |
| 3.3.5  | <i>Detection of additional biosynthetic products from the <i>ooc</i><sub>4Rx13</sub> gene cluster</i> .....                      | 50         |
| 3.3.6  | <i>Cytotoxic properties of oocydin A <b>11</b> and its analogs</i> .....   | 51         |
| 3.4  | DISCUSSION .....   | 51         |
| 3.5  | EXPERIMENTAL PROCEDURE.....  | 57         |
| 3.5.1  | <i>General</i> .....   | 57         |
| 3.5.2  | <i>Bacterial strains and culture conditions</i> .....  | 57         |
| 3.5.3  | <i>Plasmid and general DNA procedure</i> .....   | 57         |
| 3.5.4  | <i>Fermentation and extraction of <i>Serratia plymuthica</i> 4Rx13</i> .....   | 58         |
| 3.5.5  | <i>Preparative HPLC and HPLC-HRMS procedure</i> .....  | 59         |
| 3.5.6  | <i>Isolation and structure confirmation of oocydin A <b>11</b></i> .....   | 59         |
| 3.5.7  | <i>Isolation and structure elucidation of dechloro-oocydin A <b>42</b></i> .....   | 59         |
| 3.5.8  | <i>Isolation and structure elucidation of oocydin B <b>43</b></i> .....  | 60         |
| 3.5.9  | <i>NMR measurements</i> .....  | 60         |
| 3.5.10   | <i>Generation of knock-out strains and cloning of constructs</i> .....   | 60         |
| 3.5.11   | <i>Expression and purification of proteins</i> .....   | 63         |
| 3.5.1  | <i>Cross-complementation, inhibitor assay, feeding experiments and in vitro halogenation assay</i> ....                          | 63         |
| 3.5.2  | <i>Assay for cytotoxicity against HeLa cells</i> .....   | 64         |
| 3.5.3  | <i>Data analysis workflow with eMZed</i> .....   | 64         |
| 3.6  | ACKNOWLEDGMENTS .....  | 64         |
| 3.7  | SUPPLEMENTAL INFORMATION .....   | 65         |
| <b>CHAPTER 4. STUDYING RECOGNITION OF THE UNUSUAL RIPP PROTEASE POYH</b> ..... |  | <b>81</b>  |
| 4.1  | SUMMARY .....  | 81         |
| 4.2  | INTRODUCTION .....   | 81         |
| 4.3  | RESULTS .....  | 83         |
| 4.3.1  | <i>Generation of a model substrate for further cleavage studies</i> .....  | 83         |
| 4.3.2  | <i>Single point mutation screen refines recognition sequence</i> .....   | 83         |
| 4.3.1  | <i>Detection of a putative zymogen form of PoyH</i> .....  | 85         |
| 4.4  | DISCUSSION .....   | 86         |
| 4.5  | EXPERIMENTAL PROCEDURE.....  | 88         |
| 4.5.1  | <i>Strains and culture conditions</i> .....  | 88         |
| 4.5.2  | <i>Construction of expression plasmids and general DNA procedure</i> .....   | 88         |
| 4.5.3  | <i>Expression and purification of proteins</i> .....   | 89         |
| 4.5.4  | <i>Cleavage Assays</i> .....   | 90         |
| 4.5.5  | <i>Mass spectrometry procedure</i> .....   | 90         |
| 4.6  | ACKNOWLEDGEMENT .....  | 91         |
| 4.7  | SUPPLEMENTAL INFORMATION .....   | 92         |
| <b>CHAPTER 5. CONCLUSION AND FURTHER PERSPECTIVES</b> .....                    |  | <b>93</b>  |
| <b>CHAPTER 6. BIBLIOGRAPHY</b> .....   |  | <b>104</b> |
| <b>CHAPTER 7. APPENDIX</b> .....   |  | <b>A</b>   |
| 7.1  | ABBREVIATIONS USED.....  | A          |
| 7.2  | SEQUENCE ALIGNMENTS .....  | E          |

# List of Figures

|   |    |
|---|----|
| FIGURE 1-1: A PLOT OF NUMBER OF RESISTANT BACTERIAL STRAINS AND COMPANIES BEING ENGAGED IN ANTIBIOTIC RESEARCH IN DEPENDENCY OF THE TIME, MODIFIED AFTER COOPER AND SHLAES 2011. .... | 1  |
| FIGURE 1-2: STRUCTURE AND BIOACTIVITY OF SELECTED POLYKETIDES. ....   | 2  |
| FIGURE 1-3: SUBDIVISION OF POLYKETIDE SYNTHASES. ....   | 2  |
| FIGURE 1-4: COENZYME A (6) TRANSFER CATALYZED BY PPTASE. ....   | 3  |
| FIGURE 1-5: SCHEME OF POLYKETIDE BIOSYNTHESIS MODIFIED AFTER XU, QIAO, AND TANG 2012. ....  | 4  |
| FIGURE 1-6: NOMENCLATURE OF POLYKETIDE INTERMEDIATES MODIFIED AFTER KEATINGE-CLAY 2012. ....  | 5  |
| FIGURE 1-7: SCHEME OF ERYTHROMYCIN BIOSYNTHESIS BY 6-DEOXYERYTHRONOLIDE B SYNTHASE (DEBS). ....   | 5  |
| FIGURE 1-8: CRYSTAL STRUCTURE OF INDIVIDUAL POLYKETIDE SYNTHASE DOMAINS TAKEN FROM KHOSLA ET AL. 2014. ....   | 6  |
| FIGURE 1-9: SCHEME OF PING-PONG BI-BI MECHANISM OF AT-CATALYZED REACTION. ....  | 6  |
| FIGURE 1-10: SCHEMATIC REPRESENTATION OF ACYLTRANSFERASE (AT)-CATALYZED REACTION MECHANISM AFTER KEATINGE-CLAY 2012. ....   | 7  |
| FIGURE 1-11: SCHEMATIC PROPOSAL OF KETOSYNTHASE (KS)-CATALYZED REACTION MECHANISM AFTER KEATINGE-CLAY 2012. ....  | 8  |
| FIGURE 1-12: SCHEMATIC REPRESENTATION OF THIOESTERASE (TE)-CATALYZED REACTION MECHANISM MODIFIED AFTER DU AND LOU 2010. ....  | 8  |
| FIGURE 1-13: SCHEMATIC REPRESENTATIVE OF KETOREDUCTASE (KR)-CATALYZED REACTION MECHANISM AFTER KEATINGE-CLAY 2012. ....   | 9  |
| FIGURE 1-14: SCHEMATIC PROPOSAL OF DEHYDRATASE (DH)-CATALYZED REACTION MECHANISM AFTER KEATINGE-CLAY 2012. ....   | 10 |
| FIGURE 1-15: SCHEMATIC PROPOSAL OF ENOYLREDUCTASE (ER)-CATALYZED REACTION MECHANISM AFTER KEATINGE-CLAY 2012. ....  | 10 |
| FIGURE 1-16: POST-PKS MODIFICATIONS AT THE EXAMPLE OF ERYTHROMYCIN BIOSYNTHESIS. ....   | 11 |
| FIGURE 1-17: REACTION MECHANISM OF A MONONUCLEAR NON-HEME IRON (MNH) A-KETOGLUTARATE (KG)- AND O <sub>2</sub> -DEPENDENT HALOGENASE MODIFIED AFTER MATTHEWS ET AL. 2014. ....         | 13 |
| FIGURE 1-18: POSITION AND PRESENCE OF ACYLTRANSFERASES (ATs) IN TRANS-AT PKSS AFTER MUSIOL AND WEBER 2012. ....   | 14 |
| FIGURE 1-19: REPRESENTATIVES OF TRANS-ACYLTRANSFERASE (AT) POLYKETIDE SYNTHASE (PKS) PRODUCTS. ....   | 15 |
| FIGURE 1-20: POSSIBLE B-BRANCHING MECHANISM MODIFIED AFTER TILL AND RACE 2014. ....   | 16 |
| FIGURE 1-21: GENERAL TYPE II POLYKETIDE SYNTHASE (PKS) REACTION SCHEME MODIFIED AFTER HERTWECK ET AL. 2007. ....  | 17 |
| FIGURE 1-22: GENERAL TYPE III POLYKETIDE SYNTHASE (PKS) REACTION MECHANISM AFTER WEISSMAN 2009. ....  | 17 |
| FIGURE 1-23: COMPARISON OF PKS, NRPS AND PKS-NRPS-HYBRIDS MODIFIED AFTER WEISSMAN AND MÜLLER 2008. ....   | 18 |
| FIGURE 1-24: STRUCTURE OF SELECTED RIBOSOMALLY PRODUCED AND POSTTRANSLATIONALLY MODIFIED PEPTIDES (RIPPs). ....   | 19 |
| FIGURE 1-25: GENERAL SCHEME OF RIBOSOMALLY PRODUCED AND POSTTRANSLATIONALLY MODIFIED PEPTIDE (RIPP) BIOSYNTHESIS. ....  | 20 |
| FIGURE 1-26: GENERATION OF A LANTHIONINE BRIDGE. ....   | 20 |
| FIGURE 1-27: POLYTHEONAMIDES (14) ARE ABLE TO FORM UNIMOLECULAR ION CHANNELS THROUGH CELL MEMBRANES, TAKEN FROM WALSH 2014. ....  | 21 |
| FIGURE 2-1: SNAC DERIVATIVES TESTED IN THE HYDROLYSIS. ....   | 26 |
| FIGURE 2-2: PHYLOGRAM OF AT-LIKE ENZYMES FROM VARIOUS TRANS-AT PKS PATHWAYS. ....   | 27 |
| FIGURE 2-3: ACYLTRANSFERASE ASSAY USING THE PEDERIN ATs. ....   | 28 |
| FIGURE 2-4: HYDROLYSIS ASSAYS USING ACYL-SNAC. ....   | 29 |
| FIGURE 2-5: SUPPLEMENTAL FIGURE 1, RELATED TO FIGURE 2-1. ....  | 38 |
| FIGURE 2-6: SUPPLEMENTAL FIGURE 2, RELATED TO FIGURE 2-3. ....  | 38 |
| FIGURE 2-7: SUPPLEMENTAL FIGURE 3, RELATED TO FIGURE 2-4. ....  | 39 |
| FIGURE 2-8: SUPPLEMENTAL FIGURE 4, RELATED TO FIGURE 2-4. ....  | 39 |
| FIGURE 3-1: OOCYDIN A BIOSYNTHETIC GENE CLUSTER OF <i>SERRATIA PLYMUTHICA</i> 4Rx13 WITH CORRESPONDING PKS MODULES. ...   | 45 |
| FIGURE 3-2: KEY CORRELATIONS OF DECHLORO-OOCYDIN A 42. ....   | 47 |
| FIGURE 3-3: ISOTOPICALLY RESOLVED ION PATTERN OF OOCYDIN A 11. ....   | 49 |
| FIGURE 3-4: KEY CORRELATIONS OF OOCYDIN B 43. ....  | 51 |
| FIGURE 3-5: PROPOSED SCENARIOS FOR OOCYDIN A HALOGENATION. ....   | 54 |
| FIGURE 3-6: REVISED OOCYDIN A 11 BIOSYNTHESIS SCHEME. ....  | 56 |
| FIGURE 3-7: EXTRACTED ION CHROMATOGRAM (EIC) OF DIFFERENT <i>SERRATIA PLYMUTHICA</i> 4Rx13 STRAINS. ....  | 65 |

|  |     |
|--|-----|
| FIGURE 3-8: <sup>1</sup> H NMR OF DECHLORO-OOCYDIN A 42 IN METHANOL-D <sub>4</sub> OBTAINED AT 500 MHZ.....  | 69  |
| FIGURE 3-9: COSY SPECTRUM OF DECHLORO-OOCYDIN A 42 IN METHANOL-D <sub>4</sub> OBTAINED AT 500 MHZ.....   | 70  |
| FIGURE 3-10: HMBC SPECTRUM OF DECHLORO-OOCYDIN A 42 IN METHANOL-D <sub>4</sub> OBTAINED AT 500 MHZ. ....   | 71  |
| FIGURE 3-11: HSQC SPECTRUM OF DECHLORO-OOCYDIN A 42 IN METHANOL-D <sub>4</sub> OBTAINED AT 500 MHZ. ....   | 72  |
| FIGURE 3-12: NOESY SPECTRUM OF DECHLORO-OOCYDIN A 42 IN METHANOL-D <sub>4</sub> OBTAINED AT 500 MHZ. ....  | 73  |
| FIGURE 3-13: <sup>1</sup> H NMR SPECTRUM OF OOCYDIN B 43 IN METHANOL-D <sub>4</sub> OBTAINED AT 500 MHZ. ....  | 74  |
| FIGURE 3-14: COSY SPECTRUM OF OOCYDIN B 43 IN METHANOL-D <sub>4</sub> OBTAINED AT 500 MHZ. ....  | 75  |
| FIGURE 3-15: HMBC SPECTRUM OF OOCYDIN B 43 IN METHANOL-D <sub>4</sub> OBTAINED AT 500 MHZ.....   | 76  |
| FIGURE 3-16: HSQC SPECTRUM OF OOCYDIN B 43 IN METHANOL-D <sub>4</sub> OBTAINED AT 500 MHZ.....   | 77  |
| FIGURE 3-17: NOESY SPECTRUM OF OOCYDIN B 43 IN METHANOL-D <sub>4</sub> OBTAINED AT 500 MHZ. ....   | 78  |
| FIGURE 3-18: SEQUENCE ALIGNMENT OF KNOWN MNH HALOGENASES USING T-COFFEE WEB SERVER (DI TOMMASO ET AL. 2011).<br>.....                                | 79  |
| FIGURE 3-19: SEQUENCE ALIGNMENT OF KNOWN MNH HALOGENASES USING EXPRESSO WEB SERVER (ARMOUGOM ET AL. 2006). 80  | 80  |
| FIGURE 4-1: POLYTHEONAMIDE 14 GENE CLUSTER AND CORRESPONDING STRUCTURE.....  | 82  |
| FIGURE 4-2: SEQUENCE ALIGNMENT OF POYA AND PTH_2340 RECOGNITION SITE.....  | 83  |
| FIGURE 4-3: COMPARISON OF POYH-PERFORMANCE WITH SINGLE-SITE MUTANTS OF POYH RECOGNITION SITE. ....   | 84  |
| FIGURE 4-4: CLEAVAGE ASSAY OF TEV VERSUS POYH IN BAEB. ....  | 85  |
| FIGURE 4-5: COMPARISON OF POYH-PERFORMANCE AFTER PREINCUBATION IN ACETATE BUFFER WITH DIFFERENT PH VALUES. ....                                      | 85  |
| FIGURE 4-6: HPLC-HRMS CHROMATOGRAM FOR TREATED POYH IN ACETATE BUFFER WITH DIFFERENT PH VALUES RUN ON A C4-<br>COLUMN. ....                          | 86  |
| FIGURE 4-7: PSI-BLAST PSEUDO-MULTIPLE SEQUENCE ALIGNMENT TAKEN FROM PHYRE-CALCULATIONS. ....   | 92  |
| FIGURE 5-1: STRUCTURES OF 5, 8, 14 AND 18.....   | 93  |
| FIGURE 5-2: ALIGNMENT OF ACYLTRANSFERASES (ATs) WITH ACYLHYDROLASES (AHs) CONSENSUS SEQUENCE AND STRUCTURE<br>PREDICTION. ....                       | 96  |
| FIGURE 5-3: MEMBERS OF THE HATERUMALIDE FAMILY DISCUSSED IN THIS CHAPTER.....  | 98  |
| FIGURE 5-4: SECTION OF OOCYDIN A BIOSYNTHETIC GENE CLUSTER OF <i>SERRATIA PLYMUTHICA</i> 4RX13 WITH CORRESPONDING<br>ASSEMBLY LINE PARTS. ....       | 100 |
| FIGURE 7-1: SEQUENCE ALIGNMENT OF ACYLTRANSFERASES (ATs) AND ACYLHYDROLASES (AHs) TO GENERATE CONSENSUS<br>SEQUENCES APPLIED IN FIGURE 5-2A.....     | E   |
| FIGURE 7-2: TABLE OF CRYSTALLIZED MNH ENZYMES AND THEIR CORRESPONDING SEQUENCE MOTIFS AND FOLDS TAKEN FROM<br>KOEHNTOPI, EMERSON, AND QUE 2005. .... | F   |

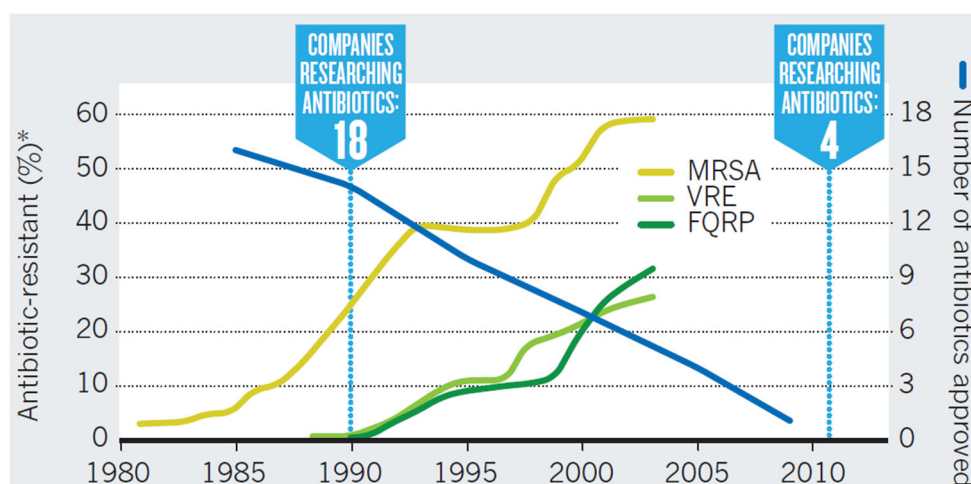
## List of Tables

|   |    |
|---|----|
| TABLE 1-1: CLASSIFICATION OF KNOWN HALOGENASES INVOLVED IN SECONDARY METABOLITE HALOGENATION.....   | 12 |
| TABLE 2-1: SUPPLEMENTAL TABLE 1. ....   | 40 |
| TABLE 2-2: SUPPLEMENTAL TABLE 2. ....   | 41 |
| TABLE 3-1: FAMILY OF HATERUMALIDES; TABLE MODIFIED AFTER KIGOSHI AND HAYAKAWA 2007. ....  | 44 |
| TABLE 3-2: PROPOSED FUNCTIONS OF <i>OOC</i> <sub>4RX13</sub> BIOSYNTHETIC GENE CLUSTER ORFS. ....   | 45 |
| TABLE 3-3: GENERATED AND USED STRAINS IN THIS STUDY.....  | 58 |
| TABLE 3-4: CONSTRUCTS USED IN THIS STUDY. ....  | 62 |
| TABLE 3-5: PRIMERS USED IN THIS STUDY.....  | 63 |
| TABLE 3-6: PROTON AND CARBON NMR DATA FOR OOCYDIN B (43) AND DECHLORO-OOCYDIN A (42) OBTAINED AT 500 MHZ FOR<br><sup>1</sup> H-NMR AND 125 MHZ FOR <sup>13</sup> C-NMR WITH SAMPLES DISSOLVED IN METHANOL-D <sub>4</sub> . .... | 66 |
| TABLE 3-7: COMPARISON OF SELECTED AMINO ACID POSITIONS OF OOCp WITH CURA AND SYRB AFTER SEQUENCE ALIGNMENT WITH<br>T-COFFEE AND EXPRESSO WEB SERVER.....  | 67 |
| TABLE 3-8: PREVIOUS AND NEW ANNOTATION OF USED <i>S. PLYMUTHICA</i> GENES.....  | 68 |
| TABLE 4-1: CONSTRUCTS USED IN THIS STUDY. ....  | 89 |
| TABLE 4-2: PRIMER USED IN THIS STUDY.....   | 89 |

# Chapter 1. Introduction

## 1.1. Specialized metabolites

Specialized metabolites (also called secondary metabolites) are synthesized by nearly every cell for various purposes. A very rough definition counts everything that is not necessary for a cell's survival as secondary metabolites (Croteau, Kutchan, and Lewis 2000). A huge diversity of compounds are encompassed by this definition including polyketides, nonribosomal peptides, ribosomally-derived peptides, alkaloids and many more. These compounds are synthesized for a wide variety of purposes by the host cell; hence, they often harbor potent activities against bacteria, fungi or cancer cells and often serve as lead structures for drug discovery or are drugs themselves (Tibrewal and Tang 2014). An impressive example of drugs developed from secondary metabolites are antibiotics. After their discovery, antibiotics were widely in the focus of pharmaceutical companies and used extensively and successfully in clinics. In the late 1960s William H. Steward, at the time US Surgeon General, said that it is "time to close the book on infectious disease" and "declare the war against pestilence won" (Spellberg 2008). Today we know that this statement was impetuous. The number of antibiotic resistant strains is increasing while the number of companies being engaged in antibiotic research is decreasing (Figure 1-1) (Cooper and Shlaes 2011). One explanation for the decrease of involved companies is the fact that biosynthetic machineries underlying the production of secondary metabolites are as diverse as the products themselves.

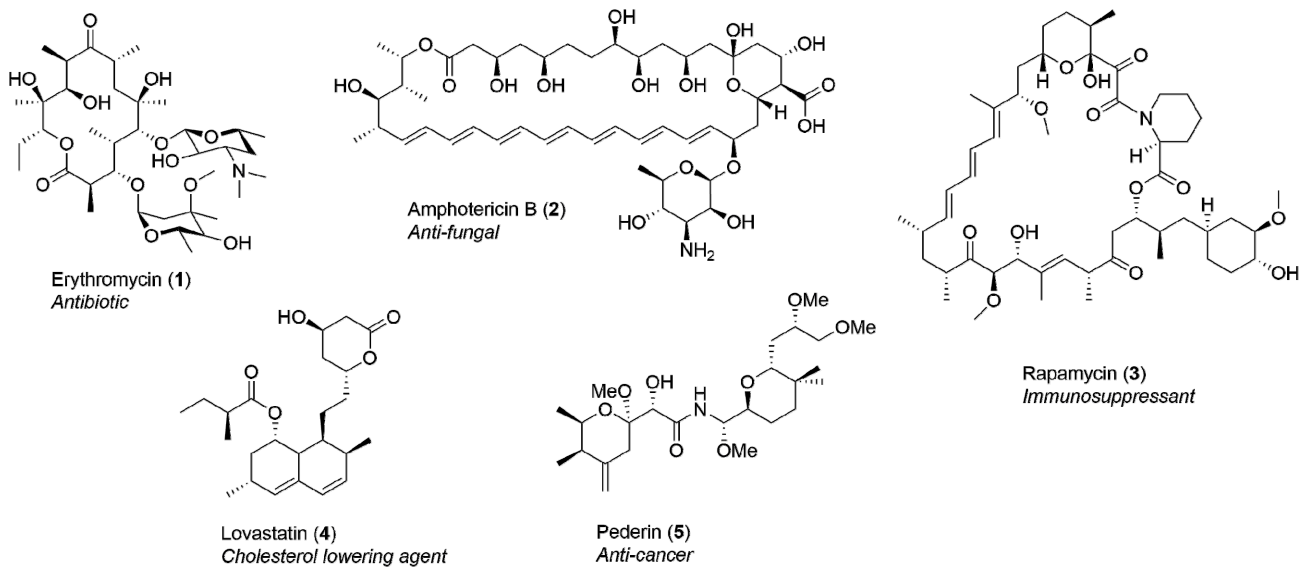


**Figure 1-1: A plot of number of resistant bacterial strains and companies being engaged in antibiotic research in dependency of the time, modified after Cooper and Shlaes 2011.**

The growing number of antibiotic resistant bacterial strains over time correlates with the decrease of pharmaceutical companies being engaged in antibiotic research in the same time scale. Abbreviations: \*, resistance is defined as proportion of clinical isolates that are resistant to antibiotic; MRSA, methicillin-resistant *Staphylococcus aureus*; VRE, vancomycin-resistant *Enterococcus*; FQRP, fluorquinolone-resistant *Pseudomonas aeruginosa*.

## 1.2. Polyketides

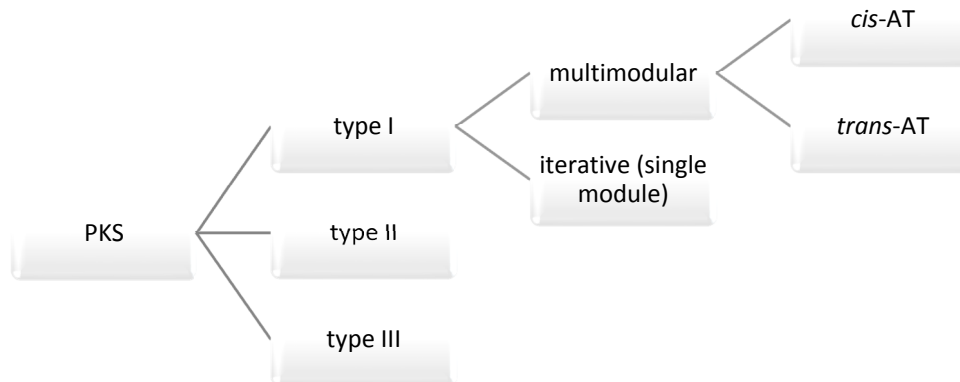
A class of secondary metabolites harboring enormously diverse activities are polyketides (Weissman 2009). It is believed, that more than 1% of the described polyketides have potential drug activity (Cummings, Breitling, and Takano 2014). This spectrum encompasses, e.g., antibiotic activity (with the very prominent example erythromycin (1)), antifungal activity (for example amphotericin B (2)), immunosuppressant activity (for example rapamycin (3)) and lipid-lowering properties (one clinically used example is lovastatin (4)) (Keatinge-Clay 2012) (Figure 1-2).



**Figure 1-2: Structure and bioactivity of selected polyketides.**

Representative polyketides: erythromycin (1), amphotericin B (2), rapamycin (3), lovastatin (4) and pederin (5).

Despite this huge diversity in chemical structure and activity, all polyketides are synthesized by enzyme complexes called polyketide synthases (PKSs) (Hill 2006). Their biosynthesis occurs through a stepwise attachment of small acyl-building blocks in a Claisen-type condensation (Staunton and Weissman 2001). PKSs can be subdivided regarding their architecture (Figure 1-3) (Weissman 2009); all subgroups will be further explained in the next chapters.



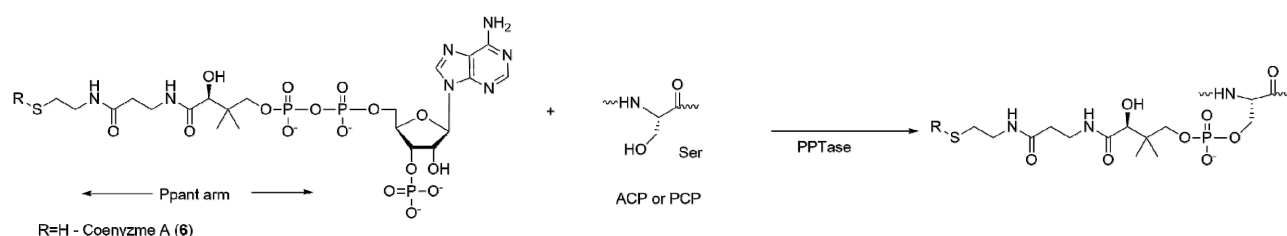
**Figure 1-3: Subdivision of polyketide synthases.**

Polyketide synthases (PKS) can be subdivided into three main types, with additional subclasses in some cases.

### 1.2.1 Type I polyketide synthase

Type I PKSs are the class with the most complicated and most versatile biosynthetic machineries and can be found in very diverse domains of life. They can be present in animals, fungi and amoebae as well as in prokaryotes (streptomycetes, mycobacteria, cyanobacteria) (Keatinge-Clay 2012). Within this class important differences can be identified. The distinction relies on the mode of enzyme-complex usage (iterative or in a row) and on the way building blocks are loaded (*cis* or *trans*).

Multimodular type I PKSs are huge mega-enzyme-complexes. These complexes can be even larger than the ribosome (~2.6 MDa) exemplified with the PKS underlying ECO-02301 biosynthesis (~4.7 MDa) (Keatinge-Clay 2012), (McAlpine, 2005). Modular type I PKSs are named after their modular architecture. Each module links an acyl-building block to the covalently attached growing chain and passes the nascent intermediate from one module to the next, with that the growing chain is shuttled from the beginning of the enzyme-complex to the end. The process is linear and resembles an assembly line. Because of this, the synonym "molecular assembly line" exists (Walsh and Fischbach 2010). This linear fashion has the advantage that the product can be easily predicted by the order of domains and vice versa. This correlation is called the "colinearity rule" (Piel 2010). One module comprises several domains and every domain is responsible for a single reaction step (Keating and Walsh 1999). A module is the smallest unit in a PKS able to catalyze a full round of monomer selection, attachment and carbon-carbon-bond formation. It harbors at least a ketosynthase (KS), an acyl carrier protein (ACP) and an acyltransferase (AT) domain. The ACP must be activated to be functional. This "priming" step converts the inactive *apo* form into the active *holo* variant by attachment of a coenzyme A (CoA) (6)-derived 4'-phosphopantetheine (Ppant) group to a conserved serine residue in a (D/E)XGxD $\underline{S}$ L motif (Figure 1-4). The priming only has to happen once as one *holo*-ACP is used repeatedly (Keating and Walsh 1999). The responsible enzyme phosphopantetheinyl transferase (PPTase) acts in a posttranslational fashion. A free thiol group at the end of this Ppant-arm serves as an anchoring point for the growing chain and is often described as a "shuttle" or "swinging arm" (Perham 2000) in the ACP. This "swinging arm" can span 20 Å and is able to channel the intermediates through all the various catalytic sites on the PKS (Sunbul, Zhang, and Yin 2009). The growing chain is tethered to the ACP throughout the whole catalytic process. Beside the ACP, acyl-building blocks are activated as CoA-thioesters as well due to the transfer of 6. In both cases thioesters and not oxygen esters are employed due to their higher reactivity. Starter moieties comprise acetyl-, benzoyl-, and propionyl-CoAs or even more varied structures, like methoxymalonyl-CoA for example can be used. The most abundant elongation acyl-CoAs are malonyl-CoA and methylmalonyl-CoA (Fischbach and Walsh 2006). Aside from these classical elongation units, ethylmalonyl-, methoxymalonyl- and even more exotic components like hydroxymalonyl- or aminomalonyl-CoA can be found (Cheng et al. 2009).



**Figure 1-4: Coenzyme A (6) transfer catalyzed by PPTase.**

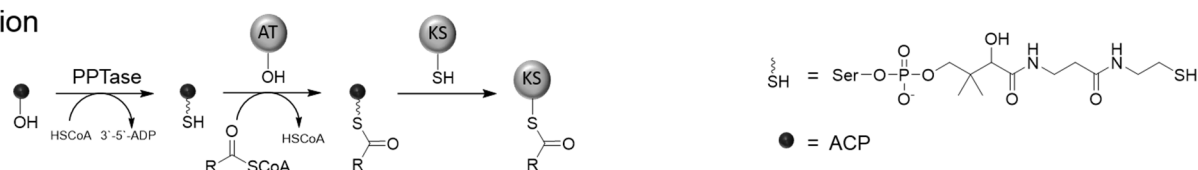
The 4'-phosphopantetheine arm of coenzyme A (6) is transferred onto a conserved serine residue of an acyl carrier protein (ACP) in the case of polyketide synthases and on a peptidyl carrier protein (PCP) in the case of nonribosomal peptide synthetases (chapter 1.3) by the enzyme phosphopantetheinyl transferase (PPTase).

The biosynthetic mechanism of modular type I PKSs resembles that of type I fatty acid synthases (FAS) from primary metabolism and both are believed to share a common ancestor, but they evolved apart from each other (Smith and Tsai 2007; Dutta et al. 2014). It is believed that within this evolution PKSs for example have inserted other domains, leading to a variation of reactions performed within a single module (Fischbach and Walsh 2006). Nevertheless, PKSs and FASs share the same basic domains (KS, malonyl-CoA/acetyl-CoA transacylase (MAT), dehydratase (DH), enoylreductase (ER), ketoreductase (KR), ACP and thioesterase (TE)) (Weissman and Müller 2008). FASs exclusively act in an iterative fashion, in contrast to multimodular type I PKSs that act in a non-iterative fashion (Weissman and Müller 2008). A huge difference between FASs and PKSs is the oxidative state of their products. In FAS the intermediates are fully reduced (Staunton and Weissman 2001), whereas in PKSs full, partial or no reductions and

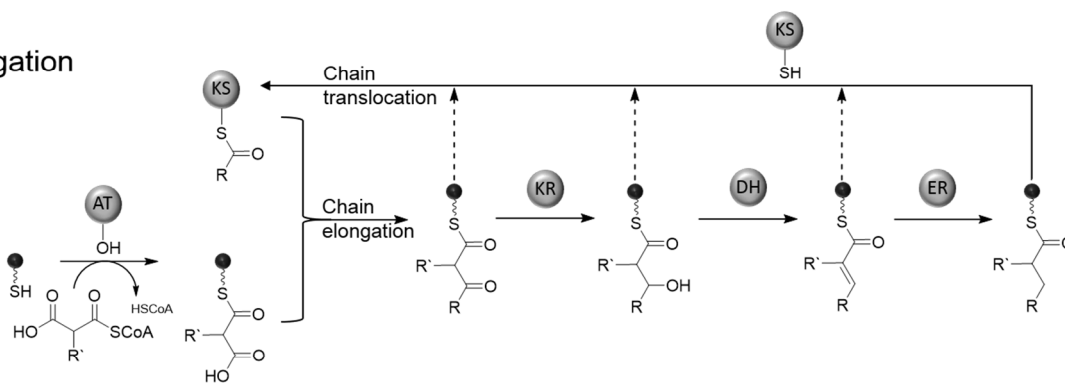
dehydrations can be present (Piel 2010). The different reductive states and the aforementioned variation in domain architecture in PKSs leads to a larger range of structural variety in polyketides than in fatty acids.

A typical reaction scheme of modular type I PKS biosynthesis is comprised of the following steps (Figure 1-5): For initiation the ACP is activated by a transfer of a Ppant group. Subsequently an acyltransferase (AT) loads the starter unit onto the *holo*-ACP. A KS domain catalyzes the transfer of this starter moiety from the *holo*-ACP onto itself. In the elongation stage an extender unit is selected by an acyltransferase (AT) domain. The AT loads an ACP with this extender moiety. The acylated ACP docks to the KS that then performs the actual decarboxylative condensation for chain elongation. After this chemical bond formation, the newly-formed  $\beta$ -ketoacyl intermediate is shuttled by the ACP to the processing enzymes and afterwards either back to the KS in case of an iterative PKS or forward to downstream modules or chain-release enzymes (Keatinge-Clay 2012). In the end a TE usually releases the growing chain from the enzyme either as linear polyketide by hydrolysis or as cyclized lactone ring.

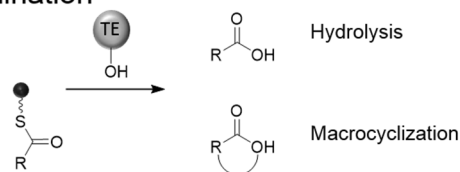
### Initiation



### Elongation



### Termination

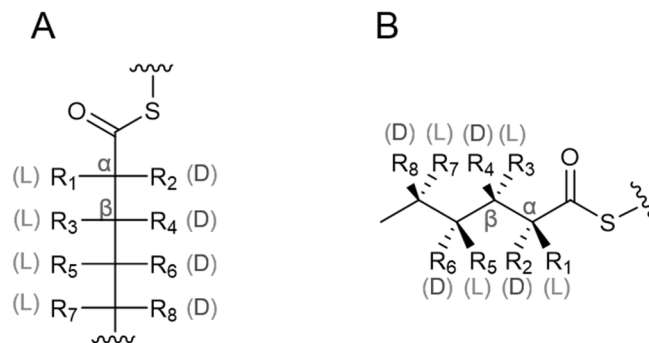


**Figure 1-5: Scheme of polyketide biosynthesis modified after Xu, Qiao, and Tang 2012.**

In the initiation step a 4'-phosphopantetheine (Ppant) group derived from coenzyme A (6) is transferred onto the acyl carrier protein (ACP) for activation. Subsequently the acyltransferase (AT) loads the starter unit onto the activated *holo*-ACP followed by a transfer of the acyl-group onto the ketosynthase (KS). In an elongation cycle an AT loads an extender moiety onto the ACP. The KS catalyzes the decarboxylative condensation of this moiety together with the already loaded unit on the ACP. Further modifications can occur with the help of ketoreductases (KR), dehydratase (DH) and enoylreductase (ER) domains. The growing and modified intermediate is translocated onto the downstream KS and a new cycle begins. In the end a thioesterase catalyzes the release of the expanding chain either by hydrolysis or macrocyclization.

Intermediates in polyketide biosynthesis are often described by the stereochemistry and location of different chemical groups. They can be referred to as having for example a  $\beta$ -hydroxy-group or a D-configuration. D- or L-configuration refers to the Fischer projection, whereas Greek letters always refer to the distance from the (thio)ester group (Figure 1-6).

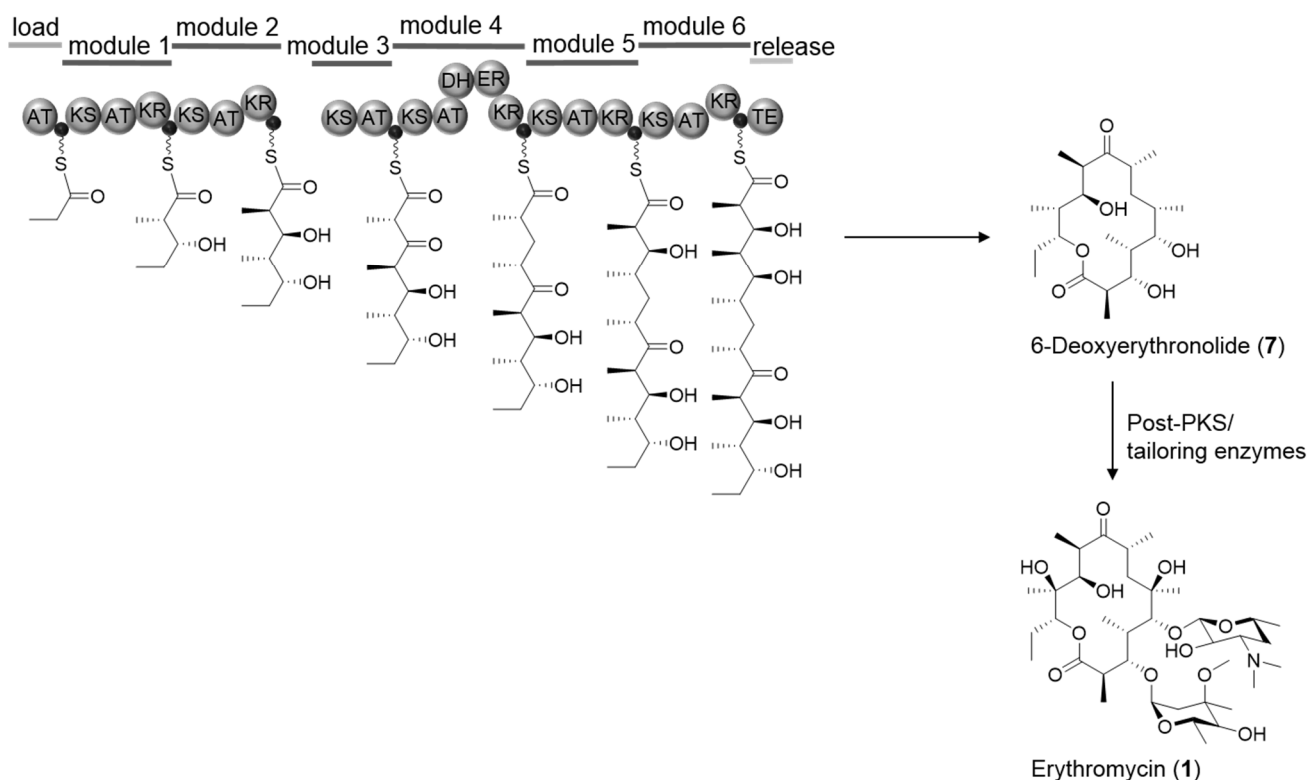




**Figure 1-6: Nomenclature of polyketide intermediates modified after Keatinge-Clay 2012.**

A representative polyketide intermediate shown in (A) Fischer projection and (B) common-chain form.

A lot of knowledge about modular type I PKSs was generated from studies on the 6-deoxyerythronolide B (6dEB) (**7**) synthase (DEBS), the aglycone of erythromycin (**1**) (Figure 1-7). In 1990 first experiments were conducted to understand the underlying biosynthesis (Cortés et al. 1990) and from then on DEBS served as the textbook example in studying modular type I PKSs. DEBS consists of six modules, encoded in three large open reading frames (ORFs) with a size of ~350 kDa each (Staunton and Weissman 2001). Every module catalyzes the decarboxylative condensation of (2S)-methylmalonyl-ACP to the inverted (2R)-2-methyl-3-ketoacyl-ACP intermediate and propionate-CoA serves as starter moiety (Khosla et al. 2014). After reactions of tailoring enzymes the final product **1** is synthesized.

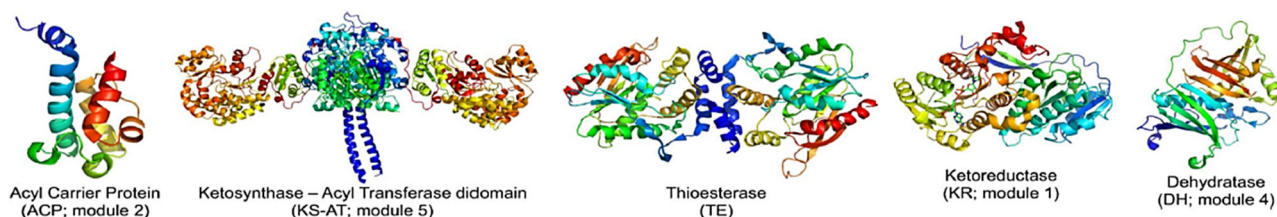


**Figure 1-7: Scheme of erythromycin biosynthesis by 6-deoxyerythronolide B synthase (DEBS).**

Propionyl-CoA serves as starter unit and is loaded by the AT onto the ACP. Every one of the six following modules is responsible for the installation and modification of a  $C_3$  moiety resulting from methylmalonyl-CoA. This elongation and alteration is accomplished by the reaction of KS, AT, ACP and optionally KR, DH and ER. The TE in the last module is responsible for a cyclic release of **7** that results in **1** after tailoring by post-PKS enzymes. Abbreviations: PKS, polyketide synthase; CoA, coenzyme A (**6**); AT, acyltransferase; ACP, acyl carrier protein; KS, ketosynthase; KR, ketoreductases; DH, dehydratase; ER, enoylreductase; TE, thioesterase.

### 1.2.1.1 Important domains employed in polyketide biosynthesis

Crystal structures of all main domains involved in type I PKS biosynthesis are visualized in Figure 1-8.

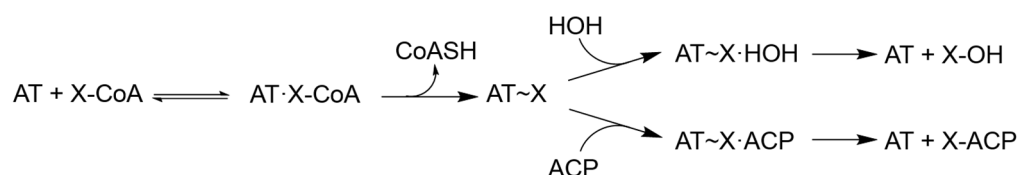


**Figure 1-8: Crystal structure of individual polyketide synthase domains taken from Khosla et al. 2014.**

All domains are obtained from 6-deoxyerythronolide B synthase (DEBS). In the case of KR the bound NADPH cofactor is depicted as well.

The ACP domain is with a size of 100 amino acids (aa) the smallest domain and usually non-catalytic. As mentioned earlier, it serves as carrier and attachment point throughout the whole biosynthetic process. It shares the same fold with carrier proteins from other pathways like the iteratively acting type II FAS or type II PKS. To ensure that the correct enzyme partner will be "delivered" with the growing chain, ACP binds to its interaction partners through different structural features (Xu, Qiao, and Tang 2012). This interaction can either be non-covalent docking or a direct fusion (Dutta et al. 2014).

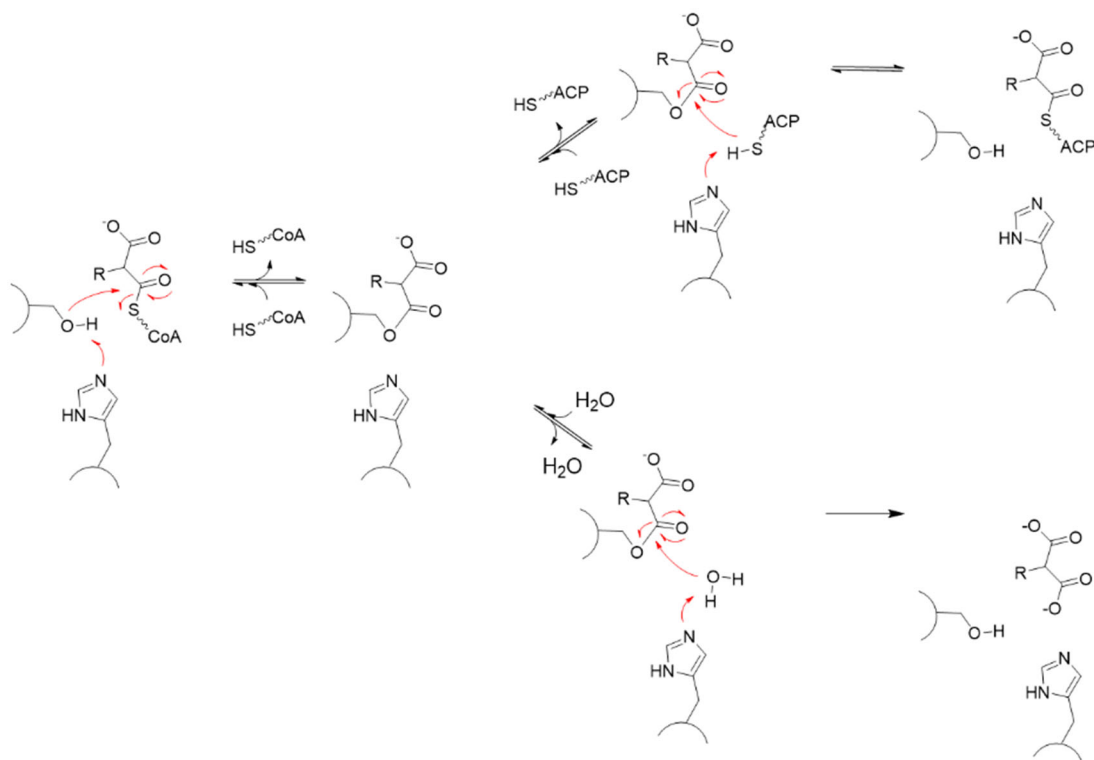
The AT domain is a protein comprising of two subdomains. The first ~240-residue-containing subdomain has an  $\alpha/\beta$ -hydrolase fold and is catalytically active and the second subdomain has a size of ~60 amino acids showing a ferredoxin-like fold. ATs in general employ a canonical Ser-His dyad located within a highly maintained GHSxG motif. The reaction mechanism resembles that of serine proteases and functions in a ping-pong bi-bi fashion (Figure 1-9) (Keatinge-Clay 2012). Competing hydrolysis might occur and it is not fully understood yet how it is prevented. Most likely the protein architecture has an influence on the outcome of the reaction (Serre et al. 1995).



**Figure 1-9: Scheme of ping-pong bi-bi mechanism of AT-catalyzed reaction.**

The translocation of an acyl group onto the acyl carrier protein (ACP) by the acyltransferase (AT) occurs in a ping-pong fashion and generates an acyl-intermediate. Competing hydrolysis is shown on the top branch. Abbreviations: CoA, coenzyme A (6); X, acyl residue.

The serine nucleophile attacks the carbonyl of the thioester substrate with the help of a neighboring histidine to form an acyl-enzyme intermediate (Figure 1-10) (Hedstrom 2002). This tetrahedral intermediate gets stabilized by an oxyanion hole (Xu, Qiao, and Tang 2012). Subsequently the ester carbonyl of the intermediate is attacked by the thiol group of the ACP phosphopantetheinyl side chain in a nucleophilic fashion. The active site His reprotonates the vacated Ser to regenerate the catalytic dyad (Smith and Tsai 2007).



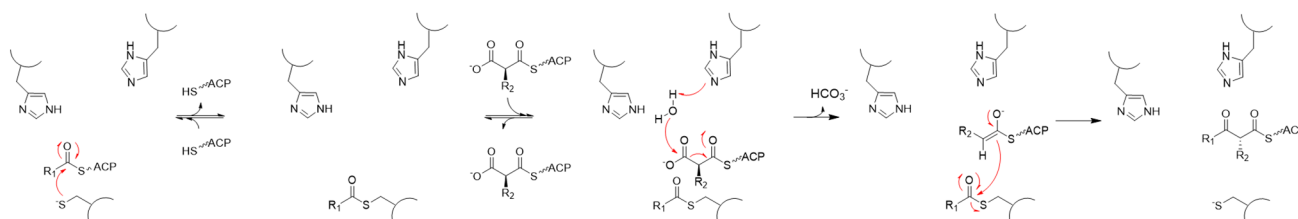
**Figure 1-10: Schematic representation of acyltransferase (AT)-catalyzed reaction mechanism after Keatinge-Clay 2012.**

In the Ser-His dyad the serine attacks the substrate's carbonyl with the help of histidine. The formed intermediate is attacked by the thiol group of an acyl carrier protein (ACP) and reprotonation of the serine leads to an ejection of the acyl-ACP. A competing hydrolysis is shown in the down branch. Abbreviation: CoA, coenzyme A (6).

The AT domains have full control over the choice of elongation moieties and a typically narrow substrate specificity (Hedstrom 2002). It is not fully elucidated how this specificity is biochemically maintained. The very rapid kinetics of this reaction makes experimental elucidation challenging. In 2013 Dunn *et al.* could show for the example of the AT from module 3 of DEBS that this substrate specificity is expressed in formation of the acyl-AT intermediate in the first half of the mechanism. The second half exhibits specificity towards the ACP (Dunn, Cane, and Khosla 2013). First hints regarding substrate extender unit specificity come from sequence comparisons. With the growing number of sequenced genomes the number of available AT sequences has increased. High sequence homology among ATs allows phylogenetic analyses that show separation into clades by preferred extender unit. One important motif (RVDVVQ) is positioned about 30 residues upstream of the catalytic serine. After comparing more than 200 AT sequences this motif was refined. An (R/Q/S/E/D)V(D/E)VVQ sequence indicates the acceptance of methylmalonyl-CoA and ZTx\$(A/T)(Q/E)\$ (Z-hydrophilic residue, \$-aromatic residue) points towards the use of malonyl-CoA (Yadav, Gokhale, and Mohanty 2003). Further information can also be gained from crystal structures (Dunn and Khosla 2013). An important role plays the direct downstream neighbor of the catalytic serine (in the GHGxG motif), which is employed in the steric binding of the substrate. ATs accepting malonyl have a branched hydrophobic amino acid at the "X-position". Other specificities are indicated by less bulky residues (Smith and Tsai 2007). Another crucial motif YASH or HAFH integrates the catalytic histidine. The first indicates methylmalonyl-CoA, the latter malonyl-CoA acceptance (Xu, Qiao, and Tang 2012). Evidence exists for a part of the C-terminal region also playing a role in substrate specificity (Smith and Tsai 2007).

The KS domain is the most conserved enzyme in PKSs and belongs to the thiolase superfamily. Because of this sequence conservation, KSs often serve as starting point for amplifying gene clusters from a metagenomic sample, for example in the case of the pederin gene cluster (Piel 2002). Although they are highly conserved, the reaction

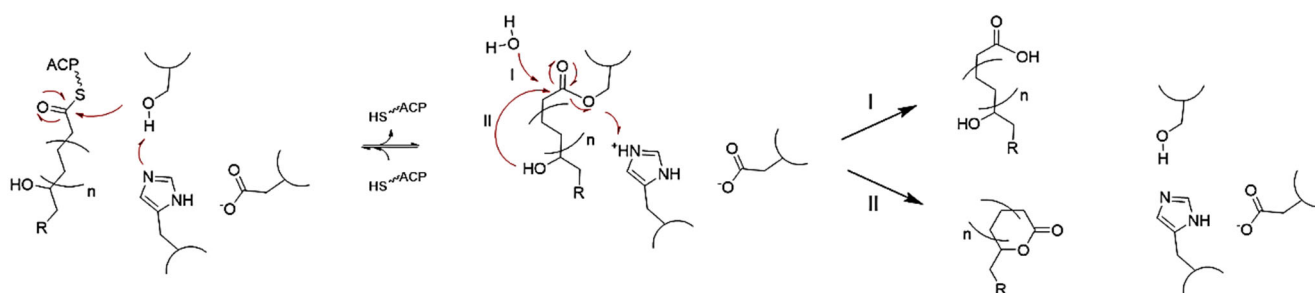
mechanism for the Claisen-like condensation reaction is not fully understood (Keatinge-Clay 2012). A proposed mechanism exhibits a two-step nature (Figure 1-11). For the first step the acyl chain of the acyl-ACP enters the KS substrate tunnel towards the reactive cysteine in a TA<sub>C</sub>SSS motif that attacks the thioester. An oxyanion hole stabilizes the negative charge of the tetrahedral intermediate via hydrogen-bonding interactions. For the second step an additional ACP that is loaded with an extender unit docks to the acylated KS. This time two histidine residues from EAHGTG and KSNIGHT motifs play an important role. One histidine is important for binding and activation of the carboxylate of the extender unit whereas the other histidine activates a water molecule for attack on this activated decarboxylate as a general base. The resulting enolate then attacks the electrophilic thioester carbonyl to elongate the polyketide by one unit (Haapalainen, Meriläinen, and Wierenga 2006), (Keatinge-Clay 2012), (Keating and Walsh 1999). In *trans*-AT PKSs, KSs have a high substrate specificity and with that a special status, this is further explained in the corresponding chapter 1.2.1.



**Figure 1-11: Schematic proposal of ketosynthase (KS)-catalyzed reaction mechanism after Keatinge-Clay 2012.**

The two-step mechanism of the Claisen-like condensation starts with the formation of an acylated cysteine intermediate. In the second step a second loaded ACP gets attacked by an activated water molecule and the resulting enolate ion attacks the thioester carbonyl. This yields an elongated polyketide intermediate.

TEs belong to the superfamily of  $\alpha/\beta$ -hydrolases and have a size of about 240-290 aa. The reaction mechanism underlying this chain release is one of the best understood ones in polyketide biosynthesis (Figure 1-12). The core of this reaction stems from the catalytic triad Ser-His-Asp. In the first half of the reaction, the carbonyl carbon of the assembled chain attached to an ACP is attacked by the nucleophilic hydroxyl group of the active site serine (positioned in a GxSxG motif). The His residue in the catalytic triad, that is stabilized by the Asp residue, acts as a general base and accepts the proton from the catalytic serine. This leads to an acyl-TE ester. In the second half of the reaction, either a macrocyclized or a linear hydrolyzed product is released, depending on the acting TE. In both cases another nucleophilic attack takes place. This happens either from an external nucleophile (usually water) leading to hydrolysis or an internal nucleophile (typically a hydroxyl on the intermediate) leading to a macrocyclized macrolide as in the case of 6dEB (**7**) (Du and Lou 2010), (Keatinge-Clay 2012).



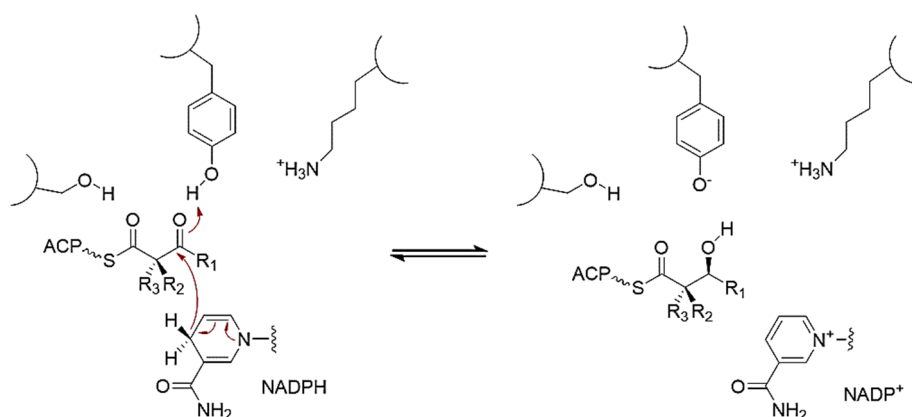
**Figure 1-12: Schematic representation of thioesterase (TE)-catalyzed reaction mechanism modified after Du and Lou 2010.**

In the beginning a serine attacks the carbonyl group of the acylated-acyl carrier protein (ACP) with the help of a histidine residue. An acyl-TE ester results from this reaction that is subsequently either hydrolyzed (I) or cyclized (II) by the attack of an either external (I) or internal (II) nucleophile.

Another large group of thioesterases, beside the here described classical ones, are the so-called type II TEs (TEII). They are usually not encoded within the assembly line (in-line) but can be found as discrete enzymes. Several experiments suggest a role in maintenance of product levels (both in polyketide or nonribosomal peptide biosynthesis), but no direct involvement in the biosynthesis (Schneider and Marahiel 1998), (Doi-Katayama and Hutchinson 2000). They were first characterized in 2001 by Heathcote *et al.* as enzymes being able to cleave off acyl units from the ACP and with this being important for PKS proofreading (Heathcote and Leadlay 2001).

Beside the mandatory KS, AT and ACP domains in a module, many additional components can be found that widen the structural variety of the product. Examples are KRs, DHs and ERs. KRs catalyze the reduction of a keto moiety to a hydroxyl group, DHs can subsequently remove a water molecule and a following reaction of ERs leads to a fully reduced product (Hopwood 2004).

The KR domain is typically monomeric and has a size of about 450 residues. Different variants of KRs are known. Most perform reduction reactions on  $\beta$ -keto groups, others can also influence the stereochemistry of the  $\alpha$ -substituent, and in addition, KRs performing epimerase activity are known as well (Keatinge-Clay 2012), (Garg *et al.* 2014). KRs belong to the superfamily of short-chain dehydrogenases/reductases and exhibit Rossmann-like folds in the C-terminal catalytic subdomain as well as in the N-terminal structural subdomains (Xu, Qiao, and Tang 2012). The latter subdomain is responsible for stabilizing the first. The overall structure creates a cleft in which the NADPH cofactor can bind. Depending on the stereochemical outcome of their product, they are subdivided into "A-type" (L-configuration) or "B-type" KRs (D-configuration). Additionally, a further subdivision is possible if the  $\alpha$ -substituent of the educt is epimerized or not. A highly conserved Ser-Tyr-Lys catalytic triad can be found in the catalytic subdomain performing the reaction (Figure 1-13). The triad activates the target  $\beta$ -carbonyl group and stabilizes the oxyanion intermediate after hydride addition. Subsequently, NADPH attacks leading to reduction (Keatinge-Clay 2012).

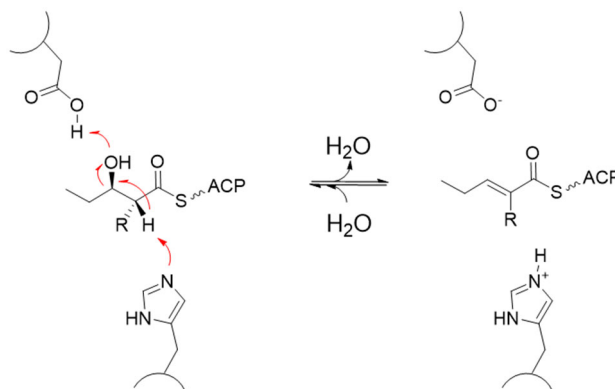


**Figure 1-13: Schematic representative of ketoreductase (KR)-catalyzed reaction mechanism after Keatinge-Clay 2012.**

The KR employs a catalytic Ser-Tyr-Lys triad and NADPH as cofactor. The cofactor attacks the  $\beta$ -carbonyl group leading to a reduction.

The DH domain is responsible for a reversible dehydration reaction on the  $\beta$ -hydroxyl-acyl polyketide intermediate resulting in an  $\alpha,\beta$ -double bond. It belongs to the family of non-metal dehydratases and possesses a double hotdog fold within its average size of 280 residues (Smith and Tsai 2007). The following reaction mechanism is proposed (Figure 1-14): Within the HPALLGD motif the catalytic aspartic acid can be located. This residue protonates the  $\beta$ -hydroxy group of the educt followed by an  $\alpha$ -proton abstraction from the catalytic His residue (positioned in an HxxxGxxxxP motif) resulting in a *syn* elimination of water. *Cis*- and *trans*-double bonds are both possible products, although the latter is formed in most cases. The performed reaction is unusual because the product is

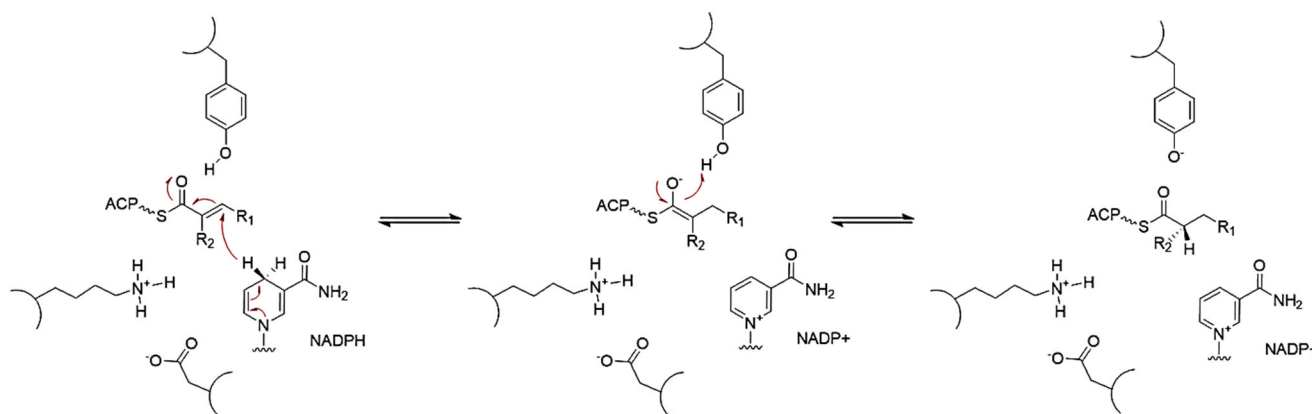
thermodynamically as stable as the educt (Keatinge-Clay 2012). Downstream enzymes select the dehydrated species over the hydrated one, pulling the equilibrium towards the dehydrated product. DH's are sometimes also able to act as isomerases, shifting the  $\alpha/\beta$ -double bond towards the  $\gamma/\delta$  position (Kusebauch et al. 2010).



**Figure 1-14: Schematic proposal of dehydratase (DH)-catalyzed reaction mechanism after Keatinge-Clay 2012.**

An aspartate protonates the  $\beta$ -hydroxy group of the acyl-acyl carrier protein (ACP). A subsequent proton abstraction by a histidine leads to a water elimination.

The ER domain is a member of the acyl-CoA reductase family of the medium-chain dehydrogenase/reductase superfamily with a size of about 310 residues. It is one of the least studied domains. The main function is a stereoselective reduction of  $\text{trans-}\alpha,\beta$ -double bonds generated by the DH with the help of NADPH (Keatinge-Clay 2012). As ER domains are in general not well described only a proposal for the reaction mechanism exists (Figure 1-15). Most likely a hydride is transferred to the  $\beta$ -carbon of the substrate from NADPH. Afterwards a protonation follows from either a general acid or solvent. A tyrosine or lysine is believed to be involved in the latter protonation (Keatinge-Clay 2012), (Xu, Qiao, and Tang 2012).

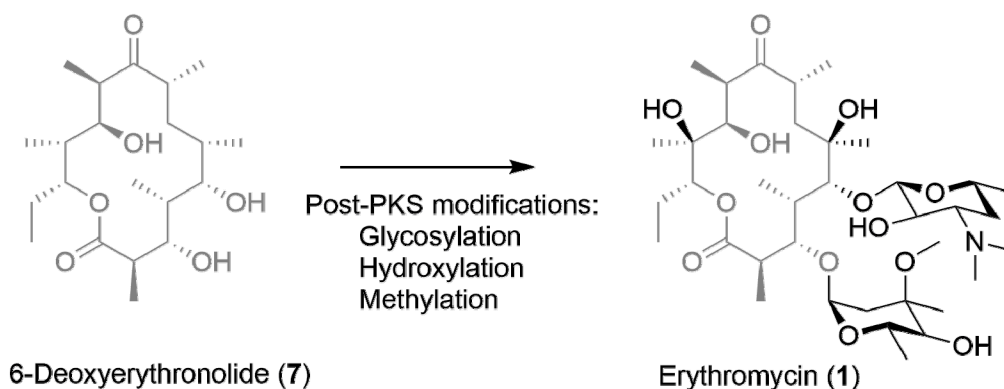


**Figure 1-15: Schematic proposal of enoylreductase (ER)-catalyzed reaction mechanism after Keatinge-Clay 2012.**

The cofactor NADPH attacks the acyl-acyl carrier protein (ACP) and a subsequent protonation leads to the reduced product.

The abovementioned domains perform their reactions while the growing chain is still tethered to the enzyme-complex. Further modifications of the polyketide intermediates can be introduced by post-PKS enzymes, also called tailoring enzymes. In these reactions the polyketide core gets diversified by, for example, attachment of sugar moieties or halogen atoms or by cyclizations (Figure 1-16) (Olano, Méndez, and Salas 2010). These modifications are in many cases necessary to generate the bioactivity. It is known for example that 6dEB (**7**), the aglycone of the antibiotic erythromycin (**1**), is less bioactive than the final product (Weissman and Leadlay 2005). Some enzymes

that catalyze these modifications are oxidoreductases, halogenases, group transferases (for example sugar moieties or methylgroups) or cyclases (Rix et al. 2002). A short overview of the most important enzymes will be given here.



**Figure 1-16: Post-PKS modifications at the example of erythromycin biosynthesis.**

In the erythromycin biosynthesis **7** undergoes glycosylation, hydroxylation and methylation to yield the final bioactive product **1**. The unmodified core is gray, all modifications are indicated in black.

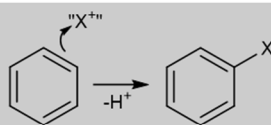
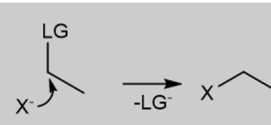
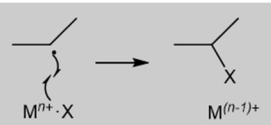
Oxidoreductases are a diverse group of enzymes and responsible for the introduction of oxygen-containing structural moieties like hydroxy, keto, epoxide and aldehyde groups. Alternatively it is possible that they modify such functionalities by addition or removal of hydrogen atoms. Members of these groups able to catalyze such reactions are called oxygenases, reductases, peroxidases and dehydrogenases (Olano, Méndez, and Salas 2010). Although these are only small functional alterations, the impact is enormous. With changes in chiral centers, solubility or hydrogen bond donors/acceptors, the physico-chemical and stereo-electronic properties of the compound are influenced (Rix et al. 2002).

The attachment of sugars is catalyzed by glycosyltransferases. As mentioned above in erythromycin biosynthesis, the attachment of sugar moieties leads to the final bioactive compound. Various decorated molecules can be found. This decoration can occur with one to more than seven sugar moieties, usually from a member of the 6-deoxyhexose family from a donor substrate to an acceptor substrate (Olano, Méndez, and Salas 2010). The shape and stereo-electronic properties of a polyketide is vastly influenced by the aforementioned modification and because of this sugar units are important for the interaction between secondary metabolites and their cellular targets (Rix et al. 2002). Predominantly the attachment of sugars happens through O-glycosidic bonds, but sometimes C- and N-glycosidic linkages can be found as well (Olano, Méndez, and Salas 2010).

Methyl groups can be transferred by methyltransferases. In most cases S-adenosylmethionine (SAM) serves as methyl donor. Nearly all functionalities in a polyketide can be methylated. This variety is exemplified with oxygen atoms in carbonyl or hydroxyl groups, nitrogen atoms in amines, sulfur atoms in thiols, carbon atoms in ring structures (Olano, Méndez, and Salas 2010). This modification can also introduce stereocentres and with that have a steric impact on the polyketide (Rix et al. 2002).

Cyclization can be achieved in several ways. Either free-standing cyclases, TEs out of type I PKSs or oxygenases can be involved in this reaction (Olano, Méndez, and Salas 2010).

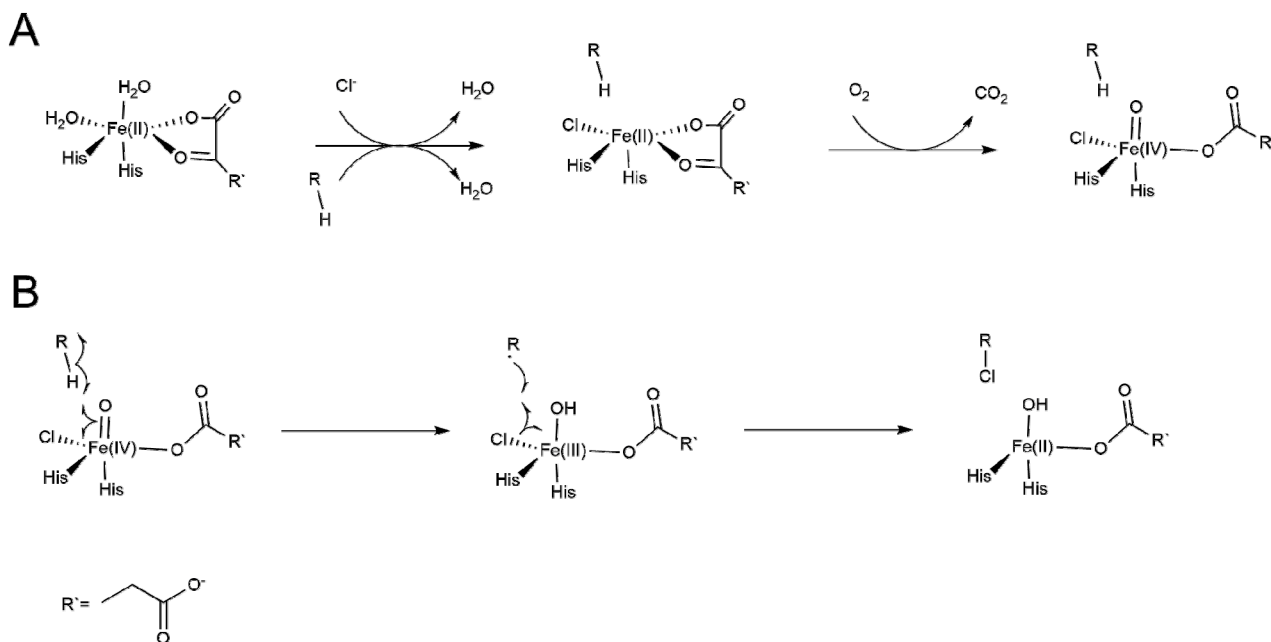
Halogenation is also one of the possible post-PKS modifications and approximately 4,700 halogenated natural products are known (van Pée 2012). Chlorine is predominantly introduced, followed by bromine, iodine and rarely fluorine (Vaillancourt et al. 2006). Surprisingly, only four classes of halogenases are known to collectively modify natural products at almost every type of chemical position. These classes can be distinguished by their underlying chemical reaction mechanisms (Table 1-1).

| Mechanism                       | Electrophilic  | Nucleophilic   | Radical   |
|---------------------------------|--|--|---|
| Form of activated halogen       | Hypohalite X <sup>+</sup>  | Halide X <sup>-</sup>  | Halogen X <sup>·</sup>  |
| Substrate requirements          | Aromatic, electron-rich  | Electrophilic, with a good leaving group   | Aliphatic, unactivated  |
|                                 |   |    |  |
| Enzyme families                 | <ol style="list-style-type: none"> <li>1) Heme-dependent haloperoxidases</li> <li>2) Vanadate-dependent haloperoxidases</li> <li>3) Flavin-dependent halogenases</li> </ol>        | <ol style="list-style-type: none"> <li>1) SAM fluorinase/chlorinase</li> <li>2) SAM-dependent halide methyl transferase</li> </ol> | Mononuclear non-heme iron/ $\alpha$ -ketoglutarate-dependent halogenase             |
| Cofactors and cosubstrates used | <ol style="list-style-type: none"> <li>1) Heme, H<sub>2</sub>O<sub>2</sub></li> <li>2) Vanadate, H<sub>2</sub>O<sub>2</sub></li> <li>3) FADH<sub>2</sub>, O<sub>2</sub></li> </ol> |  | Fe(II), O <sub>2</sub> , $\alpha$ -ketoglutarate                                    |

**Table 1-1: Classification of known halogenases involved in secondary metabolite halogenation.**

Haloperoxidases were known since the 1960s and believed for a long time to be the only existing enzymes able to halogenate (Roberts 2013). The halogenating agent, a hypohalite, acts on electron-rich substrates (Smith, Grünschow, and Goss 2013). Heme-dependent haloperoxidases and vanadate-dependent haloperoxidases can be found in plants (including marine algae), bacteria, fungi and mammals (Blasiak and Drennan 2008). In the beginning of the 2000s, more types of halogenases were discovered. FADH<sub>2</sub>-dependent halogenases also react in an electrophilic fashion. Many halogenations, among others of vancomycin and chlorotetracycline, are described to be performed by this class. They employ molecular oxygen and flavin as corresponding reductive agents. The reduced flavin is typically delivered by a partner protein reductase (Neumann, Fujimori, and Walsh 2008). They can either react on PCP-bound substrates or on free-standing ones (Blasiak and Drennan 2008). A lot about the exact reaction mechanism is known nowadays but it is still not fully revealed and the influence of certain amino acids to the reaction is different between members of the family (Smith, Grünschow, and Goss 2013). Enzymes that are able to perform fluorination have been found as well (Dong et al. 2004). Due to the high electronegativity of the fluorine atom, only a nucleophilic reaction mechanism is possible and enzymes rely on SAM. Three years later a further SAM-dependent halogenating enzyme was described. But in this case the enzyme cannot catalyze fluorinations (Eustáquio et al. 2007). For a long time it remained mysterious how halogenation at unactivated carbons can be achieved, although many compounds are known with a halogen atom at such positions. Examples are syringomycin E, barbamide and jamaicamide A. It turns out that mononuclear non-heme iron (MNH)  $\alpha$ -ketoglutarate (KG)- and O<sub>2</sub>-dependent halogenases are able to perform their reaction at such sites (Vaillancourt, Yin, and Walsh 2005), (Vaillancourt et al. 2005), (Vaillancourt, Vosburg, and Walsh 2006). Furthermore it became apparent that they are also responsible for the introduction of cyclopropane rings after chlorination as in curacin A biosynthesis (Gu et al. 2009). With one exception (Hillwig and Liu 2014), all MNH halogenases act on ACP- or PCP-tethered substrates (Smith, Grünschow, and Goss 2013).





**Figure 1-17: Reaction mechanism of a mononuclear non-heme iron (MNH)  $\alpha$ -ketoglutarate (KG)- and  $O_2$ -dependent halogenase modified after Matthews et al. 2014.**

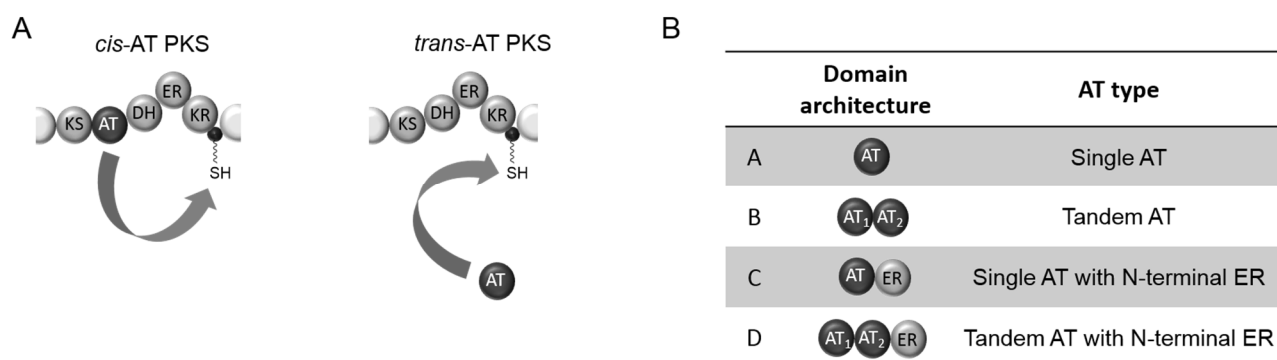
(A) Generation of the reactive Fe(IV)-peroxo species. (B) After hydrogen abstraction from the substrate by the Fe(IV)-peroxo species a halide is transferred to the substrate radical.

The underlying reaction mechanism of MNH halogenases is proposed to be similar to other MNH enzymes (Smith, Grünschow, and Goss 2013). In the first step the halogen coordinates the iron and the substrate binds. Afterwards  $O_2$  coordinated to the open site, attacks  $\alpha$ -KG and its decarboxylation leads to the reactive Fe(IV)-peroxo species (Figure 1-17A) (Wong et al. 2013), (Matthews et al. 2014). This reactive species abstracts a hydrogen atom from the substrate, leaving a radical behind. A rebound attack from the halogen onto the substrate carbon follows, yielding the halogenated product (Figure 1-17B) (Butler and Sandy 2009), (Galonić et al. 2007). The role of  $\alpha$ -KG is more diverse than only acting as a cofactor. In the example of CurA from the curacin biosynthesis, a conformational switch of the enzyme was shown after  $\alpha$ -KG and chloride binding allowed the substrate to enter the enzyme (Khare et al. 2010). This switch is one hallmark of MNH halogenases. Not all of the factors are known which distinguish an MNH halogenase from other MNH enzymes. The character of the active site definitely has an influence. In MNH enzymes, normally two histidine residues and one carboxylate residue, either from a glutamate or aspartate, build up the 2-His-1-carboxylate facial triad (Koehtop, Emerson, and Que 2005). In halogenases the carboxylate residue is exchanged against a small hydrophobic amino acid to generate space for the halogenase binding pocket resulting in a 2-His motif (Pratter et al. 2014b). More factors have to play a role since single-point mutants were not able to make an MNH halogenase out of an MNH hydroxylase or vice versa (Hangasky et al. 2013). In MNH hydroxylases the radicalic rebound occurs from an  $OH\cdot$ , this  $OH\cdot$  could also be generated in MNH halogenases. The question still remains how MNH halogenases prevent hydroxylation. The origin of some halogenations still cannot be explained by these four classes of enzymes, for example the alkynyl bromide moiety of jamaicamide is still a mystery.

### 1.2.1 *Trans*-AT PKSs

In 2002, the gene cluster of pederin (5) PKS from an uncultured bacterial symbiont of the *Paederus* beetle was found and cloned by Piel where surprisingly no acyltransferases were encoded in-line but rather separately

upstream in the gene cluster (Piel 2002). This discovery was the first time an "AT-less" PKS was described (Piel 2010). Following the pederin PKS, the Shen group described the leinamycin gene cluster, where also only a single external AT could be found (Cheng and Shen 2003). It became more and more obvious that pederin (**5**) and leinamycin are no outliers but actually the first members of a new class of PKS. This class was called *trans*-AT PKS and the canonical abovementioned modular type I PKS were named *cis*-AT in comparison (Piel 2010). In *trans*-AT PKSs the AT is exclusively encoded as separate protein whereas in *cis*-AT PKS it is always present in-line (Figure 1-18A). Typically one to three copies of ATs are encountered in a *trans*-AT PKS gene cluster. These discrete proteins can be single-standing, fused to each other or to an ER domain (Figure 1-18B) (Piel 2010), (Bumpus et al. 2008). The almost exclusively used substrate of ATs is malonyl-CoA. Because of the presence of maximal three AT domains for many modules they have to interact iteratively with all ACPs (Weissman and Müller 2008).



**Figure 1-18: Position and presence of acyltransferases (ATs) in *trans*-AT PKSs after Musiol and Weber 2012.**

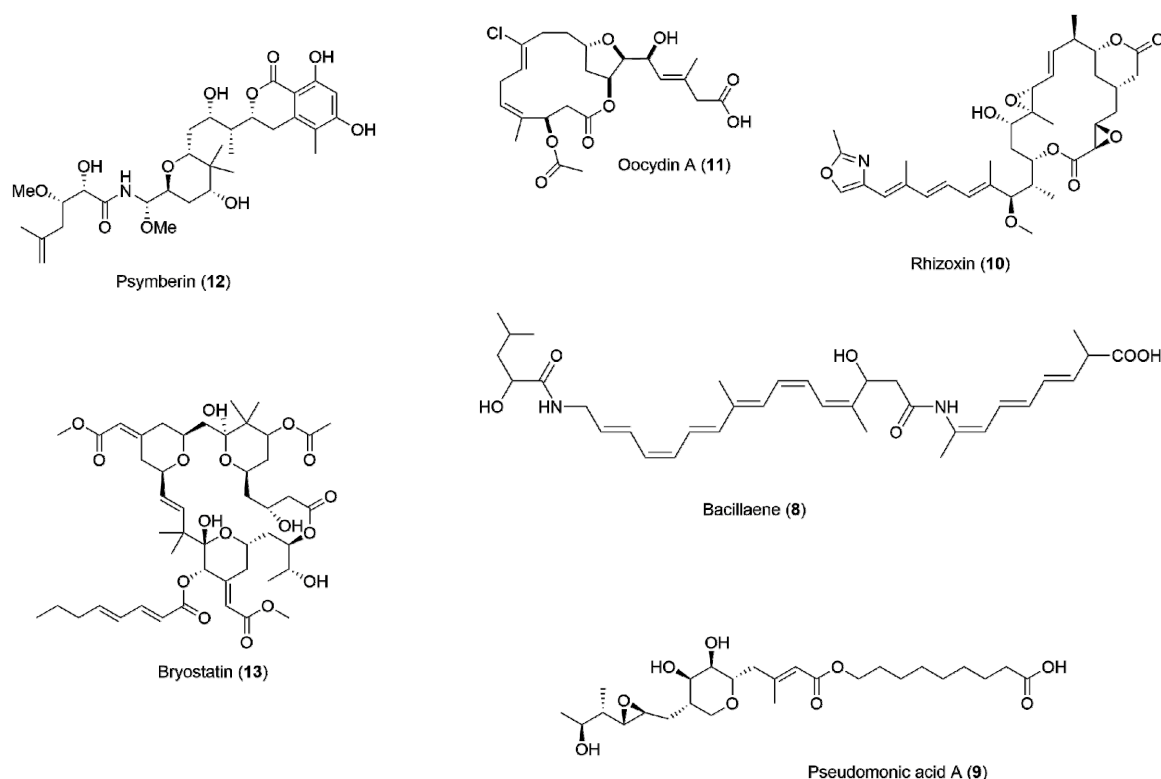
(A) Comparison of AT position in *cis*-AT versus *trans*-AT polyketide synthases (PKSs). In *cis*-AT systems the AT is encoded in-line and acts from there, whereas in contrast, the AT in *trans*-AT PKSs is encoded separately and acts in *trans*. (B) Possible domain appearance of ATs in *trans*-AT PKSs.

The first intensively studied representative of this late PKS subclass was the gene cluster responsible for bacillaene (**8**) biosynthesis (*pksX/baeX*). This gene cluster was already described in 1993 (Scotti and Albertini 1993) and was the first portrayed PKS gene cluster identified from a *Bacillus*; later it appeared that it is actually a *trans*-AT/NRPS hybrid. The *pksX* and *baeX* gene clusters can be found in *Bacillus subtilis* 168 and *Bacillus amyloliquefaciens* FZB42 respectively (Chen et al. 2006b). Jana Moldenhauer was successful in elucidating all biosynthetic steps performed by the Bae PKS (Moldenhauer, Chen, and Borriss 2007), (Moldenhauer et al. 2010). These intensive investigations resulted in an invaluable level of knowledge in this young research field.

Over time, more type I PKSs were found in which the AT is not part of a PKS module but a distinct and iteratively acting enzyme. Recent data suggest that around one quarter of known type I PKS are actually *trans*-AT PKSs (O'Brien et al. 2014) and some selected representatives of *trans*-AT PKS products can be seen in Figure 1-19. Furthermore it was shown by Nguyen *et al.* that *cis*-AT and *trans*-AT PKSs evolved via convergent evolution separately from each other (Nguyen et al. 2008). Typically gene duplication events of single ancestor modules give rise to *cis*-AT PKSs and a combination of domains gained from variable sources via horizontal gene transfer can be found in *trans*-AT PKSs (Nguyen et al. 2008), (Jenke-Kodama, Börner, and Dittmann 2006).

Although it appears on first glance that *cis*- and *trans*-AT PKSs are build up comparably the same way and perform more or less the same reactions, they are actually more distinct. *Trans*-AT PKSs harbor many different domains that are exclusively present in these systems and their modules can be split at unusual positions. This results in many peculiar module variations (eight in *cis*-AT systems versus more than 50 in *trans*-ATs). Often not every domain is used, some get skipped whereas, on the other hand, some may be used iteratively (Piel 2010). All these particularities make the colinearity rule hardly applicable. It could be shown that a prediction of the product can be

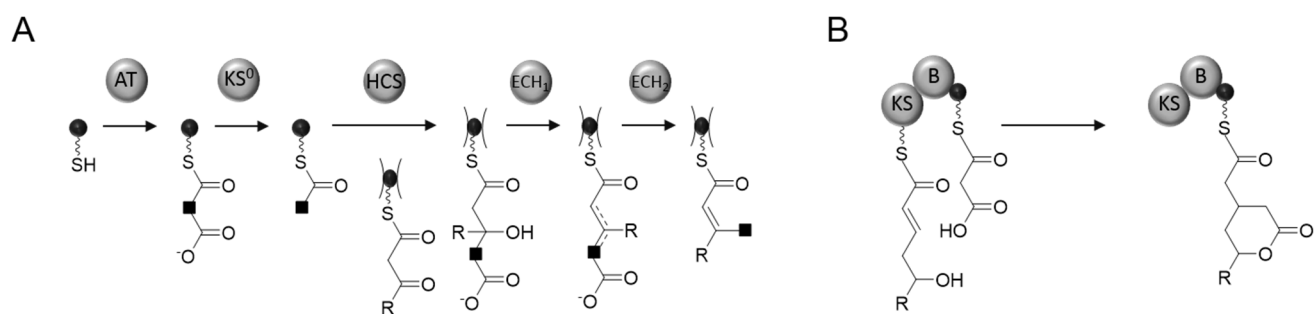
achieved nevertheless, relying on the ketosynthase domain. Contrary to the *cis*-AT where KS have loose substrate specificities, the substrate specificity in *trans*-AT systems is quite narrow and correlates with KS sequence (Nguyen et al. 2008). With the help of this correlation product predictions can be deduced from the gene cluster sequence (Jenner et al. 2013), (Kohlhaas et al. 2013).



**Figure 1-19: Representatives of *trans*-acyltransferase (AT) polyketide synthase (PKS) products.**

Polyketide representatives are bacillaene (8), pseudomonic acid A (9), rhizoxin (10), oocidin A (11), psymbenin (12) and bryostatin (13).

A further peculiarity is the presence of more complex variations at the  $\beta$ -keto position called  $\beta$ -branching. The  $\beta$ -position is electrophilic in comparison with the nucleophilic  $\alpha$ -position and accordingly needs a nucleophilic alkyl source (Calderone 2008). The  $\beta$ -branch is less common than branching at the  $\alpha$ -position and is predominantly observable with *trans*-AT PKSs, but not exclusively (Piel 2010), (Calderone 2008). The  $\beta$ -branches are introduced by an interaction of five proteins: a free-standing ACP, a non-elongating KS ( $KS^0$ ), an HMG-CoA-synthase (HCS) homolog, an enoyl-CoA hydratase (ECH) homolog such as a dehydratase and a second ECH homolog serving as a decarboxylase (Figure 1-20). This HCS cassette can normally be found in modules without KR, DH or ER and often with tandem-ACPs (Haines et al. 2013). In rhizoxin (10) a Michael-type acetyl addition was described as a further  $\beta$ -modification. A "B" domain (for branching), instead of the HCS cassette, can be found (Figure 1-20). So far it remains open what the contribution of the KS or the B-domain are for this type of  $\beta$ -modification (Kusebauch et al. 2009), (Bretschneider et al. 2013).



**Figure 1-20: Possible  $\beta$ -branching mechanism modified after Till and Race 2014.**

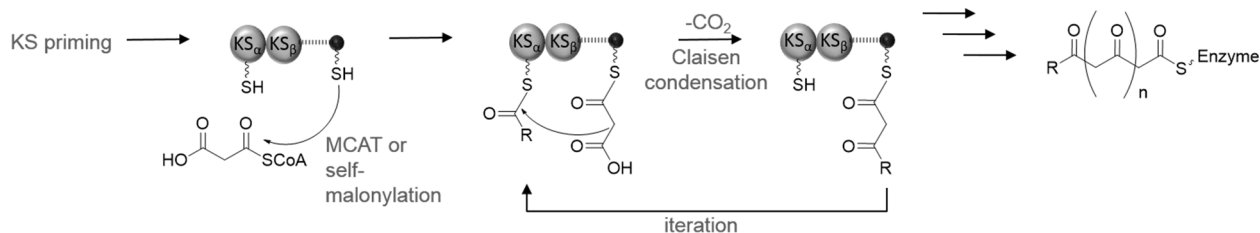
(A) Isoprenoid-like  $\beta$ -branching catalyzed by a HCS cassette. The black square indicates the carbon that will result in the branched methyl group. (B) Alternative  $\beta$ -branching in a Michael-type acetyl addition. Abbreviations: ACP, acyl carrier protein; AT, acyltransferase; KS, ketosynthase; HCS, 3-hydroxy-3-methylglutaryl-coenzyme A-synthase; ECH, enoyl-CoA hydratase; B, branching.

Very few examples of *cis/trans*-AT hybrid PKS clusters exist. The first one described is kirromycin (Weber et al. 2008), the second one is psymbirin (**12**) (Fisch et al. 2009) and enacyloxin gene cluster is a hybrid as well (Mahenthalingam et al. 2011). The assembly line underlying mupirocin (**9**) biosynthesis is an example of a diverse mixture of *trans*-AT PKS, FAS and iterative type I PKS that is able to synthesize products via two different pathways (Gao et al. 2014).

In addition to multi-modular type I PKSs, another modular PKS group exists that acts in an iterative fashion. Most iterative type I PKSs are described from filamentous fungi (for example lovastatin (**4**) nonaketide synthase from *Aspergillus terreus*), but bacterial representatives are known as well (e.g. PKS responsible for the enediynes core) (Xu, Qiao, and Tang 2012). They consist of a single set of domains that is used repeatedly like in type II and type III PKSs. The domain composition can vary and this decides about the structure of the product (Xu, Qiao, and Tang 2012).

### 1.2.2 Type II PKSs

Type II PKSs are almost exclusively employed by Gram-positive actinomycetes (Hertweck et al. 2007). Characteristic products are phenolic compounds (Staunton and Weissman 2001). Therapeutically important products are the antibiotic tetracycline and the anti-cancer agent daunorubicin (Zhang and Tang 2009). These products differ a lot from the previously described type I PKSs. They are not built up in a modular way, but possess only a single set of dissociated domains. Cyclizations occur after chain extension (Shen 2003). A "minimal PKS" is responsible for chain initiation and elongation. This minimal PKS consists of a  $KS_{\alpha}$  and  $KS_{\beta}$  (alternatively named chain length factor CLF) domain and an ACP (Figure 1-21) (Weissman 2009). All three genes can be normally found to be encoded beside each other (Hertweck et al. 2007). Every single domain is discrete and separable, but operates as a strong heterodimer (Austin and Noel 2003). The exclusively used extender unit is malonyl-CoA; starter moieties can vary. Further diversification with the help of tailoring enzymes occurs in these biosyntheses as well as in type I polyketide biosyntheses (Zhang and Tang 2009). The length of the final polyketide chain is determined by the CLF with the help of a protein cavity, but it is not known yet how a release of this fully elongated product takes place (Hertweck et al. 2007).

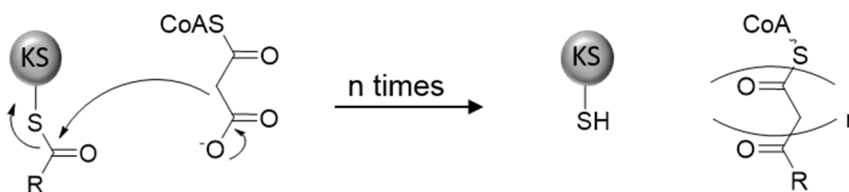


**Figure 1-21: General type II polyketide synthase (PKS) reaction scheme modified after Hertweck et al. 2007.**

The minimal PKS performs chain initiation and iterative decarboxylative condensation. Abbreviations:  $KS_{\alpha}$ , ketosynthase  $\alpha$ ;  $KS_{\beta}$ , ketosynthase  $\beta$ ; ACP, acyl carrier protein; MCAT, malonyl-CoA:ACP transferase.

### 1.2.3 Type III PKSs

Type III PKSs can be found in plants, bacteria and fungi (Weissman 2009). They are responsible for the synthesis of a large variety of products, usually mono- or bicyclic, such as chalcones, stilbenes, pyrones, and resorcinic lipids (Yu et al. 2012). This diversity of products is remarkable since these PKSs are homodimeric, relatively small in size (40-50 kDa) and do not employ ACPs (Shen 2003). The reaction takes place directly on the acyl-CoA substrates themselves without the need of ACP-bound intermediates (Figure 1-22) (Shen 2003). The biosynthesis occurs also iteratively that means that the homodimer fulfils priming, extension, and cyclization reaction (Yu et al. 2012). Intramolecular condensation and aromatization of the linear intermediates can be found very often after chain extension (Weissman 2009). The best studied example is the chalcone synthase (CHS) that was discovered in the 1970s (Austin and Noel 2003).



**Figure 1-22: General type III polyketide synthase (PKS) reaction mechanism after Weissman 2009.**

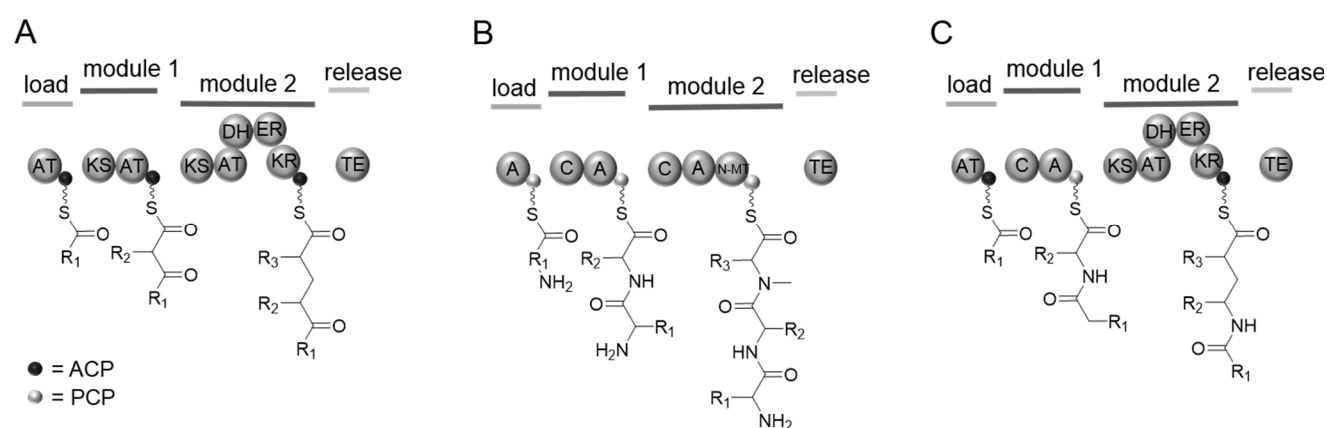
Chain elongation happens at a single multifunctional active site that performs all necessary steps directly on the acyl-CoA substrate to assemble a polyketide. Abbreviations: KS, ketosynthase; CoA, Coenzyme A (6).

## 1.3 Nonribosomal peptide synthetases

Nonribosomal peptides (NRP) are often highly active peptides not synthesized by the ribosome, but by nonribosomal peptide synthetases (NRPS). The large and well-known antibiotic family of penicillins, and the immunosuppressive cyclosporin are just two examples. NRPs have a size ranging from 2-23 aa with a complex structure (Caboche et al. 2010). Biosynthesis happens in the same logic as for modular PKSs (chapter 1.2.1), following a thio-template mechanism (Figure 1-23A/B). The different use of building blocks is the biggest contrast between PKS and NRPS. No acyl-CoA building blocks are employed by the NRPSs, but instead small adenosine triphosphate (ATP)-activated amino acids as adenosine monophosphate (AMP)-esters (Keating and Walsh 1999). The high complexity comes from more than 300 different amino acids that are used as building blocks, including many unnatural ones (Du and Lou 2010). NRPSs are modular assembly lines like multi-modular type I PKSs with a minimal module consisting of condensation (C), adenylation (A) and thiolation (T) domains responsible for one round of chain elongation (Fischbach and Walsh 2006). The A domain is responsible for building block selectivity and activates them as amino acyl adenylate with the consumption of ATP (Finking and Marahiel 2004). As for the

KS domain in PKS systems, the C domain performs the actual amide-bond formation (Donadio, Monciardini, and Sosio 2007). The growing peptide is tethered throughout the whole assembly line process to a PCP (peptidyl carrier protein) or thiolation (T) domain that gets activated by attachment of a Ppant group (Lai and Walsh 2006). Chain release is catalyzed by a TE domain leading either to a linear peptide or a lactam ring (Du and Lou 2010). Further domains for modification are possible as well; the most common are epimerization (E), methylation (M) and reduction (R) domains (Fischbach and Walsh 2006).

Since the biosynthetic logic between PKS and NRPS is strikingly similar and also the reaction mechanisms underlying for example C-C-bond formation/C-N-bond formation or chain release are the same, it is not surprising to find hybrid biosynthetic machineries comprised of both NPRS and PKS modules (Du and Lou 2010). Here the NRPS-bound growing intermediate gets further elongated from the PKS module or vice versa (Figure 1-23c). One example of such a hybrid is bacillaene (**8**), already mentioned several times in the first part of this introduction (Chen et al. 2006b). The antitumor agent leinamycin, immunosuppressant FK506 and the antitumor agent bleomycin are hybrids as well (Nikolouli and Mossialos 2012), (Sattely, Fischbach, and Walsh 2008).



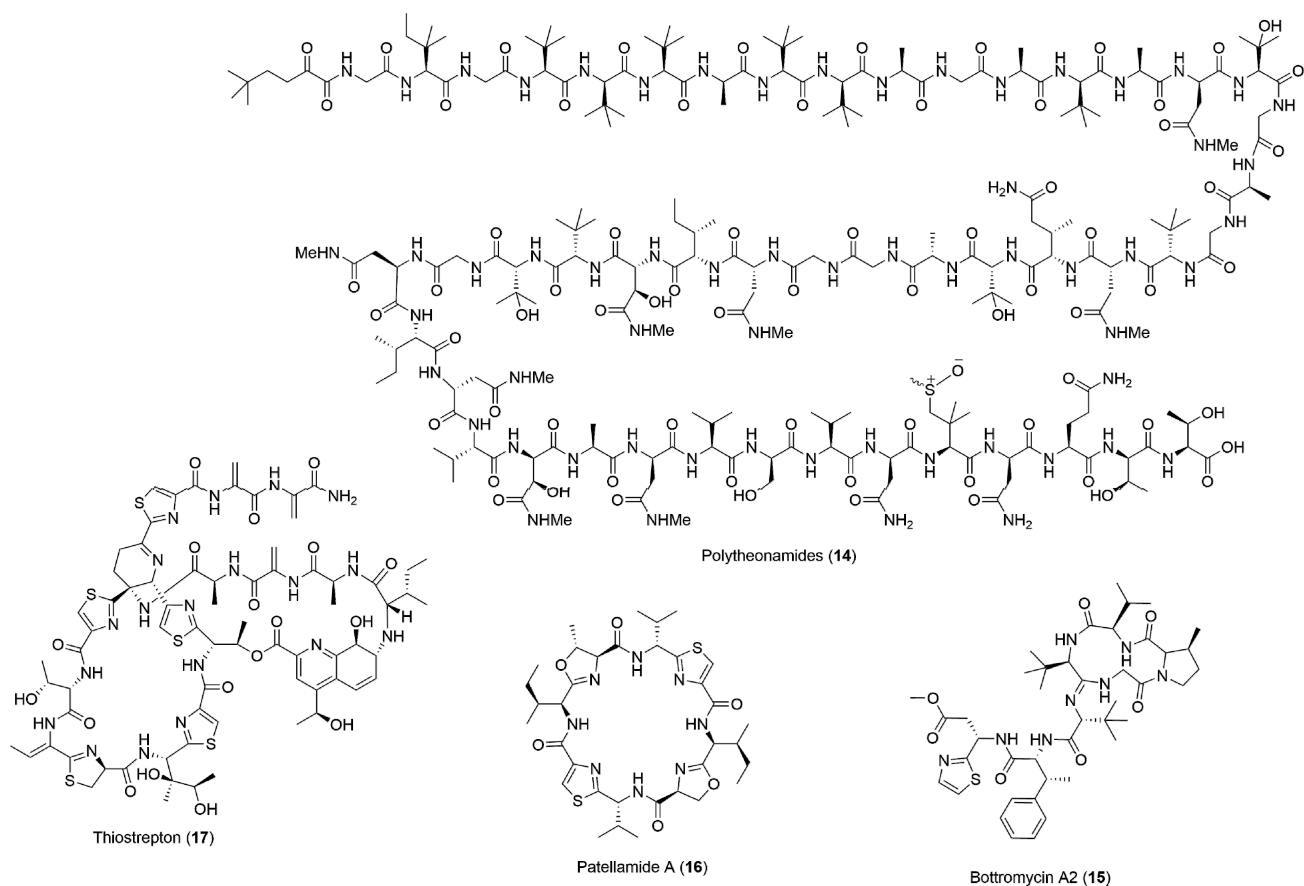
**Figure 1-23: Comparison of PKS, NRPS and PKS-NRPS-hybrids modified after Weissman and Müller 2008.**

(A) Representative modular *cis*-type I PKS reaction scheme. (B) Representative NRPS reaction scheme. (C) Representative PKS-NRPS hybrid reaction scheme. Abbreviations: PKS, polyketide synthase; NRPS, nonribosomal peptide synthetase; ACP, acyl carrier protein; PCP, peptidyl carrier protein; AT, acyltransferase; KS, ketosynthase; DH, dehydratase; ER, enoylreductase; TE, thioesterase; A, adenylation domain; C, condensation domain; N-MT, N-methyltransferase domain.

It becomes apparent that some hallmarks of NRPs are present in ribosomally synthesized peptides as well. The existence of D- or non-proteinogenic amino acids in a peptide historically indicated manufacture by an NRPS, however that does not directly lead to the conclusion of a nonribosomal origin any more. The distinction line between nonribosomally and ribosomally produced proteins is getting blurry (Walsh 2014).

## 1.4 Ribosomally synthesized and post-translationally modified peptides

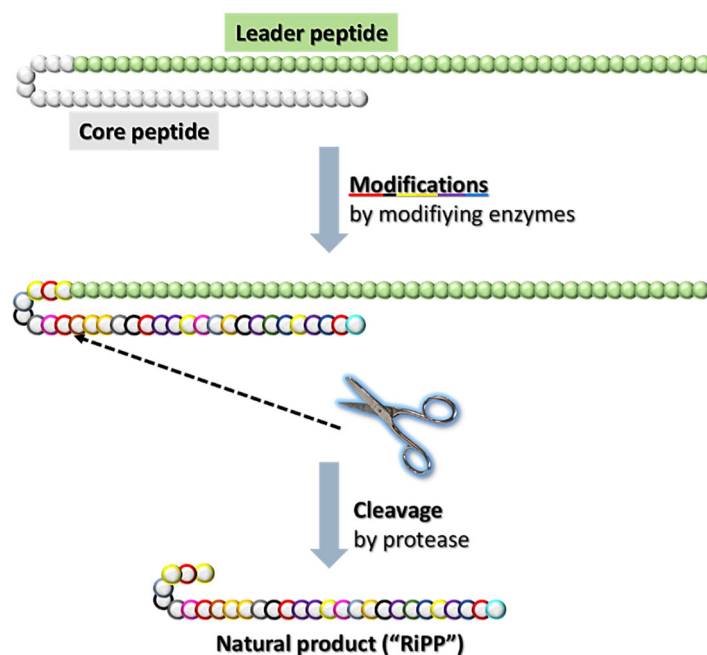
Peptidic natural products can not only arise from the action of an NRPS, but also by ribosomal synthesis as well. Representatives of this natural product class are polytheonamides (**14**), bottromycins (**15**), cyanobactins (for example patellamide (**16**)) and thiopeptides (for example thiostreptone (**17**)) (Figure 1-24) (Arnison et al. 2013).



**Figure 1-24: Structure of selected ribosomally produced and posttranslationally modified peptides (RiPPs).**

RiPP representatives are polytheonamides (14), Bottromycin A2 (15), Patellamide (16) and Thiostrepton (17). Polytheonamides A and B differ only in the configuration of the sulfoxide moiety in residue 44. The sulfoxide results a spontaneous oxidation during polytheonamide isolation.

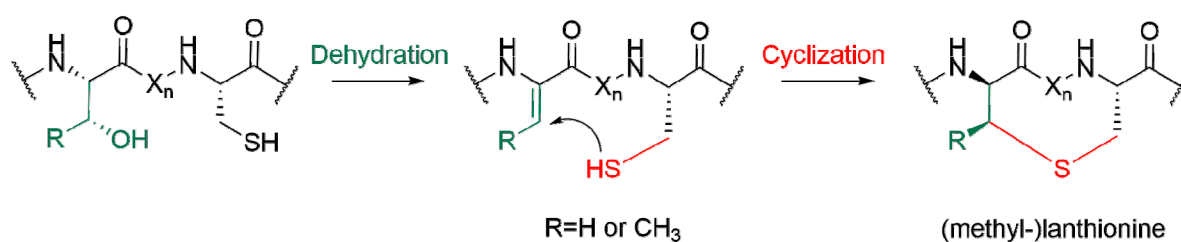
The characteristic feature of ribosomally synthesized and post-translationally modified peptides (RiPPs) is their modification after peptide synthesis on the ribosome. With the increasing use of sequencing techniques their massive appearance in all three domains of life has become more and more obvious. Irrespective diverse structures, all RiPPs have the same general underlying biosynthesis (Figure 1-25). First a ~ 20-110 aa-long precursor peptide, that is encoded by one gene, is synthesized by the ribosome. This precursor consists of a usual N-terminal leader peptide and a C-terminal core peptide that results in the later RiPP. In some cases a C-terminal leader peptide can be found as in the case of 15. The core peptide is extensively modified posttranslationally by various modifying enzymes that are often found to be common in many different RiPP biosynthetic pathways (Arnison et al. 2013). Typical modifications are C- and N-methylations, epimerizations, hydroxylations and dehydratations (Walsh 2014). Beside the aforementioned modifications, alteration at the Cys residue and macrocyclization occur frequently. The thiol group of cysteine can be converted to thioether, disulfide, sulfoxide and thiazol(in)es (Arnison et al. 2013). The vast number of posttranslational modifications lead to a large structural variety, improved target recognition and better metabolic and chemical stability (Yang and van der Donk 2013). The final step is a proteolytic cleavage conducted by a typically membrane-bound protease separating the modified core from the unmodified leader peptide resulting in the final RiPP product (Arnison et al. 2013).



**Figure 1-25: General scheme of ribosomally produced and posttranslationally modified peptide (RiPP) biosynthesis.**

A precursor peptide consisting of a leader and core peptide is modified by respective modifying enzymes. A protease cleaves off the modified core peptide and releases the final RiPP. Balls represent individual amino acids.

Until now, only little was known about the role of the leader peptide in RiPP biosynthesis. It is proposed to play a role in recognition of modifying enzymes, as secretion signal, for immunity and as support of peptide folding (Oman and van der Donk 2010), (Arnison et al. 2013). But many questions still remain open, for example about the timing and identification of the modifying enzymes. Moreover modifications without the need of a leader peptide are known (Dunbar and Mitchell 2013). More working groups start to focus on these open questions. Morinaka *et al.* studied the radical SAM epimerase PoyD as an example of novel post-translational modifications (Morinaka et al. 2014). Garg *et al.* could shed light onto the reaction sequence of the dehydratase NisB as an example for a long known enzyme class catalyzing so far not fully understood post-translational modification (Garg, Salazar-Ocampo, van der Donk 2013). In general the field of RiPP biosynthesis reveals a lot of unusual bio-chemistry that is just waiting to get explored. One textbook family of RiPPs are the lanthipeptides that comprise a large group. Members of this group having antibacterial activity are called lantibiotics. Their name stems from the lanthionine-bridge, an intramolecular thioether crosslink (Yang and van der Donk 2013). This crosslink is generated by a dehydration of a serine or threonine residue to dehydroalanine (Dha) or dehydrobutyrine (Dhb) respectively. Subsequently a cyclization takes place by a Michael-type addition by a cysteine residue onto the dehydroamino acid (Figure 1-26) (Arnison et al. 2013).



**Figure 1-26: Generation of a lanthionine bridge.**

Dehydration of a serine or threonine is followed by a thioether crosslink to generate the lanthionine bridge.



Lanthipeptides are known since the end of the 1920s and their ribosomal origin was already verified in 1988 (Arnison et al. 2013). In contrast are the polytheonamides (**14**). They were thought for a long time to be the largest NRPS products known (Inoue et al. 2010), but in 2012 Freeman *et al.* proved their ribosomal origin with up to 49 posttranslational modifications to yield the final RiPP (Freeman et al. 2012). These modifications contain C- and N-methylations, a dehydration, hydroxylations and epimerizations resulting in an almost perfect alteration of D- and L-amino acids. They are the first characterized members of a new RiPP family called the proteusins. This family is distinguished from others by their unusually long 5'-leader sequence that shows homology to nitrile hydratases and was proposed by Haft *et al.* from an *in silico* study (Haft, Basu, and Mitchell 2010). Polytheonamide A and B were first described in 1994 as cytotoxic polypeptides isolated from the marine sponge *Theonella swinhoei* (Hamada et al. 1994). They have a size of 48 amino acids and vary only in the conformation of the sulfoxide moiety at residue 44 (Hamada et al. 2005). The extremely high cytotoxicity is a result of the  $\beta$ -helical structure that is able to span the cell membrane (Figure 1-27) and with that disturb the ion equilibrium, leading to cell death (Iwamoto et al. 2010).

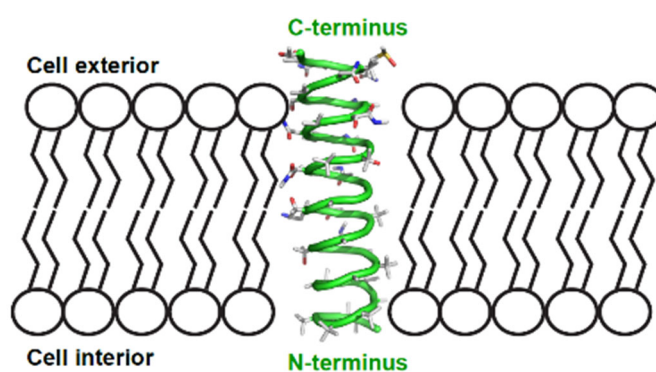


Figure 1-27: Polytheonamides (**14**) are able to form unimolecular ion channels through cell membranes, taken from Walsh 2014.

## Aim of this Thesis

---

Specialized (also called secondary) metabolites are an expansive class of molecules that contains many promising drug lead structures. Some of the class members, such as erythromycin (**1**), are known and studied for decades but others have only been discovered in recent years. Within this short time frame many peculiarities and unusual biochemistry have become apparent within this class. It therefore presents a treasure trove for drug discovery and tool development.

The work in this thesis focused on the release and cleavage mechanisms involved in the biosynthesis of the cytotoxic compounds pederin (**5**) and that of polytheonamides (**14**), a group of compounds that also exhibits cytotoxic activity. Chain release is a central biosynthetic step involved in generating final bioactive compounds. In addition, halogenation of the cytotoxic compound oocydin A (**11**) was studied in this thesis. Halogenation influences bioactivity and halogenations at vinyl moieties as observed here is typically difficult to obtain during chemical synthesis. Enzymes that are able to catalyze halogenation or chain release might be employed as a tool in secondary metabolite research or even unrelated fields.

**5** and **11** are synthesized by a *trans*-AT PKSs. *Trans*-AT PKSs do not encode their acyltransferase (AT) in-line but as a distinct protein and ATs act in *trans*, unlike the canonical modular type I PKS which are called *cis*-AT PKSs (Piel 2002). *Trans*-AT PKSs apply the same biosynthetic logic as *cis*-AT PKSs but domain types, their mode of usage, architecture and additional features differ. This demonstrates that not every principle valid for the well-studied class of *cis*-AT PKSs can be transferred directly to *trans*-AT PKS (Piel 2010). Gaining a further basic understanding in this new class of mega-enzyme-complexes is of notable interest as it could be shown that around one third of all modular type I PKSs are *trans*-AT PKSs (O'Brien et al. 2014).

One key aspect of *trans*-AT PKSs was the observation that several AT gene copies could be detected in a large number of underlying gene clusters. This seems contrary to the finding that one copy of this enzyme would be sufficient for serving the entire PKS (Cheng and Shen 2003), (Lopanik et al. 2008). Previous results from Dr. Jana Moldenhauer obtained for the bacillaene (**8**) biosynthetic gene cluster *baeX* suggested an impact in proofreading (Moldenhauer, Chen, and Borriss 2007). This hypothesis was supported by the findings of Dr. Katrin Zimmermann and Dr. Holger Niederkrüger that only one of two AT-copies (PedD) out of the pederin (**5**) pathway is able to transfer acyl-units. The other copy (PedC) does not perform this reaction. In this study the role of PedC as a potential proofreading enzyme was tested with the help of *in vitro* experiments. A potential application of PedC as tool in studying biosynthetic pathways was taken into account.

Another part of this work dealt with **11**, a very potent chlorinated macrolactone with a diverse range of activities, e.g., against oomycetes and cancer cells (Kigoshi and Hayakawa 2007). As no halogenase candidates were known from the characterization of the gene cluster (Matilla et al. 2012), the enzymes responsible for installing the unusual vinyl chloride-moiety were assigned and studied. Newly discovered halogenases might be used in chemoenzymatic synthesis, as the introduction of halogen atoms at a vinyl moiety is hard to achieve synthetically. Furthermore, the influence of chlorination on cytotoxicity of **11** was tested. Oocydin A (**11**) is a member of a large group of congeners in conjunction with the structurally similar and potential cytotoxic haterumalides and biselides (Kigoshi and Hayakawa 2007). The possibility of a similar biosynthetic origin was elucidated.

A family of secondary metabolites that is even younger than the *trans*-AT PKSs are the proteusins, which belong to the class of ribosomally synthesized and posttranslationally modified peptides (RiPPs). In RiPP biosynthesis, a protease is responsible for cleavage of the final bioactive compound from its leader peptide and thus enables the transport of the RiPP to its target. Little is currently known about protease substrate recognition on the amino acid sequence and protein levels. We therefore studied this process *in vitro* at the example of the protease PoyH. PoyH belongs to the *poy* gene cluster that is responsible for the biosynthesis of the proteusin polytheonamides (**14**). The

protease PoyH is unique in the field of RiPPs as it is not membrane-bound. While this *in vitro* study leads to a more general understanding of substrate recognition by PoyH, a potential application of PoyH as a more general protease for studying other RiPP pathways or general molecular biology approaches was tested.

## Chapter 2. Polyketide Proofreading by an Acyltransferase-Like Enzyme

---

Katja Jensen,<sup>1</sup> Holger Niederkrüger,<sup>1</sup> Katrin Zimmermann,<sup>1</sup> Anna L. Vagstad,<sup>2</sup> Jana Moldenhauer,<sup>1</sup> Nicole Brendel,<sup>3</sup> Sarah Frank,<sup>1</sup> Petra Pöplau,<sup>1</sup> Christoph Kohlhaas,<sup>1</sup> Craig A. Townsend,<sup>2</sup> Marco Oldiges,<sup>4</sup> Christian Hertweck,<sup>3</sup> Jörn Piel<sup>1</sup>

<sup>1</sup>Kekulé Institute of Organic Chemistry and Biochemistry, University of Bonn, 53121 Bonn, Germany; <sup>2</sup>Department of Chemistry, Johns Hopkins University, Baltimore, MD 21218, USA; <sup>3</sup>Leibniz Institute for Natural Product Research and Infection Biology, HKI, 07745 Jena, Germany; <sup>4</sup>Institute of Biotechnology 2, Forschungszentrum Jülich, 52425 Jülich, Germany

K.J., K.Z., H.N., A.L.V., N.B., and M.O. expressed proteins, J.M. constructed the *B. amyloliquefaciens* mutants and analyzed metabolites of the bacillaene pathway, K.Z. and H.N. conducted the acyl transfer assays, K.J. conducted acyl-SNAC hydrolysis assays and kinetics, A.L.V. conducted acyl-ACP hydrolysis assays, S.F., P.P., and C.K. synthesized SNAC derivatives, J.P. performed the phylogenetic analysis, all authors designed the experiments, analyzed the data, and wrote the manuscript.

This work was published in *Chemistry & Biology*, 19, 329–339, March 23, 2012 and highlighted by (Kwan and Schmidt 2012).

### 2.1 Summary

*Trans*-acyltransferase polyketide synthases (*trans*-AT PKSs) are an important group of bacterial enzymes producing bioactive complex polyketides. One difference to textbook PKSs is the presence of one or more free-standing AT-like enzymes. While one homolog loads the PKS with malonyl units, the function of the second copy (AT2) was unknown. We studied the two ATs PedC and PedD involved in pederin biosynthesis in an uncultivated symbiont. PedD displayed malonyl-, but not acetyltransferase activity towards various acyl carrier proteins (ACPs). In contrast, the AT2 PedC efficiently hydrolyzed acyl units bound to *N*-acetylcysteamine or ACP. It accepted substrates with various acyl chain lengths and functionalizations, but did not cleave malonyl-ACP. These data are consistent with a role of PedC in PKS proofreading, suggesting a similar function for the other members of the AT2 group and providing new strategies for polyketide titer improvement and biosynthetic investigations.

### 2.2 Introduction

Bacterial complex polyketides are biosynthesized on giant multifunctional enzymatic assembly lines called modular polyketide synthases (PKSs) (Hertweck 2009), (Staunton and Weissman 2001), (Piel 2010). These enzymes contain numerous catalytic domains, the order of which typically determines the structure of the polyketide chain. In addition, multiple acyl carrier protein (ACP) domains serve as anchors that covalently bind intermediates of increasing chain length via thioester linkages. After full elongation, a C-terminal thioesterase (TE) domain usually cleaves the thioester bond by hydrolysis or lactonization to release the polyketide (Du and Lou 2010). In textbook PKSs, all biosynthetic components needed to generate the polyketide core structure are domains incorporated in the megaenzyme. In contrast, the recently discovered group of *trans*-acyltransferase (*trans*-AT) PKSs diverge from

this architecture in using multiple covalently and non-covalently joined components (Piel 2010). *Trans*-AT PKSs produce diverse bioactive polyketides, including the antibiotic bacillaene (**8**) from bacilli (Butcher et al. 2007), (Moldenhauer, Chen, and Borriss 2007), (Chen et al. 2006b) and the cytotoxins pederin (**5**) (Piel 2002), psymberin/irciniastatin A (**12**) (Cichewicz, Valeriote, and Crews 2004), (Fisch et al. 2009), (Pettit et al. 2004), and rhizoxin (**10**) (Iwasaki, Namikoshi, and Kobayashi 1986), (Partida-Martinez and Hertweck 2005) from various endosymbiotic bacteria.

*Trans*-acting components identified in these poorly understood systems are ATs needed for loading the PKS with biosynthetic building blocks (Cheng and Shen 2003), (Wong and Khosla 2011), enoylreductases (ER) for reduction of  $\alpha,\beta$ -unsaturated intermediates (Bumpus et al. 2008), and enzymes to convert  $\beta$ -keto groups to carbon branches (Calderone et al. 2008). The polyketide products are almost (Musiol et al. 2011) exclusively constructed from malonyl units, and a single, free-standing AT resembling those from bacterial fatty acid synthases (here termed AT1) has been shown to be sufficient for loading multiple *trans*-AT PKS modules with these units (Cheng and Shen 2003), (Lopanik et al. 2008). It is therefore peculiar that more than half of all characterized *trans*-AT PKS systems, including those of **5**, **8**, and **10**, contain another AT homolog (AT2). These enzymes are either freestanding or fused with the malonyl-loading AT and/or the *trans*-ER. Except for the fact that their inactivation can result in reduced polyketide titers (Lopanik et al. 2008), nothing is known about the function of these ubiquitous enzymes.

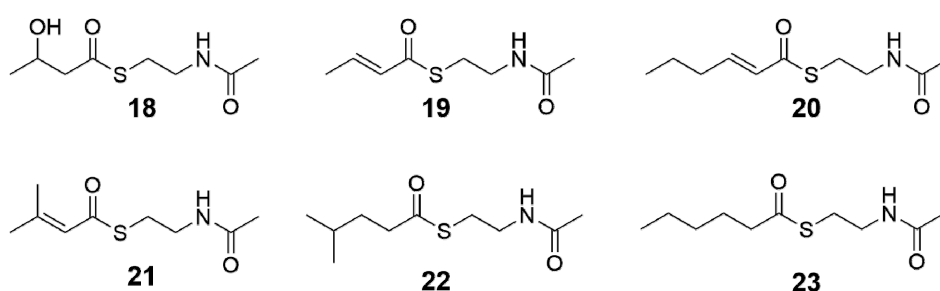
Much of the current knowledge about *trans*-AT PKS functionality stems from studies on the polyketide antibiotic bacillaene (**8**) in *Bacillus amyloliquefaciens* and *Bacillus subtilis* (Chen et al. 2009), (Bumpus et al. 2008), (Calderone et al. 2008), (Calderone et al. 2006), (Dorrestein and Kelleher 2006), (Moldenhauer, Chen, and Borriss 2007), (Moldenhauer et al. 2010). Functional dissection of the bacillaene (*bae*) PKS was greatly facilitated by the discovery that mutants with an inactivated TE domain produce polyketides of virtually all elongation stages (Moldenhauer, Chen, and Borriss 2007), (Moldenhauer et al. 2010). This phenomenon, which has with one exception (Yu and Floss 1999) not been reported for textbook (*cis*-AT) PKSs, suggests the existence of an as-yet uncharacterized proofreading mechanism that releases stalled intermediates from the enzyme. Such truncated products were also observed for the rhizoxin (**10**) (Kusebauch et al. 2009) and, for some elongation stages, the mupirocin (Wu et al. 2008) pathways, indicating that intermediate hydrolysis is a more general phenomenon for *trans*-AT PKSs. Here, we report genetic, biochemical and chemical studies of a member of the AT2 group, identifying it as a thioester hydrolase for structurally diverse *N*-acetylcysteamine- or ACP-bound acyl units. Our data suggest that members of the AT2 group are not involved in acylation of PKS modules but act as proofreading factors that reactivate PKS modules by releasing stalled acyl units.

## 2.3 Results

### 2.3.1 Study of BaeB, a hydrolase candidate from bacillaene biosynthesis

Suspecting that the release factor is clustered with other PKS genes, we first focused on the previously uncharacterized gene *baeB* of the bacillaene (**8**) pathway. Its translated product resembles thioester hydrolases, such as glyoxalase II, and therefore represented a good candidate for intermediate release. To investigate its function *in vivo*, we constructed the *B. amyloliquefaciens* mutant JM157 lacking *baeB*. Analysis of the extracts by liquid chromatography-high resolution mass spectrometry (LC-HRMS) showed that **1** was produced at significantly reduced (68-87%) titers as compared to the wild-type strain FZB42 (Figure 2-5/Figure S1), indicating that BaeB does not catalyze an essential biosynthetic step. Since no intermediates are observed in strains harboring an intact TE domain, the influence of BaeB on thioester hydrolysis could not be examined in this mutant. Therefore, we created the double mutant JM157-54 that lacked both the TE and BaeB to allow for direct metabolic comparison with the

TE single mutant JM54-2. While JM54-2 produced the full series of intermediates, only traces of two compounds lacking three and one elongation units, respectively, were detected in extracts of the double mutant. This result may suggest that BaeB is the off-loading factor. However, we noticed that for unknown reasons all *baeB* deletion mutants exhibited strongly diminished growth. Thus, although relative polyketide yields from both TE deletion strains had been normalized to an internal standard (the PKS-independent compound bacillibactin), we could not exclude that the decreased biosynthetic activity of the *bae* pathway was an indirect effect of the altered physiology. This issue was addressed by constructing a pHis8-based expression plasmid containing *baeB* for *in vitro* enzymatic studies. Expression resulted in soluble, N-terminally octahistidyl-tagged BaeB that was incubated with a range of *N*-acetylcysteamine (SNAC) conjugates of short-chain carboxylic acids (**18–22**, Figure 2-1). SNAC thioesters are known to be accepted as surrogates of ACP-bound intermediates by various PKS components (Cane and Khosla 2002), (Pohl et al. 1998), (Vergnolle et al. 2011). To detect thioester cleavage, hydrolyzed free thiols were monitored in a continuous assay using Ellman's reagent (DTNB, 5,5'-dithio-bis(2-nitrobenzoic acid)), which reacts with thiols to the yellow cleavage product 2-nitro-5-thiobenzoate. However, no conversion was detected for any tested compound or condition.

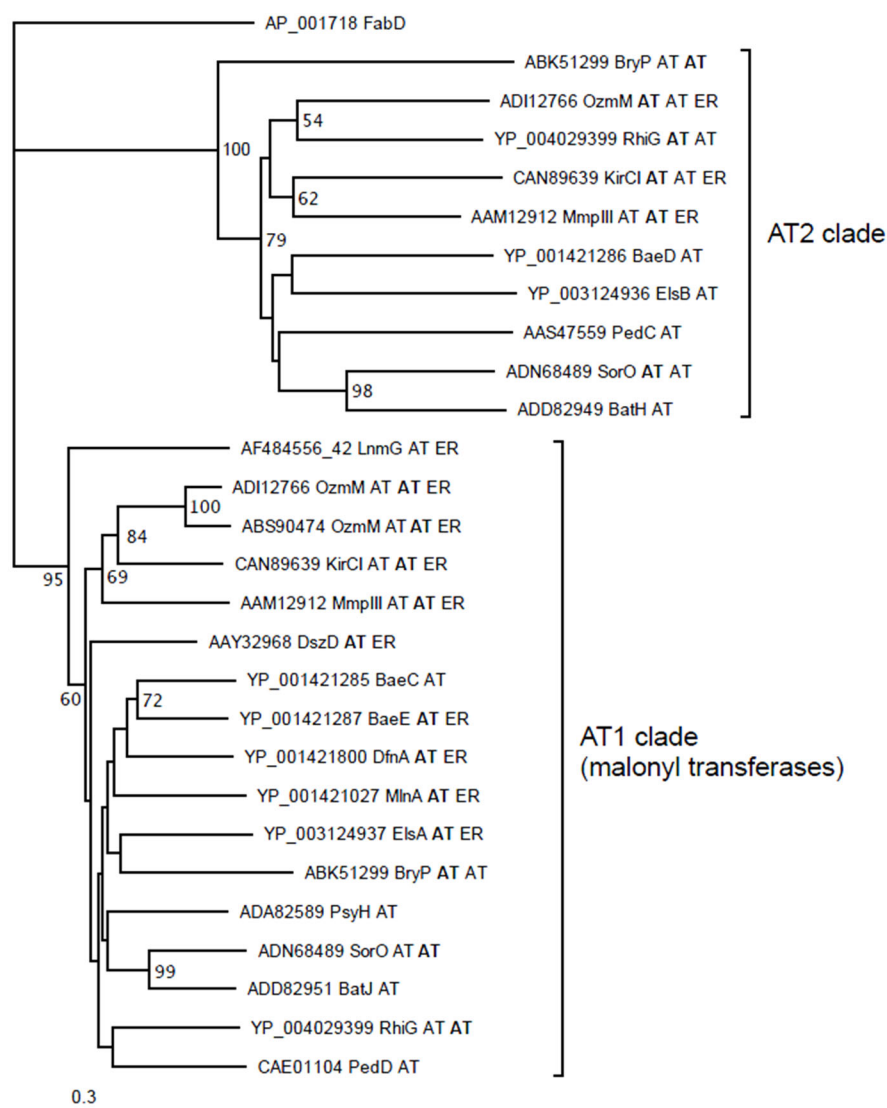


**Figure 2-1: SNAC derivatives tested in the hydrolysis.**

For the preparation of compounds **18–22** see Experimental Procedures. See also Figure 2-5/Figure S1.

### 2.3.2 Differential roles of the pederin AT-like enzymes in malonyl transfer

During these studies it was found that TE deletion in the rhizoxin (**10**) producer *Burkholderia rhizoxinica* B1 resulted in a similar release of intermediates as known for bacillaene (Kusebauch et al. 2009). The fact that neither the rhizoxin (*rhi*) PKS cluster nor the entire genome (Lackner et al. 2011b), (Lackner et al. 2011a) encoded a homolog of BaeB, combined with the negative results of the BaeB hydrolysis assay, suggested that a different enzyme is responsible for thioester cleavage. The AT2 present in many *trans*-AT systems was a reasonable candidate, because homologs also occur in the bacillaene (**8**) and rhizoxin (**10**) pathways. Moreover, thioester hydrolysis conforms to the mechanism of ATs, which involves transfer of a thioester acyl moiety to a serine residue followed by nucleophilic displacement (Wong and Khosla 2011). Since the bacillaene AT2 from *B. amyloliquefaciens* could not be expressed as a soluble protein, and the presence of five AT-like genes within this strain complicates any *in vivo* knock-out studies, we selected the AT homologs PedC and PedD of the pederin (**5**) pathway from a *Pseudomonas* sp. symbiont of *Paederus* spp. beetles (Piel 2002) for further studies. These enzymes offered the distinct advantage of being among the very few known ATs that are not fused to additional domains, thus simplifying functional analysis.

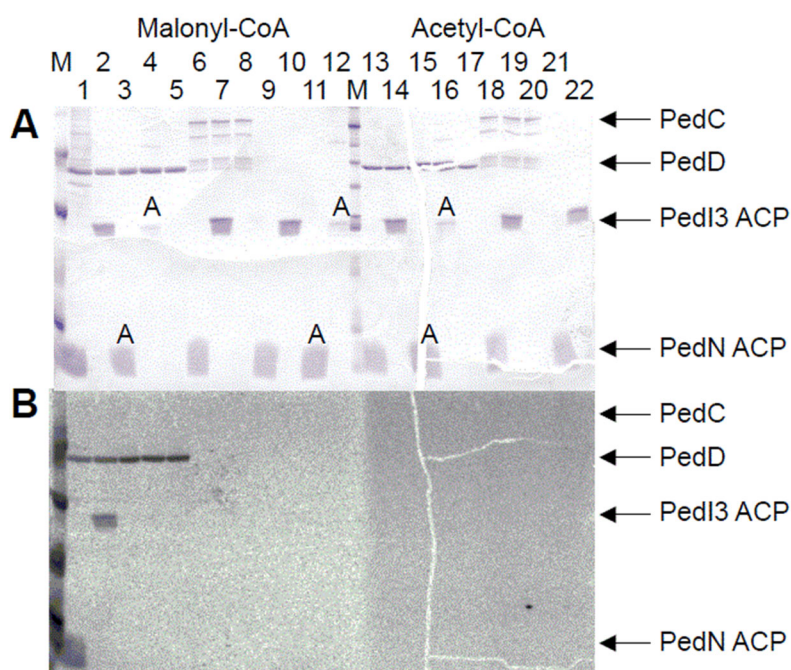


**Figure 2-2: Phylogram of AT-like enzymes from various *trans*-AT PKS pathways.**

Tip labels consist of accession numbers, protein name, and domain architecture. For multidomain proteins, the relevant domain is shown in bold. Sequences from the following pathways were used: Bae, bacillaene (*B. amyloliquefaciens*); Bat, batumin/kalimantacin (*Pseudomonas fluorescens*); Bry, bryostatin ("*Candidatus Endobugula sertula*"); Dfn, difficidin (*B. amyloliquefaciens*); Dsz, disorazol (*Sorangium cellulosum*); Els, elansolid (*Chitinophaga pinensis*); Kir, kirromycin (*Streptomyces collinus*); Lnm, leinamycin (*Streptomyces atrolivaceus*); Mmp, mupirocin (*Pseudomonas fluorescens*); Ozm, oxazolomycin (ADI12766 is from *Streptomyces bingchenggensis* and ABS90474 from *Streptomyces albus*); Psy, psymberin (uncultivated symbiont of sponge *Psammocinia aff. bulbosa*); Rhi, rhizoxin (*Burkholderia rhizoxina*); Sor, soraphen (*Sorangium cellulosum*). Outgroup is the *E. coli* AT from fatty acid biosynthesis (FabD).

In a phylogenetic analysis (Figure 2-2), PedD grouped with members of the AT1 clade, including the previously characterized malonylating ATs LnmG of the leinamycin route (Cheng and Shen 2003) and BryP (N-terminal domain) of the bryostatin pathway (Lopaniak et al. 2008). PedC fell into the AT2 group. To functionally test the roles of *ped* ATs suggested by the phylogenetic data, both enzymes were produced for *in vitro* analyses. Initial difficulties to obtain soluble PedC by a range of expression strategies were overcome by expressing the protein with an N-terminally fused maltose-binding protein or a C-terminal Strep-tag in the presence of the chaperones GroES and GroEL, which yielded low amounts of soluble enzyme. PedC and PedD were then subjected to a transfer assay employing [2-<sup>14</sup>C]-malonyl- or [1-<sup>14</sup>C]-acetyl-CoA as test substrate with an ACP as acceptor. To investigate whether the two ATs might be specific for different types of ACPs, the following proteins were expressed: the free-standing variant PedN (Piel et al. 2004a), the integrated tandem domain PedI-ACP3 from module 3 of the pederin PKS (Piel

et al. 2004a), and the integrated monodomain ACPs PsyA-ACP3 and PsyD-ACP1 from elongation modules 3 and 4 of the psymberin PKS (Figure 2-3 and Figure 2-6/Figure S2) (Fisch et al. 2009). PedN is expected to be involved in the generation of the exomethylene carbon in pederin, PedI3 and PsyA3 to participate in the chain elongation process, and PsyD1 to belong to a non-elongating module with an inactive ketosynthase domain. ACPs were expressed in the *apo*- as well as in the *holo*-form carrying the 4'-phosphopantetheinyl moiety that serves as anchor for the acyl units. After incubating the ATs, ACPs, and radiolabeled acyl-CoA species in different combinations, labeling of proteins was monitored by electrophoretic separation and autoradiography. The analysis revealed that PedD self-loads the malonyl radiolabel and efficiently transfers it to all types of *holo*-ACPs, but not to their *apo* forms. In contrast, ACPs remained unlabeled in the PedC assay. Using [1-<sup>14</sup>C]-acetyl-CoA as test substrate, neither AT or ACP was labeled. These results confirmed our *in silico* prediction that PedD is the AT responsible for substrate loading and identified PedC as a candidate for intermediate release.



**Figure 2-3: Acyltransferase assay using the pederin ATs.**

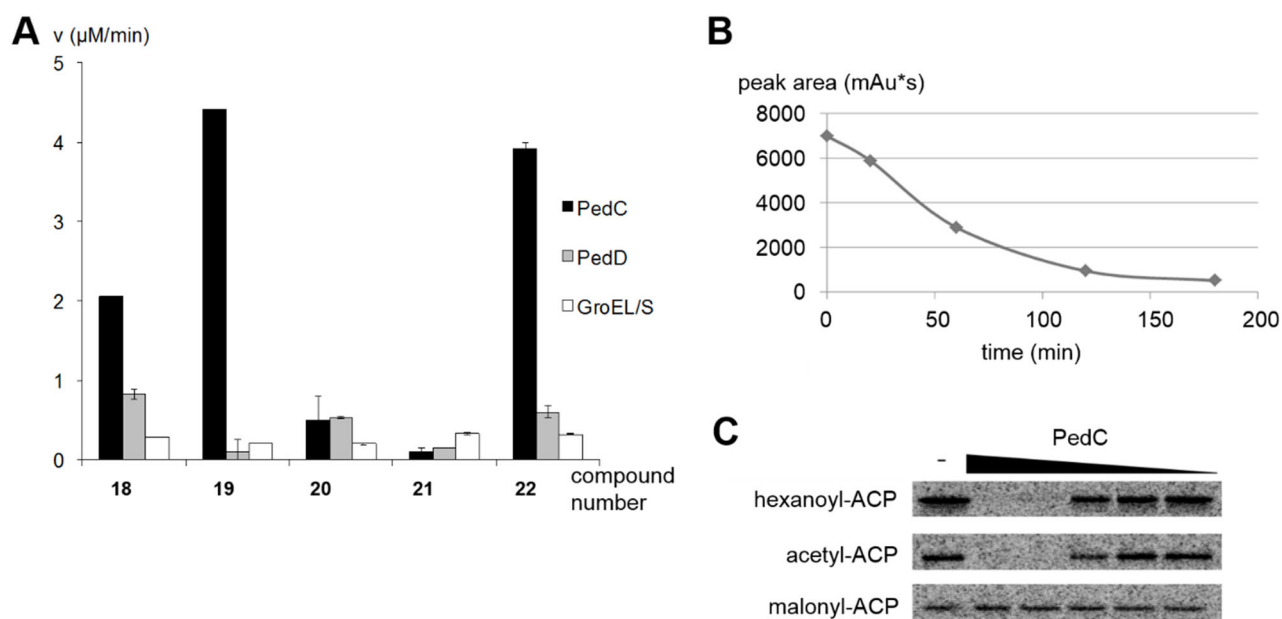
(A) SDS-PAGE analysis of incubations containing PedD (lanes 1–5, 13–17) or PedC (lanes 6–8, 18–20), radiolabeled malonyl-CoA (lanes 1–12) or acetyl-CoA (lanes 13–21), and the ACPs of PedI3 (tandem ACP) or PedN. ACPs were in the *holo* form or the *apo* form, the latter indicated by an A. Protein markers are indicated with an M. (B) Autoradiogram of the same gel. See also Figure 2-6/Figure S2.

### 2.3.3 The AT2 PedC hydrolyzes acyl N-acetylcysteamine thioesters

We next investigated hydrolysis of short-chain intermediate mimics by PedC and PedD using the SNAC derivatives **18–22** and Ellman's reagent. In agreement with its function as a malonyl transferase, PedD did not have significant hydrolysis activity towards any of the thioesters tested (Figure 2-4A). In contrast, assays with the AT2 PedC showed conversion within minutes for reactions with the acyl donors **18**, **19**, and **22**. On the other hand, hydrolysis was negligible for **20** and **21**. We also tested hydrolysis of the hexanoyl thioester **23** by removing aliquots at different time points and quantifying the SNAC derivative after high-pressure liquid chromatography (HPLC) separation. This analysis revealed that **23** is hydrolyzed by PedC, albeit at a lower rate than **18**, **19**, and **22** (Figure 2-4B). Expression levels of PedC were very low, and even a 20 L culture provided just sufficient enzyme amounts to determine kinetic parameters for substrate **18** ( $K_M$  of 88.48 mM and a  $k_{cat}/K_M$  of  $14.41 \text{ M}^{-1} \text{ s}^{-1}$ ) (Figure 2-8/Figure S4). These should be regarded as preliminary, since SNAC derivatives instead of PKS-bound acyl units were used and the true acyl



substrates could be fully elongated intermediates. All incubations with PedC also contained the chaperones GroEL/ES, from which the enzyme could not be separated (Figure 2-7/Figure S3). Prior to kinetic characterization, background hydrolysis activity assays were therefore conducted using purified protein from parallel expressions containing the chaperones with and without PedC coexpression. The full range of substrate concentrations used for kinetic characterization were tested with both enzyme purifications, where minimal background hydrolysis was detected in the protein preparation lacking PedC. These control experiments link the SNAC hydrolysis activity solely to PedC. These results thus support a role for PedC as a hydrolase for acyl-SNAC thioesters carrying various functional groups that typically occur during polyketide biosynthesis, i.e.,  $\beta$ -hydroxyl groups (**18**),  $\alpha,\beta$ -double bonds (**19**), and  $\alpha,\beta$ -reduced units (**22,23**). It should be noted that the  $\beta$ -keto variant could not be tested, since it was spontaneously hydrolyzed in the assay medium. Using the same assay, we also studied a PedC homolog from the rhizoxin (**10**) pathway. In this PKS system, the AT2 is fused to a C-terminal AT1 on the bidomain protein RhiG (Partida-Martinez and Hertweck 2006). To uncouple both activities, three protein variants were expressed, in which none or either one of the AT domains carried an inactivating Ser-to-Ala mutation at the acyl-binding site. However, neither protein displayed activity with any of the SNAC derivatives **18-22**. One reason for this unexpectedly differential behavior of the *ped* and *rhi* AT2s might be that inactive forms of RhiG were obtained during protein expression or purification. Since SNAC mimics were used instead of acyl-ACPs, we could not exclude the possibility that the observed PedC activity was an artifact unrelated to PKS offloading.



**Figure 2-4: Hydrolysis assays using acyl-SNAC**

(A) Reaction rates observed for various SNAC derivatives. Error bars specify standard deviations. (B) Consumption of hexanoyl-SNAC by PedC. After treatment of the thioester with PedC, aliquots were removed and separated by HPLC. The diagram shows the amount of hexanoyl-SNAC determined by manual integration at 232 nm at different reaction time points. (C) Phosphor image of PksA acyl-ACP hydrolysis assays with PedC. The signal is emitted from radiolabeled acyl-ACP after electrophoretic separation of proteins. Acyl-ACP (10 mM) was treated for 10 min with 0 (minus sign indicates negative control), 5, 0.83, 0.14, 0.023, or 0.005 mM PedC. See also Figure 2-7/S3 and Figure 2-8/S4.

#### 2.3.4 The AT2 PedC releases ACP-bound acyl units

This issue was addressed in a further series of experiments employing ACP-bound  $^{14}\text{C}$ -radiolabelled acyl variants. Based on the ability of PedC to hydrolyze hexanoyl-SNAC (**23**), [1- $^{14}\text{C}$ ]-hexanoyl-ACP was selected as an initial test substrate. Hexanoyl is the native starter unit used in aflatoxin biosynthesis by *Aspergillus* spp. fungi (Minto and

Townsend 1997), thus allowing us to test an extended acyl unit with its cognate ACP. Moreover, since this ACP is derived from a fungal iterative PKS, we hoped to obtain insights into whether PedC would act on enzymes other than *trans*-AT PKSs. To generate the hexanoyl-loaded protein, the N-terminally His<sub>6</sub>-tagged ACP mono-domain from the iterative megasynthase PksA of *Aspergillus parasiticus* was expressed in its *apo* form and loaded using [1-<sup>14</sup>C]-hexanoyl-CoA by the phosphopantetheinyl transferase Svp (Crawford et al. 2006). In an analogous way, we also prepared [1-<sup>14</sup>C]-acetyl- and [2-<sup>14</sup>C]-malonyl-ACP as further test substrates. These acyl-ACP variants were individually incubated with a serial dilution of PedC in a fixed time point assay. Subsequently, aliquots from incubations were separated by polyacrylamide electrophoresis, and radiolabeled proteins were visualized by phosphorimaging (Figure 2-4C). The data clearly show that hexanoyl- and acetyl-ACP are completely hydrolyzed within 10 min at higher PedC concentrations, while malonyl-ACP remained intact.

## 2.4 Discussion

Our studies on SNAC and ACP derivatives demonstrate that PedC hydrolyzes structurally diverse thioesters. Acyl units accepted as substrates include completely reduced,  $\alpha,\beta$ -unsaturated,  $\beta$ -hydroxylated and branched moieties of chain lengths between two and at least six carbon atoms. None of the tested compounds were cleaved by PedD. Furthermore, PedD, but not PedC, exhibited malonyl transfer activity. These data are consistent with differential roles for the two AT homologs in pederin biosynthesis, during which PedD loads all ACPs with malonyl units, and PedC acts as a proofreading enzyme that removes stalled acyl units from blocked modules. In agreement, PedC did not cleave malonyl-ACP, which is an essential precursor for polyketide biosynthesis. As PedC does not display detectable AT activity, we propose to rename this enzyme acyl hydrolase (AH). PedC homologs occur in many *trans*-AT PKS systems (Piel 2010), suggesting that AH-mediated proofreading is widespread in these pathways.

To date, AT-like editing enzymes are only known from few type II PKSs, where they participate in the selection of non-acetyl starter units (Tang, Koppisch, and Khosla 2004). A different type of editing enzyme is known for *cis*-AT PKSs (Heathcote and Leadlay 2001) and nonribosomal peptide synthetases (Schwarzer et al. 2002), (Yeh et al. 2004). Here, free-standing (type II) TEs act as release factors for the removal of stalled aberrant units. These units are likely short-chain (amino-)acyl moieties that can arise from premature decarboxylation of PKS extenders (Heathcote and Leadlay 2001) or from misloading of carrier proteins by either phosphopantetheinyl transferases (Schwarzer et al. 2002) or building block-selecting domains (Yeh et al. 2004). In agreement, almost all examined type II TEs preferred short-chain reduced acyl units in *in vitro* assays, while more extended moieties and those carrying hydroxyl or double bond functionality were poor or not substrates (Schwarzer et al. 2002), (Yeh et al. 2004), (Heathcote and Leadlay 2001), (Koglin et al. 2008). This selectivity appears in contrast to PedC, which cleaves acyl donors ranging from two (acyl-ACP) to six (**22** and hexanoyl-ACP) carbon atoms and exhibits tolerance for various functional groups occurring in polyketide biosynthesis. The data therefore suggest that PedC and other AHs might not only restore PKS modules blocked with decarboxylated monomers, but also play a role in the release of stalled oligoketides. In agreement, all cases of intermediate release involving *trans*-AT PKSs, i.e., for bacillaene (Moldenhauer, Chen, and Borriss 2007), rhizoxin (Kusebauch et al. 2009), and mupirocin (Wu et al. 2008), have been reported from assembly lines containing AH-type enzymes. In *trans*-AT systems, this function could be of particular significance, since phylogenetic (Irschik et al. 2010), (Nguyen et al. 2008), (Teta et al. 2010) and preliminary unpublished functional data indicate that KSs are highly selective for the  $\alpha,\beta$ -functionalization of incoming substrates. Thus, if  $\beta$ -processing is faulty, further elongation of incorrectly processed intermediates would not occur, resulting in biosynthetic blocks. This issue should be of less importance for *cis*-AT PKSs, in which KS domains are rather promiscuous (Chen et al. 2006a), (Menzella, Carney, and Santi 2007). Nevertheless, hydrolysis of long-chain intermediates was also reported for a small number of *cis*-AT systems, an important example being rifamycin biosynthesis (Yu and Floss 1999). In

this case, however, functional data argued against an involvement of the type II TE, leaving the underlying biochemistry unclear (Floss and Yu 1999).

In conclusion, our data show that PedC, a member of a clade of previously uncharacterized AT-like enzymes ubiquitously encoded in *trans*-AT PKS gene clusters, is a thioester hydrolase that could play a key role in biosynthetic proofreading. Such a function provides interesting perspectives for engineering and studying PKS pathways. Addition of AHs to *trans*-AT PKS pathways lacking these enzymes could result in enhanced polyketide yields. With PKSs carrying engineered loss-of-function mutations in their TE domains, AHs could be effective tools to study entire PKS pathways by observing released intermediates in a similar way as previously performed for bacillaene and rhizoxin biosynthesis. Such potential applications are currently under investigation in our laboratory.

## 2.5 Significance

*Trans*-AT PKSs are an important, but still poorly understood group of biosynthetic enzymes involved in the production of bioactive bacterial polyketides. About half of all known *trans*-AT PKS gene clusters encode, in addition to the *trans*-acting AT, another AT-like protein with previously unknown function. We have identified for the first time an enzymatic function for a member of this group. Our data show that the pederin homolog from an uncultivated beetle symbiont hydrolyzes SNAC- and ACP-bound acyl units with various functional groups and chain lengths. In contrast, neither the malonyl-transferring AT nor the bacillaene-specific enzyme BaeB with as-yet unknown function displayed measurable hydrolase activity *in vitro*. These results suggest the existence of a novel group of editing hydrolases, here termed AHs, that remove aberrant or stalled acyl units from blocked *trans*-AT PKS modules and might be responsible for intermediate release observed in several polyketide routes. The pederin AH or its homologs could be useful tools to study PKS functionality or, as inactivation of AHs has been shown to reduce polyketide titers (Lopanik et al. 2008), to increase compound yields by adding AH genes to routes lacking them. In addition to attributing function to a core component of *trans*-AT PKSs, this study represents one of the few examples (Lopanik et al. 2008), (Zimmermann et al. 2009) that provide biochemical insights into polyketide pathways of uncultivated bacterial symbionts associated with invertebrates. As *trans*-AT PKSs are a major group of enzymes responsible for natural product biosynthesis in these organisms (Piel 2009), knowledge about their enzymology is an important prerequisite to develop heterologous expression systems for rare animal-derived polyketides.

## 2.6 Experimental procedure

### 2.6.1 General

[1-<sup>14</sup>C]-hexanoyl-CoA, [1-<sup>14</sup>C]-acetyl-CoA, and [2-<sup>14</sup>C]-malonyl-CoA were purchased from American Radiolabeled Chemicals, Perkin-Elmer or NEN Radiochemicals, and Morovek respectively. Unlabeled acyl-CoAs were purchased from Sigma.

### 2.6.2 Bacterial strains and culture conditions

*E. coli* BL21(DE3) (Invitrogen) or *E. coli* Rosetta-gami 2 (DE3) pLysS (Novagen) served as the host strain for protein expression experiments. *E. coli* strain XL1 blue (Stratagen) served as host for routine subcloning. *E. coli* strains were grown in Luria-Bertani (LB) medium (Sambrook and Russell 2001). Detailed information regarding antibiotics,

additions and inducer of the single strains can be found in Table 2-1/Table S1. Because it was not possible to obtain soluble proteins without chaperones in some cases, pKJE7, pGro7 or pG-Tf2 (Takara) were used that produced DnaK-DnaJ-GrpE, GroEL-GroES, and GroES-GroEL-Tig respectively.

### 2.6.3 Plasmids and general DNA procedures

DNA isolation, plasmid preparation, restriction digests, gel electrophoresis, and ligation reactions were conducted according to standard methods (Sambrook and Russell 2001). pBluescript II SK(+) was used for subcloning. All primer sequences are listed in Table 2-2/table S2. Final constructs were verified by sequencing. If not mentioned otherwise, PCR products were cloned into pBluescript II SK(+) by using the method of Marchuk et al. (Marchuk, Drumm, and Saulino 1991).

### 2.6.4 Generation of *Bacillus amyloliquefaciens* mutants and polyketide analysis

Generation of *B. amyloliquefaciens* strain JM54-2 (the TE single mutant) and the conditions for polyketide analysis were published previously (Moldenhauer et al. 2010). Sequences of PCR primers can be seen in Table 2-2/Table S2.

To generate JM157-54 (lacking the TE and BaeB), we used a more complex cloning strategy. A PCR product carrying the chloramphenicol resistance gene was amplified from pDG268 (Antoniewski, Savelli, and Stragier 1990) using primers FP-cat-BamHI and RP-cat-XbaI. This PCR product was cloned into pBluescript II SK(+) to obtain pJM128. This plasmid was digested with *XbaI/BamHI*, and the insert was ligated into the corresponding restriction sites of pJM135 to obtain pJM140. To construct pJM135, a PCR product was amplified from *B. amyloliquefaciens* genomic DNA using primers RP-H-vor-Promotor-XbaI and FP-H-vor-Promotor-NotI and ligated into pBluescript II SK(+) to yield pJM133. This plasmid was digested with *NotI/XbaI* and the resulting insert was ligated into the corresponding restriction sites of pBluescript II SK(+) to obtain pJM135. For the generation of pJM137, a PCR product was amplified from *B. amyloliquefaciens* genomic DNA using primers RP-nach-BaeB-2-KpnI and FP-nach-BaeB-2-ApaI. This PCR product was ligated into pBluescript II SK(+) to obtain pJM137. pJM137 was then digested with *ApaI/KpnI* and the insert ligated into pJM140 to obtain pJM155. A PCR product was amplified from *B. amyloliquefaciens* genomic DNA using primers RP-Promotor-ApaI and FP-Promotor-BamHI and ligated into pBluescript II SK(+) to give pJM139. The insert of pJM139 was digested with *BamHI/ApaI* and inserted into the corresponding site of pJM155 to obtain pJM157. After linearization of pJM157 with *KpnI*, the DNA was integrated into the genome of *B. amyloliquefaciens* JM161 (construction described below) by double crossover to yield the *baeB* deletion mutant JM157. In the last step, pJM54 (Moldenhauer et al. 2010) was linearized with *SacI* and transformed into *B. amyloliquefaciens* JM157 to obtain the *baeB*-TE double mutant JM157-54.

The following strategy was used to generate JM161: A PCR product was amplified from pUCemr (gift from R. Borriss, Humboldt University, Berlin) using primers FP-erm-XbaI and RP-erm-ApaI and ligated into pBluescript II SK(+) to yield pJM136. The plasmids pJM135 and pJM136 were both digested with *XbaI* and *ApaI*, and the insert of pJM136 was ligated into pJM135 to obtain pJM143. pJM137 was digested with *ApaI* and *KpnI*, and this insert was ligated into the corresponding restriction sites of pJM143 to obtain pJM145. The erythromycin antibiotic cassette was replaced with the pIC333 (Steinmetz and Richter 1994) spectinomycin cassette through a blunt-end/*XbaI* ligation into pJM145 to obtain pJM161. The resulting construct was linearized and transformed into the *B. amyloliquefaciens* wild type FZB42 to obtain strain JM161.

## 2.6.5 Construction of expression plasmids

All primer sequences to generate inserts of expression plasmids are provided in Table 2-2/Table S2.

To construct the BaeB expression plasmid pKJ7, a 0.7 kb PCR product was amplified from *B. amyloliquefaciens* genomic DNA using primers BaeB\_low-GC\_pHis8-3\_for and BaeB\_low-GC\_pHis8-3\_rev. This PCR-product was ligated into pBluescript II SK(+) to obtain pKJ1. pKJ1 was then digested with *Bam*HI and *Hind*III, and the 0.7 kb fragment was inserted into the corresponding sites of pHis8 (Jez et al. 2000).

To generate the *pedC* C-terminal strep-tagged fusion construct, a 1 kb PCR product was amplified from pPD23E7 (Piel 2002) using primers pedC\_FP and pedC\_RP. This PCR-product was ligated into pBluescript II SK(+) to obtain pHN37. pHN37 was then digested with *Bsa*I, and the 1 kb fragment inserted into the corresponding site of pASK-IBA3 (IBA Biotagnology) to yield pHN38.

The *pedC* C-terminal hexahistidine-tagged fusion construct pET-pedC was generated by PCR amplification of the gene using pHN38 as template DNA and pedC-5 and pedC-3 as primer. The PCR product was ligated into pET-24a(+) (Novagen) at the *Nde*I and *Xho*I restriction sites after digestion of the insert with the same enzymes.

To construct the PedD expression plasmid pKZ178, a 1.1 kb PCR product was amplified from pPD23E7 (Piel et al. 2004a) using primers pedD\_FP and pedD\_RP. This PCR-product was ligated into pBluescript II SK(+) to obtain pKZ177. pKZ177 was then digested with *Eco*RI and *Hind*III, and the 1.1 kb fragment was inserted into the corresponding sites of pHis8 to yield pKZ178.

To construct the *holo*-PedI3 expression plasmid pKZ123, a 550 bp PCR product was amplified from pS9D2 (Piel et al. 2004b) using primers ped I3\_pRSETf and pedI3\_pRSETr. This PCR-product was ligated into pBluescript II SK(+) to obtain pKZ57. pKZ57 was then digested with *Bgl*II and *Hind*III and the 550 bp fragment inserted into the corresponding site of pHis8 *svp* to yield pKZ123. To construct the *apo*-PedI3 expression plasmid pKZ176, pKZ123 was cut with *Not*I to remove the phosphopantetheinyl transferase gene *svp* and then religated to obtain pKZ176.

To construct the *holo*-PedN expression plasmid pKZ124, a 240 bp PCR product was amplified from *Paederus fuscipes* metagenomic DNA using primers pedN\_pRSETf and pedN\_pRSETr. This PCR-product was ligated into pBluescript II SK(+) to obtain pKZ32. pKZ32 was then digested with *Bgl*II and *Hind*III, and the 240 bp fragment was inserted into the corresponding site of pHis8 *svp* to yield pKZ124. To construct the *apo*-PedN expression plasmid pKZ161, pKZ124 was cut with *Not*I to remove the *svp* gene and then religated to obtain pKZ161.

To construct the *holo*-PsyA expression plasmid pHN77, a 300 bp PCR product was amplified from pPSKF1 (Fisch et al. 2009) using primers PsyA-ACP3 for and PsyA-ACP3 rev. This PCR-product was ligated into pBluescript II SK(+) to obtain pHN75. pHN75 was then digested with *Bam*HI and *Hind*III, and the 300 bp fragment was inserted into the corresponding site of pHis8 *svp* to yield pHN77. To construct the *apo*-PsyA expression plasmid pHN60, a 300 bp PCR product was amplified from pPSKF1 (Fisch et al. 2009) using primers PsyA-ACP3 for and PsyA-ACP3 rev. This PCR-product was ligated into pBluescript II SK(+) to yield pHN59. pHN59 was then digested with *Bam*HI/*Eag*I and the 300 bp fragment inserted into the corresponding site of pHis8 to yield pHN60.

To construct the *holo*-PsyD expression plasmid pHN78, a 320 bp PCR product was amplified from pPSKF1 using primers PsyD-ACP1 for and PsyD-ACP1 rev. This PCR-product was ligated into pBluescript II SK(+) to obtain pHN76. pHN76 was then digested with *Bam*HI and *Hind*III, and the 320 bp fragment inserted into the corresponding site of pHis8 *svp* to yield pHN78. To construct the *apo*-PsyD expression plasmid pHN66, a 320 bp PCR product was amplified from pPSKF1 using primers PsyD-ACP1 for and PsyD-ACP1 rev. This PCR-product was ligated into pBluescript II SK(+) to obtain pHN65. pHN65 was then digested with *Bam*HI and *Eag*I, and the 320 bp fragment was inserted into the corresponding site of pHis8 to yield pHN66.

Construction of the expression plasmid for the PksA ACP was reported earlier (Crawford et al. 2006).

The gene coding for RhiG was amplified by PCR from cosmid 20G04 (Partida-Martinez and Hertweck 2006) with the primers Expr\_tAT\_F/Expr\_tAT\_R. Ligation into the *EcoRV* site of pGEM-T Easy cloning vector (Promega) yielding pNB118. The *NdeI/HindIII* fragment of pNB118 was ligated into the *NdeI/HindIII* sites of pET28a(+) expression vector, which carries an N-terminal His<sub>6</sub>-Tag, resulted in pNB121.

To introduce point mutations into active site of RhiG, the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene) was used on pNB121 (AT1°AT2 – S106A and AT1AT2° – S468A). The plasmid pNB146 contains a *rhiG* variant coding for AT1°AT2, and the plasmid pNB147 contains a *rhiG* variant coding for AT1AT2°.

## 2.6.6 Expression and purification of proteins

To prepare sufficient amounts of PedC for kinetic studies, large-scale fermentation was needed. A colony of *E. coli* cells freshly transformed with pHN38 and pGro7 was cultivated up to an optical density (OD<sub>600</sub>) of about 4 at 37 °C in 2 mL LB medium. The grown biomass was used to inoculate the second preculture in 1 L baffled shake flasks with 150 mL culture volume at 120 rpm/min and 37 °C. The second preculture medium contained 5 g/L yeast extract, 10 g/L tryptone, 2.5 g/L NaCl, 0.25 g/L MgSO<sub>4</sub> × 7 H<sub>2</sub>O, 1.26 g/L NaH<sub>2</sub>PO<sub>4</sub>, and 10 g/L glucose, adjusted to pH = 7.2 with 4 N NaOH. Main cultivation was performed in a 30 L scale bioreactor (Chemap) with 20 L initial culture volume at 37 °C. Dissolved oxygen concentration was controlled at ≥ 40% of air saturation, and the pH was kept at 7.0 with 25% ammonia solution. Medium was similar as for the second preculture, except that 20 g/L glucose and 5 g/L arabinose were present. A total of 900 mL of the second preculture were used to inoculate the main culture resulting in an initial OD<sub>600</sub> = 0.25. At OD<sub>600</sub> = 4.5 (after about 3 hr) 4 mg of anhydrotetracycline (IBA Biotagnology) was added (resulting in 200 ng/mL), and the temperature was decreased to 25 °C in less than 10 minutes. After 7 hr cultivation time, a glucose feed (600 g L<sup>-1</sup> glucose monohydrate) was started to keep the glucose concentration in a range of 13-17 g/L throughout the cultivation process. Glucose measurement was performed using an AkkuCheck glucose analyzer (Roche Diagnostics). After 24 hr of cultivation time the bioreactor was cooled to 10 °C, and the cell suspension (OD<sub>600</sub> = 35) was harvested. Biomass separation was achieved by centrifugation at 8,000 rpm (rotor type JLA8.1000, Avanti J-20 XP, Beckman), and the biomass was stored at -80 °C prior to further protein purification. The wet cell weight was 500 g. For kinetic measurements of PedC, elution fractions were dialyzed 3 times for at least 3 hr against 200 mM potassium phosphate buffer (KPi) pH 7.4 containing 10% (v/v) glycerol. If necessary, fractions were concentrated with Vivaspins columns (VWR). Protein concentration was determined either by Bradford assay or, if expressed with chaperones with SDS-PAGE-absorption measurements. Specifically, SDS-PAGE was performed with all protein samples along with BSA samples of known concentration. The band intensities were integrated using the program GeneTools (Syngene). A calibration curve was established, and the concentration of the protein samples was determined. Proteins were flash frozen in liquid nitrogen after adding 10% (v/v) glycerol and stored at -80 °C.

For the overexpression of all other proteins, except for pET-PedC, 1 l cultures were grown to an OD<sub>600</sub> of 0.7 (37 °C, 200 rpm), induced with either isopropyl β-D-1-thiogalactopyranoside (IPTG) or anhydrotetracycline, and grown for approximately 20 hr at 16 °C. Cells of expression strains were harvested by centrifugation (11,000 × g, 7 min) at 4 °C. For pET-PedC, 1 l expression cultures were supplemented with 50 ng/mL tetracycline prior to inoculation with a 10 mL saturated starter culture. Cells were grown at 37 °C 250 rpm and growth was followed by monitoring OD<sub>600</sub>. The cultures, at OD<sub>600</sub> ~0.4, were cooled on ice for 15 min, induced with 1 mM IPTG, and expressed at 18 °C for 16 hrs.

Cells expressing proteins with His<sub>6</sub>- or His<sub>8</sub>-Tag were resuspended in lysis buffer (25 mM Tris-HCl, pH 7.6, 0.5 M NaCl, 10 mM imidazole and 0.1% (v/v) Triton X-100). The pH was adjusted to 7.6 in the case of PedD, RhiG, and their derivatives. A pH of 8 was chosen for all other proteins. The cell suspension was kept on ice and sonicated in 10 s intervals. The cell debris was removed by centrifugation at 23,500 × g for 30 min at 4 °C. The supernatant was

incubated with 1.0 mL Ni-NTA agarose (Qiagen) for at least 1 hr at 4 °C and washed with 10 mL lysis buffer. His-tagged proteins were eluted with lysis buffer containing increasing concentrations of imidazole up to 250 mM. For the expression using pET-pedC, a modified protocol was used. Cells were harvested by centrifugation (4,100 x g, 20 min) at 4 °C and resuspended (1 g per 4 mL) in Ni-buffer (50 mM KPi pH 7.6, 300 mM NaCl, 10 % glycerol, 20 mM imidazole) and lysed by treatment with 1 mg/mL lysozyme (Sigma) for 30 min, followed by sonication in 10 s intervals. The lysate was clarified by centrifugation (27,000 x g, 30 min), decanted, and bound to 1 mL Ni-NTA resin (Qiagen) for 1 hr rotating at 4 °C. The resin was washed three times with 10 mL Ni-buffer and eluted three times with 0.5 mL Ni-buffer containing 250 mM imidazole. Fractions were analyzed by SDS-PAGE. The flowthrough contained a large amount of unbound PedC, which was subsequently collected by a second round of Ni<sup>2+</sup> affinity chromatography. Fractions containing PedC were pooled, concentrated to ~7 mg/mL using a 10,000 MWCO Amicon Ultra Centrifugal Filter Unit (Millipore), and dialyzed against storage buffer containing 100 mM KPi pH 7.0, 10% glycerol, and 2 mM DTT. The concentration of PedC was determined by Bradford assay in duplicate and adjusted for GroEL content.

Cells containing expression constructs with Strep-tags were resuspended in Buffer W (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA). After sonication and centrifugation as described above, the supernatant was applied to a column with 1 mL Strep-Tactin-Sepharose (IBA Biotagnology). The column was washed twice with 5 mL Buffer W and eluted with six fractions of 0.5 mL Buffer E (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin). Columns were used up to five times and regenerated according to manufacturer's specification.

### 2.6.7 Acyltransferase assays

To study acyl transfer from ATs to ACPs, we performed assays with [2-<sup>14</sup>C]-malonyl-CoA and [1-<sup>14</sup>C]-acetyl-CoA (Perkin-Elmer). Assay mixtures consisted of 4 µL dithiothreitol (DTT) (10 mM), 5 µL AT, 9 µL ACP, and either 2 µL [malonyl-2-<sup>14</sup>C-CoA] (0.01 mCi/mL) or 1 µL acetyl-CoA [acetyl-1-<sup>14</sup>C-CoA] (0.02 mCi/mL). Lysis buffer, instead of buffer containing the enzyme, was used in the negative controls. Everything was mixed on ice and the reaction was started with the addition of the substrate. Samples were incubated on ice for various times, followed by the addition of equal volumes of 2x SDS-sample buffer (0.09 M Tris-HCl, pH 6.8, 20% (v/v) glycerol, 2% (v/v) SDS, 0.02% (v/v) bromophenol blue, 0.1 M DTT) and incubation at 99 °C for 3 min. After separation by SDS-PAGE the gels were dried and exposed to X-ray film. Band quantification was performed with an Image Reader and the software Bas Reader (Raytest). The autoradiogram was edited with the software TINA 2.09d (Raytest). To differentiate between self-loading of ACPs and AT-mediated loading, we also performed measurements without AT.

### 2.6.8 Acyl-SNAC hydrolysis assays

After desalting in 200 mM KPi, pH 7.4 and 10% (v/v) glycerol, proteins were incubated with different test substrates at a range of concentrations. In these assays, PedC was used only as purified, strep-tagged protein (Figure 2-7/Figure S3). A typical assay, adapted from Heathcote et al. (Heathcote and Leadlay 2001) consisted of 250 nM AT, 200 mM KPi, pH 7.4, 0.2 mM DTNB (Alfa Aesar) and 3% (v/v) dimethylsulfoxide (DMSO) at 30 °C in a total volume of 1 mL. All test substrates were dissolved as 2 M stocks in DMSO except compound **18**, which was dissolved in 200 mM KPi, pH 7.4. A 20 mM DTNB stock solution was resuspended in 200 mM KPi, pH 7.4. Samples containing **20** and **22** were incubated for 2 min at room temperature (RT) prior to measurement, allowing better solubilization. Enzymatic activities were monitored by measuring the absorbance increase at 412 nm resulting from hydrolysis of Ellman's reagent ( $\epsilon=14,150 \text{ M}^{-1} \text{ cm}^{-1}$ ) (Frank et al. 2007) for 18 min using a BioMate 3 photometer (Thermo). All measurements were carried out in at least duplicate and corrected for the background rate of hydrolysis without enzyme. For the determination of kinetic parameters, initial rates were plotted versus the substrate concentration

and fitted to the Michaelis-Menten equation with GraphPad Prism 4.0 (GraphPad Software). For those enzymes a Michaelis-Menten fit was not possible, only the initial velocities were plotted.

Hydrolysis of hexanoyl-SNAC (mimicking the hexanoyl starter unit of PksA from *Aspergillus parasiticus* involved in norsolorinic acid biosynthesis) was determined using an HPLC-based assay. Hexanoyl-SNAC (1 mM) was treated with 5  $\mu$ M PedC in 100 mM KPi pH 7.0, 10% glycerol, and 1 mM DTT. At each time point, 150  $\mu$ L reaction aliquots were passed through a 10,000 MWCO Microcon Centrifugal Filter Device (Millipore) at 14,000 x g for 10 min to remove the protein. Fifty microliter injections of microcon flowthrough were separated over a Prodigy 5 $\mu$  ODS3 100 Å, 250 x 4.6 mm column (Phenomenex) using an Agilent Technologies 1200 series HPLC and a 40:60:0.1 acetonitrile:water:formic acid isocratic method at 1 mL/min. The peak area of hexanoyl-SNAC (RT for 12.0 min) was determined by manual integration at 232 nm wavelength. Controls with no enzyme or BSA showed no loss in hexanoyl-SNAC signal, indicating that hydrolysis was catalyzed by PedC.

### 2.6.9 Acyl-ACP hydrolysis assays

To determine whether PedC can hydrolyze acyl-ACP intermediates, we used a radiochemical hydrolysis assay. The PksA ACP monodomain was activated with various  $^{14}$ C-labeled CoAs and treated with a serial dilution of PedC in a fixed time point assay.

The N-terminally hexahistidine-tagged PksA *apo*-ACP monodomain (excised from the six domain iterative megasynthase) was purified as previously described and activated by an amended protocol described below (Crawford et al. 2006).  $^{14}$ C-labeled acyl-ACP was generated by treating 100  $\mu$ M *apo*-ACP with 5  $\mu$ M Svp (phosphopantetheinyl transferase from *Streptomyces verticillus*), and 400  $\mu$ M total acyl-CoA (mixture of unlabeled and  $^{14}$ C-labeled to achieve 5.5  $\mu$ Ci/mL) in 100 mM KPi, 10 mM MgCl<sub>2</sub>, 1 mM TCEP, and 10% glycerol for 45 min at RT. Acyl-ACP can be frozen for later use or diluted to be used immediately.

PedC was thawed and serially diluted (1:5) to five different concentrations ranging from 20 to 0.02  $\mu$ M 4x stock solutions in 100 mM KPi pH 7.0 with 10% glycerol. Twelve microliters of each PedC stock was pre-aliquoted and the assay was initiated by addition of 36  $\mu$ L of 13.3  $\mu$ M  $^{14}$ C-acyl-ACP. After 10 min at room temperature, 20  $\mu$ L aliquots of each reaction were quenched into 10  $\mu$ L of 4x SDS loading dye. Final reaction conditions were 10  $\mu$ M acyl-ACP treated with 0 (negative control), 5, 0.83, 0.14, 0.023, or 0.005  $\mu$ M PedC. Samples were separated by 16% SDS-PAGE, dried on filter paper using a slab gel dryer (Hoeffer Scientific), exposed overnight to a storage phosphor screen (Amersham), and imaged using a Typhoon 9410 Imager (Johns Hopkins Integrated Imaging Center).

### 2.6.10 Phylogenetic analysis

Sequences of AT-like proteins were aligned using MUSCLE. Phylogenetic analysis was performed by Maximum Likelihood inference using PhyML with a Jones-Taylor-Thornten (JTT) amino acid replacement model.

### 2.6.11 Synthesis of acyl-SNAC derivatives **18-23**

Synthesis of the known compounds **18** and **23** was performed as published previously (Crawford et al. 2006), (Grüschow et al. 2007). The general procedure for compounds **19-22** was as follows. A solution of the corresponding acid (5 mmol) in dichloromethane (15 mL) was cooled down to 0 °C for 15 minutes. 4-dimethylaminopyridine (DMAP, 1 mmol), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide (EDC, 6 mmol) and N-acetylcysteamine (6 mmol) were added, and the mixture was stirred over night at 25 °C. The reaction mixture was quenched with saturated



aqueous ammonium chloride and extracted with dichloromethane. The combined organic phases were dried with MgSO<sub>4</sub>, filtered, concentrated and subjected to silica gel column chromatography as eluant to provide the respective thioester.

S-Crotonoyl-N-acetylcysteamine (**19**). Yield was as follows: 54%, column chromatography with ethyl acetate. <sup>1</sup>H-NMR: (400 MHz, CDCl<sub>3</sub>, RT) δ [ppm] = 6.95 (dq, 1H, <sup>3</sup>J<sub>H-H</sub> = 15.4 Hz, <sup>3</sup>J<sub>H-H</sub> = 6.9 Hz), 6.16 (dq, 1H, <sup>3</sup>J<sub>H-H</sub> = 15.4 Hz, <sup>4</sup>J<sub>H-H</sub> = 1.6 Hz), 5.95 (brs, 1H), 3.46 (q, 2H, <sup>3</sup>J<sub>H-H</sub> = 6.6 Hz), 3.08 (t, 2H, <sup>3</sup>J<sub>H-H</sub> = 6.6 Hz), 1.93 (s, 3H), 1.88 (dd, 3H, <sup>3</sup>J<sub>H-H</sub> = 6.9 Hz, <sup>4</sup>J<sub>H-H</sub> = 1.6 Hz); <sup>13</sup>C-NMR: (100 MHz, CDCl<sub>3</sub>, RT) δ [ppm] = 190.5, 170.6, 142.1, 130.1, 40.1, 28.4, 23.4, 18.2; MS: (ESI, 10 eV, *m/z* (%)) 224.1 (100) [M + Na]<sup>+</sup>; HR-MS (ESI, 10 eV, *m/z*): C<sub>8</sub>H<sub>13</sub>NO<sub>2</sub>SH 188.0740; measured: 188.0743 [M + H]<sup>+</sup>.

(E)-S-(Hex-2-enoyl)-N-acetylcysteamine (**20**): Yield was as followed: 43%, column chromatography with cyclohexane/ethyl acetate 2:1. <sup>1</sup>H-NMR: (400 MHz, CDCl<sub>3</sub>, RT) δ [ppm] = 6.92 (dt, 1H, <sup>3</sup>J<sub>H-H</sub> = 15.5 Hz, <sup>3</sup>J<sub>H-H</sub> = 7.0 Hz), 6.13 (td, 1H, <sup>3</sup>J<sub>H-H</sub> = 15.5 Hz, <sup>4</sup>J<sub>H-H</sub> = 1.5 Hz), 5.88 (bs; 1H), 3.49-3.44 (pq, 2H, <sup>3</sup>J<sub>H-H</sub> = 6.5 Hz), 3.09 (t, 2H, <sup>3</sup>J<sub>H-H</sub> = 6.5 Hz), 2.19 (ddt, 2H, <sup>4</sup>J<sub>H-H</sub> = 1.5 Hz, <sup>3</sup>J<sub>H-H</sub> = 7.0 Hz, <sup>3</sup>J<sub>H-H</sub> = 7.3 Hz), 1.96 (s; 3H; H-1), 1.50 (tq, 3H, <sup>3</sup>J<sub>H-H</sub> = 7.4 Hz), 0.94 (t, 6H, <sup>3</sup>J<sub>H-H</sub> = 7.4 Hz); <sup>13</sup>C-NMR: (100 MHz, CDCl<sub>3</sub>, RT) δ [ppm] = 190.4, 170.2, 146.5, 128.4, 39.8, 34.2, 28.2, 23.2, 21.2, 13.7; MS: (ESI, 10 eV, *m/z*) 120.0 (8) [NAC + H]<sup>+</sup>, 216.1 [M + H]<sup>+</sup>, 238.1 [M + Na]<sup>+</sup>.

S-(3-Methylcrotonyl)-N-acetylcysteamine (**21**). Yield was as followed: 40%, column chromatography with ethyl acetate. <sup>1</sup>H-NMR: (400 MHz, CDCl<sub>3</sub>, RT) δ [ppm] = 5.98 (s, 1H), 5.98 (brs, 1H), 3.45 (q, 2H, <sup>3</sup>J<sub>H-H</sub> = 6.0 Hz), 3.04 (t, 2H, <sup>3</sup>J<sub>H-H</sub> = 6.0 Hz), 2.13 (s, 3H), 1.93 (s, 3H), 1.86 (s, 3H); <sup>13</sup>C-NMR: (100 MHz, CDCl<sub>3</sub>, RT) δ [ppm] = 189.8, 170.5, 155.1, 123.2, 40.2, 28.6, 27.5, 23.4, 21.5; MS: (ESI, 10 eV, *m/z* (%)) 210.1 (100) [M + Na]<sup>+</sup>; HR-MS (ESI, 10 eV, *m/z*): C<sub>9</sub>H<sub>15</sub>NO<sub>2</sub>SH 202.0896; measured: 202.0893 [M + H]<sup>+</sup>.

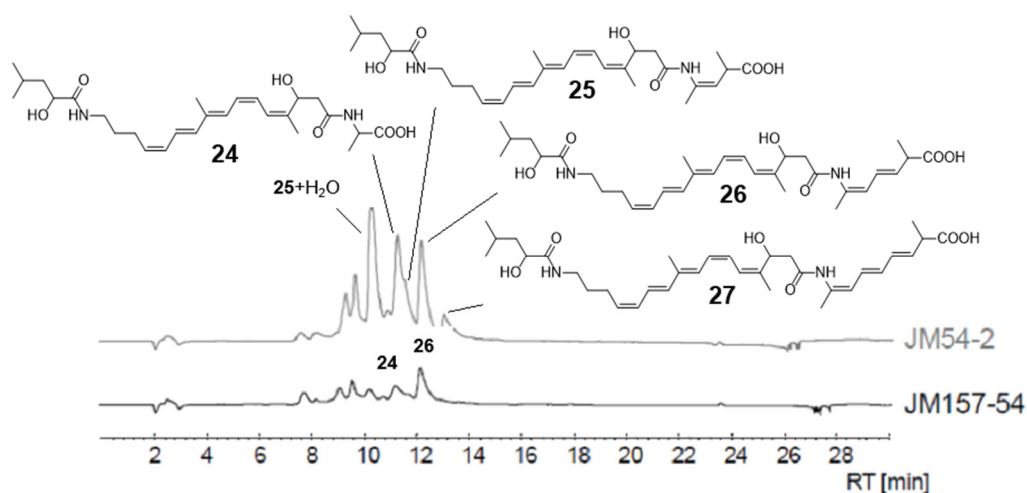
S-(4-Methylvaleryl)-N-acetylcysteamine (**22**): Yield: 90%, column chromatography with cyclohexane/ethyl acetate; gradient: 1:1 to 1:3. <sup>1</sup>H-NMR: (400 MHz, CDCl<sub>3</sub>, RT) δ [ppm] = 5.82 (bs, 1H), 3.43 (pq, 2H, <sup>3</sup>J<sub>H-H</sub> = 6.4 Hz), 3.02 (t, 2H, <sup>3</sup>J<sub>H-H</sub> = 6.4 Hz), 2.60-2.56 (m, 2H), 1.96 (s, 3H), 1.62-1.53 (m, 3H), 0.90 (d, 6H); <sup>13</sup>C-NMR: (100 MHz, CDCl<sub>3</sub>, RT) δ [ppm] = 200.3, 170.2, 42.2, 39.7, 34.3, 28.4, 27.6, 23.2, 22.2; MS: (ESI, 10 eV, *m/z*) 218.1 [M + H]<sup>+</sup>, 240.1 [M + Na]<sup>+</sup>.

For details, see Supplemental information.

## 2.7 Acknowledgement

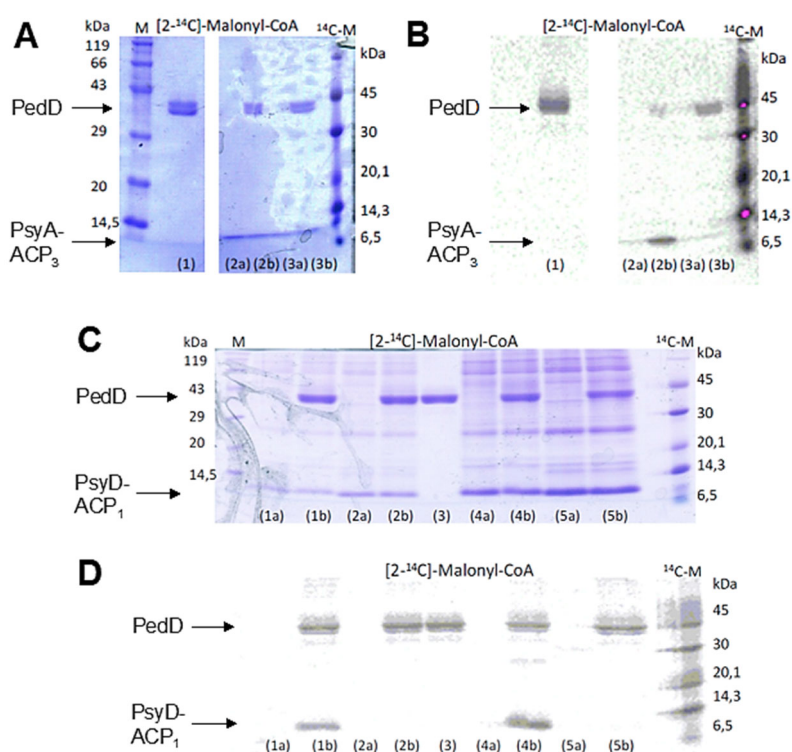
We thank H. Chang, M. Engeser, C. Sondag, and H.-J. Brandt for technical assistance, R. Borriss, B. S. Moore, and M. Marahiel for plasmids, and M. F. Freeman for experimental suggestions. This work was financially supported by the DFG (PI 430/1-3, PI 430/6-1, PI 430/9-1, SFB 624 and GRK 804 to J.P.).

## 2.8 Supplemental Information



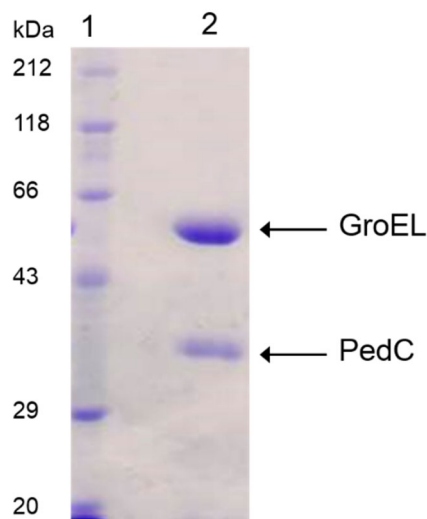
**Figure 2-5: Supplemental Figure 1, related to Figure 2-1.**

HPLC traces of crude extracts of *B. amyloliquefaciens* JM54-2 (bacillaene TE deletion mutant) in comparison to JM157-54 (TE/*baeB* double mutant). Compounds were identified by LC-HRMS and LC-NMR analysis (Moldenhauer, Chen, and Borriss 2007), (Moldenhauer et al. 2010). In JM157-54 extracts only small amounts of compounds **24** and **26** could be detected.



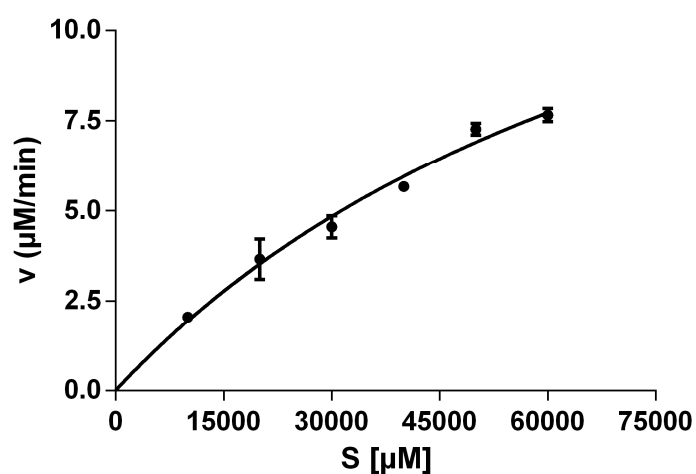
**Figure 2-6: Supplemental Figure 2, related to Figure 2-3.**

Acyltransferase assay using PedD, radiolabeled malonyl-CoA, and two integrated ACPs of the multimodular psymberrin PKS. (A): SDS-PAGE analysis of assays containing the ACP of PsyA, module 3. Lane 1, PedD only; lane 2a, *holo*-ACP; lane 2b, PedD + *holo*-ACP; lane 3a, *apo*-ACP; lane 3b, PedD + *apo*-ACP. All assays contained [2-<sup>14</sup>C]-malonyl-CoA. (B): Autoradiogram of the same gel. (C): SDS-PAGE analysis of assays containing the ACP of PsyD, module 1. Lane 1a, *holo*-ACP; lane 1b, PedD + *holo*-ACP; lane 2a, *apo*-ACP; lane 2b, PedD + *apo*-ACP; lane 3, PedD; lane 4a, *holo*-ACP; lane 4b, PedD + *holo*-ACP, lane 5a, *apo*-ACP; lane 5b, PedD+ *apo*-ACP. For assays of lanes 4a to 5b, the ACP was concentrated 10x using Vivaspin ultrafiltration columns (molecular weight cut-off 5000). (D): Autoradiogram of the same gel. All assays contained [2-<sup>14</sup>C]-malonyl-CoA.



**Figure 2-7: Supplemental Figure 3, related to Figure 2-4.**

SDS-PAGE analysis of purified PedC used for the hydrolysis assays. Samples were stained with Coomassie brilliant blue for visualization. Lane 1, molecular weight marker; lane 2: purified PedC with GroEL after affinity chromatography.



**Figure 2-8: Supplemental Figure 4, related to Figure 2-4.**

Rate versus substrate concentration for the hydrolysis of 3-hydroxybutyryl-SNAC (**18**) by PedC. Calculated kinetic parameters are  $K_m = 88.48 \pm 27.97$  mM,  $k_{cat} = 76.48$  min<sup>-1</sup>,  $v_{max} = 0.01912 \pm 0.0041$  mM min<sup>-1</sup>, and  $k_{cat}/K_m = 14.41$  M<sup>-1</sup>s<sup>-1</sup>.

| Construct                                  | Plasmid              | Additive                          | Inducer                            | Expression host                   |
|--|----------------------|-----------------------------------|------------------------------------|-----------------------------------|
| PedC<br>(C-terminal Strep-tag)             | pHN38<br>+ pGro7     | arabinose<br>(0.5 mg/mL)          | anhydrotetracycline<br>(200 ng/mL) | <i>E. coli</i><br>BL21DE3         |
| PedC<br>(N-terminal MBP tag)               | pH66                 | glucose<br>monohydrate<br>(2 g/L) | IPTG<br>(1 mM)                     | <i>E. coli</i><br>BL21DE3         |
| PedC<br>(C-terminal His <sub>6</sub> -tag) | pET-pedC<br>+ pG-Tf2 | tetracycline<br>(50 ng/mL)        | IPTG<br>(1 mM)                     | <i>E. coli</i> BL21-<br>Gold(DE3) |

|  |                              |                          |                   |   |
|--|------------------------------|--------------------------|-------------------|---|
| PedD<br>(N-terminal His <sub>8</sub> -tag)             | pKZ178                       | -                        | IPTG<br>(0.5 mM)  | <i>E. coli</i><br>BL21DE3                       |
| GroES/L  | pGro7                        | arabinose<br>(0.5 mg/mL) | -                 | <i>E. coli</i><br>BL21DE3                       |
| RhiG and mutants<br>(N-terminal His <sub>6</sub> -tag) | pNB121,<br>pNB146,<br>pNB147 | -                        | IPTG<br>(0.5 mM)  | <i>E. coli</i><br>BL21DE3                       |
| PsyD-ACP1<br>(N-terminal His <sub>8</sub> -tag)        | pHN66,<br>pHN78              | -                        | IPTG<br>(0.75 mM) | <i>E. coli</i><br>BL21DE3                       |
| PsyA-ACP3<br>(N-terminal His <sub>8</sub> -tag)        | pHN60,<br>pHN77              | -                        | IPTG<br>(0.75 mM) | <i>E. coli</i><br>BL21DE3                       |
| PedN<br>(N-terminal His <sub>8</sub> -tag)             | pKZ124, pKZ161               | -                        | IPTG<br>(0.4 mM)  | <i>E. coli</i><br>BL21DE3                       |
| PedI3<br>(N-terminal His <sub>8</sub> -tag)            | pKZ123, pKZ176               | -                        | IPTG<br>(1 mM)    | <i>E. coli</i> Rosetta<br>gami 2 (DE3)<br>pLysS |
| BaeB<br>(N-terminal His <sub>8</sub> -tag)             | pKJ7                         | -                        | IPTG<br>(0.5 mM)  | <i>E. coli</i><br>BL21DE3                       |

**Table 2-1: Supplemental Table 1.**

Constructs and their respective antibiotic resistance cassettes, additives, inducer and expression strains.

| Primer name                     | Sequence (5'-3')                        |
|---------------------------------|---|
| FP-cat- <i>Bam</i> HI           | AAAGGATCCGACAGCTTATCATCGGCAATA          |
| RP-cat- <i>Xba</i> I            | AAATCTAGAGGCGTAGAGGATCTGGAGC            |
| RP-H-vor-Promotor- <i>Xba</i> I | AAATCTAGATTTCTCCATCTCATACTGTGGTG        |
| FP-H-vor-Promotor- <i>Not</i> I | AAAGCGGCCGCCATCCGATTACGTTTATCGAAATTACG  |
| RP-nach-BaeB-2- <i>Kpn</i> I    | AAAGGTACCAGAAGGCGGTCAAATGGATCA          |
| FP-nach-BaeB-2- <i>Apa</i> I    | AAAGGGCCCGTTCTATCTAACTAGCTTTTCTTTGAGG   |
| RP-Promotor- <i>Apa</i> I       | AAAGGGCCCTGTCACCATTCCCATTAAAAGAT        |
| FP-Promotor- <i>Bam</i> HI      | AAAGGATCCTAACAACGTTTATGTGAGACTAAACC     |
| FP-erm- <i>Xba</i> I            | AAATCTAGACGAGGAATTTGTATCGATAAGAAATAG    |
| RP-erm- <i>Apa</i> I            | AAAGGGCCCATATAAGGAATTGAAGTTAAATTAGATGCT |
| BaeB_low-GC_pHis8-3_for         | GGATCCATGGATCATAATATGAAGTGCATCAA        |

|                         |   |
|-------------------------|---|
| BaeB_low-GC_pHis8-3_rev | <b>AAGCTTTT</b> ACTTAAAATGAAACAGCCCTTTTTG                                 |
| pedC_FP                 | ATGGT <b>AGGTCT</b> CAAATGAAAGACTTGCAAAATATACAGAACAC                      |
| pedC_RP                 | ATGGT <b>AGGTCT</b> CAGCGCTACGTTGGTCGAGTTGAGCAGA                          |
| pedC-5                  | GAT <b>ACATATG</b> AAAGACTTGCAAAATATACAGAAC                               |
| pedC-3                  | GATT <b>CTCGAG</b> ACGTTGGTCGAGTTGAGC                                     |
| pedD_FP                 | AA <b>GAATT</b> C AAATCGTACCTTTTTCCCGG                                    |
| pedD_RP                 | AAA <b>AGCTTT</b> CACCACACCTTTTCAACTAAA                                   |
| pedI3_pRSETf            | AAA <b>AGATCT</b> GAGCAGAAGGTA TATGCGGTCATT                               |
| pedI3_pRSETr            | AAA <b>AGATCTT</b> CATATCAGGCTCCGTACGTACTG                                |
| pedN_pRSETf             | AAA <b>AGATCT</b> ATCCGCGAACGCATTTTCAATGTGATTGCGAGA<br>AATACGCTTGAAGTCCTT |
| pedN_pRSETr             | AAA <b>AGATCTT</b> CACTGTACATACTGGCTGAG                                   |
| PsyA-ACP3 for           | <b>GGATCC</b> ACGTCGAGCGGGAACTTGCGACAGTGG                                 |
| PsyA-ACP3 rev           | <b>CGGCCG</b> TAAACGCATACCGCTTCGAGCTGCTGCGC                               |
| PsyD-ACP1 for           | <b>GGATCC</b> ACTTCGTCGCCAAAGGGCAATCTGACCG                                |
| PsyD-ACP1 rev           | <b>CGGCCG</b> TCAAGGTATACACGCTTCGATGTGGGCAGCCAA                           |
| Expr_tAT_F              | <b>GAATTC</b> AGGACGAAATCTATGCCGTA  |
| Expr_tAT_R              | <b>AAGCTTT</b> ACGATCAGCGGCTTTGTT   |

**Table 2-2: Supplemental Table 2.**

Sequence of PCR primers used for cloning of obtained constructs. Introduced restriction sites are marked with **bold** letters.

# Chapter 3. Deciphering Halogenation in Oocydin and Haterumalide Biosynthesis

---

Katja Jensen,<sup>1</sup> Reiko Ueoka,<sup>1</sup> Brandon I. Morinaka,<sup>1</sup> Nancy Magnus,<sup>2</sup> Birgit Piechulla,<sup>2</sup> Jörn Piel<sup>1</sup>

<sup>1</sup> Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zurich, 8093 Zurich, Switzerland

<sup>2</sup> University of Rostock, Institute for Biological Sciences, 18059 Rostock, Germany

K.J. and N.M. generated *Serratia* knock out strains, K.J., R.U. and B.I.M. isolated compounds, performed NMR measurements and structure elucidation, K.J. cloned constructs, performed rescue-experiments and cross-complementation, K.J. and R.U. conducted cytotoxicity assays, K.J., B.I.M. and J.P. wrote the manuscript, all authors designed the experiments and analyzed the data.

*In preparation*

## 3.1 Summary

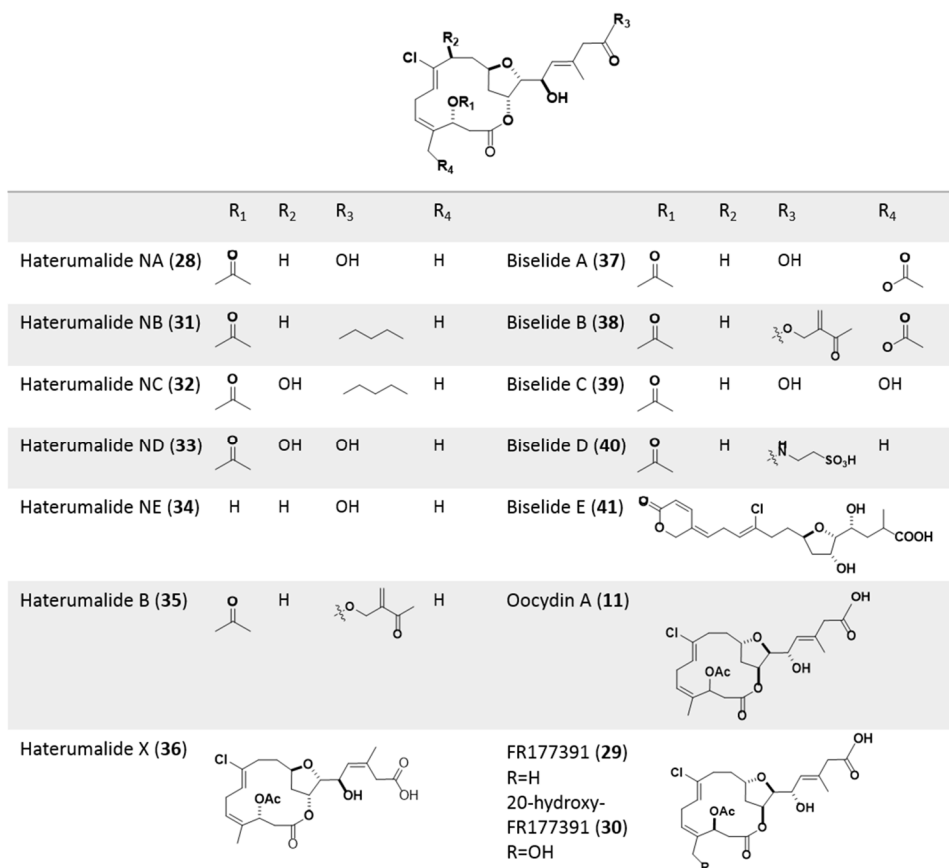
Mononuclear non-heme (MNH) iron,  $\alpha$ -ketoglutarate- and O<sub>2</sub>-dependent halogenases are described to act individually and almost exclusively on protein-bound substrates generating a saturated, halogenated moiety. During the course of our study on the biosynthetic pathway of oocydin A an anti-cancer and agricultural important polyketide macrolide we identified two proteins involved in chlorination at a vinyl moiety of oocydin A. One protein exhibits distant similarity to MNH halogenases, the other does not exhibit homology to any described protein. This two-component halogenase represents a new type. Unchlorinated oocydin A was isolated together with oocydin B a new analog with an extended side chain. A revised biosynthetic model was proposed explaining the biosynthesis of a compound with a longer side chain in parallel with a short-chain compound.

## 3.2 Introduction

To date around 4700 halogenated natural products are known (Gribble 2010). Halogen atoms can be introduced into molecules by different enzymes and mechanisms. For around 35 years haloperoxidases were the only known halogenating enzymes. The halogenating agent, a hypohalite, acts on electron-rich substrates such as pyrroles or phenols (Neumann, Fujimori, and Walsh 2008), (Galonić et al. 2007). Haloperoxidases exhibit only a very low regioselectivity as the halogenating agent is freely diffusible (Wagner, el Omari, and König 2009). Subsequently, flavin-dependent halogenases were shown to halogenate electron-rich substrates with a hypohalite as well (Blasiak and Drennan 2008). These halogenases react either on carrier protein (CP)-bound or on free-standing substrates (Blasiak and Drennan 2008), typically indoles, pyrroles or phenol rings (Smith, Grüschow, and Goss 2013). Molecular oxygen and flavin are employed by flavin-dependent halogenases. The reduced flavin is typically delivered by a partner protein reductase (Neumann, Fujimori, and Walsh 2008). In addition SAM-dependent halogenases were described with the underlying mechanism of a nucleophilic substitution with a halide on electrophilic substrates (Eustáquio et al. 2007). The introduction of halogen atoms on aliphatic unactivated substrates as in syringomycin E, barbamide, coronatine, and curacin A have remained enigmatic until mononuclear non-heme iron (MNH)  $\alpha$ -ketoglutarate (KG)- and O<sub>2</sub>-dependent halogenases were shown to carry out these unique transformations (Vaillancourt, Yin, and Walsh 2005), (Vaillancourt et al. 2005), (Vaillancourt, Vosburg, and Walsh 2006). All known MNH halogenases were believed to perform their reaction on acyl CP (ACP)- or peptidyl CP (PCP)-tethered

substrates with WelO5 out of the welwitindolinone biosynthetic pathway as the only exception (Hillwig and Liu 2014). The reaction mechanism is proposed to be analogous to the related MNH oxygenases (Smith, Grüşchow, and Goss 2013). A reactive oxo-ferryl species is formed after binding of the substrate and molecular oxygen (Butler and Sandy 2009). Following hydrogen abstraction from the substrate, a transfer of the halide to the substrate radical follows (Buongiorno and Straganz 2013). In curacin A biosynthesis the halogenase CurA shows a conformational switch induced upon chloride and  $\alpha$ -KG binding (Khare et al. 2010). One main distinguishing feature between MNH oxygenases and halogenases is the active site. In oxygenases a 2-His-1-carboxylate facial triad is present as common metal binding motif (Koehtop, Emerson, and Que 2005). The iron(II) is coordinated by two histidine residues and a carboxylate, typically represented by an aspartate or glutamate residue. This stands in contrast to halogenases where the iron(II) is coordinated by two histidine residues, bearing a chloride binding pocket (Buongiorno and Straganz 2013). Here the Asp/Glu is exchanged against small hydrophobic residues to provide space for the chlorine atom to coordinate the iron (Pratter et al. 2014b). Single-point mutations in these sites did not change the fate of enzyme reactivity, as no switch between hydroxylase and halogenase or vice versa could be induced, leading to the assumption that more factors determine the reaction outcome (Hangasky et al. 2013).

Oocydin A **11** is a chlorinated macrolactone highly active against oomycetes, agricultural pathogens, and is among a larger family of structurally related molecules which includes the haterumalides and biselides (Table 3-1) (Kigoshi and Hayakawa 2007). Notable is not only the structurally related macrolactone ring of almost all family member as skeletal structure, but also the similarity in more unusual side chains as the esterified but-3-en-2-one-moiety in **35** and **38**. Member of the haterumalide family were isolated in different combinations from diverse organisms and habitats. Up to now they were isolated from marine ascidian (Ueda and Hu 1999), (Teruya et al. 2005), marine sponge (Takada et al. 1999) and various bacteria (Strobel and Hess 1999), (Sato et al. 2005), (Levenfors et al. 2004), (Thaning et al. 2001), (Matilla et al. 2012). Oocydin A **11** was isolated from the bacterium *Serratia marcescens* in 1999 (Strobel and Hess 1999). Haterumalide NA **28** was isolated from the marine sponge *Ircinia* sp. in the same year as **11** and more over exhibits the same structure as **11** (Takada et al. 1999). A compound with the same structure was additionally isolated from *Serratia liquefaciens* and named FR177391 **29** (Sato et al. 2005). Furthermore **11** and **28** were isolated from the marine ascidian of the family *Didemnidae* and *Serratia plymuthica* (Levenfors et al. 2004), (Teruya et al. 2005). **11**, **28**, and **29** share the same structure and relative stereochemistry but final assignment of absolute stereochemistry remains ambiguous. The activity profile of **11**, **28** and **29** comprises in addition to anti-oomycetes also bioactivity against other fungi and hyperlipidemia, suppression of apothecial formation of *Sclerotinia sclerotiorum* and cytotoxicity in low micromolar concentration (Sato et al. 2005), (Levenfors et al. 2004), (Thaning et al. 2001), (Matilla et al. 2012).



**Table 3-1: Family of haterumalides; table modified after Kigoshi and Hayakawa 2007.**

Preliminary assignment of stereochemistry was adopted from Kigoshi and Hayakawa 2007.

The gene cluster responsible for the biosynthesis of **11** was identified in 2012 and revealed a *trans*-AT polyketide synthase (PKS) origin (Matilla et al. 2012). This gene cluster could be found in the plant-associated bacteria *Serratia marcescens* MSU97, *S. plymuthica* A153, *S. plymuthica* 4Rx13 and *Dickeya dadantii* Ech703 and production of **11** was verified for all strains (Matilla et al. 2012). The published model for biosynthesis of **11** was not fully conclusive. A reasonable prediction could only be achieved by Matilla *et al.* after more than 1/3 of present domains were hypothesized to be inactive and one full module was hypothesized to be used iteratively. Furthermore, halogenation could not be explained.

*Trans*-acyltransferase (AT) PKSs are modular mega-enzyme-complexes responsible for the step-wise attachment of small, activated acyl-coenzyme A (CoA) building blocks (Keating and Walsh 1999). One PKS module is responsible for one round of elongation, consisting of at least a ketosynthase (KS) domain for chain elongation and an ACP to which the growing chain is tethered throughout the entire elongation (Staunton and Weissman 2001). In the end, the intermediate is usually released from the PKS by the action of a thioesterase (TE) (Du and Lou 2010). Additional domains can be present (Weissman 2009). Further modifications can occur post-PKS like cyclization, glycosylation or halogenation (Hertweck 2009). The main characteristics of *trans*-AT PKSs are the ATs that act as separate proteins instead of the *cis*-AT PKSs, the great diversity of module and domain variations and the presence of unusual domains (Piel 2010).

Here, we report genetic and chemical studies of the gene cluster encoding for the biosynthesis of **11**, identifying genes responsible for chlorination and the isolation of dechloro-oocydin A (**42**). The halogenating enzymes represent an unusual type of halogenase. Furthermore a new analog oocydin B (**43**) was isolated and with that a revision of the published biosynthetic model occurred explaining the biosynthesis of products with different length.



### 3.3 Results

#### 3.3.1 Confirmation of **11** biosynthesis by *S. plymuthica* 4Rx13

This study was performed on the *S. plymuthica* 4Rx13-homolog (Figure 3-1, Table 3-2) of the already published *ooc* gene cluster out of *S. plymuthica* A153 (Matilla et al. 2012). For the following use, host organisms are indicated with either "A153" specifying *S. plymuthica* A153 or "4Rx13" indicating *S. plymuthica* 4Rx13.

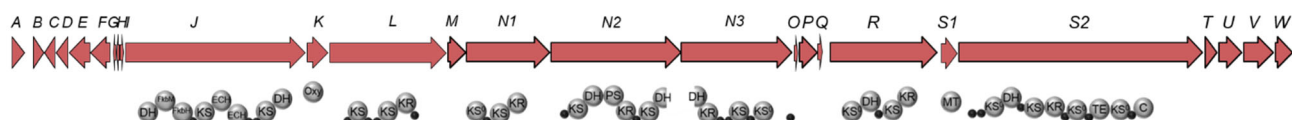


Figure 3-1: Oocydin A biosynthetic gene cluster of *Serratia plymuthica* 4Rx13 with corresponding PKS modules.

| Locus-Tag     | Gene name    | Protein function  |
|---------------|--------------|---|
| SOD_c22810    | <i>oocA</i>  | hypothetical hydrolase  |
| SOD_c22820    | <i>oocB</i>  | hypothetical efflux-protein   |
| SOD_c22830    | <i>oocC</i>  | hypothetical enoyl-CoA hydratase/isomerase involved in $\beta$ -branching           |
| SOD_c22840    | <i>oocD</i>  | hypothetical enoyl-CoA hydratase/isomerase involved in $\beta$ -branching           |
| SOD_c22850    | <i>oocE</i>  | hypothetical 3-hydroxy-3-methylglutaryl-CoA synthase involved in $\beta$ -branching |
| SOD_c22860    | <i>oocF</i>  | hypothetical KS involved in $\beta$ -branching                                      |
| SOD_c22870    | <i>oocG</i>  | hypothetical ACP involved in $\beta$ -branching                                     |
| SOD_c22870    | <i>oocH</i>  | unknown function  |
| SOD_c22870    | <i>oocI</i>  | unknown function  |
| SOD_c22880    | <i>oocJ</i>  | PKS   |
| SOD_c22890    | <i>oocK</i>  | hypothetical flavin-dependent Baeyer-Villiger monooxygenase                         |
| SOD_c22900    | <i>oocL</i>  | PKS   |
| SOD_c22910    | <i>oocM</i>  | hypothetical flavin-dependent monooxygenase   |
| SOD_c22920    | <i>oocN1</i> |   |
| SOD_c22930    | <i>oocN2</i> | PKS   |
| SOD_c22940    | <i>oocN3</i> |   |
| not annotated | <i>oocO</i>  | ACP   |
| SOD_c22950    | <i>oocP</i>  | hypothetical halogenase   |
| SOD_c22960    | <i>oocQ</i>  | hypothetical halogenase   |
| SOD_c22970    | <i>oocR</i>  | PKS   |
| SOD_c22980    | <i>oocS1</i> |   |
| SOD_c22990    | <i>oocS2</i> | PKS   |
| SOD_c23000    | <i>oocT</i>  | unknown function  |
| SOD_c23010    | <i>oocU</i>  | hypothetical enoylreductase   |
| SOD_c23020    | <i>oocV</i>  | fused acyltransferase/acylhydrolase   |
| SOD_c23030    | <i>oocW</i>  | acyltransferase   |

Table 3-2: Proposed functions of *ooc*<sub>4Rx13</sub> biosynthetic gene cluster ORFs.

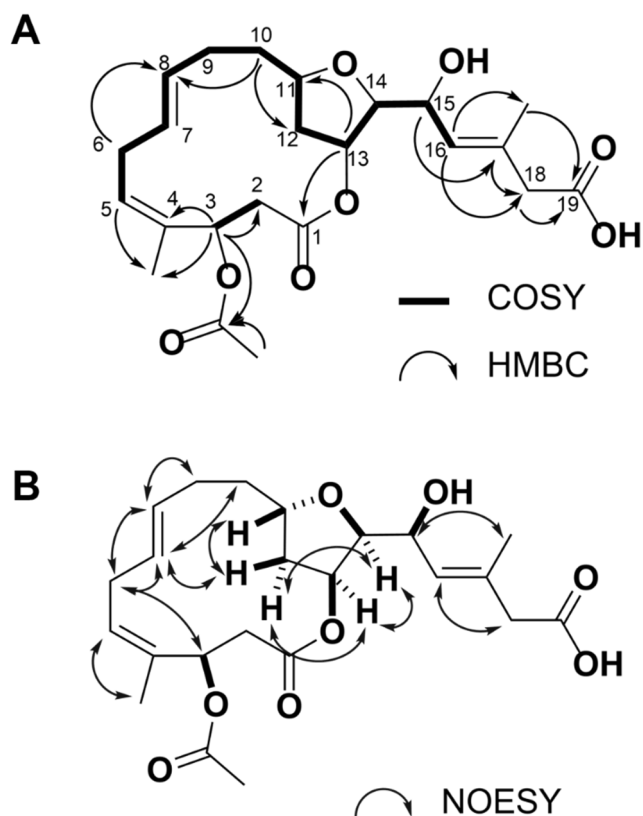
The gene *oocO* was not annotated in the *S. plymuthica* 4Rx13 genome, but is present. The genes *oocN1-oocN3* and *oocS1/oocS2* are annotated as separate genes, but most likely gene borders are frame shifts and they encode for one PKS protein each as it is the case in *ooc* gene cluster from other *Serratia* and *Dickeya*.

We first validated detection of **11** in *S. plymuthica* 4Rx13 as published by Matilla *et al.* (Matilla *et al.* 2012). After optimizing growth conditions a low production (~0.5 mg/L of **11**) could be obtained by preparative HPLC. High-resolution MS (HRMS) showed the isotopic chlorine pattern ( $m/z$  493.1600 predicted, 493.1565 obtained for  $[M+Na^+]^+$ , Figure 3-3A). For verification 1.2 mg **11** ( $C_{23}H_{31}ClO_8$ ) were isolated and confirmed by NMR analyses (data not shown). To unequivocally confirm that *oocR*<sub>4Rx13</sub> is responsible for **11** production a PKS encoding gene *oocR*<sub>4Rx13</sub> knock out strain 4Rx13 $\Delta$ *oocR* was generated by  $\lambda$ RED mediated replacement of the gene *oocR*<sub>4Rx13</sub> with an apramycin antibiotic-cassette. As expected, a complete abolishment of **11** production was detected with HPLC-HRMS.

### 3.3.2 Assigning Function to Halogenase Candidates by Gene Inactivation

We were intrigued by the vinyl chloride at C<sub>8</sub> of the macrolide ring, an unusual structural feature for polyketides. Previously, the two proposed flavin-dependent proteins *oocK*<sub>A153</sub> and *oocM*<sub>A153</sub> were suggested to be involved in chlorination because corresponding halogenases react on electron-rich substrates (Blasiak and Drennan 2008) but knock out experiments could not confirm this hypothesis as dechloro **11** was not detected (Matilla *et al.* 2012). In addition the typical FAD-dependent halogenase sequence motif GWxWxIP is not present in either of these proteins (Zehner *et al.* 2005). We hypothesized that one of the proteins with so far unknown function as annotated by Matilla *et al.* (Matilla *et al.* 2012) might be involved in halogenation reaction as no homology of any of the other proteins out of this gene cluster to known halogenases is reported. We speculated that rather *oocP*<sub>4Rx13</sub>, *oocQ*<sub>4Rx13</sub> or *oocT*<sub>4Rx13</sub> are responsible for halogenation as *oocH*<sub>sp</sub> or *oocI*<sub>sp</sub> are of very small size (30 and 51 aa respectively). To test our hypothesis *oocP*<sub>4Rx13</sub>, *oocQ*<sub>4Rx13</sub> and *oocT*<sub>4Rx13</sub> genes were knocked out individually by  $\lambda$ RED-mediated replacement with a kanamycin antibiotic-cassette and cultures of corresponding strains were extracted. We compared HPLC-HRMS runs of wild type and knock out culture extracts with the open-source software eMZed (Kiefer, Schmitt, and Vorholt 2013). This python-based software generated tables of  $m/z$  values extracted from the HPLC-HRMS runs and compared tables for runs of wild type and knock out strain with each other to generate candidate  $m/z$  values representing differences in metabolite production. This procedure is similar to principal component analysis (PCA). When culture extracts of 4Rx13 $\Delta$ *oocT* were analyzed, no difference was observed when compared to the wild type. These results suggest that *oocT*<sub>4Rx13</sub> is not employed in chlorination and might not catalyze an essential step in the biosynthesis of **11**. Analysis of 4Rx13 $\Delta$ *oocP* allowed detection of a mass correlated to dechloro **11** ( $m/z$  459.1989 predicted, 459.1964 obtained for  $[M+Na^+]^+$ ) as proposed above. The compound with the mass  $m/z$  459.1989 eluted at the same retention time as the chlorinated compound (Figure 3-7). It does not show an isotope pattern characteristic for a chlorine atom but an isotope pattern of non-halogenated compounds. Unexpectedly the same mass was detected in 4Rx13 $\Delta$ *oocQ*. Functional complementations were performed to exclude downstream effects of the antibiotic cassette in the knock out strains. For these experiments rescue plasmids were cloned. The resulting plasmids carried either *oocP*<sub>4Rx13</sub> or *oocQ*<sub>4Rx13</sub> or both genes (subsequently named pKJ55). Each plasmid was transformed into 4Rx13 $\Delta$ *oocP* and 4Rx13 $\Delta$ *oocQ*, the empty vector served as control. Chlorination was restored in 4Rx13 $\Delta$ *oocQ* + *oocQ*<sub>4Rx13</sub>, 4Rx13 $\Delta$ *oocP* + *oocP*<sub>4Rx13</sub>, 4Rx13 $\Delta$ *oocP* + pKJ55 and in 4Rx13 $\Delta$ *oocQ* + pKJ55. However if *oocP*<sub>4Rx13</sub> was transformed into 4Rx13 $\Delta$ *oocQ* or vice versa no chlorination was detectable. No functional complementation was observed in the control experiments, thus a polar effect was unlikely. In addition, the observed results showed the necessity of both proteins for activity. Since an ion with a mass consistent with dechloro-oocydin A was detected in 4Rx13 $\Delta$ *oocP* and 4Rx13 $\Delta$ *oocQ* we isolated this compound from 1.2 L fermentation broth of 4Rx13 $\Delta$ *oocP* after ethyl acetate extraction and two subsequent rounds of HPLC to give 0.6 mg compound and 1D and 2D NMR spectra were obtained (Figure 3-8 - Figure 3-12). The NMR spectra were nearly identical to those of **11** (Roulland 2008), (Thaning *et al.* 2001), but a vinyl proton ( $\delta_H$  5.35) was present at C<sub>8</sub> ( $\delta_C$  128.3) instead of no signal due to the attached chloride ion in **11**. The structure of the proposed dechloro-oocydin A **42** (Figure 3-2) was consistent with the NMR data (Table 3-6). The stereochemistry of the double bonds

were assigned based on NOESY correlations to have *4Z*, *7E* and *16E* configuration. A relative stereochemistry of the protons at the tetrahydrofuran ring was proposed based on NOESY correlations (Figure 3-2B) and is comparable to earlier proposals for haterumalide NA (Takada et al. 1999). Additionally, relative stereochemistry of C<sub>3</sub>, C<sub>11</sub>, C<sub>13</sub>, C<sub>14</sub> and C<sub>15</sub> was deduced from described relative stereochemistry (Kigoshi et al. 2003) in comparison with the above mentioned NOESY correlations.



**Figure 3-2: Key correlations of dechloro-oocydin A 42.**

(A) Key COSY and key HMBC correlations of 42. (B) Key NOESY correlations of 42 with deduced relative stereochemistry.

### 3.3.3 Bioinformatic analysis of OocP<sub>4Rx13</sub> and OocQ<sub>4Rx13</sub>

To gain more information about previously unknown proteins OocP<sub>4Rx13</sub> and OocQ<sub>4Rx13</sub>, they were subjected to bioinformatic analysis. OocP<sub>4Rx13</sub> was reported to exhibit 25% identity to the hypothetical protein BuboB\_26000 (GenBank accession number ZP\_02381200) (Matilla et al. 2012). BuboB\_26000 was also reported to be the closest protein for Bamb\_5931 (GenBank accession number YP\_777810.1) of the enacyloxin biosynthetic pathway and hypothesized to be responsible for halogenation with the putative function as a MNH chlorinase (Mahenthalingam et al. 2011). This indicates a weak homology of OocP<sub>4Rx13</sub> to MNH halogenases. Additionally, the conserved domain analysis of BLAST (Marchler-Bauer et al. 2011) detected a cupin-like domain and with that putatively classifies OocP<sub>4Rx13</sub> into the JmjC superfamily. Cupin-like domains are characteristic of MNH halogenases, but not haloperoxidases, SAM-, or flavin-dependent halogenases (Hangasky et al. 2013). In order to further elucidate the homology of OocP<sub>4Rx13</sub> to MNH halogenases, a sequence comparison was performed with described MNH halogenases SyrB2, CurA, CytC3, KtzD, BarB2, DysB2, JamE and HctB (Vaillancourt, Yin, and Walsh 2005), (Khare et al. 2010), (Galonić et al. 2007), (Galonić, Vaillancourt, and Walsh 2006), (Gu et al. 2009), (Pratter et al. 2014a), (Fujimori et al. 2007), (Ridley et al. 2005) and human protein hydroxylase Jmjd5 as a representative of other JmjC-domain containing MNH enzymes (Wang et al. 2013), (del Rizzo, Krishnan, and Trievel 2012). A standard sequence

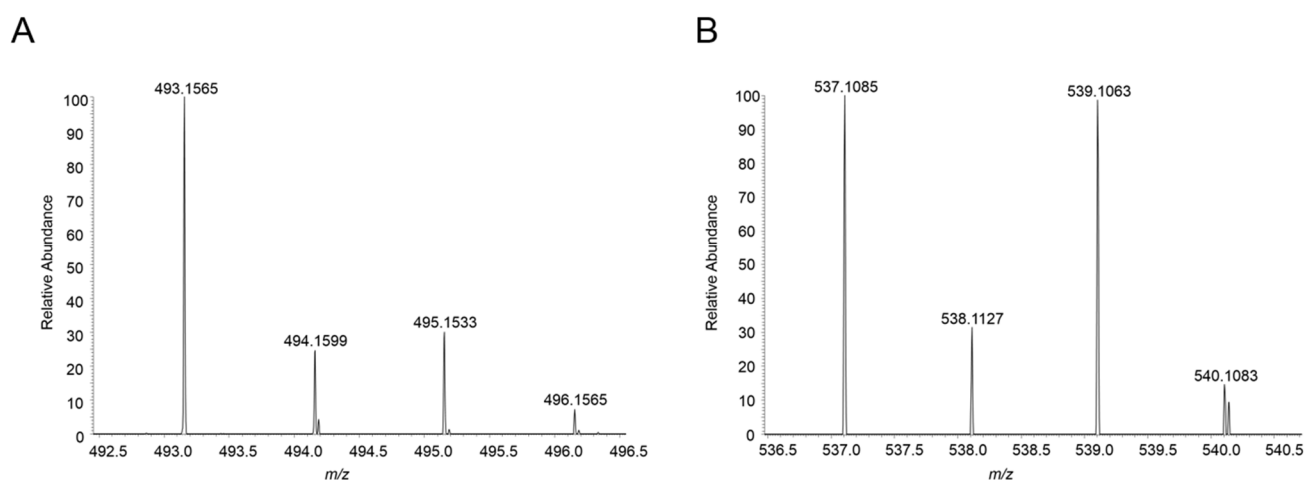
alignment using the T-Coffee web server (Di Tommaso et al. 2011) and a structural alignment applying the Expresso server (Armougom et al. 2006) were carried out (Figure 3-18, Figure 3-19). For a better comparison of residues involved in reaction and cofactor binding the results of the alignments were annotated in a similar manner as Hangasky *et al.* (Hangasky et al. 2013) suggested for positions of interest. Additional important residues for reaction or cofactor binding known from X-ray crystal structure of the MNH halogenases SyrB2 and CurA (Blasiak et al. 2006), (Khare et al. 2010), (Wong et al. 2013) were taken into account. It is worth noting that OocP<sub>4Rx13</sub> does not align very well in either sequence or structural alignments and shows the lowest overall scores. In structural alignment it showed a lower score than in the sequence alignment (Figure 3-18 and Figure 3-19). Nevertheless, the facial triad (position F1-5, F1-7, F2-3 in the annotation after Hangasky *et al.*, Table 3-7) is nearly identical in all enzymes. For MNH halogenases one of the most important residues is described to be F1-7 as this amino acid is mutated in all described MNH halogenases from E or D to small hydrophobic amino acids in comparison to MNH oxygenases. At this position OocP<sub>4Rx13</sub> is consistent with the other halogenases aligned here. As can be seen in Table 3-7 few other residues show similarity as well, but divergence can also be detected in various positions. These results together with the conserved domain detection by BLAST indicate the existence of a homology of OocP<sub>4Rx13</sub> to known MNH halogenases although the overall alignment shows weak correlation and the influence of diverging residues onto the function of OocP<sub>4Rx13</sub> needs to be clarified.

Matilla *et al.* reported the highest similarity of OocQ<sub>4Rx13</sub> to SpiBuddy\_0991 (NCBI accession number YP\_004247012.1) from the spirochaete *Sphaerochaeta globus* st. Buddy with 25 % identity. SpiBuddy\_0991 is annotated in the NCBI database as a condensation-domain containing protein. Additional DELTA-BLAST analysis (Boratyn et al. 2012) revealed a weak homology to a glycerophosphodiester phosphodiesterase (NCBI accession number WP\_027529496.1; E-value: 5e<sup>-11</sup>; 18% identity) from *Bradyrhizobium* sp. WSM3983. Conserved domain analysis detected two potential domain hits. The first hit is GDPD\_SpGDE\_like-domain, containing uncharacterized bacterial glycerophosphodiester phosphodiesterase and similar proteins (NCBI accession number cd08567) (Tommassen et al. 1991) with an E-value of 0.02. The second hit is a provisional inorganic polyphosphatase/ATP-NAD kinase domain (NCBI accession number PRK02645) (Garavaglia et al. 2004), (Kawai et al. 2000) with an E-value of 0.04. The protein sequence of OocQ<sub>4Rx13</sub> analyzed by the KEGG SSDB ortholog search (Kanehisa et al. 2002) revealed 30.8 % identity (SW-score: 107) with a prolycopene isomerase, a *cis-trans* isomerase from *Oryza brachyantha* (NCBI accession number XP\_006663539.1). Due to the very low homologies no reasonable hypothesis can be postulated regarding function of OocQ<sub>4Rx13</sub>.

### 3.3.4 Clarification of OocP<sub>4Rx13</sub> and OocQ<sub>4Rx13</sub> influence on the halogenation reaction

To obtain further insight into the role of OocP<sub>4Rx13</sub> and OocQ<sub>4Rx13</sub> in halogenation of **11** it was of interest to know if  $\alpha$ -KG is required for OocP<sub>4Rx13</sub> by employing an  $\alpha$ -KG inhibitor. At present, all described MNH halogenases employ  $\alpha$ -KG as a cofactor. A structural mimic of  $\alpha$ -KG, 2,4-pyridinecarboxylic acid (2,4-PDCA, also known as lutidinic acid) (Rose et al. 2011) was chosen. The inhibitor was added at various concentrations (5-333 mM) to cultures of the wild type and incubated for 48 hours. However, no inhibition was detected and production of **11** was neither abolished nor diminished. A possible explanation might be that the inhibitor is not able to enter the cell or the enzyme is too divergent to allow for inhibition.

To test the substrate range of the halogenase, *S. plymuthica* 4Rx13 was individually cultured in media supplemented with different halogen sources (sodium bromide, sodium iodide or sodium fluoride). Only cultures grown in the presence of sodium bromide showed a *m/z* value with isotope pattern corresponding to brominated **11** (*m/z* 537.1095 predicted, 537.1085 obtained for [M+Na]<sup>+</sup>; Figure 3-3B). As a control, PKS knock out strain 4Rx13 $\Delta$ oocR was grown under the same conditions and no brominated compound could be detected. No incorporation of either iodine or fluorine was detectable in the wild type and in 4Rx13 $\Delta$ oocR. Small amounts of chlorinated **11** was observed in all extracts most likely due to unavoidable contamination with trace chloride. This phenomenon is also described in other studies (Pratter et al. 2014a), (Vaillancourt, Vosburg, and Walsh 2006).



**Figure 3-3: Isotopically resolved ion pattern of oocydin A 11.**

(A) Mass spectrum of **11** ( $C_{23}H_{31}ClO_8$ ) with the typical chlorine pattern 3:1 ( $m/z$  493.1600 predicted, 493.1565 obtained for  $[M+Na]^+$ ). (B) Mass spectrum of brominated **11** ( $C_{23}H_{31}BrO_8$ ) with the typical bromine pattern 1:1 ( $m/z$  537.1095 predicted, 537.1085 obtained for  $[M+Na]^+$ ).

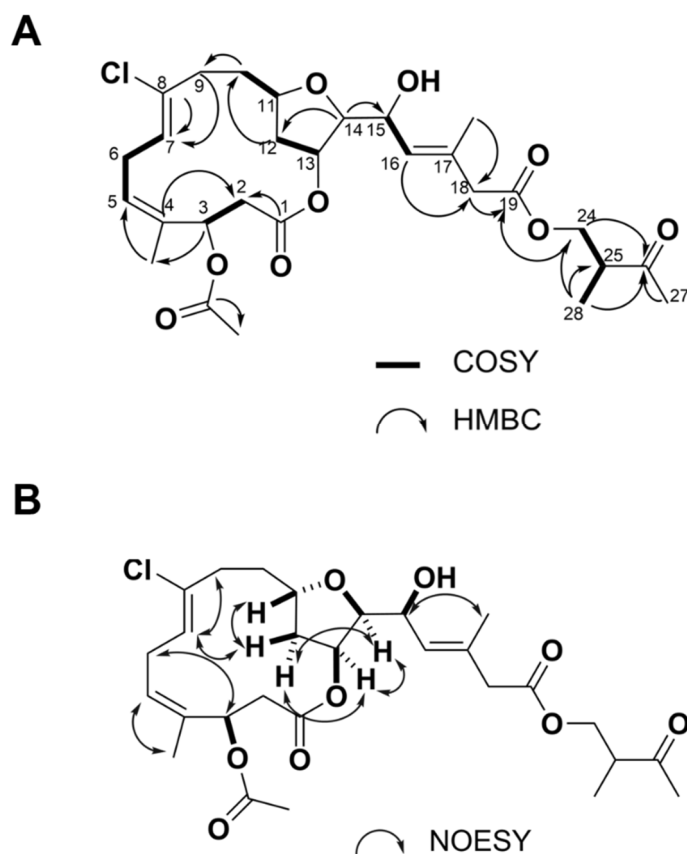
To clarify whether  $OocP_{4Rx13}$  and  $OocQ_{4Rx13}$  act in parallel or successively culture extracts of  $4Rx13\Delta oocP$  and  $4Rx13\Delta oocQ$  were subjected to HPLC-HRMS measurements and corresponding data were compared with using the eMZed software to detect differences between  $4Rx13\Delta oocP$  and  $4Rx13\Delta oocQ$  culture extracts. In case of a subsequent action of  $OocP_{4Rx13}$  and  $OocQ_{4Rx13}$  potential reaction intermediates should be detectable in the culture extracts of one knock out strain but not within the other. No difference between both culture extracts would be expected in case of a parallel action of  $OocP_{4Rx13}$  and  $OocQ_{4Rx13}$ . No significant differences between both knock out strains were observed. Cross-complementation experiments were carried out for further clarification. For this purpose the supernatant of a  $4Rx13\Delta oocP$  culture was added to  $4Rx13\Delta oocQ$  cells and reverse. However, no chlorinated species were detected in the corresponding culture extracts. The same assays performed with  $4Rx13\Delta oocR$ ,  $4Rx13\Delta oocP$  and  $4Rx13\Delta oocQ$  did not yield detectable chlorinated products either. Ions corresponding to **42** could be detected in the HPLC-HRMS data of the cell fraction leading to the assumption that a transport of precursor compounds from the supernatant into the cells takes place. It cannot fully be excluded that the detection of **42** in the cell fraction results from compounds sticking to the cell membrane rather than entering the cell. Nevertheless the data suggest that both proteins have to be present at the same time for halogenation to occur.

Further assays were conducted to further elucidate the interplay between  $OocP_{4Rx13}$  and  $OocQ_{4Rx13}$ . A substrate feeding assay was carried out in which  $OocP_{4Rx13}$  and  $OocQ_{4Rx13}$  were cloned with either an N-terminal hexahistidine-tag or a C-terminal Strep-tag (Table 3-4). Two additional constructs were generated carrying both genes in the same vector. One construct harbors  $OocP_{4Rx13}$  with an N-terminal hexahistidine-tag and  $OocQ_{4Rx13}$  with a C-terminal Strep-tag and vice versa in the second construct. For whole-cell feeding experiments proteins were expressed but not purified. Instead of purification, culture extract of  $4Rx13\Delta oocP$  (5-15 mg) were added to the *E. coli* expression cultures. These supplemented cultures were incubated for 48 hrs prior to extraction and HPLC-HRMS analysis. This feeding assay was performed with either tagged  $OocP_{4Rx13}$ , or tagged  $OocQ_{4Rx13}$  or both double constructs each carrying both tagged proteins. A combination of cultures of expressed tagged  $OocP_{4Rx13}$  and  $OocQ_{4Rx13}$  proteins in a 1:1 ratio was obtained as well. In all extracts neither **11** nor any chlorinated compound could be detected, but **42**, as fed substrate, was present in every case. To exclude that the fed culture extract cannot enter the cell and resulting from that no subsequent chlorination can occur, the same assays were conducted with lysed cells. Therefore aliquots of induced cultures were transferred into new tubes, also supplemented with  $4Rx13\Delta oocP$

culture extract, tightly sealed and subsequently sonicated in a water bath. After sonication, assays were incubated for 48 hrs as well and extracted for HPLC-HRMS analysis. The results were identical to those for the whole-cell feeding assays meaning no chlorination was detectable. To exclude an influence of media components as expression was carried out in LB medium but cultures were fermented in EPB medium, *OocP*<sub>4Rx13</sub> and *OocQ*<sub>4Rx13</sub> were separately expressed, purified and applied in an *in vitro* assay modified after Eustáquio *et al.* (Eustáquio *et al.* 2014). The culture extract of *4Rx13ΔoocP* served as potential substrate material. Again no halogenation was detectable but the detection of **42** was observed in every assay. The negative results in the *in vitro* assays might be resulted from a potential oxygen lability of *OocP*<sub>4Rx13</sub> and probably *OocQ*<sub>4Rx13</sub>. MNH enzymes are often known to be oxygen sensitive and therefore purified under anaerobic conditions (Vaillancourt *et al.* 2005), but not exclusively (Pratter *et al.* 2014b). Our experiments were conducted under aerobic conditions. This oxygen sensitivity does not explain negative results in the whole-cell feeding experiments. Further experiments have to be conducted to establish a system that halogenation can occur outside of the native host. This is a prerequisite for more targeted studies of the role of both proteins in halogenation and understanding of this reaction.

### 3.3.5 Detection of additional biosynthetic products from the *ooc*<sub>4Rx13</sub> gene cluster

Several members of the haterumalide family (Table 3-1) are often extracted together from the same producer. Teruya *et al.* reported the isolation of **28** together with **37-41** from the marine ascidian *Didemnidae* sp. and Levenfors *et al.* isolated **28**, **34**, **35** and haterumalide X (**36**), the *Z* isomer of **28** from *S. plymuthica* A153 (Levenfors *et al.* 2004), (Teruya *et al.* 2005). In order to test if *S. plymuthica* 4Rx13 is able to produce additional members of the haterumalide family and if the production is linked to the *ooc*<sub>4Rx13</sub> gene cluster, HPLC-HRMS data of the general PKS knock out strain *4Rx13ΔoocR* culture extracts were compared with corresponding HPLC-HRMS data of wild type culture extracts with the help of the eMZed software. As expected, **11** production was abolished in the knock out. Beside **11**, an additional mass peak with a chlorinated isotopic pattern ( $m/z$  577.2167 [M+Na<sup>+</sup>]<sup>+</sup>) was not apparent in *4Rx13ΔoocR* strongly indicating that the compound is also associated with the *ooc*<sub>4Rx13</sub> gene cluster (Figure 3-7). We were able to isolate the compound from 1.8 L fermentation broth via two subsequent rounds of HPLC to gain 0.5 mg for 1D and 2D NMR analysis (Figure 3-4). The NMR spectra revealed the close structural relation to **11** as signals for the macrolactone- and tetrahydrofuran ring were nearly identical. The attached acetyl-moiety at C<sub>3</sub> was present in the newly isolated compound as well. The linear side chain was the same until C<sub>19</sub>. Additional peaks in NMR spectra revealed the presence of an additional esterified side chain. The new isolated compound has only a mass difference of 2 Da to the known congener haterumalide B **35** (C<sub>28</sub>H<sub>37</sub>ClO<sub>9</sub>,  $m/z$  575.2018 predicted for [M+Na<sup>+</sup>]<sup>+</sup>) indicating a structural similarity that could be confirmed by the NMR data. The only difference in NMR spectra between the new isolated compound and **35** is the carbon shift of C<sub>28</sub>. In **35** a δ<sub>H</sub> 6.13/6.34 and δ<sub>C</sub> 126.9 for the exomethylen-group is reported (Levenfors *et al.* 2004). At this position δ<sub>H</sub> 1.13 and δ<sub>C</sub> 13.3 were detected in the new isolated compound verifying a methyl group at this position and saturation of the double bond. The new compound was named oocydin B **43** (C<sub>28</sub>H<sub>39</sub>ClO<sub>9</sub>,  $m/z$  577.2175 predicted for [M+Na<sup>+</sup>]<sup>+</sup>; corresponding 1D- and 2D-NMR spectra: Figure 3-13 - Figure 3-17, Table 3-6). It exhibits almost the same side chain as **35** and **38**. A relative stereochemistry of the protons at the tetrahydrofuran ring was proposed based on NOESY correlations (Figure 3-4B) and is comparable to **42**. Additionally, the relative stereochemistry of C<sub>3</sub>, C<sub>11</sub>, C<sub>13</sub>, C<sub>14</sub> and C<sub>15</sub> was deduced from **42** in comparison with the above mentioned NOESY correlations. The double bonds were assigned based on NOESY correlations to have *4Z*, *7Z* and *16E* configuration. HPLC-HRMS showed  $m/z$  values correlating to dechloro **43** ( $m/z$  543.2564 predicted, 543.2537 obtained for [M+Na<sup>+</sup>]<sup>+</sup>) in *4Rx13ΔoocP* and *4Rx13ΔoocQ* culture extracts. In addition, masses corresponding to brominated **43** could be detected after culturing the wild type strain in sodium bromide supplemented media ( $m/z$  621.1669 predicted, 621.1638 obtained for [M+Na<sup>+</sup>]<sup>+</sup>). No other masses corresponding to additional members of the haterumalide family could be detected in *4Rx13ΔoocR* HPLC-HRMS data.



**Figure 3-4: Key correlations of oocydin B 43.**

(A) Key COSY and key HMBC correlations of **43**. (B) Key NOESY correlations of **43** with deduced relative stereochemistry.

### 3.3.6 Cytotoxic properties of oocydin A **11** and its analogs

Since cytotoxic activities for **11**, **31** and other members of the haterumalide family were previously reported (Ueda et al. 2009), (Ueda and Hu 1999), (Teruya et al. 2005), (Takada et al. 1999), the compounds isolated here **11**, **42** and **43** were subjected to cytotoxicity assays against HeLa cells and  $IC_{50}$  values were determined. Oocydin A **11** exhibited highest activity ( $IC_{50}=0.38 \mu M$ ), followed by **42** ( $IC_{50}=1.54 \mu M$ ) and **43** ( $IC_{50}=7.05 \mu M$ ). The chlorine atom appears to play only a minor role in cytotoxic activity. An impact of the aliphatic side chain and its length was reported before, but in previous studies at least **11** and **35** had almost the same cytotoxic effect. Taken that into account, the exomethylen-group seems to be important for anti-cancer activity.

## 3.4 Discussion

Our studies on the *ooc*<sub>4Rx13</sub> gene cluster showed production of **11** in *S. plymuthica* 4Rx13 and revealed **42** and **43** as a new member of the haterumalide family. We were able to assign the two genes *oocP*<sub>4Rx13</sub> and *oocQ*<sub>4Rx13</sub> as involved in chlorination of **11** by generation of individual knock out strains and functional complementation experiments. It was demonstrated that both genes have to be present for the chlorination reaction. By substituting media components with a bromine source, bromine it could be introduced into **42**, however fluorine and iodine could not. The non-chlorinated compound **42** was isolated and the structure elucidated. Cytotoxicity of **42** and **43** was shown.

OocP<sub>4Rx13</sub> was identified as distant homolog of MNH enzymes. No information regarding function of OocQ<sub>4Rx13</sub> could be obtained.

We could show that both proteins OocP<sub>4Rx13</sub> and OocQ<sub>4Rx13</sub> need to be present for halogenation as confirmed by functional complementation experiments. For some halogenation reactions it is known that two proteins are required. In the case of flavin-dependent halogenases enzyme pairs are described where both proteins are necessary. In RebF/RebH one component is the halogenase, the other is responsible for reduction of the FADH cofactor (Yeh, Garneau, and Walsh 2005). In barbamide biosynthesis a halogenase pair is described, but both proteins are MNH halogenases. Here BarB2 performs a double chlorination and BarB1 installs the third chlorine atom (Galonić, Vaillancourt, and Walsh 2006). Our data suggest that OocP<sub>4Rx13</sub> is a homolog of MNH enzymes. An alignment of OocP<sub>4Rx13</sub> with known MNH halogenases shows comparable amino acid residues within positions most important for the reaction but low overall conservation. To date, described MNH enzymes responsible for halogenation are single enzymes and depend on  $\alpha$ -KG as a cofactor. Here in the case of halogenation of **11** a second protein has to be present, otherwise no halogen atom is incorporated into the molecule. The requirement for a second protein for an MNH-like halogenase homolog is unprecedented. Only OocQ<sub>4Rx13</sub> which shows distant homology to other proteins, e.g., glycerophosphodiester phosphodiesterase, or isomerase allows us to speculate about its function and the scenario of OocP<sub>4Rx13</sub>-OocQ<sub>4Rx13</sub> interaction.

Determination of the cognate substrate that gets halogenated was not possible yet. It is likely that **42** is the substrate, but a saturated species of **42** might also be the conceivable substrate (Figure 3-5A). Most of the knowledge regarding MNH halogenases has been acquired from enzyme assays, protein crystallization or computational studies. The isolation of dechloro-oocydin A (**42**) as a potential substrate of an MNH enzyme homolog with the double bond present between C<sub>7</sub>-C<sub>8</sub> was surprising. MNH enzymes are known to react via a radical mechanism (Figure 3-5E2) and in the case of oocydin A (**11**), a radical would be formed at a sp<sup>2</sup> carbon. These vinyl radicals are highly energetic and unstable (Hioe and Zipse 2011). Nevertheless it is possible that double bonds are substrates for MNH enzymes (Ishikawa et al. 2014) and we suggest a reaction mechanism (Figure 3-5E3) for this scenario. We were able to observe traces of **42** in culture extracts of the wild type and 4Rx13 $\Delta$ ooc7 suggesting that **42** is the substrate. An alternate substrate is possible as well, for example a C<sub>7</sub>-C<sub>8</sub> saturated species. This would imply that this saturated species is halogenated and subsequently desaturated. A similar scenario was reported in the case of jamaicamide A & B and curacin A biosynthesis (Gu et al. 2009) and hectochlorine biosynthesis (Pratter et al. 2014a). In the case of jamaicamide A & B and curacin A biosynthesis an enzyme sequence containing the domain series ...AT-Hal-ACP-ACP-ACP-ACP-KS-HCS-ECH<sub>1</sub>-ECH<sub>2</sub>... is responsible for halogenation or cyclopropane formation where desaturation takes place after halogenation. In this case, the components usually involved in  $\beta$ -branching (namely ACP-KS-HCS-ECH<sub>1</sub>-ECH<sub>2</sub>) are able to catalyze additional reactions than the classical  $\beta$ -branching but are not engaged in any further  $\beta$ -branching reaction. In the case of oocydins a jamaicamide-type reaction sequence is less probable. Here the ACP-KS-HCS-ECH<sub>1</sub>-ECH<sub>2</sub> cassette is most likely responsible for the  $\beta$ -branching at C<sub>17</sub> but the cassette might act twice and with that performs desaturation after halogenation as well. The jamaicamide-like scenario would mean for **42** halogenation that the halogenation is not prerequisite for a subsequent desaturation event but that both reactions can occur in parallel as we were able to detect traces of **42** in the wild type culture extract as well as mentioned above. But no mass peaks corresponding to saturated versions of **11** or **42** were detected. In the case of hectochlorin a vinyl-chloride moiety can be generated by the MNH halogenase HctB in a postulated two-step reaction. The first step would be a halogenation and the second step a desaturation. The proposed reaction sequence of HctB cannot explain the presence of **42** as desaturation is hypothesized to occur post halogenation. The hypothesis that **42** is the real substrate is further supported by KS substrate specificity. KSs in *trans*-AT PKSs are highly substrate specific (Nguyen et al. 2008), (Jenner et al. 2013). It is unlikely that KSs accept two different moieties resulting in a saturated and a desaturated double bond.

One question regarding OocP<sub>4Rx13</sub>-OocQ<sub>4Rx13</sub> interplay is whether both proteins catalyze the halogenation reaction together or if one protein generates the substrate for the second protein in a sequence of events (Figure 3-5B). By



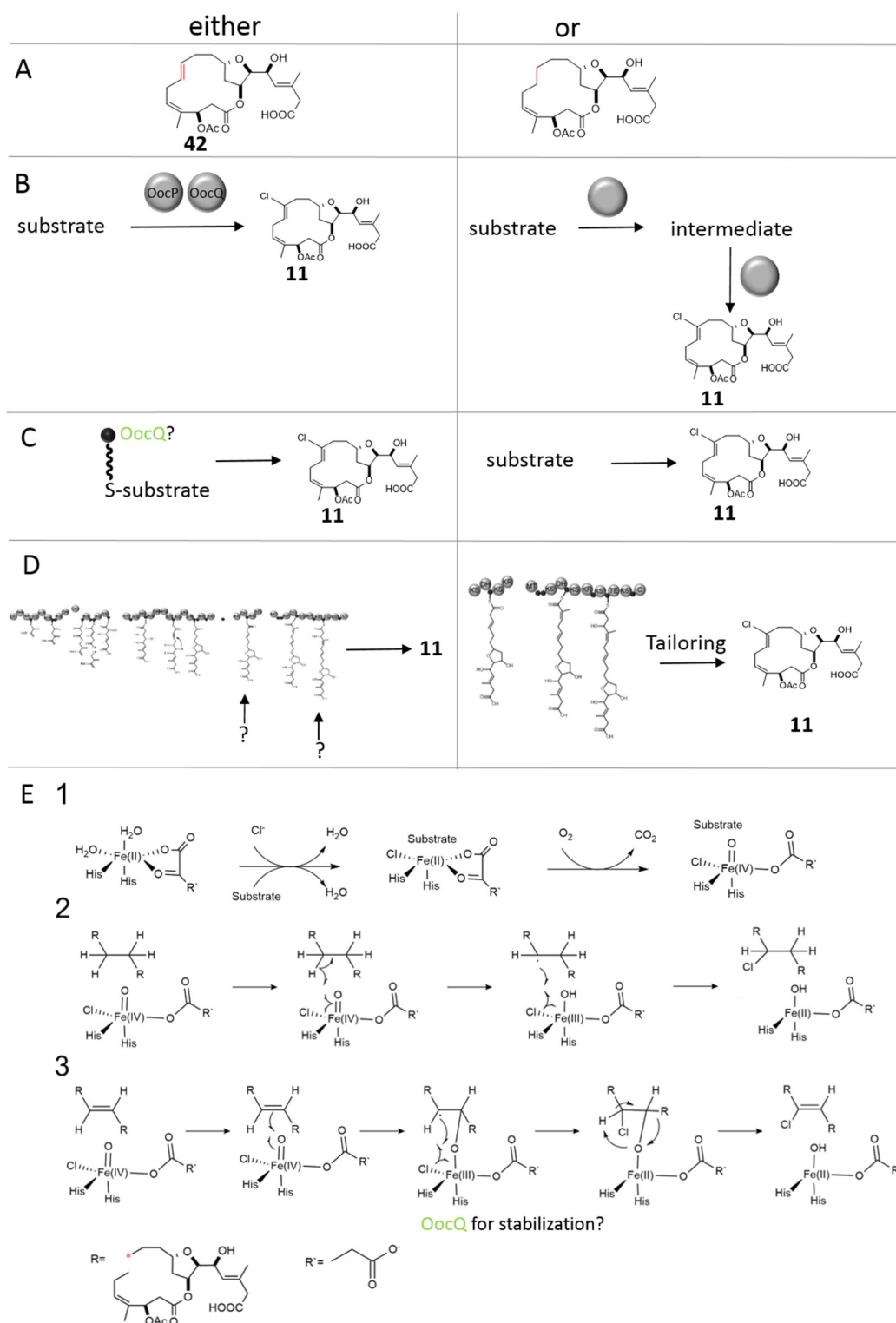
cross-complementation and comparison of 4Rx13 $\Delta$ *oocP* and 4Rx13 $\Delta$ *oocQ* culture extracts it does not seem to be a sequential reaction. No differential products of a first reaction step could be detected in culture extract comparison. Additionally, we would expect to observe halogenation in cross-complementation experiments as intermediates should have been processed.

We can only speculate if the halogenation reaction is carried out on a free or a protein bound substrate (Figure 3-5C). All described MNH halogenases react on a CP-tethered substrate and only in the case of the alkaloid welwitindolinone no CP is present for halogenation by WelO5 (Hillwig and Liu 2014). In the case of HctB no separate ACP is necessary as well but the CP-moiety is included in the three-domain halogenating enzyme (Pratter et al. 2014a). Our data do not suggest an inclusion of a CP moiety neither in *OocP*<sub>4Rx13</sub> nor in *OocQ*<sub>4Rx13</sub> as no homologies of *OocQ*<sub>4Rx13</sub> or domains of *OocP*<sub>4Rx13</sub> towards any CPs could be established. Without any CP homology a chlorination on a free substrate as in the case of WelO5 is imaginable. If halogenation takes place on a free or a CP-bound substrate, it is unknown if the event occurs as a tailoring step in the end or while the growing chain is still tethered to the PKS (Figure 3-5D). Beside WelO5, all other characterized MNH halogenases perform their reaction on the elongating intermediate and predominantly in the beginning of the biosynthesis. We could not detect halogenated intermediates in culture extracts of the general PKS knock out strain 4Rx13 $\Delta$ *oocR* that would have been likely in an early chlorination. It is also possible that intermediates were not released from the PKS in 4Rx13 $\Delta$ *oocR*. It might be possible that halogenation would happen downstream of *oocR*<sub>4Rx13</sub>-action and no chlorination would be detectable in 4Rx13 $\Delta$ *oocR* anyways.

Our data suggest that either the halogenase acts in one of the last biosynthetic steps. The interplay between *OocP*<sub>4Rx13</sub> and *OocQ*<sub>4Rx13</sub> remains enigmatic so far. We cannot exclude an alternate way of halogen attachment as homologies of both proteins were very weak.

Further experiments are needed to exclude some of the aforementioned scenarios and shed light on the functions of both proteins. A first step could be the analysis of probable *OocP*<sub>4Rx13</sub> cofactors. All MNH halogenases characterized so far depend on  $\alpha$ -ketoglutarate. But not every MNH enzyme relies on  $\alpha$ -KG (other possibilities are pterin, ascorbate, Rieske cluster or absence of a cofactor (Buongiorno and Straganz 2013)). All previous mentioned MNH enzymes show homology on primary sequence level and exhibit the same facial triad (Koehtop, Emerson, and Que 2005). Our first trials to inhibit  $\alpha$ -KG with 2,4-PDCA did not result in a detectable inhibition of the enzyme and with that the usage of  $\alpha$ -KG as cofactor can not be concluded from this experiments. A functioning *in vitro* halogenation would be a good system to test for several potential cofactors. So far we were only able to restore halogenation in the native host with rescue plasmids but not heterologously with expression of proteins or whole-cell halogenation approaches. One or both of the proteins might be oxygen labile as described for many, but not all MNH enzymes (Vaillancourt et al. 2005), (Pratter et al. 2014b), (Ishikawa et al. 2014). Anaerobic experiments might help to overcome halogenation hurdles *in vitro*. It is also possible that additional factors are required for the halogenation which are usually delivered by the native host *Serratia*. In whole-cell halogenation trials, oxygen sensitivity of proteins cannot explain why halogenation did not occur. Perhaps alternate substrates are necessary for the proteins and no subsequent halogenation was possible but it is reasonable to estimate a substrates presence in 4Rx13 $\Delta$ *oocP* culture extracts.

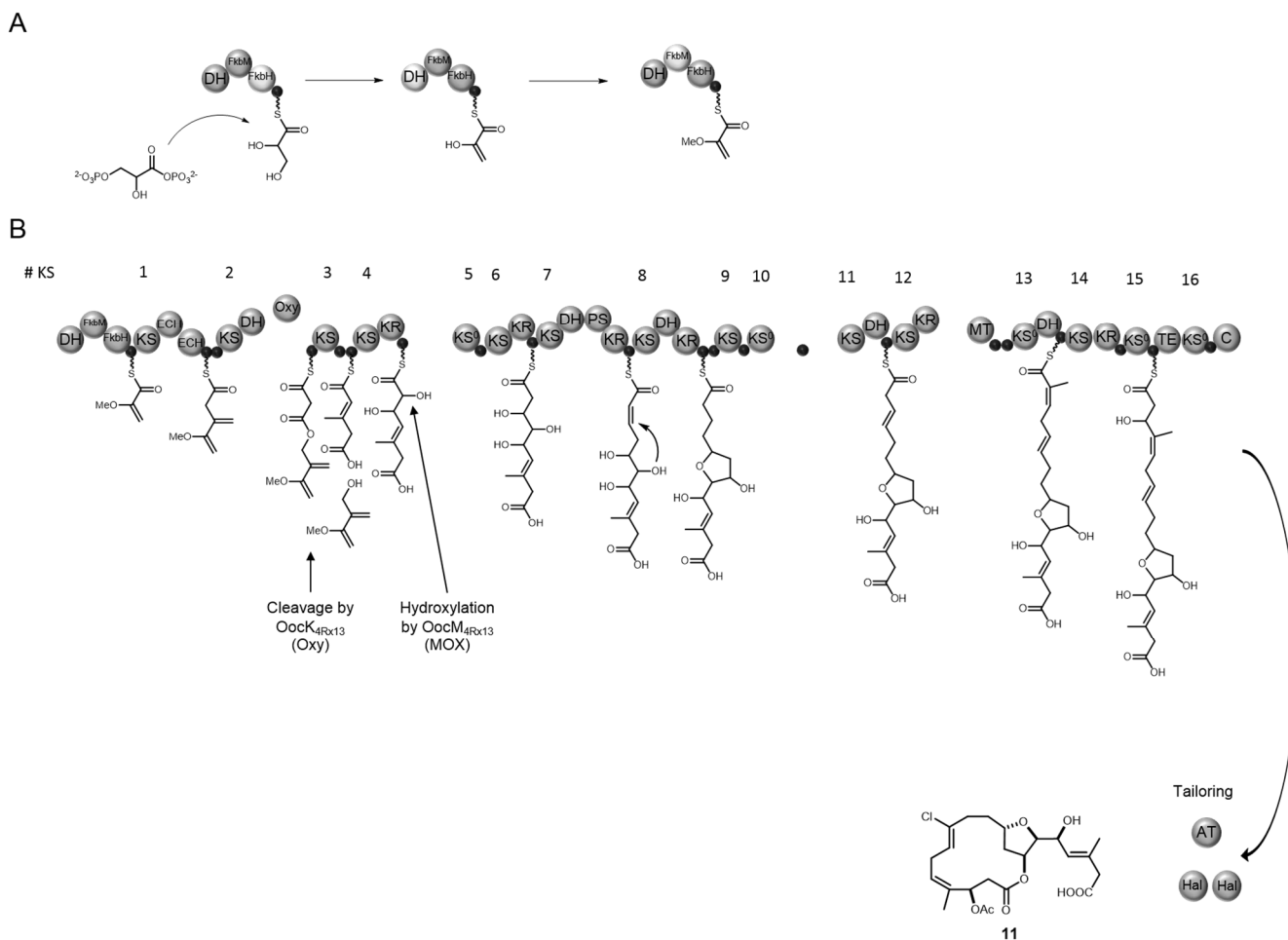
Further experiments have to clarify the role of *OocQ*<sub>4Rx13</sub>. Due to the very low homology to any known protein we can only speculate about the function within this halogenation reaction.



**Figure 3-5: Proposed scenarios for oocydin A halogenation.**

In every case (A-E) a fully different option is possible as well but not depicted here. Thinkable functions of OocQ are indicated in the figure in green. (A) The substrate could either be dechloro-oocydin A (**42**) or a **42**-analog with a desaturation between C<sub>7</sub>-C<sub>8</sub>. Red indicates the bond where halogenation will occur. (B) OocP and OocQ could either catalyze halogenation in parallel or in a subsequent reaction to yield oocydin A (**11**). (C) The substrate could either be carrier protein-bound (indicated with the little black ball) or free for halogenation. (D) Timing of halogenation could be either during chain elongation at the PKS or in a tailoring step. Arrows are an exemplary indication of possible halogenation time points. (E) Reaction mechanisms for halogenation of **11**. (1) First reaction steps necessary to form the reactive Fe(IV)-species with the help of  $\alpha$ -ketoglutarate (or other cofactors, see corresponding text). (2) Described halogenation mechanism for a saturated substrate (Matthews et al. 2014). (3) Possible halogenation mechanism for a substrate with a double bond. The red asterisk indicates the carbon atom which gets halogenated.

Having the biosynthesis of **11** and previously unknown **43** unambiguously linked to the *ooc*<sub>4Rx13</sub> gene cluster we propose a revised biosynthesis model (Figure 3-6). The existing model (Matilla et al. 2012) did not explain the biosynthesis of a product with an extended chain. The starter domains DH-FkbM-FkbH are known to be responsible for an introduction of modified 1,3-biphosphoglycerate into the biosynthetic pathway (Dorrestein et al. 2006) and were hypothesized to be inactive in **11** biosynthesis (Matilla et al. 2012). Our data suggest that these domains are active and catalyze the generation of the starter moiety from 1,3-biphosphoglycerate (Figure 3-6A). The proposed starter moiety is so far not described in a different biosynthetic pathway. This starter moiety gets modified in a way as it appears in **43** side chain. The intermediate during starter moiety generation which exhibits a hydroxyl and exomethylen group at the  $\beta$ -position (Figure 3-6A) can rearrange to an intermediate carrying a keto and a methyl group at the  $\beta$ -position. The subsequent methylation by FkbH would not occur. The structural difference between haterumalide B (**35**) and **43** is one saturation in an exomethylen group at C<sub>25</sub>. This saturation might be catalyzed by the enoylreductase *OocU*<sub>4Rx13</sub>. Structural variety in the side chain of **11** and **43** might result from a reaction of *OocK*<sub>4Rx13</sub>. The proposed function of this protein is a flavin-dependent monooxygenase and was hypothesized to be involved in chlorination of **11** (Matilla et al. 2012). As this was ruled out by Matilla *et al.* and we could show that two other proteins, namely *OocP*<sub>4Rx13</sub> and *OocQ*<sub>4Rx13</sub>, are responsible for the installation of a halogen atom. We suggest that *OocK*<sub>4Rx13</sub> is rather responsible to catalyze a cleavage of the growing intermediate in **11** biosynthesis. This hypothesis is supported by the fact that *OocK*<sub>4Rx13</sub> closest homolog known is PedG from the pederin biosynthetic pathway with 61 % identity (Matilla et al. 2012). PedG was already hypothesized to be involved in a cleavage occurring during pederin biosynthesis (Piel 2002). The domain architecture in the region where cleavage is hypothesized to take place is also comparable between pederin and oocydin PKS. In both cases a KS-DH-Oxy-ACP domain architecture can be found. *OocK*<sub>4Rx13</sub> might be therefore a flavin-dependent Baeyer-Villiger monooxygenase (BVMO) (van Berkel, Kamerbeek, and Fraaije 2006). BVMOs are known to be able to catalyze bond cleavages as in the case of MtmIV (Bosserman et al. 2013) and GilOII (Tibrewal et al. 2012). This supports our hypothesis of a cleavage by *OocK*<sub>4Rx13</sub>. In addition to **43** the structural moiety of an extended side chain can also be found in haterumalide B (**35**) and biselides B (**38**), suggesting the involvement of a cleaving enzyme also in the biosynthesis of haterumalide NA (**28**)/**35** and biselides A (**37**)/**38**. The  $\beta$ -branch at C<sub>4</sub> is most likely installed at the second module between KS3 and KS4 (Figure 3-6) contrary to the earlier proposed KS2 by proteins encoded by *oocC*<sub>4Rx13</sub>, *oocD*<sub>4Rx13</sub>, *oocE*<sub>4Rx13</sub>, *oocF*<sub>4Rx13</sub> and *oocG*<sub>4Rx13</sub>. Hydroxylation of the  $\alpha$ -position after KS4 might result from the *trans*-acting hypothetical monooxygenase *OocM*<sub>4Rx13</sub>. This hydroxyl-group gets incorporated in the tetrahydrofuran-ring. An involvement of DH and PS domain in the formation of cyclic ethers was reported recently and we assume the DH and PS domain here to be responsible for tetrahydrofuran-ring cyclization (Pöplau et al. 2013). We propose a generation of the double bond between C<sub>7</sub>-C<sub>8</sub> between KS11 and KS12. A very unusual domain architecture can be found in the last module. Normally either a TE or a C-domain are responsible for chain release. A TE-KS<sup>0</sup>-ACP-C architecture is quite remarkable. One other related example is the gene cluster underlying FR901464 biosynthesis. It was characterized to catalyze a dehydration and *cis*-double bond formation prior to cyclization (He et al. 2014). In **11** biosynthesis the only *cis*-double bond rather stems from  $\beta$ -branching than these four domains. BLAST analysis of all remaining genes with unknown function, namely *oocT*<sub>4Rx13</sub>, *oocl*<sub>4Rx13</sub> and *oocJ*<sub>4Rx13</sub> do not produce any plausible homolog candidates.



**Figure 3-6: Revised oocydin A 11 biosynthesis scheme.**

(A) Generation of starter moiety from 1,3-bisphosphoglycerate. Respective acting domains are colored in light gray. (B) Oocydin A biosynthesis. KSs are numbered. Abbreviations: little black circles, acyl carrier proteins (ACP); DH, dehydratase; Fkb, Fkb-like enzymes; KS, ketosynthase; ECH, enoyl-CoA hydratase; Oxy, oxygenase; KR, ketoreductases; KS<sup>0</sup>, non-elongating KS; PS, pyran synthase; MT, methyltransferase; TE, thioesterase; C, condensation domain; AT, acyltransferase; Hal, halogenase; MOX, monooxygenase.

Further experiments have to reveal if the biosynthetic model holds true. In this studies also the nature of the starter moiety might get investigated as well as the hypothesized cleavage reaction by  $OocK_{4Rx13}$ .

The structure elucidation of **43** adds an additional member to the haterumalide family. Other members of the haterumalide family including biselides could not be detected by HPLC-HRMS. In a typical extraction, the amounts of product **11** are low, **43** is produced in smaller amounts and additional compounds beneath our detection limit may be possible.

The absolute stereochemistry of **11**, **28** and **29** is not fully assigned yet, although total synthesis was achieved by several groups (Ueda et al. 2009), (Roulland 2008), (Kigoshi and Hayakawa 2007), (Hoye and Wang 2005), (Gu and Snider 2003), (Kigoshi et al. 2003). According to Kigoshi and Roulland, **11** and **29** exhibit the same stereochemistry (Kigoshi and Hayakawa 2007), (Roulland 2008), Sato and Hoye stated **11** and **28** to be identical based on optical rotation (Sato et al. 2005), (Hoye and Wang 2005). Further studies have to clarify if all products synthesized by the  $ooc_{4Rx13}$  gene cluster have the same stereochemistry or not. We performed NOESY measurements and are able to comment on the relative but not the absolute stereochemistry. It might be possible to hypothesize an absolute configuration if beside relative stereochemistry also KR domains were taken into account. KRs are known to catalyze reactions in a stereospecific manner and sequence motifs from their primary sequence allow prediction of the

stereochemistry of the product (Keatinge-Clay 2007). All six KR domains present in the oocydin A assembly line do not exhibit any of the conserved sequence motif and with that no further assumptions regarding an absolute stereochemistry can be deduced. A biosynthesis of both enantiomers by the same gene cluster would be rather unusual, since normally enzymes introducing stereo centers catalyze their reaction with a high stereospecificity (Kwan and Schulz 2011). Oocydin A-type compounds might serve as important secondary metabolite for the producer otherwise it would not be present in a range of diverse habitats (ocean, soil). It is very likely that haterumalides isolated from marine sponges or ascidians are of bacterial origin as well (Piel 2009).

Cytotoxicity assays showed a minor involvement of the chlorine in activity. The reduced bioactivity of **43** might be because of the longer side chain. When **28** and **35** were tested in parallel they showed nearly the same IC<sub>50</sub>. However, if only a part of the macrolactone ring together with the side chain were tested, huge differences appear. Interestingly the side chain of **35** is reported to have a higher activity in comparison to **43** and an important influence of the conjugated ketone moiety in target interaction was proposed (Ueda et al. 2009).

In conclusion we were able to assign two genes to the halogenation reaction in **11** biosynthesis and with that describe a new two-component halogenase. We were able to isolate the dechloro analog from corresponding knock out strains and showed that bromine can be introduced as a halogen as well. A revised model for **11** biosynthesis is proposed explaining the biosynthesis of **11** and **43** with its extended side chain by one gene cluster. A further understanding of this gene cluster might lead to a targeted production of different haterumalides and aid its agricultural application.

## 3.5 Experimental procedure

### 3.5.1 General

Solvents used in HPLC-HRMS measurements were LC-MS grade (Fisher Chemicals), solvents applied in preparative HPLC purifications were HPLC grade (Sigma).

### 3.5.2 Bacterial strains and culture conditions

*E. coli* strain XL1 blue (Stratagen) served as host for routine subcloning, *E. coli* BL21DE3 (Invitrogen) served as host for protein expression. *E. coli* and *S. plymuthica* strains were grown in Luria-Bertani (LB) medium if not mentioned else (Sambrook and Russell 2001) and in TB medium (Tartoff and Hobbs 1987) for protein expression. Generated and used strains can be seen in Table 3-3. *Serratia plymuthica* 4Rx13 (NCBI accession number NC\_021591.1) (originally *S. odorifera* 4Rx13) was renamed and reannotated during the performance of our study (Table 3-8).

### 3.5.3 Plasmid and general DNA procedure

DNA isolation, plasmid preparation, restriction digests, gel electrophoresis, and ligation reactions were conducted according to standard methods (Sambrook and Russell 2001). pBluescript II SK(+) was used for subcloning by using the method of Marchuk et al. (Marchuk, et al., 1991). All primer sequences are listed in Table 3-5. Final constructs were verified by sequencing and final strains by PCR. For electroporation of *S. plymuthica* regular *E. coli* protocols can be applied.

### 3.5.4 Fermentation and extraction of *Serratia plymuthica* 4Rx13

For fermentation, *S. plymuthica* 4Rx13 or corresponding knock out strains were grown for at least 48 hrs, 250 rpm, 27 °C in a baffled flask in enriched potato dextrose broth (EPB) (24 g/L potato dextrose broth (Difco), 6 g/L bactopectone, 4 g/L yeast extract, 100 mg/L NaCl). When appropriate antibiotics were used (Table 3-3, Table 3-4). Extraction was carried out with Amberlite XAD 16N (20-60 mesh, Sigma Aldrich). For this purpose, 2 % (v/v) wet-activated XAD 16N was added to the culture at the time of inoculation.

| Strain   | Antibiotic                                       |
|--|--|
| <i>Serratia plymuthica</i> 4Rx13                     | -  |
| 4Rx13 $\Delta$ <i>oocR</i>                           | Apramycin (50 µg/mL)                             |
| 4Rx13 $\Delta$ <i>oocP</i>                           | Kanamycin (12.5 µg/mL)                           |
| 4Rx13 $\Delta$ <i>oocQ</i>                           | Kanamycin (12.5 µg/mL)                           |
| 4Rx13 $\Delta$ <i>oocT</i>                           | Kanamycin (12.5 µg/mL)                           |
| 4Rx13 $\Delta$ <i>oocP</i> + pKJ42                   | Kanamycin (12.5 µg/mL)<br>Ampicillin (100 µg/mL) |
| 4Rx13 $\Delta$ <i>oocP</i> + pKJ43                   |  |
| 4Rx13 $\Delta$ <i>oocP</i> + pKJ55                   |  |
| 4Rx13 $\Delta$ <i>oocP</i> + pBAD/ <i>Myc</i> -His C |  |
| 4Rx13 $\Delta$ <i>oocQ</i> + pKJ42                   |  |
| 4Rx13 $\Delta$ <i>oocQ</i> + pKJ43                   |  |
| 4Rx13 $\Delta$ <i>oocQ</i> + pKJ55                   |  |
| 4Rx13 $\Delta$ <i>oocQ</i> + pBAD/ <i>Myc</i> -His C |  |

**Table 3-3: Generated and used strains in this study.**

If not mentioned differently the general procedure for extraction was as follows. If cells were fermented with XAD 16N the supernatant was decanted, washed with dH<sub>2</sub>O (~10 mL/L culture) and extracted with methanol (~15 mL/L culture). The mixture was stirred 30-60 min at room temperature (RT) and the organic layer evaporated under reduced pressure. In a culture without XAD 16N the whole culture was extracted three times with ethyl acetate. The organic layer was dried with sodium sulfate and the solvent was removed under reduced pressure.

For small-scale fermentation the cells were fermented in 1 mL EPB in a 2 mL snap-cap tube with a hole in the lid for at least 48 hrs at 27 °C, 990 rpm in a heating Thermo Shaker (Ditabis). The culture was transferred into a fresh 2 mL snap-cap tube, 1 mL ethyl acetate was added, everything mixed properly and extracted for at least 1 hr on a rotating wheel. The organic layer was transferred and evaporated under reduced pressure. If a separate extraction of cells and supernatant was conducted, both portions were extracted separately as described above.

### 3.5.5 Preparative HPLC and HPLC-HRMS procedure

Compound isolation occurred with two subsequent HPLC preparations with different instruments. The first HPLC preparation was performed on an Agilent 1260 Infinity HPLC containing preparative pump, multiple wavelength detector, and preparative-scale fraction collector split to an Agilent 6100 single quadrupole mass spectrometer for low resolution measurements. Second HPLC preparation was performed on an Agilent 1260 Infinity HPLC containing autosampler, quaternary pump, thermostatted column compartment, variable wavelength detector and analytical scale fraction collector.

HPLC-HRMS experiments we performed on a Dionex Ultimate 3000 UHPLC coupled to a Thermo Scientific Q Exactive mass spectrometer. Samples were analyzed by reversed phase HPLC (column: Phenomenex, Kinetex, C18, 2.6  $\mu$ , 150 x 4.6 mm; temperature: 27 °C; flow rate 0.5 mL/min; mobile phase/gradient: 5% CH<sub>3</sub>CN + 0.1% formic acid/H<sub>2</sub>O + 0.1% formic acid for 5 min then ramped to 95% CH<sub>3</sub>CN + 0.1% formic acid/H<sub>2</sub>O + 0.1% formic acid over 12 min, kept at 95% CH<sub>3</sub>CN + 0.1% formic acid/H<sub>2</sub>O + 0.1% formic acid for 3 min constant, ramped back to 5% CH<sub>3</sub>CN + 0.1% formic acid/H<sub>2</sub>O + 0.1% formic acid over 2 min and kept for 3 min; detection: ESI-MS, positive ion mode).

### 3.5.6 Isolation and structure confirmation of oocydin A **11**

To isolate oocydin A **11**, the wild type fermentation broth was extracted with XAD 16N. The crude extract (3.53 g) was dissolved in water and extracted with chloroform, afterwards the aqueous layer was extracted with ethyl acetate and the organic layer evaporated under reduced pressure. The extract was subjected to HPLC purification (column: Phenomenex, Luna, C18(2), 100 Å, 5  $\mu$ , 250 x 21.20; flow rate 15 mL/min; mobile phase/gradient: 35% CH<sub>3</sub>CN + 0.1% trifluoroacetic acid (TFA)/H<sub>2</sub>O + 0.1% TFA for 60 min then ramped to 100% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA over 10 min, kept at 100% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA for 10 min; detection: 210 nm). The fractions between 48-52 minutes were collected and the solvent was removed under reduced pressure. The dried fractions were subjected to a second round of HPLC purification (column: Phenomenex, Luna Phenyl-Hexyl, 100 Å, 5  $\mu$ , 250 x 10.00; flow rate 2 mL/min; mobile phase/gradient: 37.5% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA for 60 min then ramped to 100% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA over 10 min, kept at 100% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA for 10 min; detection: 210 nm). The solvent was evaporated under reduced pressure and 1.2 mg of pure fraction subjected to NMR measurements.

### 3.5.7 Isolation and structure elucidation of dechloro-oocydin A **42**

To isolate dechloro-oocydin A **42**, the wild type fermentation broth was extracted with ethyl acetate and the organic layer evaporated under reduced pressure. The crude extract (281.4 mg) was subjected to HPLC purification (column: Phenomenex, Luna, C18(2), 100 Å, 5  $\mu$ , 250 x 21.20; flow rate 15 mL/min; mobile phase/gradient: 37% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA for 60 min then ramped to 100% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA over 20 min, kept at 100% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA for 2 min and ramped back to 37% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA over 3 min; detection: 210 nm). The fractions between 24-30 minutes were collected and the solvent was removed under reduced pressure. The dried fractions were subjected to a second round of HPLC purification (column: Phenomenex, Luna Phenyl-Hexyl, 100 Å, 5  $\mu$ , 250 x 10.00; flow rate 2 mL/min; mobile phase/gradient: 37.5% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA for 60 min then ramped to 100% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA over 10 min, kept at 100% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA for 10 min and ramped back to 37% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA over 3 min; detection: 210 nm). The solvent was evaporated under reduced pressure and 0.6 mg of pure fraction subjected to NMR measurements.

### 3.5.8 Isolation and structure elucidation of oocydin B **43**

To isolate oocydin B **43**, the wild type fermentation broth was extracted with XAD 16N. The crude extract (3.53 g) extracted with chloroform, afterwards the aqueous layer extracted with ethyl acetate and the organic layer evaporated under reduced pressure. The extract was subjected to HPLC purification (column: Phenomenex, Luna, C18(2), 100 Å, 5  $\mu$ , 250 x 21.20; flow rate 15 mL/min; mobile phase/gradient: 35% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA for 60 min then ramped to 100% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA over 10 min, kept at 100% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA for 10 min; detection: 210 nm). The fractions between 70-73 minutes were collected and the solvent was removed under reduced pressure. The dried fractions were subjected to a second round of HPLC purification (column: Phenomenex, Luna Phenyl-Hexyl, 100 Å, 5  $\mu$ , 250 x 10.00; flow rate 2 mL/min; mobile phase/gradient: 60% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA for 60 min then ramped to 100% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA over 10 min, kept at 100% CH<sub>3</sub>CN + 0.1% TFA/H<sub>2</sub>O + 0.1% TFA for 10 min; detection: 210 nm). The solvent was evaporated under reduced pressure and 0.5 mg of pure fraction subjected to NMR measurements.

### 3.5.9 NMR measurements

All NMR spectra were recorded on a Bruker Avance III spectrometer equipped with a 5 mm cold probe at 500 MHz for <sup>1</sup>H NMR and 125 MHz for <sup>13</sup>C NMR. Chemical shifts were referenced to the solvent peak at  $\delta_{\text{H}}$  3.31 and  $\delta_{\text{C}}$  49.15 for methanol-*d*<sub>4</sub>.

### 3.5.10 Generation of knock-out strains and cloning of constructs

Knock-out strains were generated with the help of  $\lambda$ RED technique (Datsenko and Wanner 2000). The following strategy was used to generate 4Rx13 $\Delta$ *oocR*: A PCR product was amplified from *S. plymuthica* 4Rx13 genomic DNA using primers So\_ko\_vorne\_BglII-for and So\_ko\_vorne\_BamHI-rev and ligated into pBluescript II SK(+) to yield pKJ24. The same procedure was performed with the primers So\_ko\_hinten\_MfeI-for and So\_ko\_hinten\_EcoRI-rev to yield pKJ25. To obtain pKJ26 the plasmid pIJ773 carrying apramycin-resistance cassette, *oriT* and FLP recognition target sites (Gust et al. 2003) was cut with *Bam*HI and the apramycin-cassette was ligated into corresponding sites of pKJ24. The plasmid pKJ26 was digested with *Pvu*II and the generated insert ligated into *Hin*clI linearized pKJ25 to yield pKJ27. The plasmid pIJ790 (Gust et al. 2003) was electroporated into *S. plymuthica* 4Rx13 and cells of the resulting strain employed in  $\lambda$ RED recombination. Therefore the protocol of Rossi (Rossi et al. 2003) was used with the *Pvu*II generated and dephosphorylated insert of pKJ27 to obtain 4Rx13 $\Delta$ *oocR*.

The knock-out strains of *oocP*<sub>4Rx13</sub>, *oocQ*<sub>4Rx13</sub> and *oocT*<sub>4Rx13</sub> were generated with the *E. coli* Gene Deletion Kit (GeneBridges) and a modified manufacturer's protocol. For preparing electrocompetent cells for  $\lambda$ RED homologous recombination, following modifications were applied: 300 mL culture of *S. plymuthica* 4Rx13 carrying pRedET (from the kit) were grown in LB medium supplemented with tetracycline (10  $\mu$ g/mL) to an OD<sub>600</sub> of 0.15-0.2 (30 °C, 170 rpm), subsequently the pRedET plasmid was induced with a final concentration of 0.4% L-arabinose, after induction cultures were grown to an OD<sub>600</sub> of 0.35-0.45 (37 °C, 170 rpm). The cultures were cooled on ice for 15 min. Cells were washed and centrifuged (5,000 x g, 20 min) at 4 °C several times prior to electroporation. Washing was performed twice with 40 mL dH<sub>2</sub>O each, four times with 25 mL 8.5% (v/v) glycerol each and the final cell pellet was resuspended in 500  $\mu$ L 8.5% (v/v) glycerol. For further experiments 40  $\mu$ L aliquots of cells can be stored at -70 °C and then electroporated. The homology cassette including a FRT-flanked antibiotic cassette for *oocP*<sub>4Rx13</sub> was generated with primer pair KO\_SOD\_b01020\_fwd/rev to obtain strain 4Rx13 $\Delta$ *oocP*. The homology cassette including a FRT-flanked antibiotic cassette for *oocQ*<sub>4Rx13</sub> was generated with primer pair KO\_SOD\_b01030\_fwd/rev



to obtain strain 4Rx13 $\Delta$ oocQ. The homology cassette including a FRT-flanked antibiotic cassette for *oocT*<sub>4Rx13</sub> was generated with primer pair KO\_SOD\_b01070\_fwd/rev to obtain strain 4Rx13 $\Delta$ oocT.

To generate the *oocP*<sub>4Rx13</sub> rescue-construct pKJ43 a 1.2 kb PCR product was amplified from *S. plymuthica* 4Rx13 genomic DNA using primers OocP-NcoI-for and OocP-BglII-rev. To generate the *oocQ*<sub>4Rx13</sub> rescue-construct pKJ42 a 0.35 kb PCR product was amplified from *S. plymuthica* 4Rx13 genomic DNA using primers OocQ-NcoI-for and OocQ-BglII-rev. To generate the *oocP*<sub>4Rx13</sub>-*oocQ*<sub>4Rx13</sub> double construct pKJ55 a 1.5 kb PCR product was amplified from *S. plymuthica* 4Rx13 genomic DNA using primers OocP-NcoI-for and OocQ-BglII-rev. Every PCR-product was digested with *NcoI* and *BglII* and inserted into the corresponding sites of pBAD/*Myc*-His C (Invitrogen) to yield the final plasmid. Every construct was electroporated in 4Rx13 $\Delta$ oocP and 4Rx13 $\Delta$ oocQ. As control empty pBAD/*Myc*-His C vector was transformed in 4Rx13 $\Delta$ oocP and 4Rx13 $\Delta$ oocQ as well. All eight resulting strains and 4Rx13 $\Delta$ oocP and 4Rx13 $\Delta$ oocQ as control were fermented in 20 mL-scale including XAD 16N. After 48 h XAD 16N was extracted and culture extracts analyzed with HPLC-HRMS.

To generate the *oocP*<sub>4Rx13</sub> N-terminal hexahistidine-tagged fusion construct, a 1.2 kb PCR product was amplified from *S. plymuthica* 4Rx13 genomic DNA using primers OocP\_BamHI\_for and OocP\_NotI\_rev. The PCR product was ligated into pCDF-DUET (Novagen) at the *BamHI* and *NotI* restriction sites after digestion of the insert with the same enzymes to yield pKJ57.

To generate the *oocQ*<sub>4Rx13</sub> N-terminal hexahistidine-tagged fusion construct, a 0.35 kb PCR product was amplified from *S. plymuthica* 4Rx13 genomic DNA using primers OocQ\_BamHI\_for and OocQ\_NotI\_rev. The PCR product was ligated into pCDF-DUET (Novagen) at the *BamHI* and *NotI* restriction sites after digestion of the insert with the same enzymes to yield pKJ59.

To generate the *oocQ*<sub>4Rx13</sub> C-terminal strep-tagged fusion construct, a 0.35 kb PCR product was amplified from *S. plymuthica* 4Rx13 genomic DNA using primers OocQ\_BglII\_for and OocQ\_XhoI\_rev. The PCR product was ligated into pET-DUET (Novagen) at the *BglII* and *XhoI* restriction sites after digestion of the insert with the same enzymes to yield pKJ61.

To generate the *oocP*<sub>4Rx13</sub> C-terminal strep-tagged fusion construct, a 1.2 kb PCR product was amplified from *S. plymuthica* 4Rx13 genomic DNA using primers OocP\_BglII\_for and OocP\_XhoI\_rev. The PCR product was ligated into pCDF-DUET (Novagen) at the *BglII* and *XhoI* restriction sites after digestion of the insert with the same enzymes to yield pKJ62.

To generate the *oocP*<sub>4Rx13</sub> N-terminal hexahistidine-tagged and *oocQ*<sub>4Rx13</sub> C-terminal strep-tagged double construct, a 0.35 kb PCR product was amplified from *S. plymuthica* 4Rx13 genomic DNA using primers OocQ\_BglII\_for and OocQ\_XhoI\_rev. The PCR product was ligated into pCDF-DUET (Novagen) at the *BglII* and *XhoI* restriction sites after digestion of the insert with the same enzymes to yield pKJ60. pKJ57 was digested with *BamHI* and *NotI* and the insert ligated into corresponding sites of pKJ60 to yield pKJ64.

To generate the *oocQ*<sub>4Rx13</sub> N-terminal hexahistidine-tagged and *oocP*<sub>4Rx13</sub> C-terminal strep-tagged double construct, a 1.2 kb PCR product was amplified from *S. plymuthica* 4Rx13 genomic DNA using primers OocP\_BglII\_for and OocP\_XhoI\_rev. The PCR product was ligated into pET-DUET (Novagen) at the *BglII* and *XhoI* restriction sites after digestion of the insert with the same enzymes to yield pKJ56. pKJ59 was digested with *BamHI* and *NotI* and the insert ligated into corresponding sites of pKJ56 to yield pKJ65.

| <b>Construct</b>  | <b>Plasmid</b> | <b>Antibiotic</b>        |
|---|----------------|--------------------------|
| OocP <sub>4Rx13</sub> (N-terminal His <sub>6</sub> -tag)  | pKJ57          | Spectinomycin (25 µg/mL) |
| OocQ <sub>4Rx13</sub> (N-terminal His <sub>6</sub> -tag)  | pKJ59          | Spectinomycin (25 µg/mL) |
| OocQ <sub>4Rx13</sub> (C-terminal Strep-tag)  | pKJ61          | Ampicillin (100 µg/mL)   |
| OocP <sub>4Rx13</sub> (C-terminal Strep-tag)  | pKJ62          | Spectinomycin (25 µg/mL) |
| OocP <sub>4Rx13</sub> (N-terminal His <sub>6</sub> -tag) and OocQ <sub>4Rx13</sub> (C-terminal Strep-tag) | pKJ64          | Spectinomycin (25 µg/mL) |
| OocQ <sub>4Rx13</sub> (N-terminal His <sub>6</sub> -tag) and OocP <sub>4Rx13</sub> (C-terminal Strep-tag) | pKJ65          | Ampicillin (100 µg/mL)   |

**Table 3-4: Constructs used in this study.**

| <b>Primer name</b>     | <b>Sequence (5'-3')</b>  |
|------------------------|--|
| So_ko_vorne_BglII-for  | AAAAGATCTCGTTCATAAAAAATGAAGTCCT  |
| So_ko_vorne_BamHI-rev  | AAAGGATCCAGGCTCAGACTTTCCGCT  |
| So_ko_hinten_MfeI-for  | AAACAATTGCCTCTTTGCAGGCGGCC   |
| So_ko_hinten_EcoRI-rev | AAAGAATTCAGTTCATTGATAATCACCATGCC   |
| KO_SOD_b01020_fwd      | <u>AGTCCTACGATTTCTCAAATATGGATCGGCATACATTCCAGCACATGATCAA</u><br><u>TTAACCCCTACTAAAGGGCGG</u>  |
| KO_SOD_b01020_rev      | <u>TGTTGAACGGCTTTCTGTGTCGTTGAGAAAACCCTTGGCAATAATCGGCGGT</u><br><u>AATACGACTCACTATAGGGCTC</u> |
| KO_SOD_b01030_fwd      | <u>CCGGCAGACAAAGCCTATTACATCCTTAATGATGATCGGGCTTCGACGGAA</u><br><u>ATTAACCCTACTAAAGGGCGG</u>   |
| KO_SOD_b01030_rev      | <u>TAAACCACAGGGGCTTAAGCCGGTGGTACTGCACCTTCTCACCGTCCAGAT</u><br><u>AATACGACTCACTATAGGGCTC</u>  |
| KO_SOD_b01070_fwd      | <u>TGGGCTCCGGCGTCATTCATATCACCAAGAGTGAACCTTTCCCGGGGTGA</u><br><u>ATTAACCCTACTAAAGGGCGG</u>    |
| KO_SOD_b01070_rev      | <u>TCATTAAATCCCATATATTGCGTGTAGAAATAGTCACTGGTGCATAAGGT</u><br><u>AATACGACTCACTATAGGGCTC</u>   |
| OocP-NcoI-for          | GTACCATGGCGAAAAGTATGACATTCAAAGT  |
| OocP-BglII-rev         | GTAAGATCTTCATTGTTTATCCCCATTG   |
| OocQ-NcoI-for          | GTACCATGGCGAGCGAGTACCTGATAAATTC  |
| OocQ-BglII-rev         | GTAAGATCTCTAACGGACGGACATTTTT   |
| OocP_BamHI_for         | GTAGGATCCGGAAAAGTATGACATTCAA   |
| OocP_BglII_for         | GTAGATCTCGAAAAGTATGACATTCAAAG  |

|                |                                       |
|----------------|---------------------------------------|
| OocP_NotI_rev  | GT <b>CGGCGC</b> GCTCATTGTTTATCCCCATT |
| OocP_XhoI_rev  | GTAC <b>CTCGAG</b> TTGTTTATCCCCATTGG  |
| OocQ_BamHI_for | GTAG <b>GATCC</b> GAGCGAGTACCTGATAAA  |
| OocQ_BglII_for | GTAG <b>ATCTC</b> AGCGAGTACCTGATAAA   |
| OocQ_NotI_rev  | GT <b>CGGCGC</b> CCTAACGGACGGACATTTT  |
| OocQ_XhoI_rev  | GTAC <b>CTCGAG</b> ACGGACGGACATTTT    |

**Table 3-5: Primers used in this study**

Restriction sites are marked **bold**, Homology arms were underlined, FRT-flanked antibiotic-cassette is indicated with italic letters.

### 3.5.11 Expression and purification of proteins

For overexpression of all proteins 20 mL cultures were grown to an OD<sub>600</sub> of 1.5-2 in TB medium (37 °C, 250 rpm), induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and grown for approximately 20 hrs at 16 °C. Cells of expression strains were harvested by centrifugation (7500 x g, 10 min) at 4 °C.

Cells expressing proteins with His<sub>6</sub>-Tag were resuspended in lysis buffer (25 mM Tris-HCl, pH 7.4, 0.5 M NaCl, 10 mM imidazole, 0.2% (v/v) Triton X-100 and 10% (v/v) glycerol). The cell suspension was kept on ice and sonicated in 10 s intervals. The cell debris was removed by centrifugation at 23,500 x g for 30 min at 4 °C. The supernatant was incubated with 1.0 mL Ni-NTA agarose (Qiagen) for at least 1 hr at 4 °C and washed with 10 mL lysis buffer. His-tagged proteins were eluted with lysis buffer containing increasing concentrations of imidazole up to 250 mM. The concentration of protein was determined by Bradford assay and fractions were analyzed by SDS-PAGE.

### 3.5.1 Cross-complementation, inhibitor assay, feeding experiments and *in vitro* halogenation assay

For cross-complementation experiments, strains were cultured in 20 mL-scale for 24 hrs. Afterwards cells and supernatant were separated by centrifugation (15 min, 8000 x g, 4 °C), cell pellets were washed with 5 mL fresh EPB medium, resuspended in 1 mL fresh EPB medium. Every possible combination of supernatant and cells including controls w/o cells and with fresh EPB medium was realized. Therefore 1 mL supernatant was supplemented with 10 μL cell suspension of different knock out strains in a fresh 2 mL snap-cap tube with hole in the lid. These combinations were fermented for additional 24 hrs at 27 °C, 990 rpm in a heating Thermo Shaker (Ditabis) and extracted with ethyl acetate. Extracts were analyzed by HPLC-HRMS.

For inhibition of the putative α-KG dependent halogenase, cells were cultured as described above in small scale in 2 mL snap-cap tube w/o hole, after 24 hrs 2,4-PDCA (10 mM in DMSO) or only DMSO as control was added to the cultures at a final concentration of 5-333 mM and fermented for further 48 hrs prior to extraction with ethyl acetate. Extracts were analyzed by HPLC-HRMS.

For feeding experiments cells were fed with 5-15 mg culture extract of 4Rx13Δ*oocP* (200 mg/mL in methanol) as substrate. For whole-cell feeding experiments cells of expression strains were grown to an OD<sub>600</sub> of 1.5-2 in TB medium (37 °C, 250 rpm), induced with 1 mM IPTG, and grown for approximately 20 hrs at 16 °C. 1 mL of culture was transferred in a fresh 2 mL snap-cap tube with hole in the lid, supplemented with substrate and cultures were fermented for additional 48 h at 27 °C, 990 rpm in a heating Thermo Shaker (Ditabis) followed by an extraction with ethyl acetate. Extracts were analyzed by HPLC-HRMS. For cell-free feeding experiments cells of expression strains

were grown to an OD<sub>600</sub> of 1.5-2 in TB medium (37 °C, 250 rpm), induced with 1 mM IPTG, and grown for approximately 20 hrs at 16 °C. 500 µL culture were transferred in a fresh 500 µL tube and supplemented with 5 - 15 mg extract of 4Rx13Δ*oocP* (200 mg/mL in methanol). Tubes were incubated for approximately 30 min in a sonication water bath (Branson) and fermented for additional 48 h at 27 °C, 990 rpm in a heating Thermo Shaker (Ditabis) and extracted with ethyl acetate. Extracts were analyzed by HPLC-HRMS.

For *in vitro* halogenation assays proteins were purified as described above and elution buffer was exchanged with Vivaspin (Sartorius Stedim, MWCO 10.000 Da) to 50 mM MOPS, pH 7.5. The assay was adapted from Eustáquio *et al.* (Eustáquio *et al.* 2014) and contained 0.1 mM substrate (culture extract of 4Rx13Δ*oocP*), 0.2 mM α-ketoglutarate, 0.02 mM ascorbic acid, 0.01 mM NH<sub>4</sub>Fe(II)SO<sub>4</sub>, 0.25 mM NaCl, 1 µM of each protein in 50 mM MOPS, pH 7.5. The assay was incubated for 48 h at room temperature without shaking and was extracted with ethyl acetate. Extracts were analyzed by HPLC-HRMS.

### 3.5.2 Assay for cytotoxicity against HeLa cells

HeLa-cells ATCC CCL-2 were cultured in Dulbecco's Modified Eagle's Medium supplemented with GlutaMAX™ (Gibco), 10% fetal bovine serum (Gibco), 1% non-essential amino acids (Gibco) and 50 µg/mL gentamycin (AppliChem) at 37 °C under an atmosphere of 5% CO<sub>2</sub>. To each well of a flat-bottom 96-well microplate containing 200 µL of HeLa cell suspension (1 x 10<sup>4</sup> cells/mL) after 24 hrs of preincubation 2 µL of sample (5 mg/mL in methanol) was added. Cells were cultured as mentioned above for 48 hrs. After addition of 50 µL Thiazolyl Blue Tetrazolium Bromide (Sigma Aldrich) (1 mg/mL in dH<sub>2</sub>O) to every well the plate was incubated 3 hrs under same conditions to stain live cells. The supernatant was carefully removed and 150 µL DMSO were added to every well. GraphPad Prism 6.0.3 (GraphPad Software) was used to calculate the IC<sub>50</sub> values.

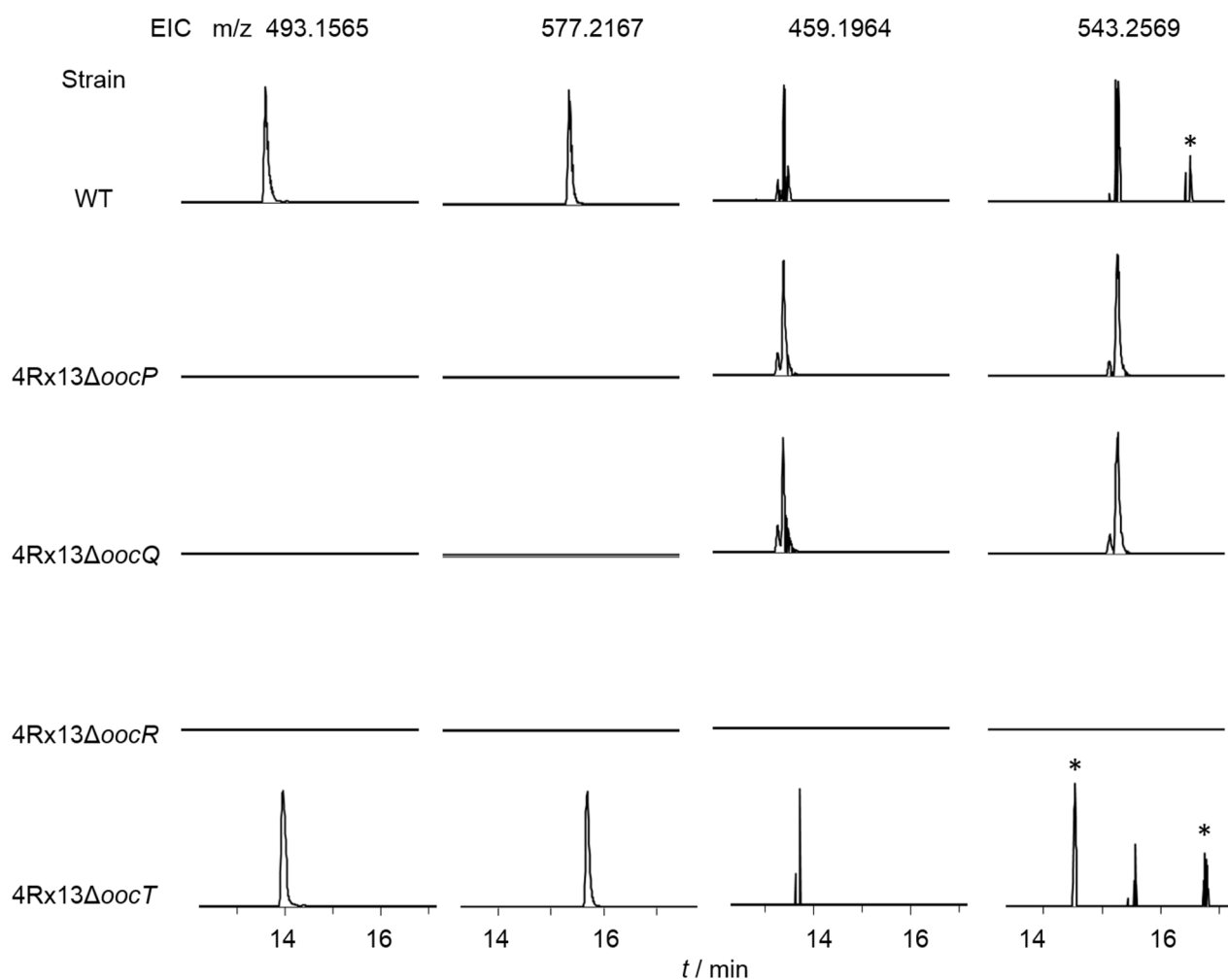
### 3.5.3 Data analysis workflow with eMZed

Data processing was carried out with eMZed using following workflow: In a first step features were extracted from data sets with integrated Feature Finder Metabo Algorithm (Griss *et al.* 2014), followed by clustering based retention time alignment using (Lange *et al.* 2007). Afterwards a sample comparison of two samples by differential analysis occurred. Both samples were further processed prior to comparison. First detected features were filtered for minimal peak quality (parameter was determined with feature finder algorithm). Second, features with same retention time and mass differences that could be explained by adducts were grouped for selected adducts ([M+Na<sup>+</sup>]<sup>+</sup>, [M+H<sup>+</sup>]<sup>+</sup>, [M+K<sup>+</sup>]<sup>+</sup>, [M+NH<sub>4</sub><sup>+</sup>]<sup>+</sup>). Finally, feature isotope ratios AM1 / AM0 and AM2 / AM0 were determined from feature peak areas. Comparison was performed defining peaks as present in both samples if mz difference was < 0.005 U and retention time was < 10 s. Results were grouped regarding if they were present in both data sets or just in one. To screen data for potential chlorated species following heuristic rule was applied: Features were selected if 1. AM2 / AM0 > 0.29 and if the AM1 / AM0 ≤ AM2 / AM0.

## 3.6 Acknowledgments

We thank Dr. Alexander O. Brachmann for help with the MS measurements and discussions. We thank Dr. Patrick Kiefer for the adjustment of python scripts in eMZed.

### 3.7 Supplemental information



**Figure 3-7: Extracted Ion Chromatogram (EIC) of different *Serratia plymuthica* 4Rx13 strains.**

EIC from oocydin A **11** ( $m/z$  493.1565  $[M+Na]^+$ ) and oocydin B **43** ( $m/z$  577.2167  $[M+Na]^+$ ) and their corresponding unchlorinated compounds dechloro-oocydin A **42** ( $m/z$  459.1964  $[M+Na]^+$ ) and dechloro-oocydin B ( $m/z$  543.2569  $[M+Na]^+$ ). Asterisks indicate compounds with the same mass but no correlation to the *ooc*<sub>4Rx13</sub> gene cluster is present as retention time and isotope pattern are different. The slight double peak in EICs of unchlorinated oocydins might be the Z-isomer of C<sub>16</sub> as reported by Levenfors *et al.*, 2004 for haterumalide NA (**28**) and haterumalide X (**36**). Isolation and structure elucidation was tried but failed due to very low production.

| Position  | Oocydin B  |   | dechloro-Oocydin A |                                   |
|-----------|------------|---|--------------------|-----------------------------------|
|           | $\delta_C$ | $\delta_H$  | $\delta_C$         | $\delta_H$                        |
| <b>1</b>  | 169.2      |   | 169.4              |                                   |
| <b>2</b>  | 38.7       | 2.78 (dd, $J = 11.6, 11.5$ )<br>2.82 (dd, $J = 11.5, 4.5$ ) | 38.6               | 2.77 (m)<br>2.75 (m)              |
| <b>3</b>  | 68.4       | 5.80 (dd, $J = 11.4, 4.5$ )                                 | 68.4               | 5.87 (dd, $J = 10.5, 5.3$ )       |
| <b>4</b>  | 134.6      |   | 133.7              |                                   |
| <b>5</b>  | 130.6      | 5.71 (m)  | 131.7              | 5.66 (t, $J = 8.8$ )              |
| <b>6</b>  | 27.5       | 2.47 (m)<br>3.50 (m)  | 30.9               | 2.58 (m)<br>3.13 (dd, $J = 5.1$ ) |
| <b>7</b>  | 126.8      | 5.32 (m)  | 130.4              | 5.14 (m)                          |
| <b>8</b>  | 132.9      |   | 128.3              | 5.35 (m)                          |
| <b>9</b>  | 35.3       | 2.29 (m)<br>2.45 (m)  | 29.6               | 1.81 (m)<br>2.32 (m)              |
| <b>10</b> | 28.8       | 1.40 (m)<br>2.29 (m)  | 30.8               | 1.40 (m)<br>2.03 (m)              |
| <b>11</b> | 77.9       | 3.94 (m)  | 79                 | 4.06 (tt, $J = 17.5, 3.9$ )       |
| <b>12</b> | 38.6       | 1.54 (m)<br>2.10 (dd, $J = 12.9, 2.8$ )                     | 37.5               | 1.41 (m)<br>2.10 (dd)             |
| <b>13</b> | 76.5       | 5.31 (m)  | 76.1               | 5.29 (m)                          |
| <b>14</b> | 84.4       | 3.89 (dd, $J = 8.8, 3.7$ )                                  | 84.4               | 3.89 (dd, $J = 8.6, 3.7$ )        |
| <b>15</b> | 66.2       | 4.53 (t, $J = 8.8$ )  | 66.4               | 4.53 (t, $J = 8.6$ )              |
| <b>16</b> | 130.7      | 5.36 (m)  | 130.5              | 5.39 (m)                          |
| <b>17</b> | 134.3      |   | 134.8              |                                   |
| <b>18</b> | 43.9       | 3.08 (s)  | 45.4               | 3.05 (d)<br>3.08 (m)              |
| <b>19</b> | 172.9      |   | 175.3              |                                   |
| <b>20</b> | 18.3       | 1.90 (brs)  | 18.6               | 1.89 (s)                          |
| <b>21</b> | 17.1       | 1.80 (brd, $J = 1.3$ )                                      | 17.1               | 1.83 (d, $J = 1.3$ )              |
| <b>22</b> | 171.2      |   | 171.3              |                                   |
| <b>23</b> | 20.8       | 2.03 (s)  | 20.9               | 2.03 (s)                          |
| <b>24</b> | 66.5       | 4.25 (dd, $J = 11.0, 7.2$ )<br>4.16 (dd, $J = 11.0, 5.7$ )  |                    |                                   |
| <b>25</b> | 46.9       | 2.95 (m)  |                    |                                   |
| <b>26</b> | 212.2      |   |                    |                                   |
| <b>27</b> | 28.6       | 2.21 (s)  |                    |                                   |
| <b>28</b> | 13.3       | 1.13 (d, $J = 7.2$ )  |                    |                                   |

Table 3-6: Proton and Carbon NMR data for Oocydin B (43) and dechloro-Oocydin A (42) obtained at 500 MHz for  $^1\text{H}$ -NMR and 125 MHz for  $^{13}\text{C}$ -NMR with samples dissolved in methanol- $d_4$ .

| CurA | OocP<br>(T-Coffee) | OocP<br>(Expresso) | SyrB2 | Annotations<br>after<br>Hangasky<br><i>et al.</i> | Proposed function  |
|------|--------------------|--------------------|-------|---|--|
| A101 | E124               | Q206               | E102  | B1-3  | close contact to C <sub>1</sub> of $\alpha$ -KG, probable involvement in substrate recognition   |
| F103 | L126               | F208               | F104  | B1-5  | probable involvement in substrate recognition  |
| V105 | E128               | G210               | K106  | B1-7  | probable involvement in substrate recognition  |
| H115 | A137               | H220               | H116  | <u>F1-5</u>                                       | facial triad-active site, non-heme   |
| G117 | G139               | G222               | A118  | <u>F1-7</u>                                       | facial triad-active site, chloride ligand in SyrB2, Asp (D)/Glu (E) is replaced with Ala (A) in halogenases  |
| R241 | R317               | R317               | R248  | B2-2  | contacting C <sub>5</sub> of $\alpha$ -KG, $\alpha$ -KG binding in SyrB2, CurA   |
| V245 | K328               | K328               | A252  | B2-6  | contact to C <sub>1</sub> of $\alpha$ -KG  |
| R247 | K330               | K330               | R254  | B2-8  | contact to C <sub>1</sub> of $\alpha$ -KG, residue is a hydrogen bond donor,<br>Arg (R) forms polar interaction with chloride in SyrB2 $\rightarrow$ most likely important in chloride binding                                   |
| H228 | H302               | H302               | H235  | <u>F2-3</u>                                       | facial triad-active site, non-heme   |
| G146 | S226               | n/a                | G156  | F3-1  | Contact to C <sub>5</sub> of $\alpha$ -KG  |
| S132 | R213               | S226               | T143  | B3-3  | close contact to C <sub>1</sub> of $\alpha$ -KG<br>through water contact to chloride in SyrB2<br>Responsible for R247 position (in turn is needed for $\alpha$ -KG positioning), Mutation $\rightarrow$ loss of function in CurA |
| W134 | W310               | A229               | W145  |   | $\alpha$ -KG binding in CurA   |
| S120 | Y140               | n/a                | F121  |   | chloride binding in CurA and SyrB2   |
| L112 | I135               | T217               | T113  |   | $\alpha$ -KG binding in SyrB2  |
| S230 | V304               | V304               | S237  |   | $\alpha$ -KG binding in CurA   |
| S122 | E142               | n/a                | N123  |   | Through water contact to chloride in SyrB2   |
| K224 | K298               | K298               | S231  |   | chloride ligand in SyrB2   |

**Table 3-7: Comparison of selected amino acid positions of OocP with CurA and SyrB after sequence alignment with T-Coffee and Expresso web server.**

Amino acid residues involved in the enzymes active site are underlined. Proteins were aligned either with T-Coffee of Expresso server. Subsequently positions known to be involved in reaction mechanism of MNH halogenases were compared. Halogenases from the following pathways were employed: CurA, curacin (*Lyngbya majuscula*); OocP, oocydin (*Serratia plymuthica*); SyrB2, syringomycin (*Pseudomonas syringae* *pv.* *syringae*); the full alignment can be seen in Figure 3-18 and Figure 3-19. Abbreviations:  $\alpha$ -KG,  $\alpha$ -ketoglutarate; n/a, no residue available.

| <b>New strain name</b>   | <b>Knock out of which gene in <i>S. plymuthica</i> 4Rx13</b> | <b>Previous annotation</b>      | <b>Corresponds to which generated strain</b> |
|--|--|---------------------------------|--|
| <i>Serratia plymuthica</i> 4Rx13 (NCBI accession number CP006250.1, NC_021591.1) | wild type  | <i>Serratia odorifera</i> 4Rx13 |  |
| 4Rx13Δ <i>oocR</i>   | <i>oocR</i> <sub>4Rx13</sub><br>( <i>SOD_c22970</i> )        | <i>SOD_b01040</i>               | So-KJ27                                      |
| 4Rx13Δ <i>oocP</i>   | <i>oocP</i> <sub>4Rx13</sub><br>( <i>SOD_c22950</i> )        | <i>SOD_b01020</i>               | So-Δ1020                                     |
| 4Rx13Δ <i>oocQ</i>   | <i>oocQ</i> <sub>4Rx13</sub><br>( <i>SOD_c22960</i> )        | <i>SOD_b01030</i>               | So-Δ1030                                     |
| 4Rx13Δ <i>oocT</i>   | <i>oocT</i> <sub>4Rx13</sub><br>( <i>SOD_c23000</i> )        | <i>SOD_b01070</i>               | So-Δ1070                                     |

**Table 3-8: Previous and new annotation of used *S. plymuthica* genes.**



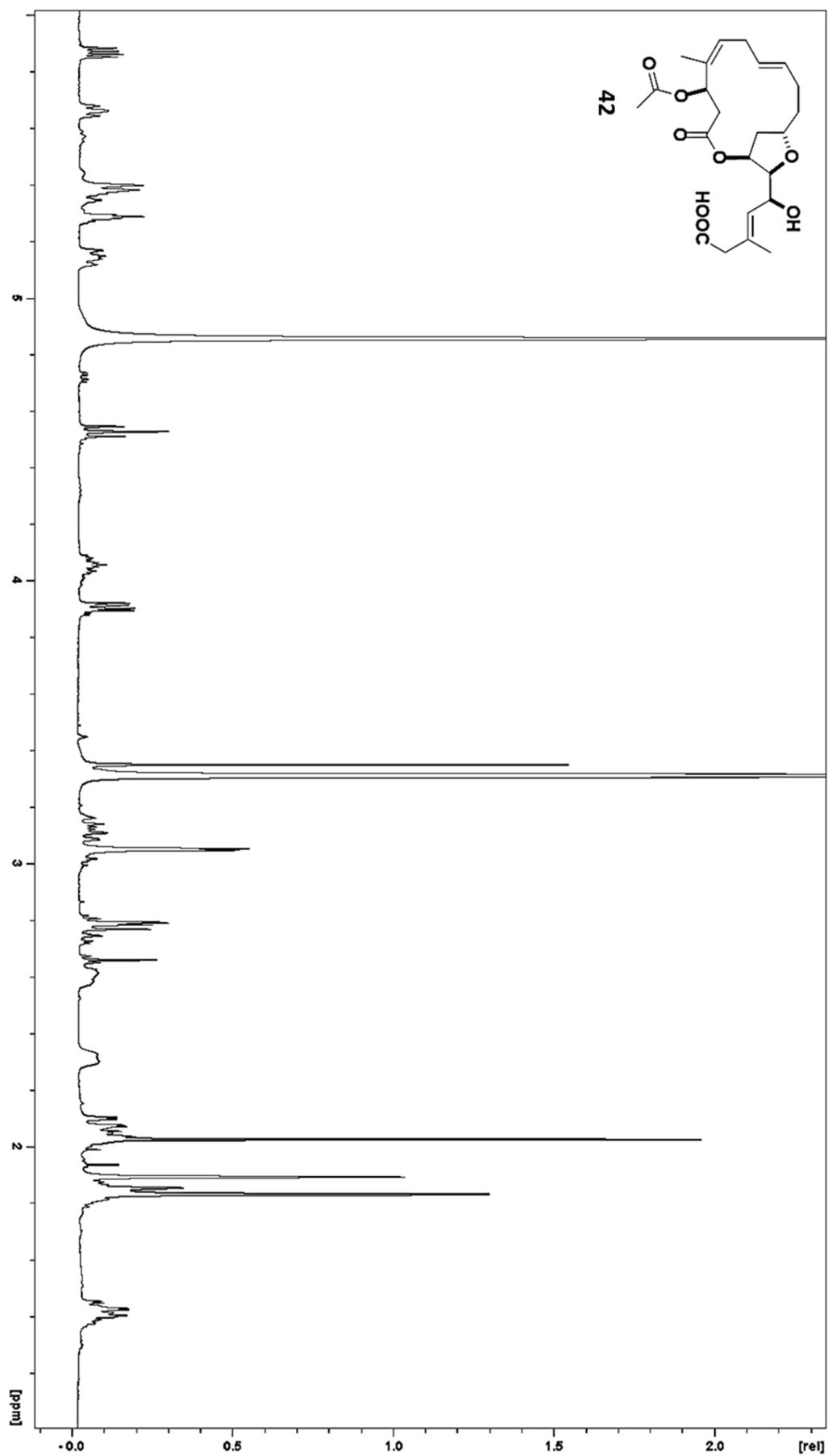


Figure 3-8:  $^1\text{H}$  NMR of dechloro-oocydin A 42 in  $\text{methanol-}d_4$  obtained at 500 MHz.

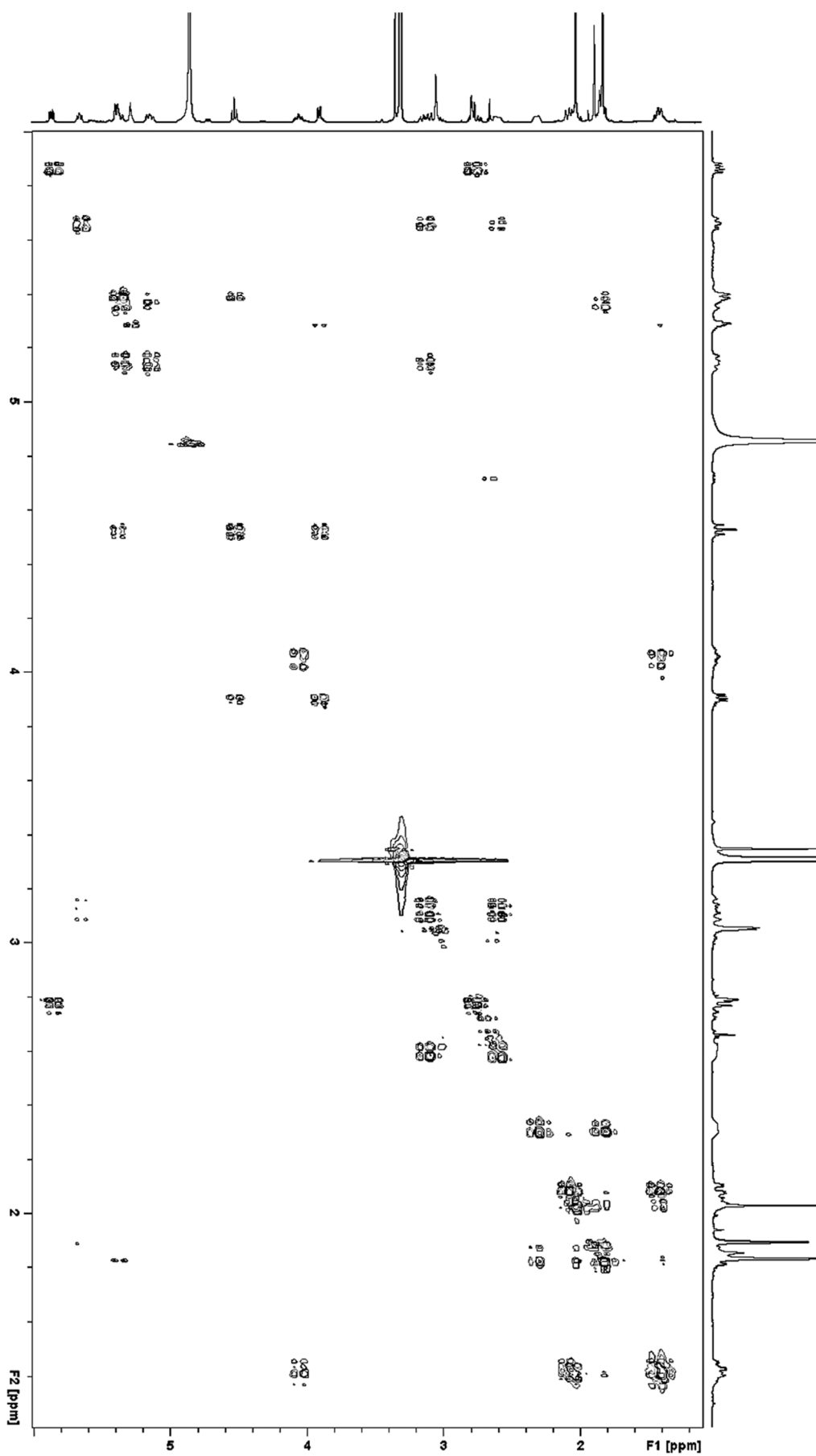


Figure 3-9: COSY spectrum of dechloro-oocycin A 42 in methanol- $d_4$  obtained at 500 MHz.

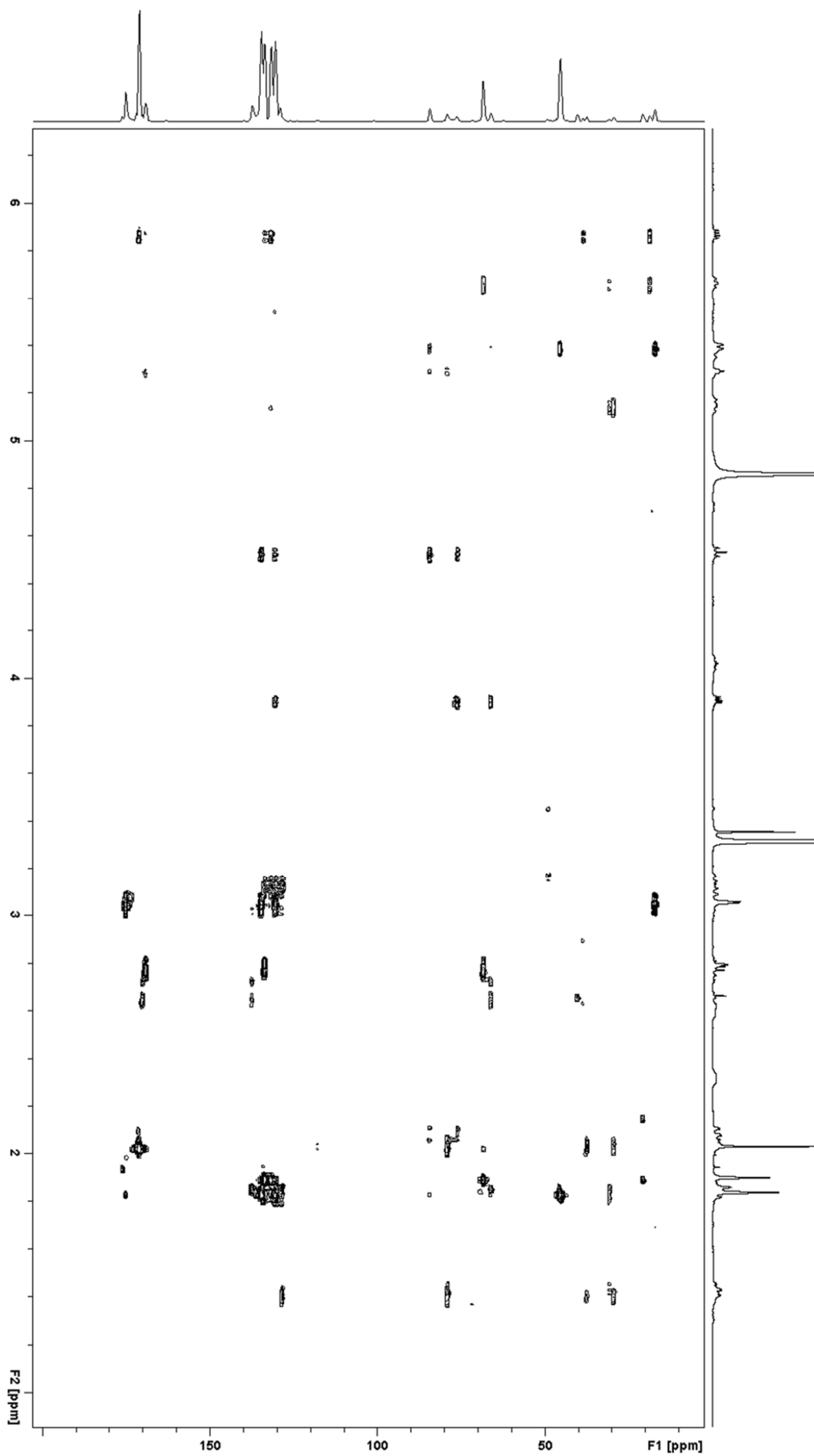


Figure 3-10: HMBC spectrum of dechloro-oocydin A 42 in methanol- $d_4$  obtained at 500 MHz.

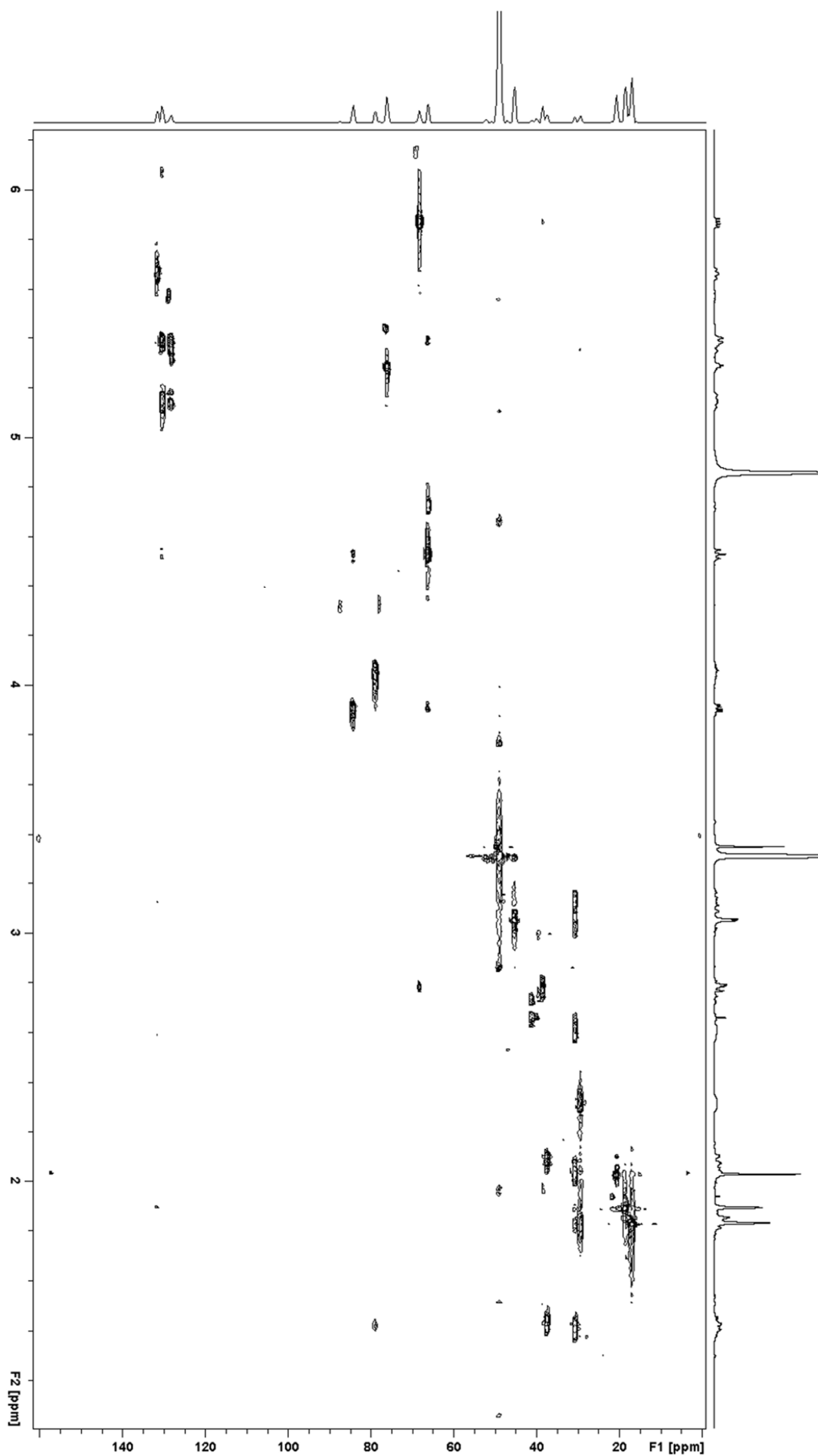


Figure 3-11: HSQC spectrum of dechloro-oocydin A 42 in methanol- $d_4$  obtained at 500 MHz.

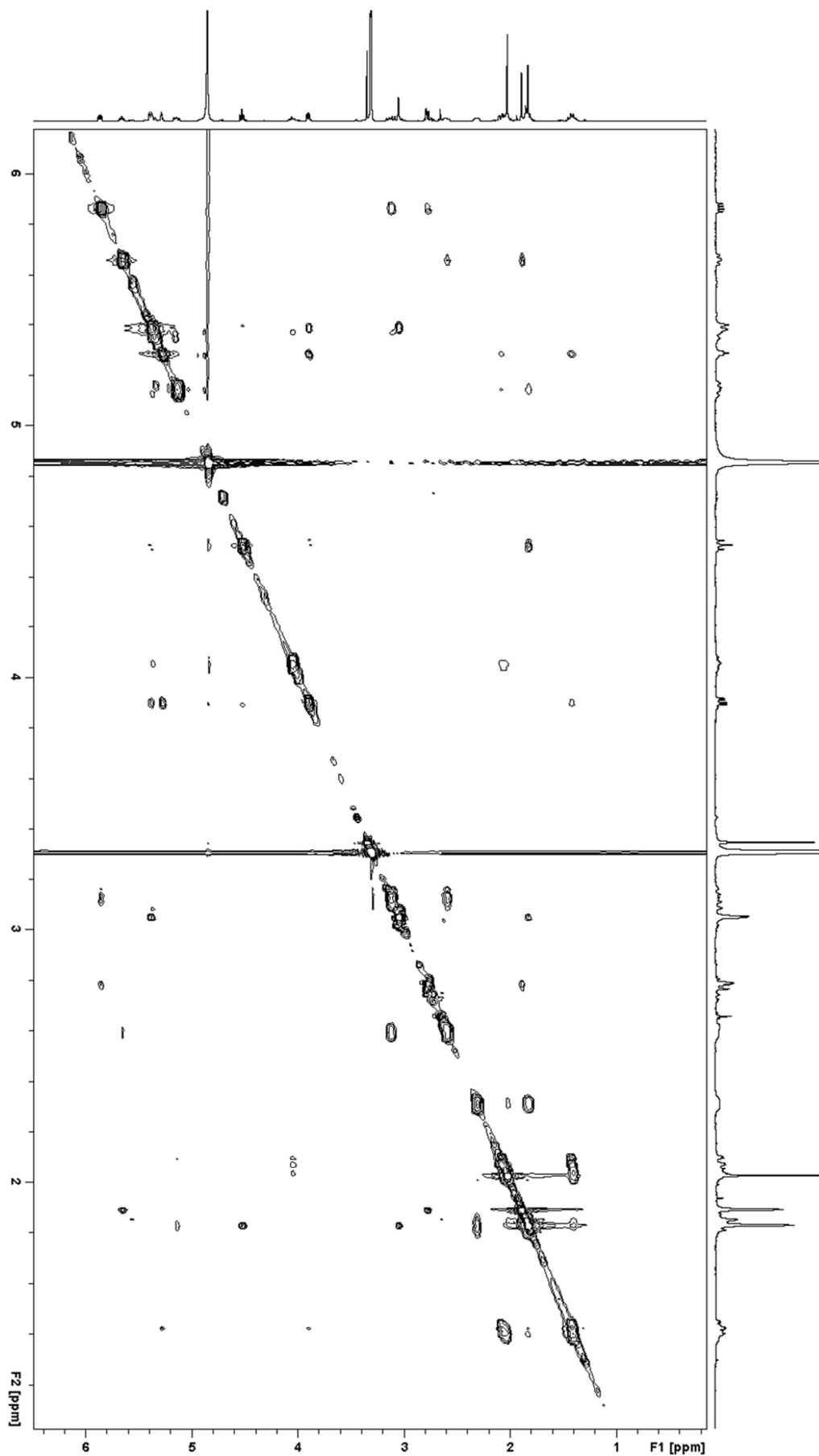


Figure 3-12: NOESY spectrum of dechloro-oocycin A 42 in methanol- $d_4$  obtained at 500 MHz.

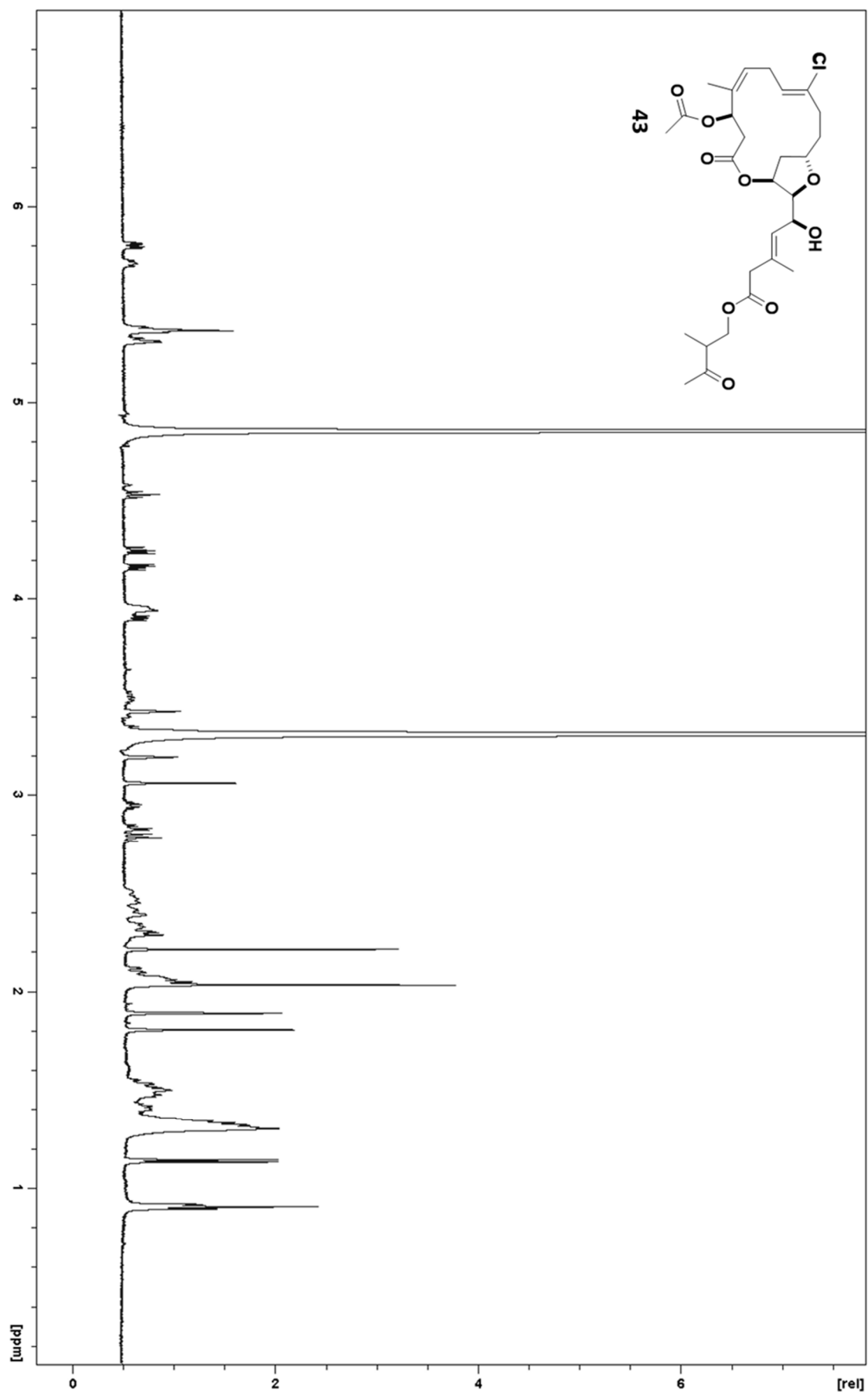


Figure 3-13:  $^1\text{H}$  NMR spectrum of oocydin B 43 in methanol- $d_4$  obtained at 500 MHz.

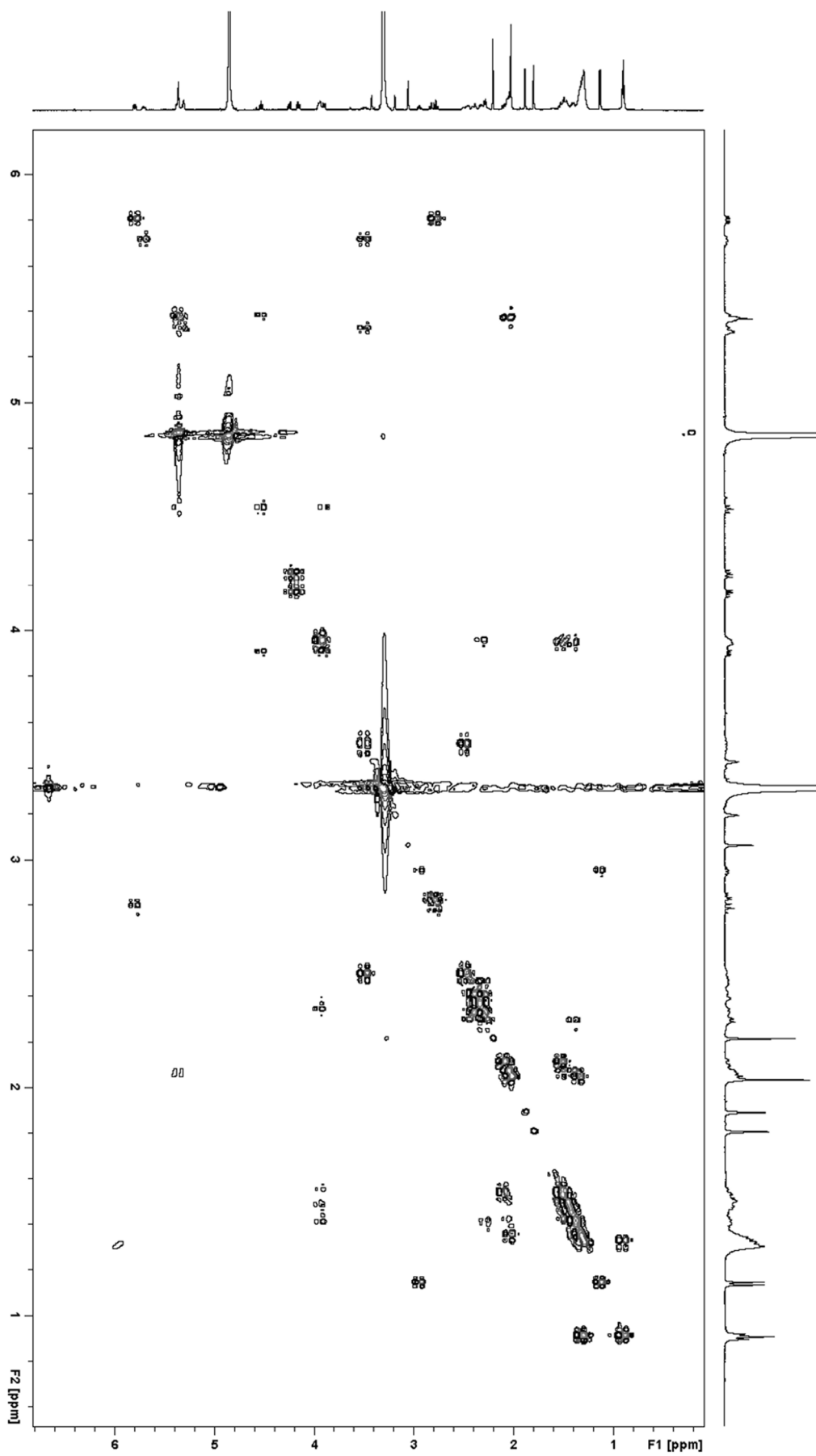


Figure 3-14: COSY spectrum of oocycin B 43 in methanol- $d_4$  obtained at 500 MHz.

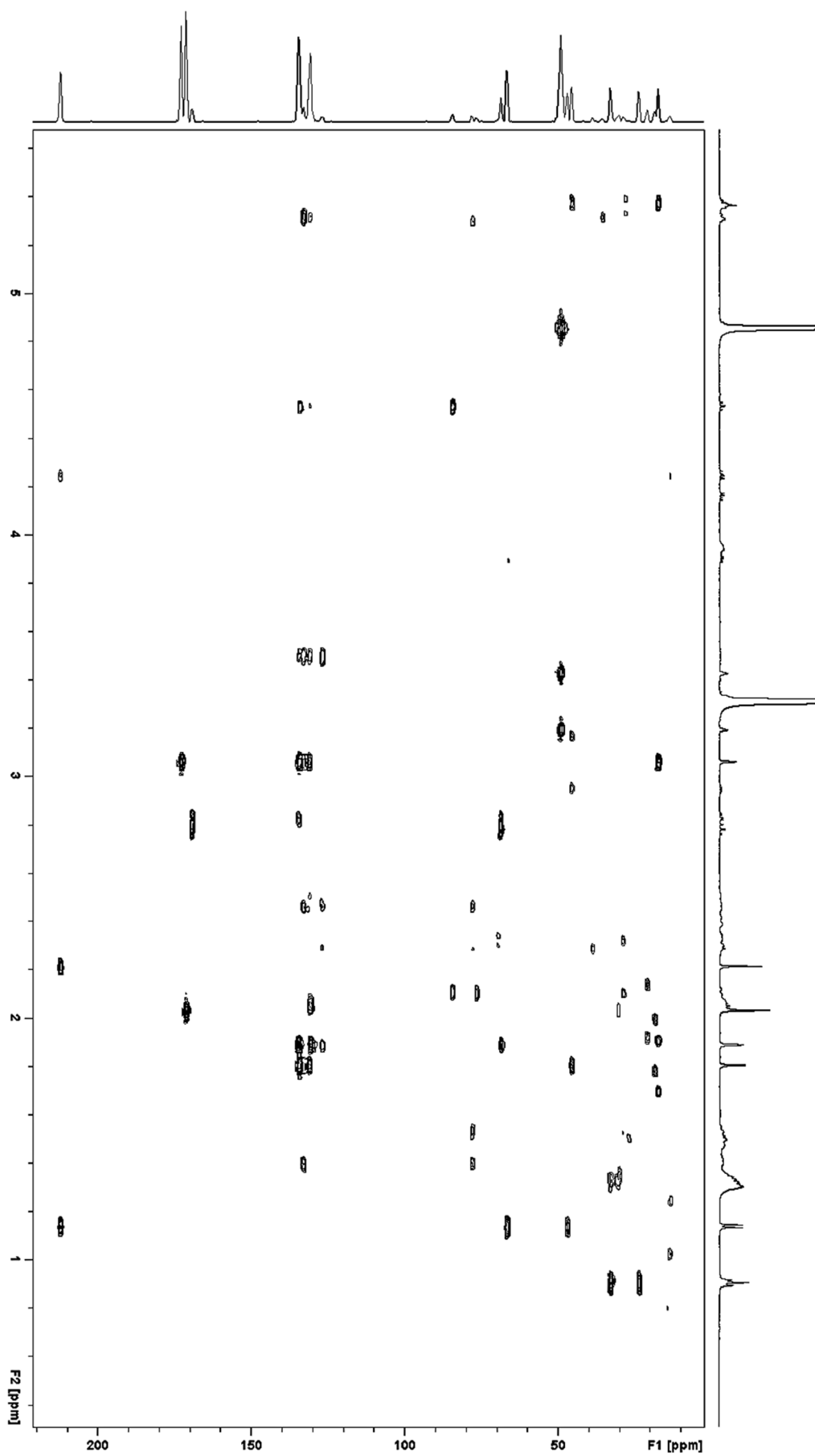


Figure 3-15: HMBC spectrum of oocydin B 43 in methanol-*d*<sub>4</sub> obtained at 500 MHz.



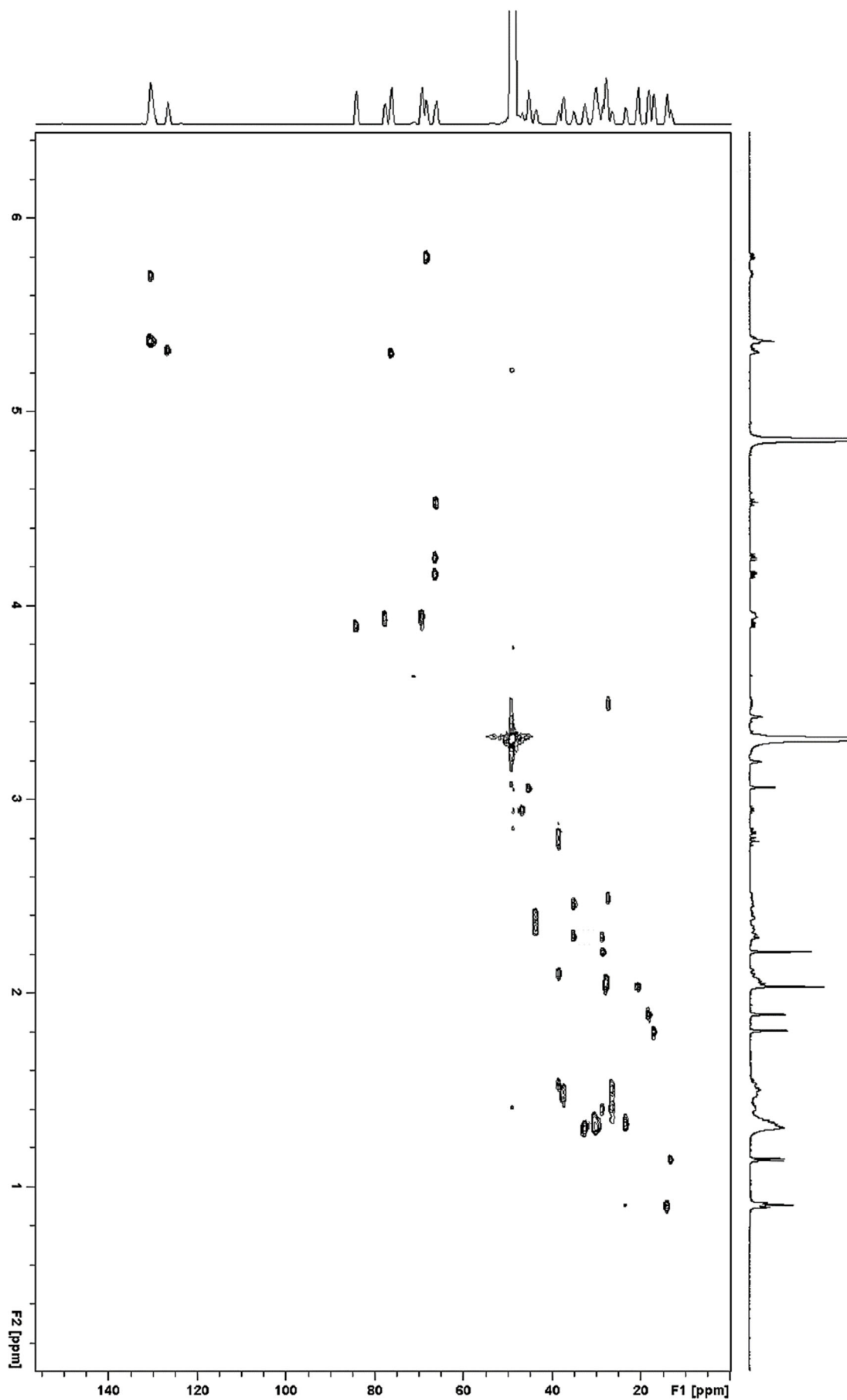


Figure 3-16: HSQC spectrum of oocycin B 43 in methanol- $d_4$  obtained at 500 MHz.

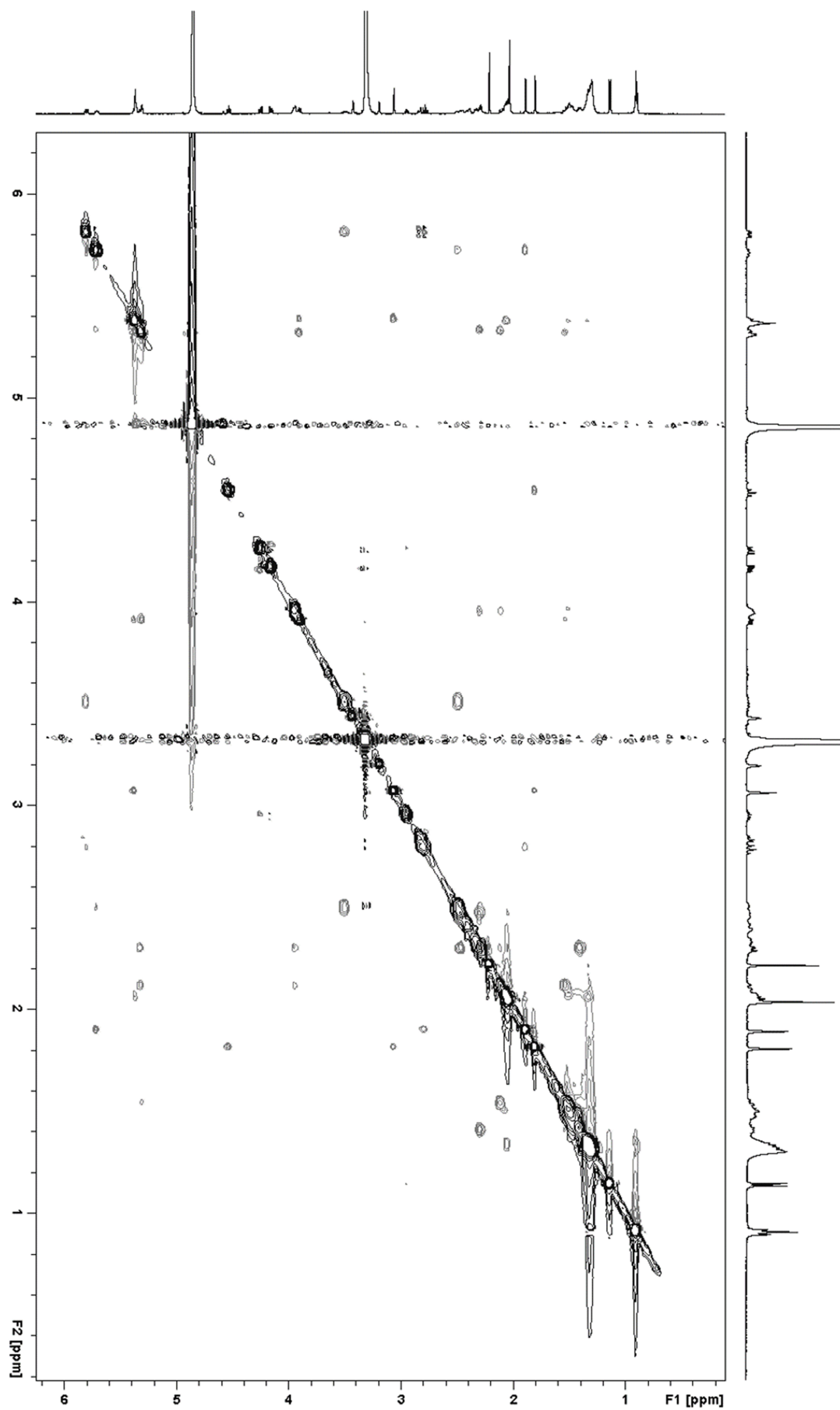


Figure 3-17: NOESY spectrum of oocydin B 43 in methanol- $d_4$  obtained at 500 MHz.

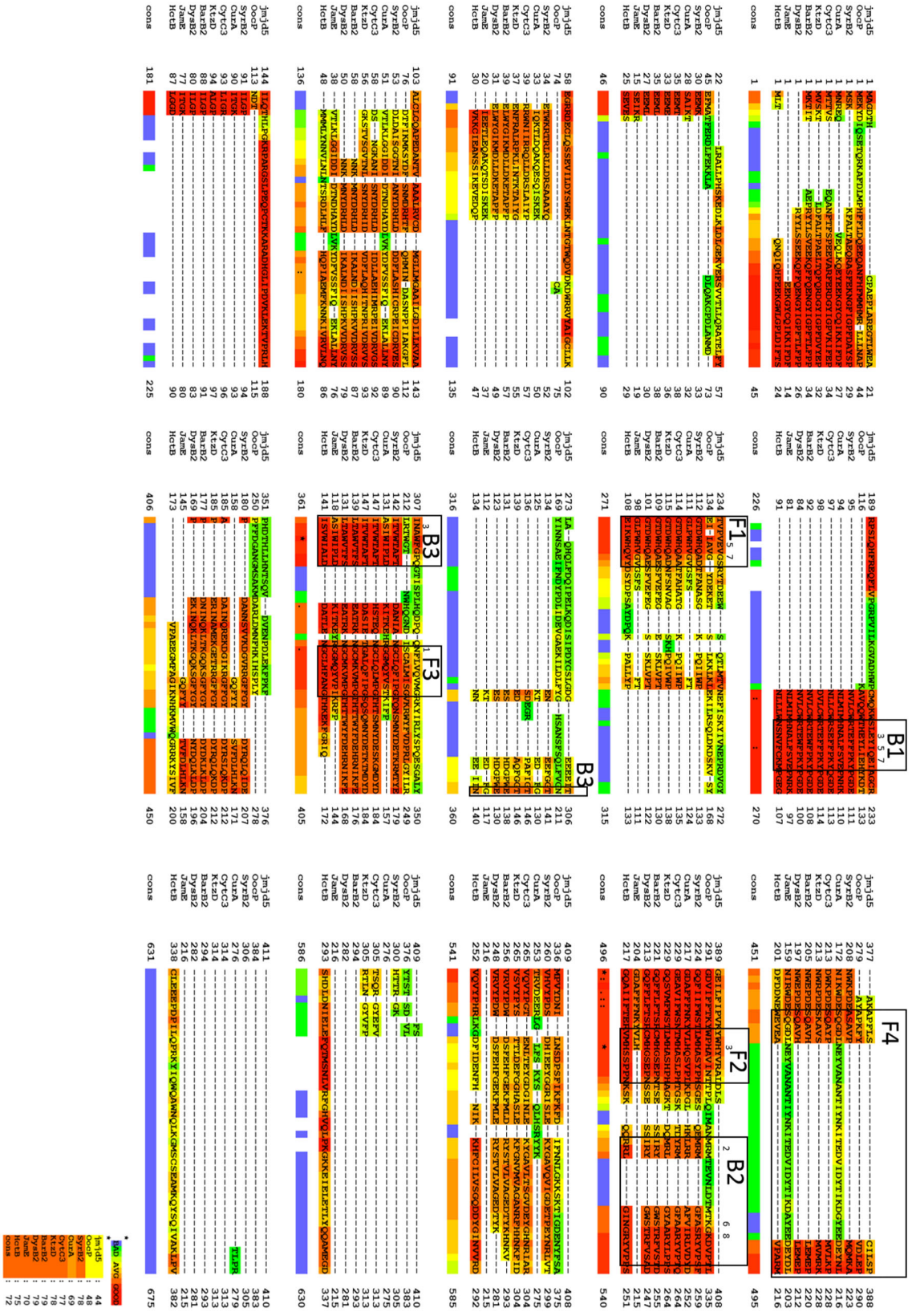


Figure 3-18: Sequence alignment of known MNH halogenases using T-Coffee web server (Di Tommaso et al. 2011).

Halogenases from the following pathways were used: BarB1, barbamide (*Lyngbya majuscula*); CurA, curacin (*Lyngbya majuscula*); CytC3, cytrotienin (*Streptomyces sp.*); DysB2, dysidenin (*Hormosilla spongelliae*); HctB, hectochlorine (*Lyngbya majuscula*); JamE, jamaidamide (*Lyngbya majuscula*); KtzD, kutzneride (*Kutzneria sp. 744*); OocP, oocydin (*Serratia plymuthica*); SyrB2, syringomycin (*Pseudomonas syringae* pv. *syringae*); for multidomain proteins only the relevant domain is shown; JmjD5, (*Homo sapiens*) served as hydroxylase comparison.

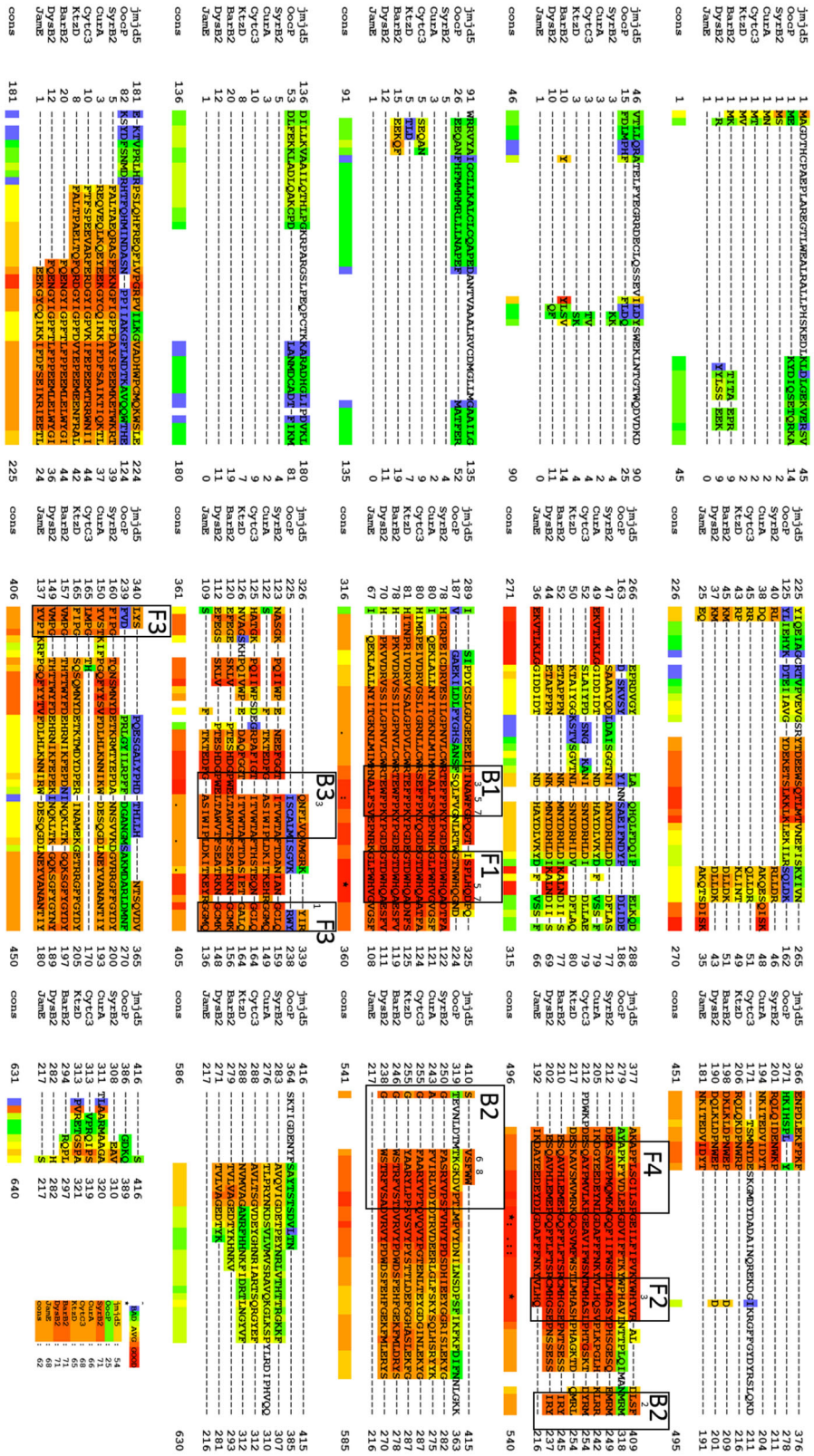


Figure 3-19: Sequence alignment of known MNH halogenases using Expresso web server (Armougom et al. 2006).

Halogenases from the following pathways were used: BarB1, barbamide (*Lyngbya majuscula*); CurA, curacin (*Lyngbya majuscula*); CytC3, cytrotienin (*Streptomyces* sp.); DysB2, dysidenin (*Hormoscilla spongelliae*); HctB, hectochlorine (*Lyngbya majuscula*); JmE, jamaidamide (*Lyngbya majuscula*); KtzD, kutzneride (*Kutzneria* sp. 744); OocP, oocydin A (*Serratia plymuthica*); SyrB2, syringomycin (*Pseudomonas syringae* pv. *syringae*); for multidomain proteins only the relevant domain is shown; JmjD5, (*Homo sapiens*) served as hydroxylase comparison.

# Chapter 4. Studying Recognition of the Unusual RiPP Protease PoyH

---

Katja Jensen, Michael Burger, Maximilian J. Helf, Jörn Piel

Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zurich, 8093 Zurich, Switzerland

K.J., M.B., and M.H. cloned constructs, M.B. expressed proteins and conducted the assays, K.J. and M.B. analyzed data, all authors planned experiments and K.J. and J.P. wrote the manuscript.

*In preparation*

## 4.1 Summary

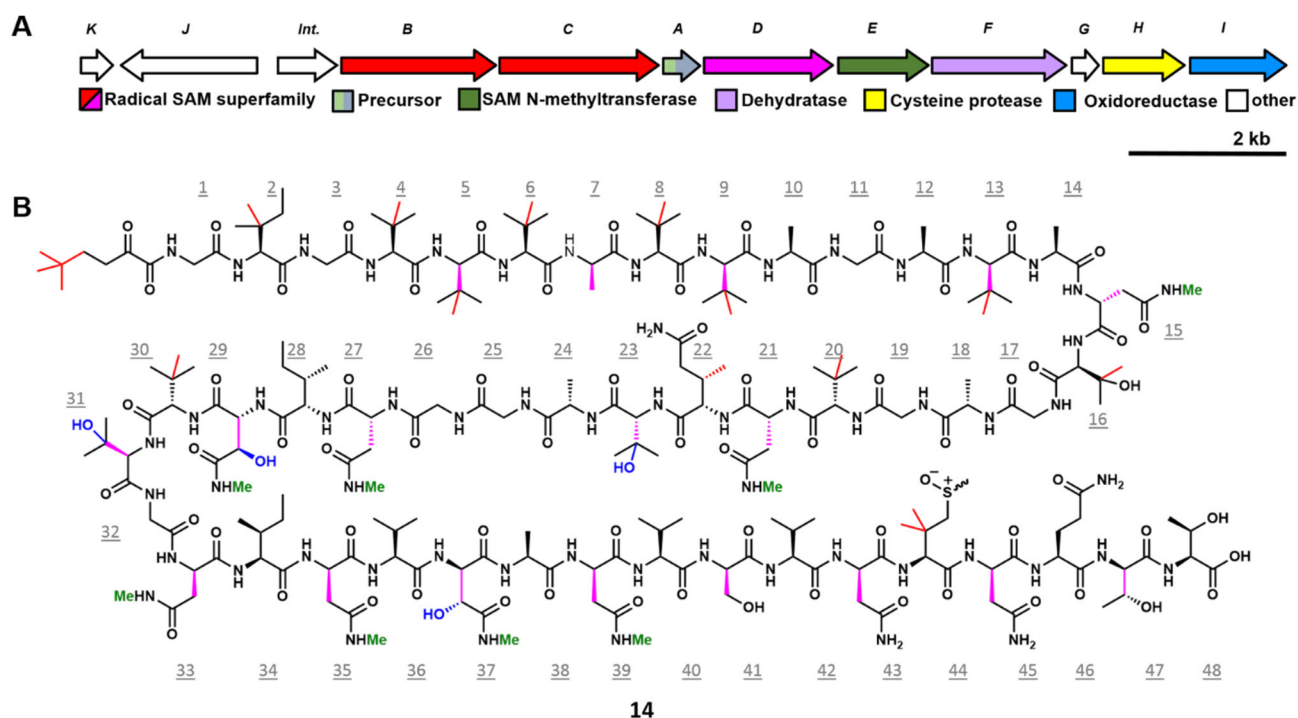
Ribosomally synthesized and posttranslationally modified peptides (RiPPs) are an important class of bioactive natural products. The proteusins are a young and extraordinary family within the RiPPs, characterized by their large nitrile hydratase or Nif11-like N-terminal leader peptide. The polytheonamide gene cluster *poy* is the first characterized member of this new RiPP family. In the *poy* gene cluster PoyH is encoded, a within RiPPs unprecedented protease. In order to investigate the substrate specificity of PoyH, we mutated the putative recognition sequence of the protease resulting in the identification of a refined recognition motif. We then explored the potential of PoyH as a general RiPP protease on a nonnative substrate. The data suggests that PoyH can be applied as general tool for RiPP synthesis. Additionally we found evidence suggesting a zymogen form of PoyH.

## 4.2 Introduction

Secondary metabolites are produced by all three domains of life and span a wide variety of compounds like polyketides, alkaloids, nonribosomal peptides or ribosomally synthesized and posttranslationally modified peptides (RiPPs). The latter is known as a major class of secondary metabolites (Oman and van der Donk 2010). Several members of the RiPP class were known and studied for long time, but the immense abundance of RiPPs were only realized with the advent of affordable high throughput genome sequencing. Recently a comprehensive report was published where all known RiPP family members were compiled and a common nomenclature for all RiPP biosynthetic pathways was suggested (Arnison et al. 2013). In general the biosynthesis starts with one ribosomally produced precursor peptide consisting of a leader and a core peptide. The core peptide is often modified by a small set of tailoring enzymes and subsequently cleaved off by a protease (Yang and van der Donk 2013).

The proteusins are a recently described subclass of RiPPs with polytheonamide A and B (**14**) from endosymbionts of the marine sponge *Theonella swinhoei* as the first characterized examples (Freeman et al. 2012), (Wilson et al. 2014). Polytheonamide A and B vary only in the configuration of the sulfoxide moiety in residue 44 that is an isolation artefact (Figure 4-1B) (Freeman et al. 2012). Polytheonamides (**14**) were reported in 1994 and until recently to be of nonribosomal origin because they contain nonproteinogenic  $\beta$ -methyl and D-amino acids, uncharacteristic for ribosomally derived peptides (Hamada et al. 1994), (Inoue et al. 2010). However, isolation and characterization of the 15 kb polytheonamide (*poy*) biosynthetic gene cluster *poy* revealed that up to 49 modifications occur on the 49-residue ribosomal peptide (Figure 4-1) (Freeman et al. 2012). These posttranslational

modifications are C- and N-methylations, hydroxylations, dehydratations and epimerizations. Due to the latter modification an almost perfect iteration of D- and L-amino acids results.



**Figure 4-1: Polytheonamide 14 gene cluster and corresponding structure.**

(A) Polytheonamide gene cluster; gene functions are indicated with corresponding colors. Int- integrase (B) Polytheonamides (**14**) only vary in the configuration of the sulfoxide moiety in residue 44 that is an isolation artefact. Posttranslational modifications are colored, analogous to genes and colors depicted in (A). Underlined numbers correspond to the residue.

Members of the proteusin family are characterized by their remarkably long N-terminal leader peptide that does not show close homology to any other precursor peptides from other characterized RiPP biosynthetic pathways (Freeman et al. 2012). The leader is homologous to nitrile hydratase or Nif11 proteins. The existence of such a family was predicted by an *in silico* study analyzing sequenced bacterial genomes by comparative genomics methods to identify secondary metabolite biosynthetic machineries (Haft, Basu, and Mitchell 2010).

Besides the unusual leader region, a further peculiarity of the polytheonamide pathway is the existence of a non-membrane bound protease. This is striking because most other RiPP proteases are membrane bound (Yang and van der Donk 2013). In polytheonamide biosynthesis, PoyH is the protease responsible for cleavage between the precursor and core peptide and does not have close homologs in other RiPP gene clusters (M.J. Helf, unpublished). PoyH exhibits close homology to cysteine peptidases of which cathepsins and papain are prominent representatives. A BLAST query against MEROPS, a database for peptidases (Rawlings, Barrett, and Bateman 2012) reveals a classification as a member of the C1A family, which contains endo- and exopeptidases and employs a Cys-His catalytic dyad. Often, members of this family contain an N-terminal propeptide that needs to be cut off for activation (Rawlings and Barrett 1994). Previous results obtained in our group showed a substrate cleavage of PoyH both *in vivo* and *in vitro*. Additionally, an inhibition of PoyH by PoyG was observed (M.J. Helf, unpublished).

Little is currently known about the function of the leader peptide and the action of the protease for release of the final RiPP product (Yang and van der Donk 2013). We have studied the recognition of the cleavage site of the protease PoyH and tested probable application potential.

## 4.3 Results

### 4.3.1 Generation of a model substrate for further cleavage studies

Because the precursor peptide PoyA is the only known substrate of PoyH, a sequence alignment of substrates to gain information regarding the recognition site was not possible. The usage of the classical GG-cleavage motif and a high conservation within other hypothetical leader peptide sequences (Haft, Basu, and Mitchell 2010) served as starting point to narrow down a potential recognition sequence. Previous functional studies suggested the preliminary recognition sequence LDQAAGGTG (M.J. Helf, unpublished). A highly similar motif was detected in a RiPP sequence from the genome of *Pelotomaculum thermopropionicum* SI (hypothetical protein PTH\_2340; NCBI accession number: YP\_001212890.1), a Gram-positive thermophile bacterium (Kosaka et al. 2008) (Figure 4-2).

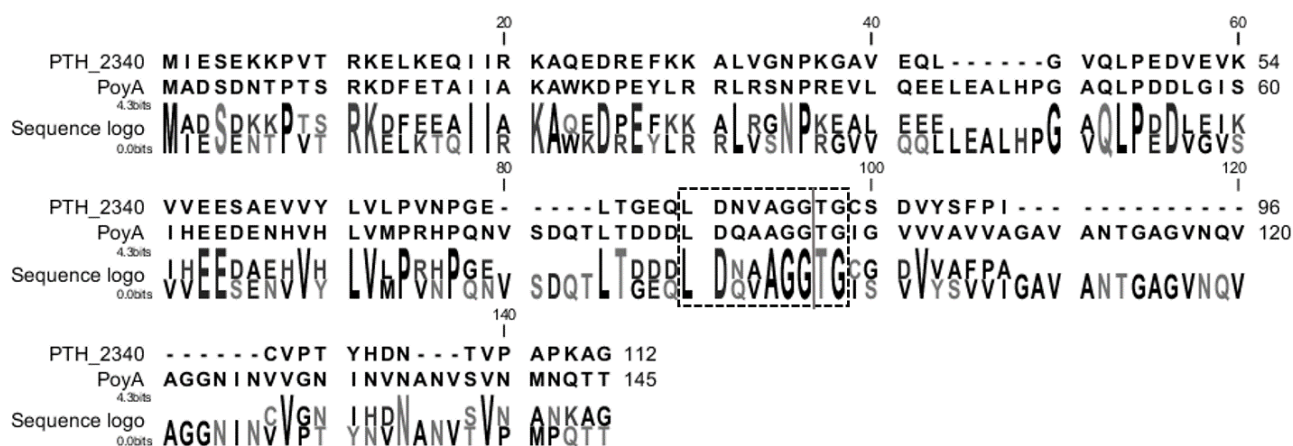


Figure 4-2: Sequence alignment of PoyA and PTH\_2340 recognition site.

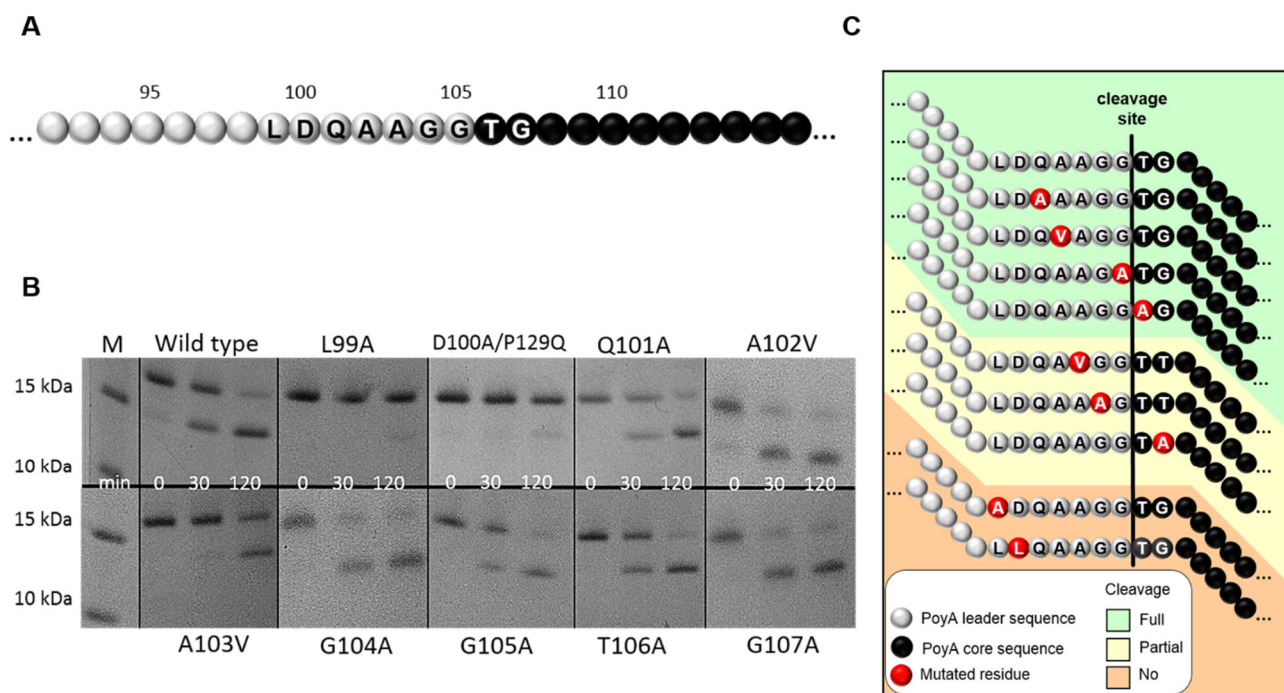
Upper sequence is the RiPP sequence from *Pelotomaculum thermopropionicum* SI, lower sequence is PoyH, the recognition site is framed. Cleavage occurs in PoyA between position 96/97 and in PTH\_2340 between position 85/86 as indicated with the grey bar.

Because recombinant PoyA expresses only in very low amount and exclusively in coexpression with modifying enzymes, we decided to conduct our studies with PTH\_2340 which expresses much better. The two differing amino acids from PTH\_2340 were mutated to match the PoyA sequence in pMH029b. Earlier experiments suggested that sequences further upstream and downstream of the recognition sequence seem to have little influence on the acceptance (M.J. Helf, unpublished).

### 4.3.2 Single point mutation screen refines recognition sequence

In order to further characterize the recognition sequence, substrates were generated in which every single amino acid within the preliminary recognition site LDQAAGGTG (Figure 4-3A) was mutated to alanine. In case an alanine was the wild-type amino acid it was exchanged with valine. All proteins were verified by high performance liquid chromatography-high resolution mass spectrometry (HPLC-HRMS) measurements with program 1 (see Experimental procedure) after tris(2-carboxyethyl)phosphine (TCEP) reduction of disulfide bonds which allows a better HPLC separation. For the D100A-mutant a P129Q mutation appeared but was accepted since it is downstream of the recognition motif, nearly at the end of the protein and believed to not influence the cleavage as suggested by previous experiments. The subsequent cleavage assay was adapted from Rascón *et al.* (Rascón et al. 2011). After optimization of the assay conditions every substrate was tested in the cleavage assay containing acetate buffer, pH 4.5, with a protease-free negative control. A quenching of the cleavage reaction occurred after

various time points. The assay was repeated several times and the following cleavage velocities were observed as exemplarily depicted in Figure 4-3B. After 30 min of incubation almost the entire wild type protein substrate was cleaved. Comparable results were seen for Q101A, A102V and T106A mutants. An impairment in cleavage velocity could be detected in G104A and G107A mutants. The cleavage of the G105A mutant was only slightly impaired. Cleavage was nearly abolished in L99A and D100A mutants and severely impaired in the A103V mutant. That shows that some residues are more crucial for recognition than others (Figure 4-3C). The recognition sequence was refined to a LDxxAG(G/A)xG motif.



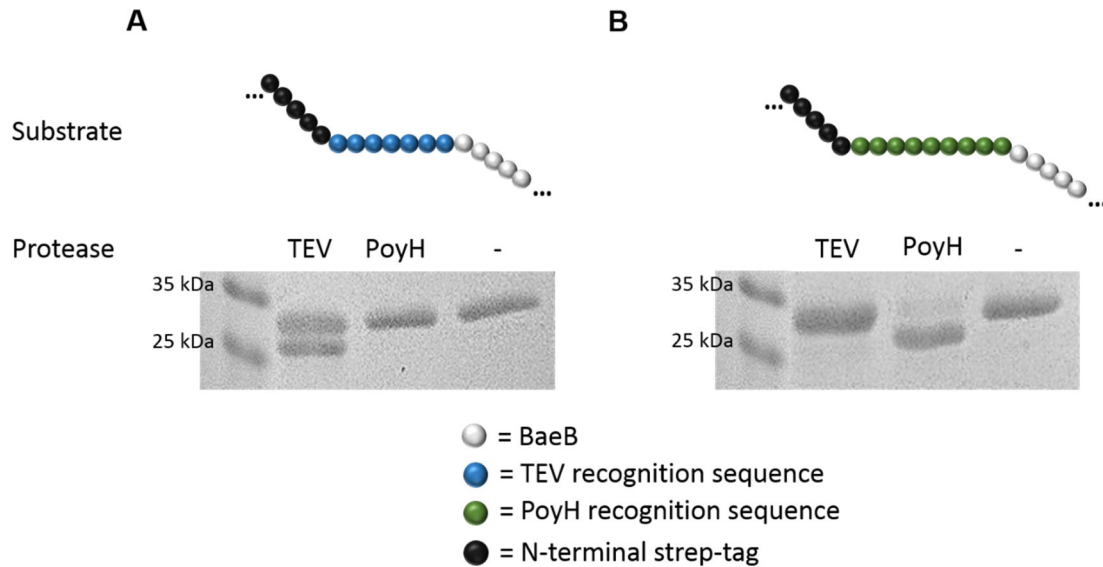
**Figure 4-3: Comparison of PoyH-performance with single-site mutants of PoyH recognition site.**

(A) Part of substrate sequence (PTH\_2340) out of pMH029b with the adjusted recognition site, numbers correspond to residue in full length protein. White indicates leader sequence and black indicates core sequence as in (C). (B) Representative SDS-PAGE of cleavage assay. Amino acids of every single position were mutated and exposed to cleavage by PoyH. The reaction was quenched directly, after 30 min and 2 hrs respectively. (C) Schematic summary of cleavage efficiency in single point mutations within the PoyH recognition sequence.

To test if PoyH is able to cleave proteins of unrelated origin we inserted the recognition sequence into BaeB from the bacillaene pathway, a member of *trans*-AT polyketide synthases (PKS) (Moldenhauer, Chen, and Borriss 2007). BaeB is a protein with so-far uncharacterized function which was incorrectly hypothesized to be involved in a correction mechanism during bacillaene biosynthesis (Jensen et al. 2012). The good protein expression properties of tagged BaeB were known from previous studies and the size of BaeB allows an adequate detection of cleavage by a SDS-PAGE analysis. The purified and tagged protein has a size of 30 kDa, whereas after cleavage the untagged protein would have only a size of 27.5 kDa using the chosen cleavage site. We generated a construct harboring BaeB with an N-terminal Strep-tag and the PoyH recognition motif and a second construct where the PoyH cleavage site was replaced with the well characterized tobacco etch virus (TEV) protease cleavage site (Parks et al. 1994). Both substrates were applied in a cleavage assay. SDS-PAGE analysis revealed two bands with a small size difference for cleavage by TEV protease of N-terminal strep-tagged BaeB carrying the TEV recognition motif, indicating a partial cleavage of the substrate (Figure 4-4A). No cleavage was detectable after PoyH treatment of the same substrate and neither in the control without any protease. If the N-terminal strep-tagged BaeB carrying the PoyH recognition motif was offered PoyH for cleavage a nearly full shift in protein size was detectable on the corresponding SDS-PAGE (Figure 4-4B) suggested that PoyH has cleaved off the Strep-tag at most of the substrate peptide molecules



offered. TEV protease does not accept this substrate with TEV recognition site and no cleavage was additionally detected in the control without any protease. These data suggest that PoyH is able to cut also in a fully RiPP unrelated system. A cleavage of the TEV recognition site by PoyH or vice versa was not visible indicating that both proteases are highly specific for their respective recognition sequence. It is further noteworthy that PoyH is seemingly more efficient in cleavage than TEV under this conditions.

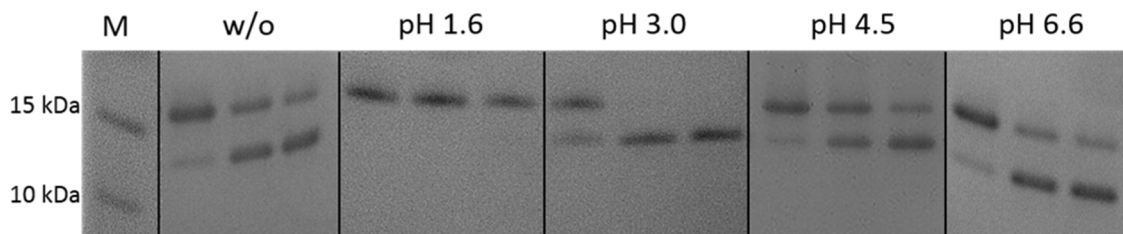


**Figure 4-4: Cleavage assay of TEV versus PoyH in BaeB.**

(A) Cleavage of Strep-tag from BaeB with an inserted TEV recognition sequence. (B) Cleavage of Strep-Tag from BaeB carrying PoyH recognition sequence. Each substrate was assayed with TEV protease, PoyH protease and without any protease. The upper band in SDS-PAGE corresponds to the non-cleaved substrate, the lower one to the cut one. Assays were conducted in buffer containing 100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin with incubation over night.

#### 4.3.1 Detection of a putative zymogen form of PoyH

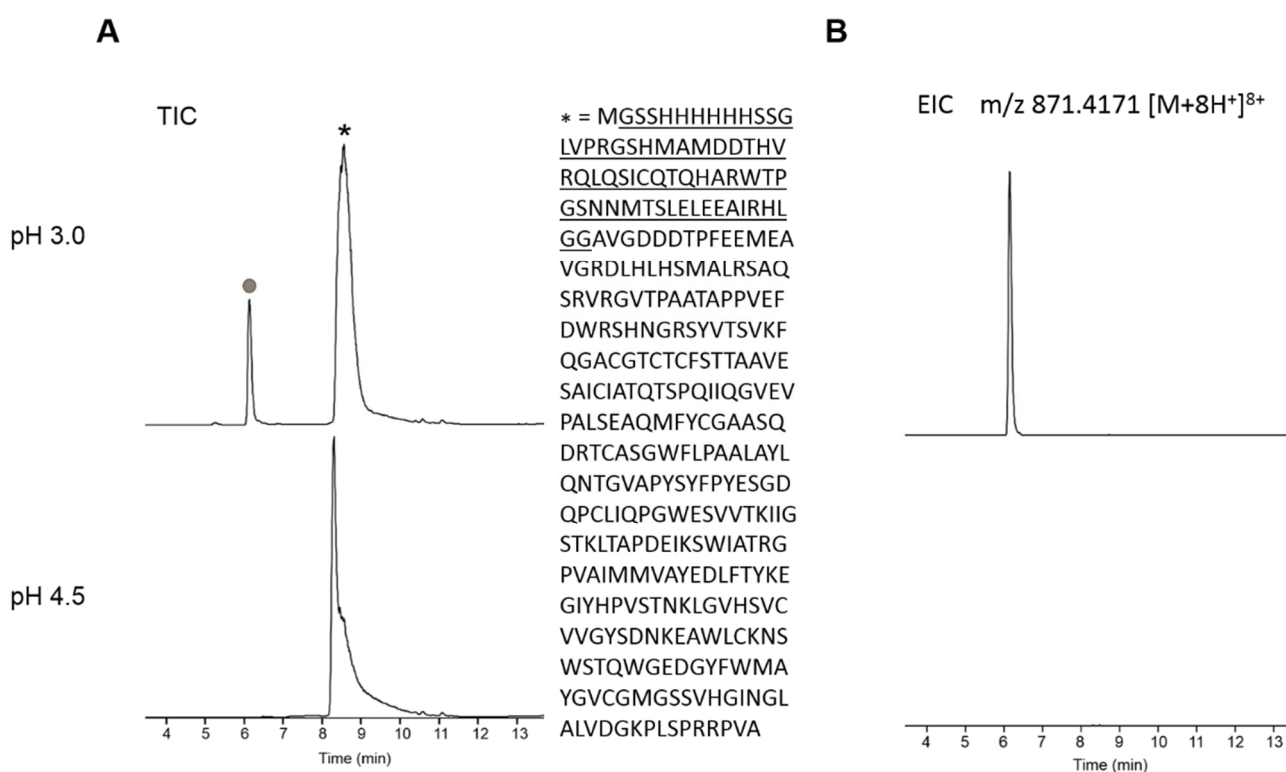
To test how different pH values of the assay buffer influence cleavage velocity of the non mutated wild type proteusin by PoyH, the protease was incubated for 6 hrs at 4 °C in acetate buffer having pH values of 1.6, 3.0, 4.5 and 6.6. Afterwards this preincubated protease was employed in a regular cleavage assay (Figure 4-5). Unincubated protease served as control. A preincubation in acetate buffer with pH 4.5 does not alter cleavage velocity compared to the control. A pH of 1.6 inactivated the protease as no cleavage at all is detectable on the SDS-PAGE. A pH of 3.0 and 6.6 seems to increase reaction speed in comparison with the control, whereas the highest velocity was detectable at pH 3.0.



**Figure 4-5: Comparison of PoyH-performance after preincubation in acetate buffer with different pH values.**

Representative SDS-PAGE of cleavage assay in acetate buffer, pH 4.5 with PoyH incubated in acetate buffer with different pH values prior to assay performance. As substrate served the non mutated wild type proteusin. Respective pH values are indicated above corresponding lanes. As control (w/o) the assay was conducted with PoyH not incubated prior to assay performance.

For cysteine proteases, self-cleavage of an inhibitory propeptide is known to yield the active form of the enzyme (Rawlings and Barrett 1994). In order to test if this self-cleavage might be responsible for the observed increased cleavage velocity, we incubated the protease for several hours in acetate buffer, pH 3.0, and as comparison in acetate buffer, pH 4.5, and subjected both samples to HPLC-HRMS measurements with program 2 and 3 (see Experimental procedure). As seen in Figure 4-6 an additional peak is detectable after incubation at pH 3.0 contrary to incubation at pH 4.5. The mass corresponding to this peak is  $m/z$  871.4171 obtained for  $[M+8H]^8+$ , in agreement with a 62 aa N-terminal portion of PoyH including the hexahistidine-tag ( $m/z$  871.4184 calculated for  $[M+8H]^8+$ ). A mass which could correspond to the C-terminal part of PoyH (the not underlined sequence in Figure 4-6A,  $m/z$  1255.8906 obtained for  $[M+24H]^{24+}$ ) can only be detected in the sample of the protease incubated in acetate buffer, pH 3.0 and not in the corresponding sample which was incubated in acetate buffer, pH 4.5. HPLC-HRMS data suggest a C-terminal peptide which is one alanine shorter than expected whereas it is not possible to determine if the N- or C-terminal alanine is missing ( $m/z$  1255.8857 calculated for  $[M+24H]^{24+}$ ). To further investigate this observed effect and exclude other possible causes of the increased reaction velocity and the occurring second peak additional experiments including MS/MS experiments have to be conducted.



**Figure 4-6: HPLC-HRMS chromatogram for treated PoyH in acetate buffer with different pH values run on a C4-column.**

(A) Total ion chromatogram (TIC) of PoyH incubated in acetate buffer, pH 3.0 and pH 4.5. Asterisk indicates full-length PoyH, the corresponding sequence can be seen on the right side. The circle indicates the cleaved part, corresponding sequence is underlined in the full-length PoyH ( $m/z$  871.4184 calculated, 871.4171 obtained for  $[M+8H]^8+$ ). (B) Extracted ion chromatogram (EIC) of  $m/z$  871.4171  $[M+8H]^8+$ .

#### 4.4 Discussion

We were able to refine the recognition sequence of PoyH to LDxxAG(G/A)xG. It was shown that especially L99A, D100A and A103V mutations have a large impact on recognition. For D100A/P129Q substrate it has to be excluded

that the effect results from the P129Q mutation rather than the aimed D100A mutation. The refinement is mostly in accordance with the detected L(E/D)xVAGG-conservation within this moiety (Haft, Basu, and Mitchell 2010).

From our experiments with BaeB it is obvious that the sequence of the original leader peptide does not play a role in substrate recognition of PoyH. By Phyre<sup>2</sup> modelling (Kelley and Sternberg 2009) we can exclude that BaeB or the N-terminal Strep-tag in the construct of BaeB with PoyH cleavage site fold the same way like the N- or C-terminal proportion of PTH\_2340 in pMH029b or mutants.

PoyH might be a widely applicable protease since we detect cleavage at various temperatures and pH conditions. We have successfully introduced the recognition motif in an unrelated protein and shown cleavage in a broad pH (pH 3.0 up to pH 8.0) and temperature (4 °C - 40 °C) range. Furthermore PoyH appears to be more efficient than the well-established TEV protease at least in the tested example. 45 times less amount of PoyH on a molar level was used in comparison to TEV and even with these much smaller amounts a more complete cleavage occurred within the same timeframe under these assay conditions. The observed robustness of PoyH strongly suggest that it might serve as a valuable tool in RiPP biosynthesis research and other fields where distinct protein cleavage is needed. An introduction of the recognition motif in further distinct proteins is currently under investigation.

The detection of an additional ion peak in the total ion chromatogram occurring after incubation in acetate buffer at pH 3.0 might be explained by the conversion of a zymogen into the active protease which can occur autocatalytically (due to e.g. pH drop) or with accessory proteins (Khan and James 1998). Further investigations are needed to exclude other underlying causes like HRMS measurement artefacts as not all data support the hypothesis of zymogen activation. On the one hand the zymogen activation is further supported beside the HPLC-HRMS-data by the increased reaction velocity after incubation in acetate buffer, pH 3.0. It is known for propapain that the optimal pH for propeptide cleavage is pH 3.3 (Khan and James 1998), a range in which PoyH also showed highest reaction velocity. The presence of a propeptide is also suggested by the MEROPS classification in C1A family since usually members possess such a zymogen form (Rzychon, Chmiel, and Stec-Niemczyk 2004). A probable propeptide would most likely be cleaved off upstream of Q<sub>107</sub> in PoyH as Q<sub>107</sub> is putatively involved together with C<sub>113</sub>, H<sub>259</sub>, N<sub>275</sub> and W<sub>277</sub> in the active site of PoyH for substrate cleaving by comparison with described active sites (Rzychon, Chmiel, and Stec-Niemczyk 2004). A PSI-BLAST Pseudo-Multiple sequence alignment calculated by the Phyre-server (Kelley and Sternberg 2009) indicates propeptide existence (Figure 4-7), which would presumably be cleaved off between residues 60-80 in a region that protrudes due to its low sequence homology within the alignment. This low homology and with that the absence of potentially conserved motifs indicates that this part of the does not encode for catalytically essential regions within the protein. On the other hand the detected additional ion in HPLC-HRMS measurements would correspond to a cleavage around position 45, meaning upstream of Q<sub>107</sub> but also upstream of the proposed self-cleavage region within residues 60-80. Additionally, several conserved sequence motifs are known to indicate the presence and position of a propeptide within the inactive protease and none of the known ERFNIN-, ERFNAQ- or G/AxNx<sub>2</sub>FxD-motifs were detected in the sequence of PoyH (Novinec and Lenarčič 2013), (Khan and James 1998). Prosegments are reported to be specific and potent inhibitors of the mature enzyme (Khan and James 1998). As stated above in the polytheonamide gene cluster an inhibitor is encoded, namely PoyG. This would presumably make an inhibitory propeptide superfluous. To sum up, the presence of a propeptide could not be clearly verified yet and needs further investigation.

In conclusion, at the example of PoyH we helped drawing a picture of RiPP protease recognition and gained information about the importance of leader sequences for protease cleavage. We were able to analyze the impact of single-point mutations in the recognition site and with that gain insights into cleavage specificity and substrate recognition of PoyH by testing a set of recognition site mutants and could demonstrate the utilization of PoyH in the cleavage of RiPP-unrelated proteins, making it more accessible as a tool for the targeted cleavage of engineered proteins and bioengineering of novel peptides.

## 4.5 Experimental procedure

### 4.5.1 Strains and culture conditions

*Escherichia coli* BL21(DE3) or *E. coli* BL21A1 (both Invitrogen) served as the host strain for protein expression experiments. *E. coli* strain XL1 blue (Stratagen) served as host for routine subcloning. *E. coli* strains were grown in Luria-Bertani (LB) medium (Sambrook and Russell 2001) and TB medium (Tartoff and Hobbs 1987) for protein expression. Detailed information regarding antibiotics, additions and inducer of the single strains can be found in Table 4-1.

### 4.5.2 Construction of expression plasmids and general DNA procedure

DNA isolation, plasmid preparation, restriction digests, gel electrophoresis, and ligation reactions were conducted according to standard methods. Final constructs were verified by sequencing. All expression plasmids employed in this study are provided in Table 4-1 and all primer sequences are listed in Table 4-2.

The plasmids pMH004 and pMH029b were a generous gift from M.J. Helf.

To generate the N-terminal histidine-tagged fusion constructs as substrate that carries single point mutations of adjusted PTH\_2340 (hereafter called "proteusin"), a 103 bp polymerase chain assembly (PCA) product was generated (1.5 u Phusion polymerase, 4  $\mu$ M forward and reverse primer, 200  $\mu$ M dNTPs, corresponding buffer) using respective forward primer and Ala\_Screen\_reverse (3 min 98 °C, 10 sec 98 °C, 30 sec 43 °C, 10 sec 72 °C, 40 iterations of step 2 to 4, 2 min 72 °C). These PCA-products were digested with *AgeI* and *KpnI* and ligated into the corresponding sites of pMH029b to obtain the single point mutants.

To generate the *baeB* N-terminal strep-tagged fusion construct with a TEV cleavage site, pKJ7 was digested with *BamHI* and *XhoI* and inserted in the corresponding restriction sited of pET28a-ST-H (gift from Dr. Anna Vagstad) to yield pKJ49. To generate the *baeB* N-terminal strep-tagged fusion construct with a PoyH cleavage site, a PCA was conducted with primers Exch.TEV-PoyH-for and Exch.TEV-PoyH-rev. This PCA product was digested with *NdeI* and *NcoI* and inserted in the corresponding sites of pET28a-ST-H. pKJ7 (Jensen et al. 2012) was digested with *BamHI* and *XhoI* to cut out *baeB* and *baeB* inserted in the corresponding restriction sited of the intermediate plasmid to yield pKJ50.

| Construct   | Plasmid   | Antibiotic                   | Inducer                          | Expression host        |
|---|---|------------------------------|----------------------------------|------------------------|
| PoyH<br>(N-terminal His <sub>6</sub> -tag)                  | pMH004  | Kanamycin<br>(50 $\mu$ g/mL) | IPTG (1 mM)<br>0.1 % L-Arabinose | <i>E. coli</i> BL21A1  |
| Proteusin (N-terminal<br>His <sub>6</sub> -tag) and mutants | pMH029b<br>pKJ47 (L99A)<br>pKJ37 (D100A)<br>pKJ38 (Q101A)<br>pKJ48 (A102V)<br>pKJ44 (A103V)<br>pKJ40 (G104A)<br>pKJ45 (G105A)<br>pKJ41 (T106A)<br>pKJ46 (G107A) | Kanamycin<br>(50 $\mu$ g/mL) | IPTG<br>(1 mM)                   | <i>E. coli</i> BL21DE3 |

|   |       |                         |                |                        |
|---|-------|-------------------------|----------------|------------------------|
| BaeB with TEV cleavage site (N-terminal Strep-tag)  | pKJ49 | Kanamycin<br>(50 µg/mL) | IPTG<br>(1 mM) | <i>E. coli</i> BL21DE3 |
| BaeB with PoyH cleavage site (N-terminal Strep-tag) | pKJ50 | Kanamycin<br>(50 µg/mL) | IPTG<br>(1 mM) | <i>E. coli</i> BL21DE3 |

**Table 4-1: Constructs used in this study.**

For N-terminal hexahistidine-tagged proteusins mutants the mutated amino acids are indicated in brackets behind the plasmid name.

| Primer name              | Sequence (5'-3')  |
|--------------------------|---|
| Ala-screen L99A forward  | AG <b>ACCGGT</b> GAGCAG <u>GCGG</u> ATCAGGCAGCGGGCGGGACCGGCTGTTCCGATGTAT<br>ATTCCTTT    |
| Ala-screen D100A forward | AG <b>ACCGGT</b> GAGCAGTT <u>GCGC</u> AGGCAGCGGGCGGGACCGGCTGTTCCGATGTAT<br>ATTCCTTT     |
| Ala-screen Q101A forward | AG <b>ACCGGT</b> GAGCAGTTGGAT <u>GCGG</u> CAGCGGGCGGGACCGGCTGTTCCGATGTAT<br>ATTCCTTT    |
| Ala-screen A102V forward | AG <b>ACCGGT</b> GAGCAGTTGGATCAG <u>GTG</u> CGGGCGGGACCGGCTGTTCCGATGTAT<br>ATTCCTTT     |
| Ala-screen A103V forward | AG <b>ACCGGT</b> GAGCAGTTGGATCAGGCAG <u>TGG</u> CGGGCGGGACCGGCTGTTCCGATGTAT<br>ATTCCTTT |
| Ala-screen G104A forward | AG <b>ACCGGT</b> GAGCAGTTGGATCAGGCAG <u>GCGG</u> CGGGACCGGCTGTTCCGATGTAT<br>ATTCCTTT    |
| Ala-screen G105A forward | AG <b>ACCGGT</b> GAGCAGTTGGATCAGGCAGCGGG <u>C</u> CGGACCGGCTGTTCCGATGTAT<br>ATTCCTTT    |
| Ala-screen T106A forward | AG <b>ACCGGT</b> GAGCAGTTGGATCAGGCAGCGGGCGGG <u>GCGG</u> GCTGTTCCGATGTAT<br>ATTCCTTT    |
| Ala-screen G107A forward | AG <b>ACCGGT</b> GAGCAGTTGGATCAGGCAGCGGGCGGGAC <u>CGC</u> TGTTCCGATGTAT<br>ATTCCTTT     |
| Ala-screen reverse       | GAG <b>GGTACCG</b> TGTTGTCATGGTAGGTGGGAACGCAGATAGGAAAGGAATATACA<br>TCGG                 |
| Exch.TEV-PoyH-for        | TAG <b>CATATG</b> GCCGGTCCCGCCCGCTGCCTGATCCAAGCCGCTGCTCTTTTCGA                          |
| Exch.TEV-PoyH-rev        | TAG <b>CCATGG</b> GCCAGCAGCTGGAGCCACCCGAGTTCGAAAAGAGCAGC                                |

**Table 4-2: Primer used in this study.**

The mutated (compared to the pMH29b sequence) DNA bases are underlined, restriction sites were marked **bold**.

### 4.5.3 Expression and purification of proteins

For overexpression of all proteins, 100 mL cultures were grown to an OD<sub>600</sub> of 1.5-2 in TB medium (37 °C, 250 rpm), induced with isopropyl β-D-1-thiogalactopyranoside (IPTG), and grown for approximately 20 hrs at 16 °C. Cells of expression strains were harvested by centrifugation (7500 x g, 10 min) at 4 °C.

Cells expressing proteins with His<sub>6</sub>-Tag were resuspended in 5 mL lysis buffer (50 mM sodium phosphate (NaPi), 0.3 M NaCl, 10 mM imidazole and 0.1% (v/v) glycerol) per gram of cells. The cell suspension was kept on ice and sonicated in 10 s intervals. The cell debris was removed by centrifugation at 14,000 x g for 35 min at 4 °C. The supernatant was incubated with 1.0 mL Ni-NTA agarose (Qiagen) for at least 1 hr rotating at 4 °C and washed with 10 mL wash buffer 1 (50 mM NaPi, 0.3 M NaCl, 20 mM imidazole and 0.1% (v/v) glycerol, pH 8) and 5 mL wash buffer 2 (50 mM NaPi, 0.3 M NaCl, 40 mM imidazole and 0.1% (v/v) glycerol, pH 8). His-tagged proteins were eluted two times with 1.5 mL elution buffer (50 mM NaPi, 0.3 M NaCl, 250 mM imidazole and 0.1% (v/v) glycerol, pH 8). The eluent was dialyzed against assay buffer containing 20 mM ammonium acetate, pH 4.5. The concentration of protein was determined by Bradford assay and fractions were analyzed by SDS-PAGE. Proteins were flash frozen in liquid nitrogen and stored at -80 °C.

Cells containing expression constructs with Strep-Tags were resuspended in Buffer W (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA). After sonication and centrifugation as described above, the supernatant was applied to a column with one column volume (CV) Strep-Tactin-Sepharose (IBA Biotagnology). The column was washed twice with five CV Buffer W and eluted with five fractions of 0.5 CV Buffer E each (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin). Columns were used up to five times and regenerated according to manufacturer's specification. Proteins were flash frozen in liquid nitrogen and stored at -80 °C.

#### 4.5.4 Cleavage Assays

If needed proteins were concentrated with Vivaspin columns (Sartorius Stedim, MWCO 10.000 Da). After thawing PoyH was centrifuged (17,000 x g, 4 °C, 30 min), the supernatant transferred in a new tube and the protein concentration measured with Bradford assay from this portion.

PoyH was diluted to 250 nM in acetate buffer (20 mM ammonium acetate, pH 4.5). For the cleavage assay 20 µM substrate and 50 nM protease were mixed in acetate buffer and incubated at 22 °C. Reaction was stopped by removing aliquots of final 3 µg protein each and quenched with 30 % 1 M ammonium acetate. Samples without protease served as negative control. The samples were analyzed on SDS-PAGE.

For acidic activation PoyH was diluted to 500 nM in 20 mM ammonium acetate buffer, pH 3.0 and subsequently incubated for at least one hour at 4 °C. Then it was applied in the cleavage assay.

In assays for cleavage of BaeB with either TEV or PoyH recognition site the proteins were incubated with different proteases in the Strep-tag elution buffer immediately after protein purification. One assay contained 3.6 µM BaeB with TEV recognition site or 5.1 µM BaeB with PoyH recognition site together with 5.128 µM TEV protease or 0.115 µM PoyH or elution buffer in case of the negative control. The samples were incubated overnight at 22 °C, quenched with SDS loading buffer and immediately analyzed on SDS-PAGE.

#### 4.5.5 Mass spectrometry procedure

HPLC-HRMS experiments we performed on a Dionex Ultimate 3000 UHPLC coupled to a Thermo Scientific Q Exactive mass spectrometer. For a size determination of expressed proteins a reduction with 5 mM tris(2-carboxy-ethyl)phosphine (TCEP) for 30 min at RT occurred. Afterwards 5 µL sample were analyzed by reversed phase HPLC-HRMS. Program 1: column: Phenomenex, Aeris Widepore, C4, 3.6 µ, 50 x 2.1 mm; column temperature: 50 °C; flow rate 0.5 mL/min; mobile phase/gradient: 5% CH<sub>3</sub>CN + 0.1% formic acid/H<sub>2</sub>O + 0.1% formic acid for 5 min then ramped to 90% CH<sub>3</sub>CN + 0.1% formic acid/H<sub>2</sub>O + 0.1% formic acid over 9 min, kept at 90% CH<sub>3</sub>CN + 0.1% formic acid/H<sub>2</sub>O + 0.1% formic acid for 4 min and ramped back to 5% CH<sub>3</sub>CN + 0.1% formic acid/H<sub>2</sub>O + 0.1% formic acid again; detection: ESI-MS, positive ion mode. Program 2: column: Phenomenex, Aeris Widepore, C4, 3.6 µ, 50 x 2.1

mm; column temperature: 50 °C; flow rate 0.5 mL/min; mobile phase/gradient: 5% CH<sub>3</sub>CN + 0.1% formic acid/H<sub>2</sub>O + 0.1% formic acid for 5 min then ramped to 95% CH<sub>3</sub>CN + 0.1% formic acid/H<sub>2</sub>O + 0.1% formic acid over 9 min, kept at 95% CH<sub>3</sub>CN + 0.1% formic acid/H<sub>2</sub>O + 0.1% formic acid for 4 min and ramped back to 5% CH<sub>3</sub>CN + 0.1% formic acid/H<sub>2</sub>O + 0.1% formic acid again ; detection: ESI-MS, positive ion mode. Program 3: same as program 2, only difference is the used column (Phenomenex, Kinetex, C18, 2.6  $\mu$ , 150 x 4.6 mm).

## 4.6 Acknowledgement

We thank Dr. Anna Vagstad for her generous gift of pET28a-ST-H and Maximilian J. Helf for pMH004 and pMH029b as gift. We are grateful to Dr. Michael F. Freeman for fruitful discussions and remarks.

## 4.7 Supplemental information

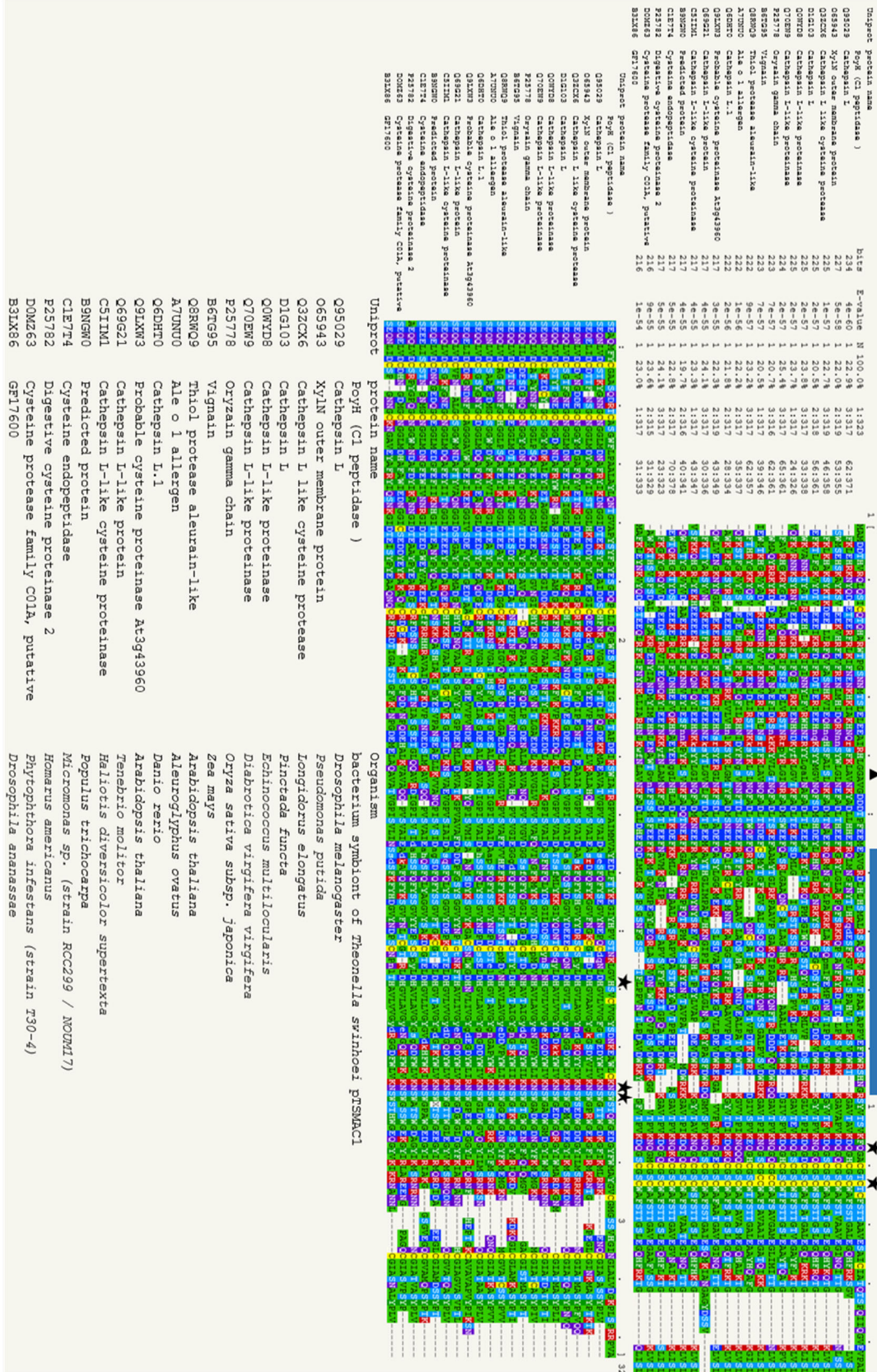


Figure 4-7: PSI-BLAST Pseudo-Multiple Sequence Alignment taken from Phyre-calculations.

Stars indicate active site residues, the bar represents the potential cutting region where propeptide gets released and triangle indicate the probable cutting site after incubation in acetate buffer, pH 3.0.



## Chapter 5. Conclusion and Further Perspectives

This thesis focused on unusual enzymatic reactions in bacterial polyketide and ribosomal peptide biosynthesis. These studies shed light on basic biosynthetic principles applied by the producer and helped gain further insight into these young classes of secondary metabolite-synthesizing machineries. In general, discovery of new pathways within the RiPP and *trans*-AT PKS families unveils an increasing number of peculiarities. A basic understanding of these unusual biosyntheses is of particular interest to the pharmaceutical industry as many medically relevant molecules are products of their biosynthesis (such as the antibacterial *trans*-AT PKS product mupirocin, trade name Bactroban® and the analgesic Prialt®, a RiPP conopeptide). Insight toward industrial application of these types of pharmaceutically relevant molecules would be desirable, for instance, a targeted modification of compounds to improve their bioactivity or an increase of production yield in a heterologous expression. Heterologous expression of pathways is still very challenging, but an important prerequisite for their study as many *trans*-AT PKS and RiPP pathways are encoded by as-of yet uncultivated bacterial symbionts associated with invertebrates. This holds most likely true for all pathways studied here. Polytheonamides (**14**) (Figure 5-1) are produced by "Entotheonella" spp. associated with the sponge *Theonella swinhoei*. The *trans*-AT PKS product pederin (**5**) is produced by a symbiont of the *Paederus* beetle, while the enantiomer of the *trans*-AT PKS product oocydin A (**11**) (Figure 5-3) was isolated from a marine sponge and an ascidian, where it is most likely produced by bacteria. Full clarification of their biosynthesis contributes to a better understanding of how heterologous expression systems might be developed. Studies of biosynthesis will also be facilitated by the expansion of the set of tools for example, acyl hydrolases (discussed in Chapter 2) for release of PKS intermediates to illuminate PKS-catalyzed biosynthetic steps and PoyH (discussed in Chapter 4) for release of proteusin cores. Many questions were answered by the work performed in this thesis, but new questions resulted from it and will be discussed in the following sections.

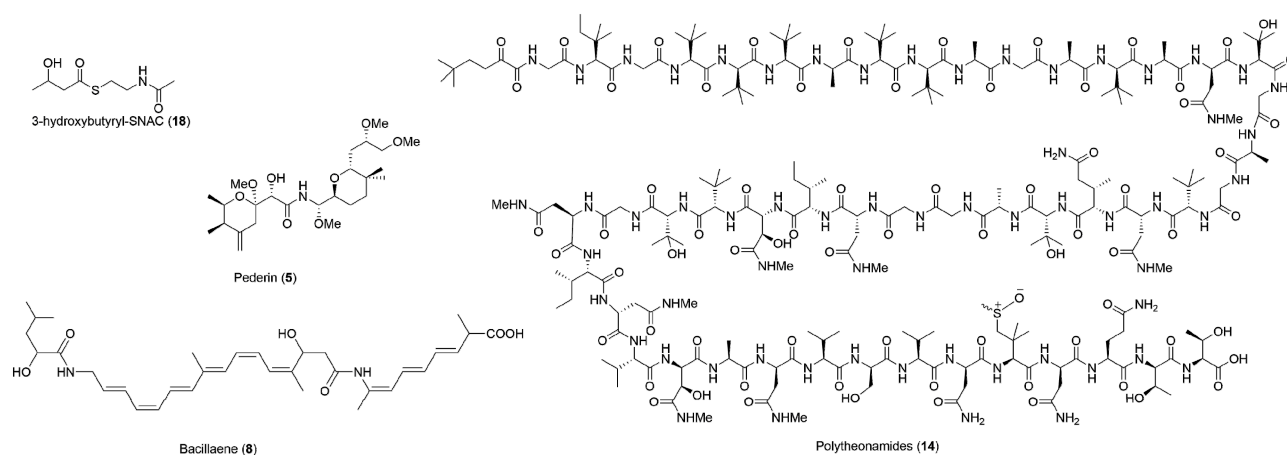


Figure 5-1: Structures of 5, 8, 14 and 18.

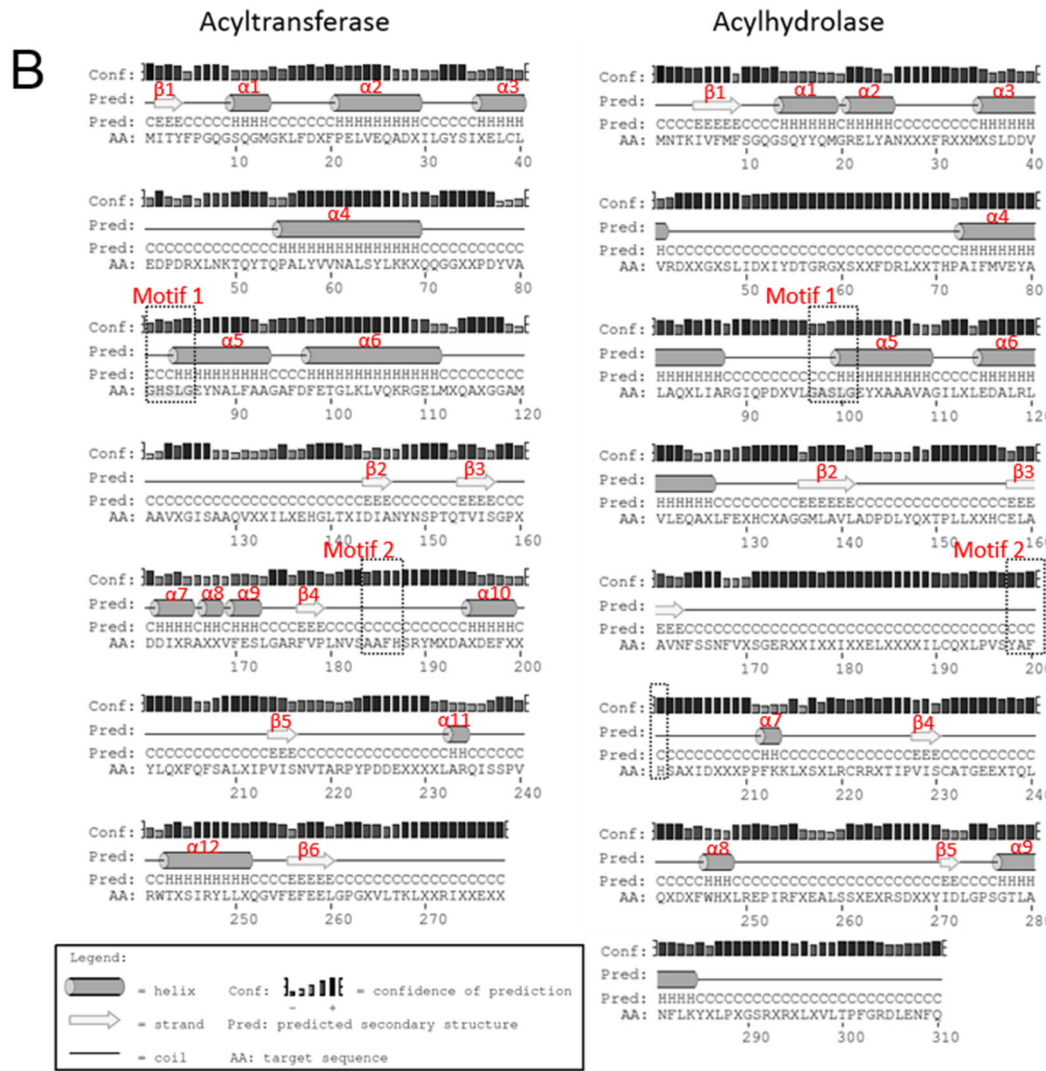
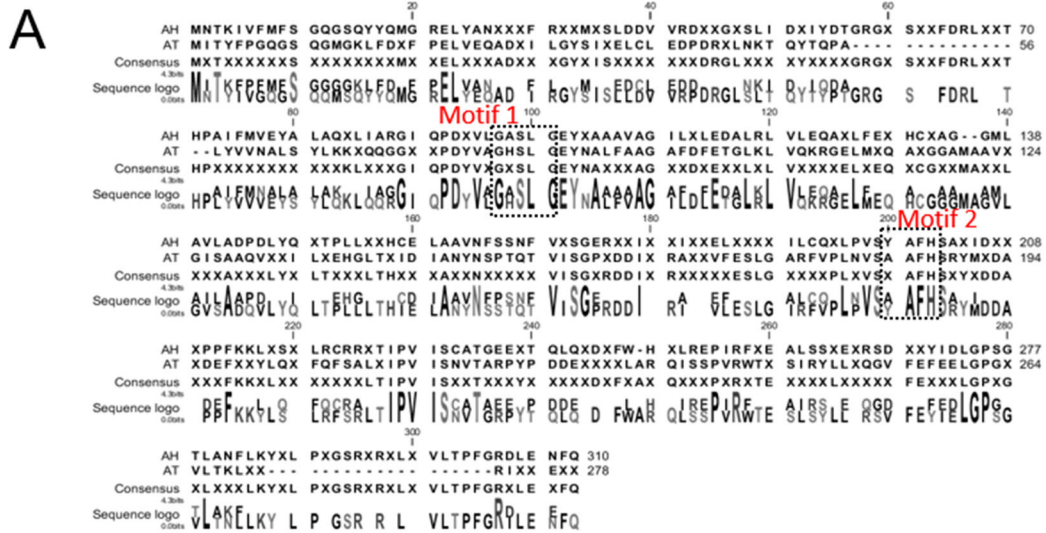
The first example of a basic underlying principle that was studied in this thesis dealt with hydrolysis of intermediates in *trans*-AT PKS pathways. We could show that PedC is able to hydrolyze *N*-acetylcysteamine bound acyl units as substrates with a variety of functionalities (fully reduced, branched,  $\beta$ -hydroxylated and  $\alpha,\beta$ -unsaturated moieties) and chain lengths (between two and six carbon atoms). PedC accepted beside SNAC substrates even ACP-bound acyl moieties from a fungal type I PKS. Malonyl-CoA as the extender unit in the biosynthesis of pederin (**5**) was not cleaved by PedC. All these observations together lead to the renaming of the PedC-like portion of ATs to acylhydrolases (AH) and suggest the involvement of PedC in an error-correcting mechanism. This editing mechanism might be employed by other *trans*-AT PKS systems as well since AH homologs are commonly found. The characterization of AHs could explain how *trans*-AT PKSs perform housekeeping during biosynthesis since no such

proofreading mechanism was previously known for *trans*-AT PKSs. The removal of stalled intermediates for error correction is important for efficient biosynthesis. Many KSs in *trans*-AT PKSs have a very narrow substrate specificity (Nguyen et al. 2008), (Jenner et al. 2013), (Kohlhaas et al. 2013). If an erroneous acyl-moiety is introduced in the growing chain or if modifications occur at incorrect positions during biosynthesis, the whole assembly line would be blocked, which would present a disadvantage for the producing cell. A proofreading mechanism is already known for *cis*-AT PKS. In this instance, type II thioesterases (TEIs) are responsible for clearing the mega-enzyme complexes. Some general concepts regarding TEI function might be transferable to AHs; however, these proofreading systems have one big difference: KSs in *cis*-AT systems are more promiscuous in their substrate acceptance than *trans*-AT KSs. Incorporation of a wrong building block in a *cis*-AT PKS is likely to be less problematic as long as functionalities for a decarboxylative condensation are present. These functionalities can disappear due to an aberrant decarboxylation of an acyl building block bound to an ACP. This can happen if the KS decarboxylates the acyl unit without condensation mainly resulting in acetyl-, propionyl- or butyryl-ACPs blocking the assembly line (Heathcote and Leadlay 2001). For this reason, typically reduced short-chain acyl-building blocks are TEI substrates. PedC, in contrast, was demonstrated to accept a large variety of different chain-length and functionalities underlining its wide substrate acceptance. This indicates that PedC does not solely act at single building block-level but on longer intermediates as well.

Several scenarios are conceivable of how AHs perform their reaction *in vivo*. They could either work side by side with ATs or independently. The first scenario seems to be more reasonable as AHs and ATs are often fused together. If one of the named scenarios holds true or if a different mechanism is present has to be further investigated. The mode of action of AHs might be comparable to the two models postulated for TEIs (Kotowska and Pawlik 2014). The high specificity model suggests that stalled intermediates are recognized with high specificities by TEI and hydrolyzed off. In contrast, the low specificity model predicts that all intermediates get cleaved off, irrespective if they are wrong or correct. A comparison of kinetic data of PedC cleaving 3-hydroxybutyryl-SNAC (**18**) (Figure 5-1) with data from TEI towards NAC-substrates showed that PedC has higher  $k_{cat}/K_m$  values and is more efficient in hydrolysis than many of the different tested PKS and NRPS TEI enzymes (Heathcote and Leadlay 2001), (Liu et al. 2008), (Yeh et al. 2004). The  $K_m$  value of PedC is also higher in comparison to values determined for PKS and NRPS TEI enzymes (Kotowska and Pawlik 2014) pointing out a lower affinity of PedC for its substrate **18**. Kinetic data suggests a low-specificity hypothesis for PedC as values show a rather slow reactivity. More experiments are necessary to elucidate if this model holds true or if PedC is slow but nevertheless specific as we were only able to obtain kinetic data for one substrate.

It would be interesting to know which structural features of ATs and AHs are responsible for the difference in their reactivity. To answer this question several alignments were conducted. In a first step, different ATs and in parallel several AHs (Figure 7-1) were aligned. Subsequently both resulting consensus sequences were aligned to each other (Figure 5-2A) and a secondary structure prediction with the PSIPRED server was completed (Buchan et al. 2013). Results can be seen in Figure 5-2B. The first striking difference detectable in Figure 5-2A is the conserved GHSxG motif (motif 1, Figure 5-2); in ATs it is invariably GHSLG and in AHs it is a GxSxG sequence. In the case of the two AHs OocV AT2 and BryP AT2 a histidine is present in the conserved GHSxG motif as can be observed in the overall alignments (Figure 7-1). Surprisingly the other five aligned AH sequences (RhiG AT1, KirCI AT1, BatH, BaeD, PedC) do show a GxSxG sequence. Recently it was shown that a His to Ala mutation at this position in the yersiniabactin AT (a mixed NRPS/*cis*-AT PKS hybrid) results in a 7-fold increased hydrolysis rate over the wild type AT (Poust et al. 2014). This was explained by a stabilizing function of the His residue within the protein, either due to a hydrogen bond with the conserved Asn (N165 in Figure 5-2A) or through a water-mediated hydrogen bond to the carbonyl oxygen of the AT-acyl ester. In the case of AH this proposed stabilization seems to be missing. A comparison with the AT and AH structure prediction conducted here reveals some more differences in comparison to the only published crystal structure of a *trans*-AT PKS AT of the disorazole PKS (DSZS AT) (Wong and Khosla 2011), especially in the small ferredoxin-like subdomain (in the AT consensus sequence position at ~127-183, in the AH consensus

sequence position at ~148-209, Figure 7-1). In the DSZS AT, three  $\beta$ -strands and three  $\alpha$ -helices can be found in this domain ( $\alpha$ 7-9 and  $\beta$ 2-4 in Figure 5-2B). Neither the three  $\alpha$ -helices nor the  $\beta$ -strand are present in the AHs (predicted with a high confidence-level). The function of this small domain is not yet clear; however, it is proposed to play a role in ACP-interaction (Wong and Khosla 2011). The Q239 in DSZS AT is suggested to stabilize the catalytic His from the second catalytic motif xAFH (motif 2, Figure 5-2). This Q is not present in AH sequences contrary to AT sequences. In addition, DSZS AT R111 is also only present in ATs but not in AHs (R/Q124 in Figure 5-2A). It is exchanged exclusively to Q in AHs. This should not have a large impact on the reaction as the  $\text{NH}_2$ -group delivered by Q could stabilize the substrate's carbonyl group as well. If the AH protein lacks stability, as suggested by the obtained sequence comparison, more water molecules would have access to the active site resulting in higher hydrolysis rates. A different composition of the hydrophobic acyl-binding cleft would also increase hydrolysis. Crystal structures of both AH and AT representatives, ideally cocrystallized with substrates are needed to elucidate which protein sequences, and consequently, which structural differences are responsible for hydrolysis and acyl transfer. It is highly likely that these small sequence changes promote the reverse hydrolytic reaction in comparison to the acyl transferring reaction.



**Figure 5-2: Alignment of acyltransferases (ATs) with acylhydrolases (AHs) consensus sequence and structure prediction.**

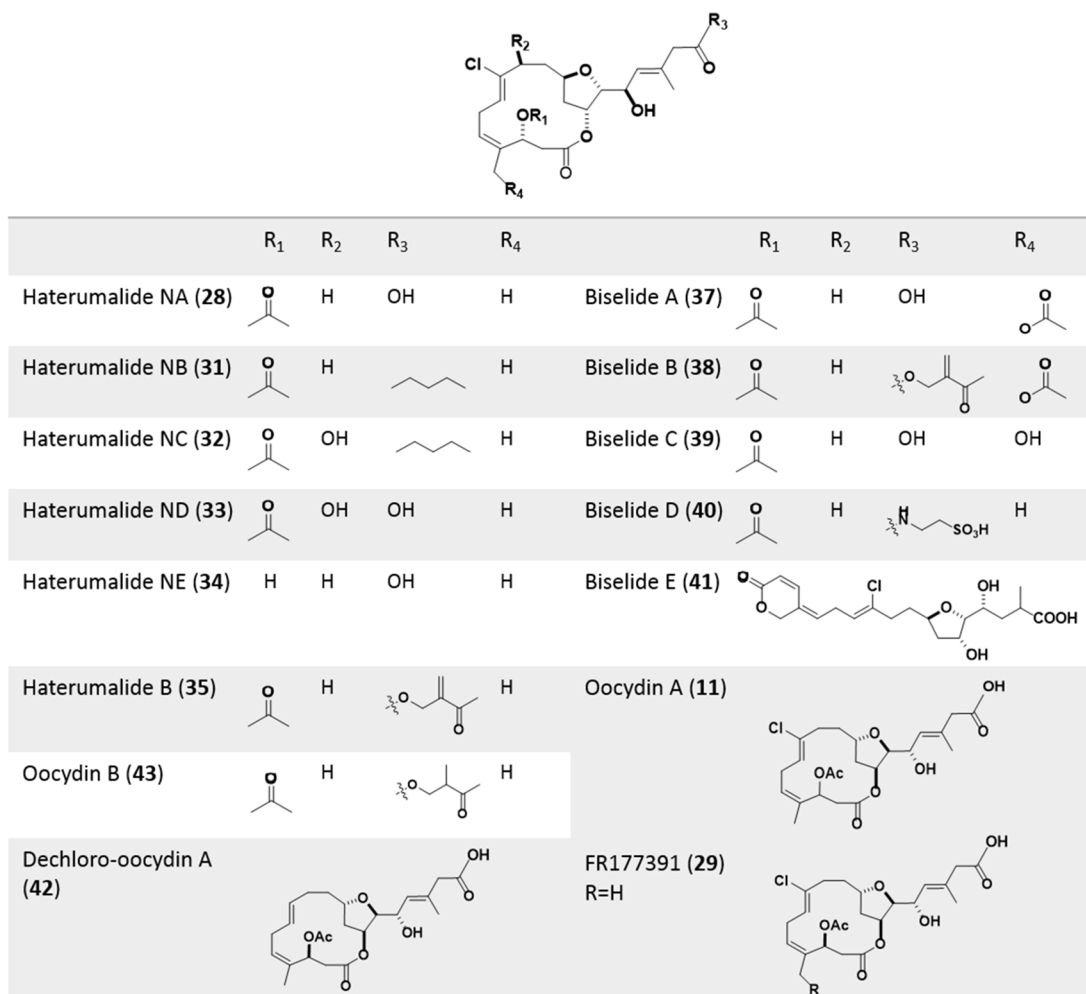
(A) Sequence alignment of AT- and AH-consensus sequences. The underlying full alignments can be seen in Figure 7-1. (B) Structure prediction obtained from the PSIPRED server of AT- and AH-consensus sequence. The numbers of  $\alpha$ -helices and  $\beta$ -strands are indicated. The active site motif 1 GxSxG and motif 2 XAFH are boxed in A and B.

Crystal structures could also answer the question of how the comparably large substrate acceptance of AHs is generated. ATs exhibit tight control of elongation moiety selection and have typically narrow substrate specificity (Hedstrom 2002). Using sequence alignments, it has been demonstrated that one important motif is positioned about 30 residues upstream of the catalytic serine. The ZTx $\phi$ (A/T)(Q/E) (Z-hydrophilic residue,  $\phi$ -aromatic residue) motif points towards the use of malonyl-CoA (Yadav, Gokhale, and Mohanty 2003). This is valid for the AT alignment conducted here, which exhibits the highly conserved xTQYTQ motif (position 58-63; Figure 7-1A). At a comparable position in the AH alignment (position 74-79; Figure 7-1B), a region of very low homology to each other is present. This could explain the wide substrate acceptance. Further information can be gained from structural comparison of their active sites (Dunn and Khosla 2013). The direct downstream neighbor of the catalytic serine in motif 1 (GHS $\phi$ G motif; Figure 5-2) plays an important role. While ATs that accept malonyl have a branched hydrophobic amino acid at the "X-position", other specificities are indicated by less bulky residues (Smith and Tsai 2007). In both AT and AH alignments an aliphatic leucine is present in this position. For ATs this is rather unusual as they almost exclusively accept malonyl-CoA in *trans*-AT PKSs. However, in AHs this allows a wide variety of possible substrates. In BatH, an AH, even the smaller alanine is present, fully reducing a possible sterical hindrance.

An AH gene can be found in more than half of all *trans*-AT PKS gene clusters characterized to date, underlining its importance. Accordingly, the question arises how the other *trans*-AT PKSs are able to perform proofreading functions and whether proofreading is accomplished in an AH-independent manner. The same peculiarity was detected for TEIIs. Not every *cis*-AT PKS gene cluster encodes for a TEII and it is not currently known why. One hypothesis states that different metabolic composition within cells lead to contrasting susceptibility to biosynthetic errors (Kotowska and Pawlik 2014). This proposal needs to be tested experimentally for AHs. An introduction of an AH copy in a strain that does not have AHs might give first hints towards an explanation.

In addition to comprehending the AH function of PedC in pederin (**5**) biosynthesis, studies of the biosynthesis of oocydin A (**11**), in particular its halogenation, is also better understood. An unprecedented two-component halogenase with similarity to MNH enzymes was detected. This two-component halogenase consists of OocP, a mononuclear non-heme (MNH) iron,  $\alpha$ -KG and O<sub>2</sub>-dependent halogenase homolog, and OocQ, a protein with no homology to enzymes with known function.

In the case of halogenation of oocydin A (**11**) the substrate is not clear yet. It might be **42** (Figure 5-3) with its double bond between C<sub>7</sub>-C<sub>8</sub> or a species with a single bond between C<sub>7</sub>-C<sub>8</sub>. A canonical MNH halogenase attacks a saturated carbon-carbon bond yielding a chloroalkane (Figure 1-17). Vinyl chloride moieties generated after halogenation are described for jamaicamide (Gu et al. 2009) and hypothesized for hectochlorine (Pratter et al. 2014a). We were able to isolate dechloro-oocydin A (**42**) with the present C<sub>7</sub>-C<sub>8</sub> double bond. If **42** is the substrate the reaction would generate a vinyl radical which is highly unstable. In case **42** is not the substrate but a species with a C<sub>7</sub>-C<sub>8</sub> single bond a subsequent desaturation has to occur to yield **11**.



**Figure 5-3: Members of the haterumalide family discussed in this chapter.**

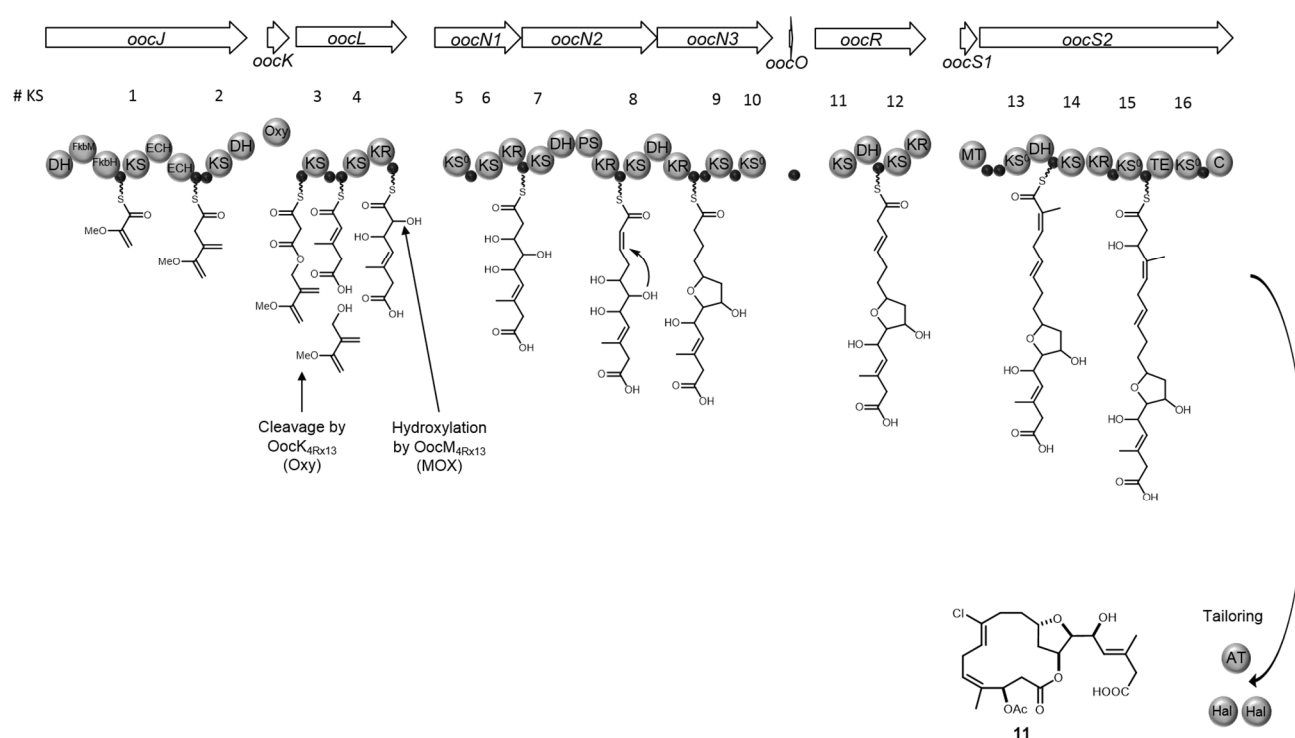
Preliminary stereochemical assignment after Kigoshi and Hayakawa 2007.

We could show that the presence of  $OocQ_{4Rx13}$  is necessary for halogenation but we were not able to clarify its function during the course of this thesis. The protein does not show homology to any studied protein and with that no function could be deduced. It might be feasible that **42** gets halogenated and  $OocQ_{4Rx13}$  might help to stabilize a radical intermediate during halogenation. If a species with a C<sub>7</sub>-C<sub>8</sub> single bond serves as substrate  $OocQ_{4Rx13}$  could theoretically also be involved in a hypothetical desaturation reaction generating the C<sub>7</sub>-C<sub>8</sub> double bond. This desaturating function of  $OocQ_{4Rx13}$  is rather unlikely because no saturated species have been detected in the knock out strain  $4Rx13\Delta oocQ$ . Generally, the differences between both halogenase knock out extracts are minute. No hints regarding any probable hydroxylated, epoxidated or saturated stable intermediates appear which would suggest a different, unknown halogenation reaction mechanism. In contrast to  $OocQ_{4Rx13}$  where no homology to known enzymes could be detected and its exact function is speculative,  $OocP_{4Rx13}$  exhibits a low homology to MNH enzymes. It would be interesting to see which cofactor is used by  $OocP_{4Rx13}$  as first trials to inhibit  $\alpha$ -KG as potential cofactor failed. Not every MNH enzyme relies on  $\alpha$ -KG (other possibilities are pterin, ascorbate, Rieske cluster or even no cofactor at all (Buongiorno and Straganz 2013)), but to date, all described MNH halogenases employ  $\alpha$ -KG. The determination of the cofactor used by  $OocP_{4Rx13}$  needs to be validated experimentally, but it might be possible to exclude some cofactor candidates by *in silico* analysis by comparing  $OocP_{4Rx13}$  with sequences and structures of already crystallized MNH enzymes. It was published that depending on the used cofactor, MNH enzymes exhibit a different structure and sequence motif (Koehntop, Emerson, and Que 2005) (Figure 7-2). A comparison of  $OocP_{4Rx13}$

with these motifs might exclude potential cofactors. First cofactor candidates can be filtered out by a structure motif comparison. We do not have a crystal structure for OocP<sub>4Rx13</sub> but as described above a conserved domain of OocP<sub>4Rx13</sub> detected by BLAST analysis is a cupin-like domain. The term cupin fold is a synonym for jelly roll fold (Aik et al. 2012). The jelly roll fold is present in half of the MNH enzymes according to Koentop *et al.*, namely in one representative of cofactor-less extradiol cleaving catechol dioxygenases, in a large portion of  $\alpha$ -KG dependent enzymes, in the cofactor-less isopenicillin N synthase (IPNS) and the ascorbate-using 1-aminocyclopropane-1-carboxylic acid oxidase (ACCO). Enzymes using Rieske cluster or pterin as cofactor do not show the jelly roll fold and can likely be excluded as potential cofactors used by OocP<sub>4Rx13</sub>. The smaller subset of candidates can be further reduced with a comparison of sequence motifs. The sequence motif of OocP<sub>4Rx13</sub> is, depending on the alignment (Table 3-7) either HxGx<sub>79</sub>H (alignment with the Espresso web server) or AxGx<sub>165</sub>H (alignment with the T-Coffee web server). Whereas the AxGx<sub>165</sub>H motif is less likely because the active site of MNH enzymes is reported to contain at least two histidine residues (Buongiorno and Straganz 2013). The same sequence motifs for the known  $\alpha$ -KG dependent MNH halogenases SyrB2 is HxAx<sub>117</sub>H and for CurA is HxGx<sub>111</sub>H. When comparing the HxDx<sub>n</sub>H sequence motifs to MNH enzymes it is only present in the jelly roll-containing portion of  $\alpha$ -KG dependent enzymes, IPNS and ACCO. A sequence alignment performed with the Espresso server using structural information (Armougom et al. 2006) was performed with the three candidates: Human Factor Inhibiting Hif (FIH-1, NCBI accession number 1MZE\_A) as an  $\alpha$ -KG dependent enzyme with the closest sequence motif match to OocP<sub>4Rx13</sub>; IPNS (NCBI accession number BAA11234.1), ACCO (NCBI accession number 1W9Y\_A) and the already aligned SyrB2 together with CurA as a control. The alignments resulted in comparable scores for all alignments (data not shown): OocP<sub>4Rx13</sub>/SyrB2/CurA: score 50; OocP<sub>4Rx13</sub>/IPNS: score 57; OocP<sub>4Rx13</sub>/Fih1: score 59; OocP<sub>4Rx13</sub>/ACCO: score 61. These alignment scores do not strongly suggest the usage of a certain cofactor. OocP<sub>4Rx13</sub> might employ either no cofactor,  $\alpha$ -KG, ascorbate or an as of yet unknown cofactor. The usage of pterin or a Rieske cluster is less likely in comparison to the previously mentioned ones after sequence analysis.

Further studies have to be conducted to clarify the exact time point of halogenation. The halogenation might occur post-translational or on the growing chain. In the latter case it unclear at which stage of biosynthesis of **11** the halogenation site is generated. The halogenation site is one of the few positions in a proposed **11** biosynthesis where the prediction of events is unclear. It is hypothesized that, in the case of OocA<sub>153</sub>, this substrate moiety is introduced within OocN<sub>A153</sub> by using the KS8-DH-KR-ACP module twice, skipping the downstream domains of OocN<sub>A153</sub> and the *trans* action of the proposed ER OocU<sub>A153</sub> (Matilla et al. 2012). The introduction of the substrate moiety is more likely to take place within OocN<sub>34Rx13</sub>/OocO<sub>4Rx13</sub>/OocR<sub>4Rx13</sub> (Figure 5-4). Five ACPs and one KS<sup>0</sup> can be found within the twelve domains of OocN<sub>34Rx13</sub>/OocO<sub>4Rx13</sub>/OocR<sub>4Rx13</sub>. This accumulation of ACPs indicates a stretch where many *trans*-acting enzymes perform their reactions. Something similar is known from  $\beta$ -branching, where a tandem ACP is suggested to prolong possible reaction time (Rahman et al. 2005), (Gu et al. 2011). In the case of curacin biosynthesis a sequence of four consecutive ACPs is proposed to provide enough time for halogenation and ring formation (Gu et al. 2009), (Gu et al. 2011). A closer look at the *in silico* hypothesized KS substrate specificity (E. J. N. Helfrich, unpublished results) of KS11 and KS12 in OocR<sub>4Rx13</sub> indicates a hydroxyl group at the KS11-substrate and a reduced moiety at the KS12-substrate. This specificity implies that reactions take place between KS10 to KS12 modifying the intermediate in a way that it gets accepted by the next KS. This substrate specificity suggests the presence of a double bond between C<sub>7</sub> and C<sub>8</sub> installed at the domains upstream of KS12. Consequently, this would also be the earliest time point at which a chlorination could appear and this would most likely be the moiety recognized by the halogenase.

Our data suggest a post-translational halogenation on the free substrate as we could detect traces of the unchlorinated compound in the wild type as well. This would be a rare feature as the only other MNH halogenase able to halogenate without the aid of a carrier protein was just recently described in an alkaloid (Hillwig and Liu 2014).



**Figure 5-4:** Section of oocydin A biosynthetic gene cluster of *Serratia plymuthica* 4Rx13 with corresponding assembly line parts.

This figure is a combination of Figure 3-1 and Figure 3-6. KS domains are numbered. Abbreviations: little black circles, acyl carrier proteins (ACP); DH, dehydratase; Fkb, Fkb-like enzymes; KS, ketosynthase; CR, crotonase; Oxy, oxygenase; KR, ketoreductases; KS<sup>0</sup>, non-elongating KS; PS, pyran synthase; MT, methyltransferases; TE, thioesterase; C, condensation-domain; AT, acyltransferase; CoA, coenzyme A 6; HCS, 3-hydroxy-3-methylglutaryl-CoA synthase; MOX, monooxygenase; Hal, halogenase; ER, enoylreductase.

The final assignment of absolute stereochemistry of oocydin A (**11**), haterumalide NA (**28**) and FR177391 (**29**) (Figure 5-3) still has to occur. All three compounds are described to have the same relative stereochemistry, meaning at least two (if not all three) have to be identical (Kigoshi and Hayakawa 2007), (Sato et al. 2005), (Roulland 2008). In addition, we were only able to assign relative stereochemistry with the help of NOESY correlations due to limited availability of **11**, **42** or **43**. An assignment of absolute stereochemistry of our isolated products is needed for a proper determination of *oocC<sub>4Rx13</sub>* products. It might be possible that products produced by *oocC<sub>4Rx13</sub>* are enantiomers to other described haterumalides. Although absolute stereochemistry is not fully proven, a correlation between extraction site and rotation was mentioned by Kigoshi and Hayakawa as the marine products **28** and **35** were levorotatory and bacterial products **11** and **29** were dextrorotatory (Kigoshi and Hayakawa 2007). However, it is very likely that in the case of isolation from a marine sponge (**28**) and an ascidian (**35**), the real producer of the compounds are bacteria as well. It has long been hypothesized that symbiotic bacteria are the true producer of many secondary metabolites isolated from marine sponges (Jensen and Fenical 1994). This was lately proven for the marine sponge *Theonella swinhoei*. Many products of the large and distinct secondary metabolite repertoire found in the sponge are actually produced by the single bacterial phylotype "*Entotheonella*" sp. (Wilson et al. 2014).

Until now members of the haterumalide family were often extracted together from one source. Isolation of **28**, **37-41** were reported from an ascidian, **28**, **31-34** were isolated together from a marine sponge and **28**, **34-36** were reported to be isolated together from the soil bacterium *S. plymuthica* A153 (Teruya et al. 2005), (Takada et al. 1999), (Levenfors et al. 2004). It was surprising that solely production of **11** was connected to the recently studied gene cluster *oocA<sub>153</sub>* although experiments were conducted with the same bacterial producer *S. plymuthica* A153 (Matilla et al. 2012), especially because of the reported common isolation of several compounds from this strain. We could show for the first time in the course of this thesis that the *oocC<sub>4Rx13</sub>* gene cluster is also responsible for the



biosynthesis of **43** and with that generating products with a short aliphatic side chain (**11**) and extended side chain (**43**). A further clarification of how the *ooc*<sub>4Rx13</sub> gene cluster is able to synthesize such a variety of compounds would be interesting. It furthermore remains to see which more members of the haterumalide family will be discovered and how these structures can be synthesized by a PKS system. It would also be intriguing to see if all compounds of the haterumalide family isolated from such diverse habitats and countries originate from *ooc*-like gene clusters or if the ability of the producer to synthesize haterumalide compounds evolved via convergent evolution. An example of one product synthesized by two different biosynthetic pathways is isopentenyl pyrophosphate synthesized either by the mevalonate or the nonmevalonate pathway (Fischbach, Walsh and Clardy 2007). The occurrence of highly structurally related compounds in diverse habitats implies an important function of the products for the producing cell as they would all produce the same set of metabolites in a range of diverse habitats.

Besides focusing on unusual enzymatic reactions in trans-AT PKS pathways also enzymatic reaction in ribosomal peptide biosynthesis was investigated. Very little is known about RiPP proteases or prokaryotic cysteine proteases in general. Our experiments, together with the previous ones from M. J. Helf were one of the first to shed light on a RiPP protease. It is surprising that not more attention was drawn onto this basic mechanism in RiPP biosynthesis as their action releases the fully bioactive compound and is almost always responsible for the transfer of the active compound out of the cell, thus enabling its translocation to the site of action. The only other published protease is McjB, belonging to the microcidin J25 biosynthetic pathway, which is ATP-dependent and reacts exclusively in presence of McjC (Yan et al. 2012). Both prerequisites are not necessary for PoyH-action. A variety of other proteases are characterized nowadays but to the best of my knowledge only one prokaryotic cysteine peptidase besides McjB has been studied up to now, namely xylellain from the Gram-negative bacterium *Xylella fastidiosa* (Nogaroto et al. 2006).

The protease PoyH seems to have a large potential for a wide range of applications. It showed reactivity at a comparably wide range of conditions (4 °C - 40 °C; pH 3.0 - pH 8.0) and accepted substrates with different up- and downstream sequences (and with that also protein structures) of the recognition motif. Further experiments with a more diverse set of potential substrates should be conducted to reveal the full range of acceptance. A broader screen of conditions will also be helpful to elucidate the full application potential of PoyH. During the testing of different conditions it should also be possible to have a closer look on the potential zymogen activation. A less active zymogen form of PoyH might get activated due to a release of a potentially inhibitory propeptide. The detected effects of a preincubation of PoyH in acetate buffer, pH 3.0 would suggest a zymogen activation. On the other hand the absence of known characteristic sequence motifs for propeptides argues against zymogen activation through a propeptide release. Therefore it would be interesting to test whether the indications for a zymogen form detected here were no artefacts and how the activation would happen in detail. The function of the potentially released propeptide form would be interesting to analyze as well. Typically, propeptides function as inhibitors, but in the **14** gene cluster a separate inhibitor named PoyG is encoded. This interplay between PoyH and PoyG and a potential propeptide needs further experimental attention.

The use of PoyH and its corresponding recognition motif in RiPP biosynthesis was previously elucidated by M. J. Helf. He incorporated the PoyH cleavage site into thiopeptide thiocillin and the lantibiotic lichenicidin VK21 precursors. In both cases, leader proteins were cleaved off after introduction of the PoyH recognition motif and exposure to PoyH. These detected cleavages indicate that PoyH might serve as a valuable tool for releasing core peptides from, for instance, different maturation states. A controlled release of core peptide would allow us to study the as of yet unexplored timing of modifications in RiPP maturation. Additionally, the GG-cleavage motif in the recognition site is common in RiPP families and this widens a potential range of applications of PoyH. The general need for proteases for molecular biology applications can be seen on the Sigma-Aldrich®-website ([www.sigmaaldrich.com](http://www.sigmaaldrich.com)). Sigma-Aldrich® currently offers more than 50 different proteases and an online tool

named "Enzyme Explorer Protease Finder". One could think that the market is saturated with proteases, but even if such a variety of proteases exists not every recognition motif is covered.

Use of PoyH as a tool for studying RiPP biosynthesis is technically undemanding. In the case of AHs and halogenases, a high potential for application exists. AHs have a potential application in biosynthetic engineering and pathway elucidation. Thanks to their size of ~30 kDa they are well-suited for protein expression and DNA manipulating techniques. It was detected in the case of bacillaene (**8**) that in a TE knock out strain, all pathway intermediates were released from the assembly line (Moldenhauer, Chen, and Borriss 2007). This finding was unexpected, since the TE itself is normally responsible for product release and no TEII are present. It was hypothesized that a second copy of a freestanding AT domain acting as an AH might be responsible for the detachment of stalled intermediates as described in Chapter 2, which *in vitro* studies further supported. The experiments with **8** open the possibility to perform similar experiments in other PKS pathways. Characterization of released truncation products helps to understand the on-PKS biosynthesis where additional tailoring residues have made the pathway intermediates cryptic. In cases where the corresponding gene cluster encodes an AH, a TE knock out should be sufficient to induce release of truncated intermediates. If no AH copy is present, an AH gene could be introduced into the strain. A set of AHs may need to be tested because it is presently unknown if they have interchangeable promiscuity, although substrate specificity is broad for PedC (the AH from the pederin pathway) and sequence alignments suggest the same for other AHs (as described above; Figure 5-2A, Figure 7-1). In addition, the ability of AHs to interact with diverse noncognate ACPs should be examined. The interaction between ATs and ACPs is not well understood in *trans*-AT PKSs. AHs will most likely interact with many ACPs. We were able to detect cleavage of a substrate from a fungal type I PKS-ACP indicating that non-cognate AH-ACP interaction can occur and ACP recognition will most likely not be a problem in AH application. A release of all biosynthetic intermediates of a PKS would facilitate pathway clarification and aid understanding of unusual chemistry within the assembly process. For example, the existence of the unusual  $\beta,\gamma$ -desaturation in **8** was identified by the release of all intermediates (Moldenhauer et al. 2010). An increase of titer levels might be possible with extra copies of AHs since a corresponding knock out leads to titer reduction (Lopanič et al. 2008). This would be a very important achievement as yields of designed and improved "unnatural" polyketides are often unsatisfactory for industrial exploitation.

The two halogenase components presented here in Chapter 3 might find application in chemoenzymatic synthesis once their *in vitro* robustness has been tested. One potential field of utilization could be drug discovery and synthesis. Halogen atoms are known to have an impact on bioavailability and bioactivity (Smith, Grüşchow, and Goss 2013). Vancomycin, a double-chlorinated antibiotic, increasingly loses its antibiotic effect as chlorine atoms are removed (Harris et al. 1985). During the drug discovery process, structure-activity relationship (SAR) studies are conducted to optimize the pharmaceutical properties of a compound. One way of improving lead compounds is the synthesis of a molecule by chemists and the test for its desired activity, followed by synthesis of a modified version and the whole process is repeated until a molecule with appropriate attributes results from it. Halogenations are a standard step in this modification procedure. However, halogenation at certain position in a molecule is hard to obtain; especially vinyl chloride moieties, as present in **11**, are a challenge for synthetic chemists. Having an assortment of available halogenating enzymes may serve as a solution in such situations and the one presented here is able to generate a vinyl chloride moiety. If OocP<sub>4Rx13</sub>/OocQ<sub>4Rx13</sub> are really able to attack and maintain a double bond during chemical synthesis, it would circumvent many steps in chemical synthesis.

In summary, we were able to study different unusual enzymatic mechanisms and with that shed light on various aspects of elemental biosynthetic steps of chain cleavage and halogenation and aid their understanding. In addition, we showed potential opportunities of how these enzymes can be applied in further secondary metabolite studies.

Our discoveries bring us a small step further towards understanding and industrial exploitation of complex biosynthesis pathways, such as those for oocydin A (**11**) or pederin (**5**).

## Chapter 6. Bibliography

---

- Aik W, McDonough MA, Thalhammer A, Chowdhury R, Schofield CJ. 2012. Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Current Opinion in Structural Biology* 22:691-700.
- Antoniewski C, Savelli B, Stragier P. 1990. The *spoIIJ* gene, which regulates early developmental steps in *Bacillus subtilis*, belongs to a class of environmentally responsive genes. *Journal of Bacteriology* 172:86-93.
- Armougom F, Moretti S, Poirot O, Audic S, Dumas P, Schaeli B, Keduas V, Notredame C. 2006. Espresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Research* 34:W604-W608.
- Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, et al 2013. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural Product Reports* 30:108-60.
- Austin MB, Noel JP. 2003. The chalcone synthase superfamily of type III polyketide synthases. *Natural Product Reports* 20:79-110.
- Blasiak L, Drennan CL. 2008. Structural perspective on enzymatic halogenation. *Accounts of Chemical Research* 42:147-15.
- Blasiak LC, Vaillancourt FH, Walsh CT, Drennan CL. 2006. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* 440:368-371.
- Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. 2012. Domain enhanced lookup time accelerated BLAST. *Biology Direct* 7:12.
- Bosserman MA, Downey T, Noinaj N, Buchanan SK, Rohr J. 2013. Molecular insight into substrate recognition and catalysis of Baeyer–Villiger monooxygenase MtmOIV, the key frame-modifying enzyme in the biosynthesis of anticancer agent mithramycin. *ACS Chemical Biology* 8:2466-2477.
- Bretschneider T, Heim JB, Heine D, Winkler R, Busch B, Kusebauch B, Stehle T, Zocher G, Hertweck C. 2013. Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* 502:124-128.
- Buchan DWA, Minneci F, Nugent TOC, Bryson K, Jones DT. 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research* 41:W340-W348.
- Bumpus SB, Magarvey NA, Kelleher NL, Walsh CT, Calderone CT. 2008. Polyunsaturated fatty-acid-like trans-enoyl reductases utilized in polyketide biosynthesis. *Journal of the American Chemical Society* 130:11614-11616.
- Buongiorno D, Straganz GD. 2013. Structure and function of atypically coordinated enzymatic mononuclear non-heme-Fe(II) centers. *Coordination Chemistry Reviews* 257:541-563.
- Butcher RA, Schroeder FC, Fischbach MA, Straight PD, Kolter R, Walsh CT, Clardy J. 2007. The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proceeding of the National Academy of Science* 104:1506-1509.
- Butler A, Sandy M. 2009. Mechanistic considerations of halogenating enzymes. *Nature* 460:848-854.
- Caboche S, Leclère V, Pupin M, Kucherov G, Jacques P. 2010. Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. *Journal of Bacteriology* 192:5143-5150.
- Calderone CT, Bumpus SB, Kelleher NL, Walsh CT, Magarvey NA. 2008. A ketoreductase domain in the PksJ protein of the bacillaene assembly line carries out both alpha- and beta-ketone reduction during chain growth. *Proceeding of the National Academy of Science* 105:12809-12814.

- Calderone CT, Kowtoniuk WE, Kelleher NL, Walsh CT, Dorrestein PC. 2006. Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis*. *Proceeding of the National Academy of Science* 103:8977-8982.
- Calderone CT. 2008. Isoprenoid-like alkylations in polyketide biosynthesis. *Natural Product Reports* 25:845-853.
- Cane DE, Khosla C. 2002. Precursor-directed biosynthesis: biochemical basis of the remarkable selectivity of the erythromycin polyketide synthase toward unsaturated triketides. *Chemistry & Biology* 9:131-142.
- Chen AY, Schnarr NA, Kim C, Cane DE, Khosla C. 2006a. Extender unit and acyl carrier protein specificity of ketosynthase domains of the 6-deoxyerythronolide B synthase. *Journal of the American Chemical Society* 128:3067-3074.
- Chen XH, Vater J, Piel J, Franke P, Scholz R, Schneider K, Koumoutsis A, Hitzeroth G, Grammel N, Strittmatter AW, et al 2006b. Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *Journal of Bacteriology* 188:4024-4036.
- Chen XH, Koumoutsis A, Scholz R, Schneider K, Vater J, Süßmuth R, Piel J, Borriss R. 2009. Genome analysis of *Bacillus amyloliquefaciens* FZB42 reveals its potential for biocontrol of plant pathogens. *Journal of Biotechnology* 140:27-37.
- Cheng Y, Shen B 2003. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proceeding of the National Academy of Science* 100:3149-3154.
- Cheng Y, Coughlin JM, Lim S, Shen B. 2009. Type I polyketide synthases that require discrete acyltransferases. *Methods in Enzymology*, Academic Press, Waltham, 459:165-186.
- Cichewicz RH, Valeriote FA, Crews P. 2004. Psymberin, a potent sponge-derived cytotoxin from *Psammocinia* distantly related to the pederin family. *Organic Letters* 6:1951-1954.
- Cooper MA, Shlaes D. 2011. Fix the antibiotics pipeline. *Nature* 472:32.
- Cortés J, Haydock SF, Roberts GA, Bevitt DJ, Leadlay PF. 1990. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* 348: 176-178.
- Crawford JM, Dancy BC, Hill EA, Udway DW, Townsend CA. 2006. Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. *Proceedings of the National Academy of Sciences* 103:16728-16733.
- Croteau R, Kutchan TM, Lewis NG. 2000. *Biochemistry & Molecular Biology of Plants*. 1<sup>st</sup> Edition: Wiley. 1250-1318.
- Natural products (secondary metabolites). *Biochemistry and Molecular Biology of Plants* 24:1250-1319.
- Cummings M, Breitling R, Takano E. 2014. Steps towards the synthetic biology of polyketide biosynthesis. *FEMS Microbiology Letters* 351:116-125.
- Datsenko K, Wanner B. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Science* 97:6640-6645.
- Del Rizzo PA, Krishnan S, Trievel RC. 2012. Crystal structure and functional analysis of JMJD5 indicate an alternate specificity and function. *Molecular and Cellular Biology* 32:4044-4052.
- Di Tommaso P, Moretti S, Xenarios I, Orobittg M, Montanyola A, Chang JH, Taly JF, Notredame C. 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research* 39:W13-W17.
- Doi-Katayama Y, Hutchinson CR. 2000. Thioesterases and the premature termination of polyketide chain elongation in rifamycin B biosynthesis by *Amycolatopsis mediterranei* S699. *The Journal of Antibiotics* 53:484-495.

- Donadio S, Monciardini P, Sosio M. 2007. Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Natural Product Reports* 24:1073-1109.
- Dong C, Huang F, Deng H, Schaffrath C, Spencer JB. 2004. Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature* 427:561-565.
- Dorrestein PC, Kelleher NL. 2006. Facile detection of acyl and peptidyl intermediates on thiotemplate carrier domains via phosphopantetheinyl elimination reactions during tandem mass spectrometry. *Biochemistry* 45:12756-12766.
- Dorrestein PC, van Lanen SG, Li W, Zhao C, Deng Z, Shen B, Kelleher NL. 2006. The bifunctional glyceryl transferase/phosphatase OzmB belonging to the HAD superfamily that diverts 1,3-bisphosphoglycerate into polyketide biosynthesis. *Journal of the American Chemical Society* 128:10386-10387.
- Du L, Lou L. 2010. PKS and NRPS release mechanisms. *Natural Product Reports* 27:255-78.
- Dunbar KL, Mitchell DA. 2013. Revealing nature's synthetic potential through the study of ribosomal natural product biosynthesis. *ACS Chemical Biology* 8:473-487.
- Dunn BJ, Khosla C. 2013. Engineering the acyltransferase substrate specificity of assembly line polyketide synthases. *Journal of the Royal Society Interface* 10:20130297.
- Dunn BJ, Cane DE, Khosla C. 2013. Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase. *Biochemistry* 52:1839-1841.
- Dutta S, Whicher JR, Hansen DA, Hale WA, Chemler JA, Congdon GR, Narayan ARH, Håkansson K, Sherman DH, Smith JL, et al 2014. Structure of a modular polyketide synthase. *Nature* 510:512-517.
- Eustáquio AS, Pojer F, Noel JP, Moore BS. 2007. Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nature Chemical Biology* 4:69-74.
- Eustáquio AS, Janso JE, Ratnayake AS, O'Donnell CJ, Koehn FE. 2014. Spliceostatin hemiketal biosynthesis in *Burkholderia* spp. is catalyzed by an iron/  $\alpha$ -ketoglutarate-dependent dioxygenase. *Proceedings of the National Academy of Sciences* 111:E3376-E3385.
- Finking R, Marahiel MA. 2004. Biosynthesis of nonribosomal peptide. *Annual Review of Microbiology* 58:453-488.
- Fisch KM, Gurgui C, Heycke N, van der Sar SA, Anderson SA, Webb VL, Taudien S, Platzer M, Rubio BK, Robinson SJ, et al 2009. Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nature Chemical Biology* 5:494-501.
- Fischbach MA, Walsh CT. 2006. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chemical Reviews* 106:3468-3496.
- Floss H, Yu T. 1999. Lessons from the rifamycin biosynthetic gene cluster. *Current Opinion in Chemical Biology* 3:592-597.
- Frank B, Knauber J, Steinmetz H, Scharfe M, Blöcker H, Beyer S, Müller R. 2007. Spiroketal polyketide formation in *Sorangium*: identification and analysis of the biosynthetic gene cluster for the highly cytotoxic spirangienes. *Chemistry & Biology* 14:221-233.
- Freeman MF, Gurgui C, Helf MJ, Morinaka BI, Uria AR, Oldham NJ, Sahl H, Matsunaga S, Piel J. 2012. Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Science* 338:387-90.
- Fujimori DG, Hrvatin S, Neumann CS, Strieker M, Marahiel MA, Walsh CT. 2007. Cloning and characterization of the biosynthetic gene cluster for kutznerides. *Proceedings of the National Academy of Sciences* 104:16498-16503.
- Galonić DP, Barr EW, Walsh CT, Bollinger JM, Krebs C. 2007. Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3. *Nature Chemical Biology* 3:113-116.

- Galonić DP, Vaillancourt FH, Walsh CT. 2006. Halogenation of unactivated carbon centers in natural product biosynthesis: trichlorination of leucine during barbamide biosynthesis. *Journal of the American Chemical Society* 128:3900-3901.
- Gao S, Hothersall J, Wu J, Murphy AC, Song Z, Stephens ER, Thomas CM, Crump MP, Cox RJ, Simpson TJ, et al 2014. Biosynthesis of mupirocin by *Pseudomonas fluorescens* NCIMB 10586 involves parallel pathways. *Journal of the American Chemical Society* 136:5501-5507.
- Garavaglia S, Raffaelli N, Finaurini L, Magni G, Rizzi M. 2004. A novel fold revealed by *Mycobacterium tuberculosis* NAD kinase, a key allosteric enzyme in NADP biosynthesis. *Journal of Biological Chemistry* 279:40980-40986.
- Garg A, Xie X, Keatinge-Clay A, Khosla C, Cane DE. 2014. Elucidation of the cryptic epimerase activity of redox-inactive ketoreductase domains from modular polyketide synthases by tandem equilibrium isotope exchange. *Journal of the American Chemical Society* 136:10190-10193.
- Garg N, Salazar-Ocampo LMA, van der Donk WA. 2013. In vitro activity of the nisin dehydratase NisB. *Proceedings of the National Academy of Sciences* 110:7258–7263.
- Gribble GW. 2010. Naturally occurring organohalogen compounds-A comprehensive update. *Progress in the Chemistry of Organic Natural Products*, Vol. 91. Wien, New York: Springer.
- Griss J, Jones A, Sachsenberg T, Walzer M, Gatto L, Hartler J, Thallinger G, Salek R, Steinbeck C, Neuhauser N, et al 2014. The mzTab data exchange format: communicating mass-spectrometry-based proteomics and metabolomics experimental results to a wider audience. *Molecular & Cellular Proteomics* 13:2765-2775.
- Grüschow S, Buchholz TJ, Seufert W, Dordick JS, Sherman DH. 2007. Substrate profile analysis and ACP-mediated acyl transfer in *Streptomyces coelicolor* type III polyketide synthases. *ChemBioChem* 8:863-88.
- Gu L, Eisman EB, Dutta S, Franzmann TM, Walter S, Gerwick WH, Skiniotis G, Sherman DH. 2011. Tandem acyl carrier proteins in the curacin biosynthetic pathway promote consecutive multienzyme reactions with a synergistic effect. *Angewandte Chemie International Edition* 50:2795-2798.
- Gu L, Wang B, Kulkarni A, Geders TW, Grindberg RV, Gerwick L, Håkansson K, Wipf P, Smith JL, Gerwick WH, et al 2009. Metamorphic enzyme assembly in polyketide diversification. *Nature* 459:731-735.
- Gu Y, Snider B. 2003. Synthesis of ent-haterumalide NA (ent-oocydin A) methyl ester. *Organic Letters* 5:4385-4388.
- Gust B, Challis GL, Fowler K, Kieser T, Chater KF. 2003. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proceedings of the National Academy of Sciences* 100:1541-1546.
- Haapalainen AM, Meriläinen G, Wierenga RK. 2006. The thiolase superfamily: condensing enzymes with diverse reaction specificities. *Trends in Biochemical Sciences* 31:64-71.
- Haft DH, Basu M, Mitchell DA. 2010. Expansion of ribosomally produced natural products: a nitrile hydratase- and Nif11-related precursor family. *BMC Biology* 8:1-70.
- Haines AS, Dong X, Song Z, Farmer R, Williams C, Hothersall J, Ploskon E, et al 2013. A conserved motif flags acyl carrier proteins for  $\beta$ -branching in polyketide synthesis. *Nature Chemical Biology* 9:685-692.
- Hamada T, Matsunaga S, Yano G, Fusetani N. 2005. Polytheonamides A and B, highly cytotoxic, linear polypeptides with unprecedented structural features, from the marine sponge, *Theonella swinhoei*. *Journal of the American Chemical Society* 127:110-8.
- Hamada T, Sugawara T, Matsunaga S, Fusetani N. 1994. Polytheonamides, unprecedented highly cytotoxic polypeptides, from the marine sponge theonella swinhoei: 1. Isolation and component amino acids. *Tetrahedron letters* 35:719-720.

- Hangasky JA, Taabazuing CY, Valliere MA, Knapp MJ. 2013. Imposing function down a (cupin)-barrel: secondary structure and metal stereochemistry in the  $\alpha$ KG-dependent oxygenases. *Metallomics* 5:287-301.
- Harris CM, Kannan R, Kopecka H, Harris TM. 1985. The role of the chlorine substituents in the antibiotic vancomycin: preparation and characterization of mono- and didechlorovancomycin. *Journal of the American Chemical Society* 107:6652-6658.
- Heathcote ML, Leadlay PF. 2001. Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chemistry & Biology* 8:207-220.
- He YH, Tang MC, Zhang F, Tang GL. 2014. Cis-double bond formation by thioesterase and transfer by ketosynthase in FR901464 biosynthesis. *Journal of the American Chemical Society* 136:4488-4491.
- Hedstrom L. 2002. Serine protease mechanism and specificity. *Chemical Reviews* 102:4501-4524.
- Hertweck C, Luzhetskyy A, Rebets Y, Bechthold A. 2007. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Natural Product Reports* 24:162.
- Hertweck C. 2009. The biosynthetic logic of polyketide diversity. *Angewandte Chemie International Edition* 48:4688-4716.
- Hill AM. 2006. The biosynthesis, molecular genetics and enzymology of the polyketide-derived metabolites. *Natural Product Reports* 23:256-320.
- Hillwig ML, Liu X. 2014. A new family of iron-dependent halogenases acts on freestanding substrates. *Nature Chemical Biology* 10:921-923
- Hioe J, Zipse H. 2011. Radical stability -thermochemical aspects. *Encyclopedia of Radicals in Chemistry, Biology and Materials*, 1<sup>st</sup> Edition, Chichester, West Sussex ; Hoboken, N.J.: Wiley, 449-476.
- Hopwood DA. 2004. Cracking the polyketide code. *PLoS Biology* 2:166-169.
- Hoye TR, Wang J. 2005. Alkyne haloallylation [with Pd(II)] as a core strategy for macrocycle synthesis: a total synthesis of (-)-haterumalide NA/(-)-oocydin A. *Journal of the American Chemical Society* 127:6950-6951.
- Inoue M, Shinohara N, Tanabe S, Takahashi T, Okura K, Itoh H, Mizoguchi Y, Iida M, Lee N, Matsuoka S. 2010. Total synthesis of the large non-ribosomal peptide polytheonamide B. *Nature Chemistry* 2:280-285.
- Irschik H, Kopp M, Weissman KJ, Buntin K, Piel J, Müller R. 2010. Analysis of the sorangicin gene cluster reinforces the utility of a combined phylogenetic/retrobiosynthetic analysis for deciphering natural product assembly by trans-AT PKS. *ChemBioChem* 11:1840-1849.
- Ishikawa N, Tanaka H, Koyama F, Noguchi H, Wang CC, Hotta K, Watanabe K. 2014. Non-heme dioxygenase catalyzes atypical oxidations of 6,7-bicyclic systems to form the 6,6-quinolone core of viridicatin-type fungal alkaloid. *Angewandte Chemie International Edition* 53:12880-12884.
- Iwamoto M, Shimizu H, Muramatsu I, Oiki S. 2010. A cytotoxic peptide from a marine sponge exhibits ion channel activity through vectorial-insertion into the membrane. *FEBS Letters* 584:3995-9.
- Iwasaki S, Namikoshi M, Kobayashi H. 1986. Studies on macrocyclic lactone antibiotics. VIII. Absolute structures of rhizoxin and a related compound. *Journal of Antibiotics* 39:424-429.
- Jenke-Kodama H, Börner T, Dittmann E. 2006. Natural biocombinatorics in the polyketide synthase genes of the actinobacterium *Streptomyces avermitilis*. *PLoS computational biology* 2:1210-1218.
- Jenner M, Frank S, Kampa A, Kohlhaas C, Pöplau P, Briggs GS, Piel J, Oldham NJ. 2013. Substrate specificity in ketosynthase domains from trans-AT polyketide synthases. *Angewandte Chemie International Edition* 52:1143-1147.



- Jensen K, Niederkrüger H, Zimmermann K, Vagstad AL, Moldenhauer J, Brendel N, Frank S, Pöplau P, Kohlhaas C, Townsend CA, et al 2012. Polyketide proofreading by an acyltransferase-like enzyme. *Chemistry & Biology* 19:329-339.
- Jensen PR, Fenical W. 1994. Strategies for the discovery of secondary metabolites from marine bacteria: ecological perspectives. *Annual Reviews in Microbiology* 48:559-584.
- Jez JM, Ferrer J, Bowman ME, Dixon RA, Noel JP. 2000. Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. *Biochemistry* 39:890-902.
- Kanehisa M, Goto S, Kawashima S, Nakaya A. 2002. The KEGG databases at GenomeNet. *Nucleic Acids Research* 30:42-46.
- Kawai S, Mori S, Mukai T, Suzuki S, Yamada T, Hashimoto W, Murata K. 2000. Inorganic polyphosphate/ATP-NAD kinase of *Micrococcus flavus* and *Mycobacterium tuberculosis* H37Rv. *Biochemical and Biophysical Research Communications* 276:57-63.
- Keating TA, Walsh CT. 1999. Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Current Opinion in Chemical Biology* 3:598-606.
- Keatinge-Clay AT. 2007. A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chemistry & Biology* 14:898-908
- Keatinge-Clay AT. 2012. The structures of type I polyketide synthases. *Natural Product Reports* 29:1050-73.
- Kelley LA, Sternberg MJ. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* 4:363-371.
- Khan A, James M. 1998. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein science* 7:815-836.
- Khare D, Wang B, Gu L, Razelun J, Sherman DH, Gerwick WH, Hakansson K, Smith JL. 2010. Conformational switch triggered by  $\alpha$ -ketoglutarate in a halogenase of curacin A biosynthesis. *Proceedings of the National Academy of Sciences* 107:14099-14104.
- Khosla C, Herschlag D, Cane DE, Walsh CT. 2014. Assembly line polyketide synthases: mechanistic insights and unsolved problems. *Biochemistry* 53:2875-2883.
- Kiefer P, Schmitt U, Vorholt JA. 2013. eMZed: an open source framework in python for rapid and interactive development of LC/MS data analysis workflows. *Bioinformatics* 29:963-964.
- Kigoshi H, Hayakawa I. 2007. Marine cytotoxic macrolides haterumalides and biselides, and related natural products. *The Chemical Record* 7:254-264.
- Kigoshi H, Kita M, Ogawa S, Itoh M, Uemura D. 2003. Enantioselective synthesis of 15-epi-haterumalide NA methyl ester and revised structure of haterumalide NA. *Organic Letters* 5:957-960.
- Koehnert K, Emerson J, Que L. 2005. The 2-His-1-carboxylate facial triad: a versatile platform for dioxygen activation by mononuclear non-heme iron (II) enzymes. *Journal of Biological Inorganic Chemistry* 10:87-93.
- Koglin A, Löhr F, Bernhard F, Rogov VV, Frueh DP, Strieter ER, Mofid MR, Güntert P, Wagner G, Walsh CT, et al. 2008. Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature* 454:907-911.
- Kohlhaas C, Jenner M, Kampa A, Briggs G, Afonso J, Piel J, Oldham N. 2013. Amino acid-accepting ketosynthase domain from a trans-AT polyketide synthase exhibits high selectivity for predicted intermediate. *Chemical Science* 4:3212-3217.

- Kosaka T, Kato S, Shimoyama T, Ishii S, Abe T, Watanabe K. 2008. The genome of *Pelotomaculum thermopropionicum* reveals niche-associated evolution in anaerobic microbiota. *Genome Research* 18:442-448.
- Kotowska M, Pawlik K. 2014. Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Applied Microbiology and Biotechnology* 98:7735-7746.
- Kusebauch B, Busch B, Scherlach K, Roth M, Hertweck C. 2009. Polyketide-chain branching by an enzymatic Michael addition. *Angewandte Chemie International Edition* 48:5001-5004.
- Kusebauch B, Busch B, Scherlach K, Roth M, Hertweck C. 2010. Functionally distinct modules operate two consecutive  $\alpha,\beta \rightarrow \beta,\gamma$  double-bond shifts in the rhizoxin polyketide assembly line. *Angewandte Chemie International Edition* 49:1460-1464.
- Kwan DH, Schulz F. 2011. The stereochemistry of complex polyketide biosynthesis by modular polyketide synthase. *Molecules* 16:6092-6115.
- Kwan J, Schmidt E. 2012. Cleaning up polyketide synthases. *Chemistry & Biology* 19:309-311.
- Lackner G, Moebius N, Partida-Martinez LP, Bol S, Hertweck C, Boland S. 2011a. Evolution of an endofungal lifestyle: deductions from the *Burkholderia rhizoxinica* genome. *BMC Genomics* 12:210.
- Lackner G, Moebius N, Partida-Martinez LP, Hertweck C. 2011b. Complete genome sequence of *Burkholderia rhizoxinica*, an endosymbiont of *Rhizopus microsporus*. *Journal of Bacteriology* 193:783-784.
- Lai JR, Walsh CT. 2006. Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis. *Biochemistry* 45:14869-14879.
- Lange E, Gropf C, Schulz-Trieglaff O, Leinenbach A, Huber C, Reinert K. 2007. A geometric approach for the alignment of liquid chromatography mass spectrometry data. *Bioinformatics* 23:i273-i281.
- Levenfors JJ, Hedman R, Thaning C, Gerhardson B, Welch CJ. 2004. Broad-spectrum antifungal metabolites produced by the soil bacterium *Serratia plymuthica* A153. *Soil Biology and Biochemistry* 36:677-685.
- Liu T, Lin X, Zhou X, Deng Z, Cane DE. 2008. Mechanism of thioesterase-catalyzed chain release in the biosynthesis of the polyether antibiotic nanchangmycin. *Chemistry & Biology* 15:449-458.
- Lopanič NB, Shields JA, Buchholz TJ, Rath CM, Hothersall J, Haygood MG, Håkansson K, Thomas CM, Sherman DH. 2008. In vivo and in vitro trans-acylation by BryP, the putative bryostatin pathway acyltransferase derived from an uncultured marine symbiont. *Chemistry & Biology* 15:1175-1186.
- Mahenthalingam E, Song L, Sass A, White J, Wilmot C, Marchbank A, Boaisha O, Paine J, Knight D, Challis G. 2011. Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic *Burkholderia ambifaria* genomic island. *Chemistry & Biology* 18:665-677.
- Marchler-Bauer A, Lu S, Anderson J, Chitasaz F, Derbyshire M, DeWeese-Scott C, Fong J, Geer L, Geer R, Ginzales N, et al 2011. CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Research* 39:D225-D229.
- Marchuk D, Drumm M, Saulino A. 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Research* 19:1154.
- Matilla MA, Stockmann H, Leeper FJ, Salmond GPC. 2012. Bacterial biosynthetic gene clusters encoding the anti-cancer haterumalide class of molecules: biogenesis of the broad spectrum antifungal and anti-oomycete compound, oocydin A. *Journal of Biological Chemistry* 287:39125-39138.
- Matthews ML, Chang W, Layne AP, Miles LA, Krebs C, Bollinger JM. 2014. Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase. *Nature Chemical Biology* 10:209-215.

- McAlpine JB, Bachmann BO, Pirae M, Tremblay S, Alarco AM, Zazopoulos E, Farnet CM. 2005. Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example. *Journal of Natural Products* 68:493-496.
- Menzella HG, Carney JR, Santi DV. 2007. Rational design and assembly of synthetic trimodular polyketide synthases. *Chemistry & Biology* 14:143-151.
- Minto R, Townsend C. 1997. Enzymology and molecular biology of aflatoxin biosynthesis. *Chemical Reviews* 97:2537-2555.
- Moldenhauer J, Chen XH, Borriss R. 2007. Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the trans-AT family. *Angewandte Chemie International Edition* 119:8348-8345.
- Moldenhauer J, Götz DC, Albert CR, Bischof SK, Schneider K, Süßmuth RD, Engeser M, Gross H, Bringmann G, Piel J. 2010. The final steps of bacillaene biosynthesis in *Bacillus amyloliquefaciens* FZB42: direct evidence for  $\beta,\gamma$  dehydration by a trans-acyltransferase polyketide synthase. *Angewandte Chemie International Edition* 122:1507-1509.
- Morinaka BI, Vagstad AL, Helf MJ, Gugger M, Kegler C, Freeman MF, Bode HB, Piel J. 2014. Radical S-adenosyl methionine epimerases: regioselective introduction of diverse D-amino acid patterns into peptide natural products. *Angewandte Chemie International Edition* 126:8643-8647.
- Musiol E, Härtner T, Kulik A, Moldenhauer J, Piel J, Wohlleben W, Weber T. 2011. Supramolecular templating in kirromycin biosynthesis: the acyltransferase KirCII loads ethylmalonyl-CoA extender onto a specific ACP of the trans-AT PKS. *Chemistry & Biology* 18:438-444.
- Musiol E, Weber T. 2012. Discrete acyltransferases involved in polyketide biosynthesis. *MedChemComm* 3:871-886.
- Neumann CS, Fujimori DG, Walsh CT. 2008. Halogenation strategies in natural product biosynthesis. *Chemistry & Biology* 15:99-109.
- Nguyen T, Ishida K, Jenke-Kodama H, Dittmann E, Gurgui C, Hochmuth T, Taudien S, Platzer M, Hertweck C, Piel J. 2008. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nature Biotechnology* 26:225-233.
- Nikolouli K, Mossialos D. 2012. Bioactive compounds synthesized by non-ribosomal peptide synthetases and type-I polyketide synthases discovered through genome-mining and metagenomics. *Biotechnology Letters* 34:1393-1403.
- Nogaroto V, Tagliavini SA, Gianotti A, Mikawa A, Barros NM, Puzer L, Carmona AK, Costa PI, Henrique-Silva F. 2006. Recombinant expression and characterization of a *Xylella fastidiosa* cysteine protease differentially expressed in a nonpathogenic strain. *FEMS Microbiology Letters* 261:187-193.
- Novinec M, Lenarčič B. 2013. Papain-like peptidases: structure, function, and evolution. *BioMolecular Concepts* 4:287-308.
- O'Brien RV, Davis RW, Khosla C, Hillenmeyer ME. 2014. Computational identification and analysis of orphan assembly-line polyketide synthases. *The Journal of Antibiotics* 67:89-97.
- Olano C, Méndez C, Salas JA. 2010. Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis. *Natural Product Reports* 27:571.
- Oman TJ, van der Donk WA. 2010. Follow the leader: the use of leader peptides to guide natural product biosynthesis. *Nature Chemical Biology* 6:9-18.
- Parks TD, Leuther KK, Howard ED, Johnston SA, Dougherty WG. 1994. Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Analytical Biochemistry* 216:413-417.

- Partida-Martinez LP, Hertweck C. 2005. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* 437:884-888.
- Partida-Martinez LP, Hertweck C. 2006. A gene cluster encoding rhizoxin biosynthesis in "*Burkholderia rhizoxina*", the bacterial endosymbiont of the fungus *Rhizopus microsporus*. *ChemBioChem* 8:41-45.
- Perham R. 2000. Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annual Review of Biochemistry* 69:961-1004.
- Piel J, Wen G, Platzer M, Hui D. 2004a. Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase. *ChemBioChem* 5:93-98.
- Piel J. 2002. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proceedings of the National Academy of Sciences* 99:14002-14007.
- Piel J. 2009. Metabolites from symbiotic bacteria. *Natural Product Reports* 26:338-362.
- Piel J. 2010. Biosynthesis of polyketides by trans-AT polyketide synthases. *Natural Product Reports* 27:996-1047.
- Piel J, Hui D, Wen G, Butzke D, Platzer M, Fusetani N, Matsunaga S. 2004b. Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proceedings of the National Academy of Sciences* 101:16222-16227.
- Pettit GR, JP Xu, JC Chapuis, RK Pettit, LP Tackett, Doubek DL, Hooper JNA, Schmidt JM. 2004. Antineoplastic agents. 520. Isolation and structure of irciniastatins A and B from the indo-pacific marine sponge *Ircinia ramosa*. *Journal of Medicinal Chemistry* 47:1149-1152
- Pohl NL, Gokhale RS, Cane DE, Khosla C. 1998. Synthesis and Incorporation of an N-acetylcysteamine analogue of methylmalonyl-CoA by a modular polyketide synthase. *Journal of the American Chemical Society* 120:11206-11207.
- Poust S, Yoon I, Adams PD, Katz L, Petzold CJ, Keasling JD. 2014. Understanding the role of histidine in the GHSxG acyltransferase active site motif: evidence for histidine stabilization of the malonyl-enzyme Intermediate. *PLoS ONE* 9:e109421.
- Pratter SM, Ivkovic J, Birner-Gruenberger R, Breinbauer R, Zangger K, Straganz GD. 2014a. More than just a halogenase: modification of fatty acyl moieties by a trifunctional metal enzyme. *ChemBioChem* 15:567-574.
- Pratter SM, Light KM, Solomon EI, Straganz GD. 2014b. The role of chloride in the mechanism of O<sub>2</sub> activation at the mononuclear nonheme Fe(II) center of the halogenase HctB. *Journal of the American Chemical Society* 136:9385-9395.
- Pöplau P, Frank S, Morinaka BI, Piel J. 2013. An enzymatic domain for the formation of cyclic ethers in complex polyketides. *Angewandte Chemie International Edition* 52:13215-13218.
- Rahman AS, Hothersall J, Crosby J, Simpson TJ, Thomas CM. 2005. Tandemly duplicated acyl carrier proteins, which increase polyketide antibiotic production, can apparently function either in parallel or in series. *Journal of Biological Chemistry* 280:6399-6408
- Rascón AA, Gearin J, Isoe J, Miesfeld RL. 2011. In vitro activation and enzyme kinetic analysis of recombinant midgut serine proteases from the Dengue vector mosquito *Aedes aegypti*. *BMC Biochemistry* 12:43-53.
- Rawlings ND, Barrett AJ, Bateman A. 2012. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research* 38:D227-D233.
- Rawlings ND, Barrett AJ. 1994. Families of cysteine peptidases. *Methods in Enzymology*, Academic Press, Waltham, 244:461-86.

- Ridley CP, Bergquist PR, Harper MK, Faulkner DJ, Hooper JN, Haygood MG. 2005. Speciation and biosynthetic variation in four dictyoceratid sponges and their cyanobacterial symbiont, *Oscillatoria spongelliae*. *Chemistry & Biology* 12:397-406.
- Rix U, Fischer C, Remsing LL, Rohr J. 2002. Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Natural Product Reports* 19:542-580.
- Roberts G. 2013. *Encyclopedia of Biophysics*. 1<sup>st</sup> Edition, Heidelberg: Springer. 674-680.
- Rose NA, McDonough MA, King O, Kawamura A, Schofield CJ. 2011. Inhibition of 2-oxoglutarate dependent oxygenases. *Chemical Society Review* 40:4364-4397.
- Rossi M, Paquelin A, Ghigo JM, Wandersman C. 2003. Haemophore-mediated signal transduction across the bacterial cell envelope in *Serratia marcescens*: the inducer and the transported substrate are different molecules. *Molecular Microbiology* 48:1467-1480.
- Roulland E. 2008. Total synthesis of (+)-oocycin A: application of the suzuki-miyaura cross-coupling of 1,1-dichloro-1-alkenes with 9-alkyl 9-BBN. *Angewandte Chemie International Edition* 120:3822-3825.
- Rzychon M, Chmiel D, Stec-Niemczyk J. 2004. Modes of inhibition of cysteine proteases. *Acta Biochimica Polonica* 51:861-873.
- Sambrook J, Russell DW. 2001. *Molecular cloning*. 3<sup>rd</sup> Edition, New York: CSHL Press.
- Sato B, Nakajima H, Fujita T, Takase S, Yoshimura S, Kinoshita T, Terano H. 2005. FR177391, a new anti-hyperlipidemic agent from *Serratia*. I. Taxonomy, fermentation, isolation, physico-chemical properties, structure elucidation and biological activities. *The Journal of Antibiotics* 58:634-9.
- Sattely ES, Fischbach MA, Walsh CT. 2008. Total biosynthesis: in vitro reconstitution of polyketide and nonribosomal peptide pathways. *Natural Product Reports* 25:757-793.
- Schneider A, Marahiel MA. 1998. Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Archives of Microbiology* 169:404-410.
- Schwarzer D, Mootz HD, Linne U, Marahiel MA. 2002. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proceedings of the National Academy of Sciences* 99:14083-14088.
- Scotti C, Albertini AM. 1993. A *Bacillus subtilis* large ORF coding for a polypeptide highly similar to polyketide synthases. *Gene* 130:65-71.
- Serre L, Verbree E, Dauter Z, Stuitje A, Derewenda Z. 1995. The *Escherichia coli* malonyl-CoA:Acyl Carrier protein transacylase at 15-Å resolution. *The Journal of Biological Chemistry* 270:12961-12964.
- Shen B. 2003. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology* 7:285-295.
- Smith DR, Gruschow S, Goss RJ. 2013. Scope and potential of halogenases in biosynthetic applications. *Current Opinion in Chemical Biology* 17:276-283.
- Smith S, Tsai S. 2007. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural Product Reports* 24:1041-1072.
- Spellberg B. 2008. Dr. William H. Stewart: mistaken or maligned? *Clinical infectious diseases* 47:294.
- Staunton J, Weissman KJ. 2001. Polyketide biosynthesis: a millennium review. *Natural Product Reports* 18:380-416.
- Steinmetz M, Richter R. 1994. Easy cloning of mini-Tn10 insertions from the *Bacillus subtilis* chromosome. *Journal of Bacteriology* 176:1761.

- Strobel G, Hess WM. 1999. Oocydin A, a chlorinated macrocyclic lactone with potent anti-oomycete activity from *Serratia marcescens*. *Microbiology* 145:3557-3564.
- Sunbul M, Zhang K, Yin J. 2009. Using phosphopantetheinyl transferases for enzyme posttranslational activation, site specific protein labeling and identification of natural product biosynthetic gene clusters from bacterial genomes. *Methods in Enzymology*, Academic Press, Waltham, 459:255-275.
- Takada N, Sato H, Suenaga K, Arimoto H, Yamada K. 1999. Isolation and structures of haterumalides NA, NB, NC, ND, and NE, novel macrolides from an okinawan sponge *Ircinia sp.* *Tetrahedron letters* 40:6309-6312.
- Tang Y, Koppisch AT, Khosla C. 2004. The acyltransferase homologue from the initiation module of the R1128 polyketide synthase is an acyl-ACP thioesterase that edits acetyl primer units. *Biochemistry* 43:9546-9555.
- Tartoff K, Hobbs C. 1987. Improved media for growing plasmid and cosmid clones. *Bethesda Research Laboratories Focus* 9:12.
- Teruya T, Suenaga K, Maruyama S, Kurotaki M, Kigoshi H. 2005. Biselides A–E: novel polyketides from the okinawan ascidian *Didemnidae sp.* *Tetrahedron* 61:6561-6567.
- Teta R, Gurgui M, Helfrich EJ, Künne S, Schneider A, van Echten-Deckert G, Mangoni A, Piel J. 2010. Genome mining reveals trans-AT polyketide synthase directed antibiotic biosynthesis in the bacterial phylum bacteroidetes. *ChemBioChem* 11:2506-2512.
- Thaning C, Welch C, Borowicz J, Hedman R. 2001. *Sclerotinia sclerotiorum* apothecial formation by the soil bacterium *Serratia plymuthica*: identification of a chlorinated macrolide as one of the causal agents. *Soil Biology and Biochemistry* 33:1817-1826.
- Tibrewal N, Pahari P, Wang G, Kharel MK, Morris C, Downey T, Hou Y, Bugni TS, Rohr J. 2012. Baeyer–Villiger C–C bond cleavage reaction in gilvocarcin and jadomycin biosynthesis. *Journal of the American Chemical Society* 134:18181-18184.
- Tibrewal N, Tang Y. 2014. Biocatalysts for natural product biosynthesis. *Annual Review of Chemical and Biomolecular Engineering* 5:347-366.
- Tomassen J, Eiglmeier K, Cole S, Overduin P, Larson T, Boos W. 1991. Characterization of two genes, *glpQ* and *ugpQ*, encoding glycerophosphoryl diester phosphodiesterases of *Escherichia coli*. *Molecular and General Genetics* 226:321-327.
- Ueda K, Hu Y. 1999. Haterumalide B: A new cytotoxic macrolide from an okinawan ascidian *Lissoclinum sp.* *Tetrahedron letters* 40:6305-6308.
- Ueda M, Yamaura M, Ikeda Y, Suzuki Y, Yoshizato K, Hayakawa I, Kigoshi H. 2009. Total synthesis and cytotoxicity of haterumalides NA and B and their artificial analogues. *European Journal of Organic Chemistry* 74:3370-3377.
- Vaillancourt F, Vosburg, Walsh C. 2006. Dichlorination and bromination of a threonyl-S-carrier protein by the non-heme FeII halogenase SyrB2. *ChemBioChem* 7:748-752.
- Vaillancourt FH, Yeh E, Vosburg DA, Garneau-Tsodikova S, Walsh CT. 2006. Nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chemical Reviews* 106:3364-3378.
- Vaillancourt FH, Yeh E, Vosburg DA, O'Connor SE, Walsh CT. 2005. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* 436:1191-1194.
- Vaillancourt FH, Yin J, Walsh CT. 2005. SyrB2 in syringomycin E biosynthesis is a nonheme FeII alpha-ketoglutarate- and O<sub>2</sub>-dependent halogenase. *Proceedings of the National Academy of Sciences* 102:10111-10116.
- van Berkel WJ, Kamerbeek NM, Fraaije MW. 2006. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *Journal of Biotechnology* 124:670-689.

- van Pée KH. 2012. Enzymatic chlorination and bromination. *Methods in Enzymology*, Academic Press, Waltham, 516:237-257.
- Vergnolle O, Hahn F, Baerga-Ortiz A, Leadlay PF, Andexer JN. 2011. Stereoselectivity of isolated dehydratase domains of the borrelidin polyketide synthase: implications for cis double bond formation. *ChemBioChem* 12:1011-1014.
- Wagner C, el Omari M, König GM. 2009. Biohalogenation: nature's way to synthesize halogenated metabolites. *Journal of Natural Products* 72:540-553.
- Walsh CT, Fischbach MA. 2010. Natural products version 2.0 - connecting genes to molecules. *Journal of the American Chemical Society* 132:2469-2493.
- Walsh CT. 2014. Blurring the lines between ribosomal and nonribosomal peptide scaffolds. *ACS Chemical Biology* 9:1653-1661.
- Wang H, Zhou X, Wu M, Wang C, Zhang X, Tao Y, Chen N, Zang J. 2013. Structure of the JmjC-domain-containing protein JMJD5. *Acta Crystallographica Section D Biological Crystallography* 69:1911-1920.
- Weber T, Laiple KJ, Pross EK, Textor A, Grond S, Welzel K, Pelzer S, Vente A, Wohlleben W. 2008. Molecular analysis of the kirromycin biosynthetic gene cluster revealed  $\beta$ -alanine as precursor of the pyridone moiety. *Chemistry & Biology* 15:175-188.
- Weissman KJ, Müller R. 2008. Protein-protein interactions in multienzyme megasynthetases. *ChemBioChem* 9:826-848.
- Weissman KJ. 2009. Introduction to polyketide biosynthesis. *Methods in Enzymology*, Academic Press, Waltham 459:3-16.
- Weissman KJ, Leadlay PF. 2005. Combinatorial biosynthesis of reduced polyketides. *Nature Reviews Microbiology* 3:925-936.
- Wilson MC, Mori T, Rückert C, Uria AR, Helf MJ, Takada K, Gernert C, Steffens UAE, Heycke N, Schmitt S, et al 2014. An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* 506:58-62.
- Wong FT, Khosla C. 2011. Structure and mechanism of the trans-acting acyltransferase from the disorazole synthase. *Biochemistry* 50:6539-6548.
- Wong SD, Srnec M, Matthews ML, Liu LV, Kwak Y, Park K, Bell CB III, Alp EE, Zhao J, Yoda Y, et al 2013. Elucidation of the Fe(IV)=O intermediate in the catalytic cycle of the halogenase SyrB2. *Nature* 499:320-323.
- Wu J, Hothersall J, Mazzetti C, O'Connell Y, Shields JA, Rahman AS, Cox RJ, Crosby J, Simpson TJ, Thomas CM, et al 2008. In vivo mutational analysis of the mupirocin gene cluster reveals labile points in the biosynthetic pathway: the "leaky hosepipe" mechanism. *ChemBioChem* 9:1500-1508.
- Xu W, Qiao K, Tang Y. 2012. Structural analysis of protein-protein interactions in type I polyketide synthases. *Critical Reviews in Biochemistry and Molecular Biology* 48:98-122.
- Yadav G, Gokhale RS, Mohanty D. 2003. Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *Journal of Molecular Biology* 328:335-363.
- Yan K, Li Y, Zirah S, Goulard C, Knappe TA, Marahiel MA, Rebuffat S. 2012. Dissecting the maturation steps of the lasso peptide microcin J25 in vitro. *ChemBioChem* 13:1046-1052.
- Yang X, van der Donk W. 2013. Ribosomally synthesized and post-translationally modified peptide natural products: new insights into the role of leader and core peptides during biosynthesis. *Chemistry-A European Journal* 19:7662-7677.

- Yeh E, Garneau S, Walsh CT. 2005. Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proceedings of the National Academy of Science* 102:3960-3965.
- Yeh E, Kohli RM, Bruner SD, Walsh CT. 2004. Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *ChemBioChem* 5:1290-1293.
- Yu D, Xu F, Zeng J, Zhan J. 2012. Type III polyketide synthases in natural product biosynthesis. *IUBMB Life* 64:285-295.
- Yu T, Floss HG. 1999. Direct evidence that the rifamycin polyketide synthase assembles polyketide chains processively. *Proceeding of the National Academy of Science* 96:9051-9056.
- Zehner S, Kotsch A, Bister B, Süßmuth R, Mendez C, Salas J, van Pée KH. 2005. A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chemistry & Biology* 12:445-452.
- Zhang W, Tang Y. 2009. In vitro analysis of type II polyketide synthase. *Methods in Enzymology*, Academic Press, Waltham, 459:367-393.
- Zimmermann K, Engeser M, Blunt JW, Munro MH, Piel J. 2009. Pederin-type pathways of uncultivated bacterial symbionts: analysis of O-methyltransferases and generation of a biosynthetic hybrid. *Journal of the American Chemical Society* 131:2780-2781.



## Chapter 7. Appendix

---

### 7.1 Abbreviations used

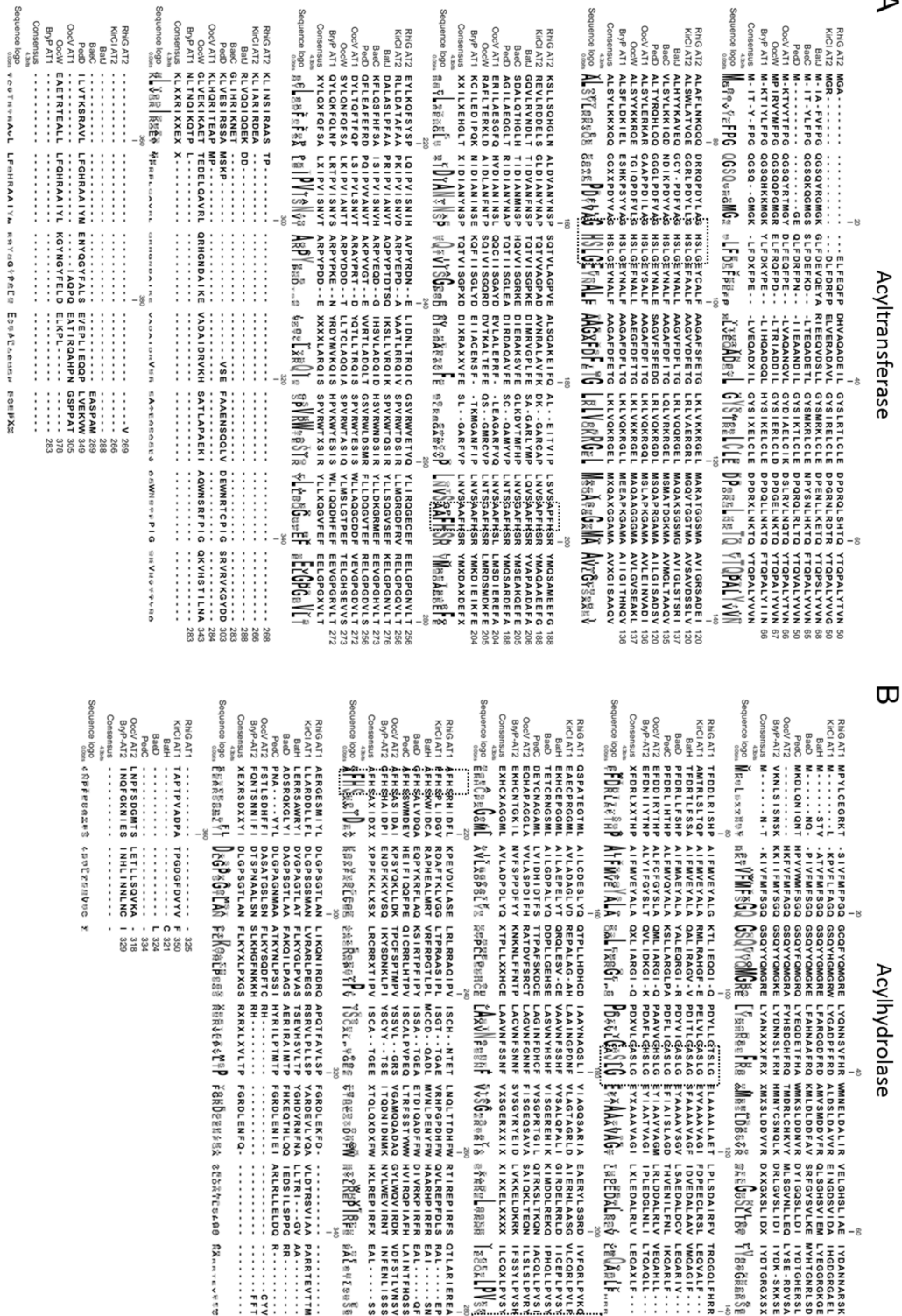
|                   |  |
|-------------------|--|
| 6dEB              | 6-deoxyerythronolide B                         |
| A                 | adenylation                                    |
| A                 | alanine  |
| aa                | amino acid                                     |
| ACCO              | 1-aminocyclopropane-1-carboxylic acid oxidase  |
| ACP               | acyl carrier protein                           |
| AH                | acylhydrolase                                  |
| Ala               | alanine  |
| AMP               | adenosinmonophosphat                           |
| Arg               | arginine                                       |
| Asn               | asparagine                                     |
| Asp               | aspartate                                      |
| ATP               | adenosintriphosphat                            |
| B                 | branching domain                               |
| <i>B.</i>         | <i>Bacillus</i>                                |
| BLAST             | basic local alignment search tool              |
| br                | broad  |
| BSA               | bovine serum albumin                           |
| BVMO              | Baeyer-Villiger monooxygenase                  |
| C                 | condensation                                   |
| C                 | cysteine                                       |
| CDCl <sub>3</sub> | fully deuterated chloroform                    |
| CHS               | chalcone synthase                              |
| Ci                | Curie  |
| CLF               | chain length factor                            |
| CoA               | coenzym A                                      |
| COSY              | correlated spectroscopy                        |
| CV                | column volume                                  |
| Cys               | cysteine                                       |
| d                 | douplet  |
| D                 | aspartate                                      |
| Da                | Dalton   |
| DEBS              | 6-deoxyerythronolide B synthase                |
| DELTA-BLAST       | domain enhanced lookup time accelerated- BLAST |
| DH                | dehydrogenase                                  |
| Dha               | $\alpha,\beta$ -didehydroalanine               |
| Dhb               | 2,3-didehydroaminobutyric acid                 |
| DMAP              | dimethylaminopyridine                          |
| dNTP              | deoxynucleoside triphosphate                   |
| DSZS              | diszorazole                                    |

|                 |  |
|-----------------|--|
| DTNB            | 5,5'-dithio-bis(2-nitrobenzoic acid)                 |
| DTT             | dithiothreitol                                       |
| E               | epimerization  |
| E               | glutamate  |
| <i>E.</i>       | <i>Escherichia</i>                                   |
| ECH             | enoyl-CoA hydratase                                  |
| EDC             | 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide        |
| EPB             | enriched potato dextrose broth                       |
| ER              | enoylreductase                                       |
| ESI             | electrospray ionisation                              |
| eV              | electronvolt   |
| F               | phenylalanine  |
| FAD             | flavin adenine dinucleotide                          |
| FAS             | fatty acid synthase                                  |
| FCS             | fetal calf serum                                     |
| FIH-1           | Human Factor Inhibiting Hif                          |
| FRT             | flippase recognition target                          |
| G               | glycine  |
| Gln             | glutamine  |
| Glu             | glutamate  |
| Gly             | glycine  |
| H               | histidine  |
| hr              | hour   |
| HCS             | HMG-CoA-synthase                                     |
| His             | histidine  |
| HMBC            | heteronuclear multiple-bond correlation spectroscopy |
| HMG             | 3-hydroxy-3-methylglutaryl                           |
| HPLC            | high-pressure liquid chromatography                  |
| HRMS            | high resolution mass spectrometry                    |
| Hz              | Hertz  |
| I               | isoleucine   |
| Ile             | isoleucine   |
| IPNS            | isopenicillin N synthase                             |
| IPTG            | isopropyl $\beta$ -D-1-thiogalactopyranoside         |
| JTT             | Jones-Taylor-Thornten                                |
| K               | lysine   |
| kb              | kilobases  |
| KG              | ketoglutarate  |
| KPi             | potassium phosphate                                  |
| KR              | ketoreductase  |
| KS              | ketosynthase   |
| KS <sup>0</sup> | non-elongating KS                                    |
| L               | liter  |
| L               | leucine  |
| LB              | Luria-Bertani  |
| Leu             | leucine  |

|            |   |
|------------|---|
| Lys        | lysine                                      |
| m          | mili  |
| m          | methylation                                 |
| M          | molar                                       |
| M          | mega  |
| m          | multiplet                                   |
| M          | methionine                                  |
| MAT        | malonyl-CoA/acetyl-CoA transacylase         |
| Met        | methionine                                  |
| min        | minute                                      |
| MNH        | mononuclear non-heme iron                   |
| MOX        | monooxygenase                               |
| MS         | mass-spectrometry                           |
| MT         | methyltransferase                           |
| MWCO       | molecular weight cut-off                    |
| <i>m/z</i> | mass to charge ratio                        |
| N          | normal                                      |
| n          | nano  |
| N          | asparagine                                  |
| n/a        | no residue available                        |
| NADP       | nicotinamide adenine dinucleotide phosphate |
| NMR        | nuclear magnetic resonance                  |
| NOESY      | nuclear Overhauser effect spectroscopy      |
| NaPi       | sodium phosphate                            |
| NRP        | nonribosomal peptide                        |
| NRPS       | nonribosomal peptide synthetase             |
| NTA        | nitrilotriacetic acid                       |
| OD         | optical density                             |
| ORF        | open reading frames                         |
| P          | proline                                     |
| PCA        | polymerase chain assembly                   |
| PCA        | principle component analysis                |
| PCP        | peptidyl carrier protein                    |
| PCR        | polymerase chain reaction                   |
| PDCA       | pyridinecarboxylic acid                     |
| Phe        | phenylalanine                               |
| PKS        | polyketide synthase                         |
| Ppant      | 4`-phosphopantetheine                       |
| ppm        | parts per million                           |
| PPtase     | phosphopantetheinyl transferase             |
| Pro        | proline                                     |
| PS         | pyran-synthase                              |
| PSI-BLAST  | position-specific iterated BLAST            |
| Q          | glutamine                                   |
| R          | reduction                                   |
| R          | Arginine                                    |

|          |  |
|----------|--|
| RiPP     | ribosomally synthesized and post-translationally modified peptide    |
| rpm      | rounds per minute  |
| RT       | room temperature   |
| RT       | retention time   |
| s        | second   |
| s        | singulet   |
| S        | serine   |
| S.       | <i>Serratia</i>  |
| SAM      | S-adenosylmethionine   |
| SAR      | structure-activity relationship                                      |
| SDS-PAGE | sodium dodecyl sulfate polyacrylamide gel electrophoresis            |
| Ser      | serine   |
| SNAC     | <i>N</i> -acetylcysteamine   |
| sp.      | species (plural)   |
| spp.     | species (singular)   |
| Svp      | phosphopantetheinyl transferase from <i>Streptomyces verticillus</i> |
| t        | triplet  |
| T        | thiolation   |
| T        | threonine  |
| TB       | Terrific broth   |
| TCEP     | tris(2-carboxyethyl)phosphine  |
| TE       | thioesterase   |
| TEII     | type II thioesterases  |
| TEV      | tobacco etch virus   |
| TFA      | trifluoroacetic acid   |
| Thr      | threonine  |
| Trp      | tryptophan   |
| Tyr      | tyrosine   |
| u        | unit   |
| UPLC     | ultra performance liquid chromatography                              |
| V        | valine   |
| v/v      | volume/volume  |
| Val      | valine   |
| W        | tryptophan   |
| w/o      | without  |
| w/v      | weight/volume  |
| x g      | times gravity  |
| Y        | tyrosine   |
| μ        | micro  |

## 7.2 Sequence alignments



**Figure 7-1: Sequence alignment of acyltransferases (ATs) and acylhydrolases (AHs) to generate consensus sequences applied in Figure 5-2A.**

(A) Sequence alignment of different ATs. (B) Sequence alignment of different AHs. Enzymes from the following pathways were used: BaeC & BaeD, bacillaene (*Bacillus amyloliquefaciens*); BatH & BatJ, batumin/kalimantacin (*Pseudomonas fluorescens*); BryP, bryostatin (*Candidatus Endobugula sertula*); KirCl, kirromycin (*Streptomyces collinus*); OocV & OocW, oocydin (*Serratia plymthical*); PedC &

PedD, pederin (uncultivated symbiont of beetle *Paederus fuscipes*); RhiG, rhizoxin (*Burkholderia rhizoxina*); for multidomain proteins only the relevant domain is shown.

**Table 1** Crystallographically characterized mononuclear non-heme Fe<sup>II</sup> enzymes containing a 2-His-1-carboxylate facial triad

| Enzyme type                              | Crystallized enzyme (PDB code)  | Sequence motif   | Structure motif   |  |
|--|---|--|---|--|
| Extradiol cleaving catechol dioxygenases | <b>BphC</b> [or <b>DHBD</b> ] (1DHY, 1EIL, 1EIQ, 1EIR, 1KW3, 1KW6, 1KW8, 1KW9, 1KWB, 1KWC) <sup>a</sup><br>(1HAN, 1KMY, 1KND, 1KNF, 1LGT, 1LKD) <sup>b</sup>  | HX <sub>63</sub> HX <sub>50</sub> E<br>HX <sub>63</sub> HX <sub>49</sub> E<br>HX <sub>60</sub> HX <sub>40</sub> E<br>HX <sub>58</sub> HX <sub>52</sub> E   | βαβ <sub>3</sub> modules [11]   |  |
|  | <b>2,3-CTD</b> [or <b>MPC</b> ] (1MPY) <sup>c</sup><br><b>HPCD</b> (1F1X, 1Q0C, 1Q0O) <sup>d</sup><br><b>MndD</b> <sup>e</sup> (1F1R, 1F1U, 1F1V) <sup>f</sup>  | HX <sub>49</sub> HX <sub>180</sub> E   | α: plate-like; β: α/β [57]  |  |
|  | <b>4,5-PCD</b> [or <b>LigAB</b> ] (1B4U, 1BOU) <sup>g</sup>   | HX <sub>5</sub> EX <sub>2</sub> H  | Jelly roll [58]   |  |
|  | <b>HGO</b> (1EY2, 1EYB) <sup>h</sup>  | HX <sub>4</sub> HX <sub>148</sub> D  | α (N-terminus): β-sheet;<br>α (C-terminus):<br>antiparallel β-pleated<br>sheet; β: twisted mixed<br>β-sheet [17]                                    |  |
|  | <b>NDO</b> (1EG9, 1NDO, 1O7G, 1O7H, 1O7M, 1O7N, 1O7P, 1O7W) <sup>a</sup>  | HXDX <sub>53</sub> H<br>HXDX <sub>153</sub> H  | Jelly roll [15]   |  |
| α-KG dependent enzymes                   | <b>ANS</b> (1GP4, 1GP5, 1GP6) <sup>i</sup><br><b>AtsK</b> (1OIH, 1OII, 1OIJ, 1OIK, 1VZ4, 1VZ5) <sup>c</sup><br><b>TauD</b> (1GQW, 1GY9, 1OS7, 1OTJ) <sup>j</sup><br><b>CarC</b> (1NX4, 1NX8) <sup>k</sup><br><b>CAS</b> (1DRT, 1DRY, 1DS0, 1DS1, 1GVG) <sup>l</sup><br><b>DAOCS</b> (1DCS, 1E5H, 1E5I, 1HJF, 1HJG, 1RXF, 1RXG, 1UNB, 1UO9, 1UOB, 1UOF, 1UOG, 1W28, 1W2A, 1W2N, 1W2O) <sup>l</sup><br><b>FIH-1</b> (1H2K, 1H2L, (1H2M) <sup>m</sup> , 1H2N, 1IZ3, 1MZE, 1MZF) <sup>h</sup><br><b>Gab</b> (1JR7) <sup>j</sup><br><b>Proline 3-hydroxylase</b> (1E5R, 1E5S) <sup>n</sup> | HXDX <sub>77</sub> H<br>HXDX <sub>129</sub> H<br>HXDX <sub>48</sub> H  | Jelly roll [15]   |  |
|  | <b>HPPD</b> (1CJX) <sup>o</sup><br>(1SP8) <sup>p</sup> (1SP9, 1SQD, 1TFZ, 1TG5) <sup>i</sup><br>(1SQJ) <sup>q</sup><br>(1T47) <sup>r</sup>  | HX <sub>78</sub> HX <sub>81</sub> E<br>HX <sub>81</sub> HX <sub>85</sub> E<br>HX <sub>82</sub> HX <sub>82</sub> E<br>HX <sub>82</sub> HX <sub>78</sub> E   | βαβ <sub>3</sub> α modules [59]   |  |
|  | Pterin dependent hydroxylases   | <b>PheOH</b> (1DMW, 1J8T, 1J8U, 1KW0, 1LRM, 1MMK, 1MMT, 1PAH, 2PAH, 3PAH, 4PAH, 5PAH, 6PAH) <sup>h</sup> (1PHZ, 2PHM) <sup>q</sup><br><b>TrpOH</b> (1MLW) <sup>h</sup><br><b>TyrOH</b> (1TOH, 2TOH) <sup>q</sup><br><b>PheOH</b> (1LTU, 1LTV, 1LTZ) <sup>r</sup> | HX <sub>4</sub> HX <sub>39</sub> E  | Catalytic: α-helical<br>basket; Tetramerization:<br>antiparallel coiled coil<br>[60]; Regulatory: α-β<br>sandwich [61] |
|  |   | <b>PheOH</b> (1LTU, 1LTV, 1LTZ) <sup>r</sup>   | HX <sub>4</sub> HX <sub>40</sub> E  | α-helical basket [62]  |
|  |   | Other oxidases   | <b>IPNS</b> (1BK0, 1BLZ, 1HB1, 1HB2, 1HB3, 1HB4, (1IPS), <sup>l</sup> 1OBN, 1OC1, 1ODM, 1ODN, 1QIQ, 1QJE, 1QJF) <sup>u</sup><br>(1UZW) <sup>v</sup> | HXDX <sub>53</sub> H   |
|  | <b>ACCO</b> <sup>w</sup>  |  | HXDX <sub>24</sub> H  | Jelly roll [20]  |

α-KG α-ketoglutarate, *BphC* and *DHBD* 2,3-dihydroxybiphenyl 1,2-dioxygenase, *2,3-CTD* catechol 2,3-dioxygenase, *MPC* meta-pyrocatechase, *HPCD* homoprotocatechuate 2,3-dioxygenase, *MndD* Mn<sup>II</sup>-dependent HPCD, *4,5-PCD* and *LigAB* protocatechuate 4,5-dioxygenase, *HGO* homogentisate 1,2-dioxygenase, *NDO* naphthalene 1,2-dioxygenase, *ANS* anthocyanidin synthase, *AtsK* alkylsulfatase, *TauD* taurine/α-KG dioxygenase, *CarC* carbapenem synthase, *CAS* clavaminic synthase, *DAOCS* deacetoxycephalosporin C synthase, *FIH-1* factor-inhibiting hypoxia-inducible factor-1, *HPPD* 4-hydroxyphenylpyruvate dioxygenase, *PheOH* phenylalanine hydroxylase, *TrpOH* tryptophan hydroxylase, *TyrOH* tyrosine hydroxylase, *IPNS* isopenicillin N synthase, *ACCO* 1-aminocyclopropane-1-carboxylic acid oxidase

<sup>a</sup>From *Pseudomonas* sp. <sup>b</sup>From *Burkholderia* sp. <sup>c</sup>From *Pseudomonas putida*. <sup>d</sup>From *Brevibacterium fuscum*. <sup>e</sup>Mn<sup>II</sup>-dependent. <sup>f</sup>From *Arthrobacter globiformis*. <sup>g</sup>From *Sphingomonas paucimobilis*. <sup>h</sup>From *Homo sapiens*. <sup>i</sup>From *Arabidopsis thaliana*. <sup>j</sup>From *Escherichia coli*. <sup>k</sup>From *Erwinia carotovora*. <sup>l</sup>From *Streptomyces clavuligerus*. <sup>m</sup>Zn<sup>II</sup> complex. <sup>n</sup>From *Streptomyces* sp. <sup>o</sup>From *Pseudomonas fluorescens*. <sup>p</sup>From *Zea mays*. <sup>q</sup>From *Rattus norvegicus*. <sup>r</sup>From *Streptomyces avermitilis*. <sup>s</sup>From *Chromobacterium violaceum*. <sup>t</sup>Mn<sup>II</sup> complex. <sup>u</sup>From *Aspergillus nidulans*. <sup>v</sup>From *Emericella nidulans*. <sup>w</sup>From *Petunia hybrida*. The PDB files corresponding to the ACCO structures were not available when this article was published.

**Figure 7-2:** Table of crystallized MNH enzymes and their corresponding sequence motifs and folds taken from KoehnTOP, Emerson, and Que 2005.

Jelly roll structure motifs and the sequence motif closest to the OocP<sub>4Rx13</sub> motif are boxed.