Structural and functional studies of cyclic depsipeptide biosynthesis

Diego A. Alonzo Muñiz
Department of Biochemistry
McGill University
Montréal, Québec, Canada
April 2019

A thesis submitted to McGill University in partial fulfilment of the requirements for the degree of Doctor of Philosophy

"Good science is no rosebed, but the romance is still there. The thrill of discovery outweighs the drudgery, the despair at one's inadequacy, the fight for financial support, the setbacks and mistakes, the long hours and the nagging fear of being overtaken. A discovery is like falling in love and reaching the top of a mountain after a hard climb all in one, an ecstasy induced not by drugs but by the revelation of a face of nature that no one has seen before, and that often turns out to be more subtle and wonderful than anyone had imagined. A true scientist derives this feeling not only from his own discoveries, but also from those of his colleagues."

Max Perutz*, "True Science", review to Advice to a Young Scientist by Peter Medawar,* The London Review of Books, March 1981

To my family and friends

Abstract

Cyclic depsipeptides, ubiquitous natural products of industrial and medical importance, are produced by nonribosomal peptide synthetases (NRPS) using a diverse repertoire of biosynthetic mechanisms. For instance, cereulide (Ces) and valinomycin (Vlm) depsipeptide synthetases generate a linear precursor by the alternate condensation of hydroxy and amino acids. Hydroxy acids are added by specialized modules consisting of adenylation (A), ketoreductase (KR) and peptidyl carrier protein (PCP) domains, while an iterative thioesterase (TE) domain catalyzes the oligomerization/cyclization of the linear precursor to release the mature product.

In order to perform a biochemical characterization of the specialized modules that add hydroxy acids in the Ces system, I expressed and purified its intact components, performed biochemical assays on the relevant domains and reconstituted the assembly of the linear depsipeptide precursor *in vitro*, showing that the 15 domains in the system were active and working in concert.

A and KR domains adenylate and reduce α-ketoacids to generate the constituting α-hydroxy acids in cyclic depsipeptides. I and collaborators determined a di-domain A-KR structure in the adenylation conformation, showing how a novel structural element connects the KR and PCP domains while participating in domain swapping. I also determined the high-resolution structure of an excised A domain bound to the product of its first half-reaction, demonstrating how the ketone group interacts with the A domain for the first time.

Finally, I determined the structures of Vlm TE bound to intermediates of the oligomerization and cyclization reactions, using a genetic code expansion strategy and substrate analogues developed by our collaborators. This work presents near-native TE-substrate structures for the first time and provides information on the role of the active site lid, a structural element present in all TE domains. Furthermore, these results serve as a proof of concept for the genetic code expansion strategy, opening the possibility of its use in other biological systems.

In general, this dissertation provides novel information on depsipeptide biosynthesis, which could be extrapolated to the machinery producing other depsipeptides such as the piscicide antimycin, the anticancer agent cryptophycin and the antifungal kutzneride.

Resumé

Les depsipeptides cycliques sont des produits naturels d'importance industrielle et médicale générés par des enzymes de synthèse de peptides non ribosomiques (NRPS), qui emploient un vaste répertoire de mécanismes biosynthétiques. Par exemple, les synthétases des depsipeptides céreulide (Ces) et valinomycine (Vlm) produisent des précurseurs linéaires par la condensation alternée d'acides aminés et d'acides hydroxylés. Ces derniers sont incorporés par des modules spécialises contenant des domaines d'adénylation (A), cétoréductase (KR) et protéine porteuse de peptide (PCP). Un domaine thioestérase itératif (TE) catalyse ensuite la cyclisation ou l'oligomérisation du précurseur linéaire et libère le produit final.

Afin de caractériser au niveau biochimique les modules spécialises qui incorporent les acides hydroxylés dans le système Ces, j'ai exprimé et purifié ses composantes intactes, effectué des analyses sur les domaines en question, et reconstitué l'assemblage du précurseur de depsipeptide linéaire in vitro, démontrant que les quinze domaines du système sont actifs et agissent de concert.

Les domaines A et KR adénylent et réduisent les acides alpha-cétoniques, produisant les acides alpha-hydroxylés présents dans les depsipeptides cycliques. Avec mes collaborateurs, j'ai déterminé la structure de la portion A-KR adoptant la conformation d'adénylation, illustrant la connexion des domaines KR et PCP par un nouvel élément structurel engagé dans l'échange de domaines. J'ai aussi obtenu la structure à haute résolution du domaine A lié au produit de sa première demi-réaction, démontrant pour la première fois comment le groupe cétone interagit avec ce domaine.

Enfin, j'ai déterminé les structures du domaine TE de Vlm lié aux intermédiaires des réactions d'oligomérisation et de cyclisation, en employant une approche d'expansion du code génétique et des analogues de substrats, développée par mes collaborateurs. Cette étude est la première

visualisation de la structure quasi-native de TE avec substrats, nous informant du rôle du « couvercle » du site actif – un élément structurel présent dans tous les domaines TE. De plus, ces résultats démontrent la faisabilité de l'approche d'expansion du code génétique, offrant la possibilité de son application à d'autres systèmes biologiques.

En résumé, cette thèse étend nos connaissances de la synthèse des depsipeptides. Ces nouvelles informations pourraient être applicables aux machines moléculaires produisant d'autres depsipeptides tels le piscicide antimycine, l'agent anticancéreux cryptophycine et l'agent antifongique kutzneride.

Table of contents

List of Figures and Tables

List of abbreviations

α-KIC: α-ketoisocaproic acid, substrate of CesA1 and StrA1.
α-KIV: α-ketoisovaleric acid, substrate of CesB1 and Vlm1-1.
α-KILE: α-ketoisoleucine, somewhat selected by CesB1.
α-HIC: α-hydroxyisocaproic acid, product of reduction of α-ketoisocaproic acid.
α-HIV: α-hydroxyisovaleric acid, product of reduction of α-ketoisovaleric acid.
α-HIL: α-hydroxyisoleucine, product of reduction of α-ketoisoleucine.
A domain: adenylation domain, catalyzes selection and adenylation of acyl monomers.
A_{core}: N-terminal subdomain of the adenylation domain.
ANL: ANL superfamily of adenylating enzymes.
Ant: antimycin synthetase
A_{sub}: C-terminal subdomain of the adenylation domain.
C domain: condensation domain, catalyzes amide or ester bond formation.
CDA: calcium-dependent antibiotic.
Ces: cereulide synthetase, produces the emetic toxin cereulide
CesA: one of the two polypeptide chains that constitute Ces.
CesB: one of the two polypeptide chains that constitute Ces.
CoA: coenzyme A
Crp: cryptophycin synthetase
C_T domain: terminal condensation domain, catalyzes chain release.
Cy domain: heterocyclization domain, catalyzes oxazoline, thiazoline and methyloxazoline formation.
DAP: 2,3-diaminopropionic acid, unnatural amino acid used in chapter 4.
E domain: epimerization domain, catalyzes L- to D- epimerization.
Ent: enterobactin synthetase.
F domain: formylation domain, catalyzes formyl transfer.
Grs: gramicidin synthetase.
KR domain: ketoreductase domain, catalyzes reduction of a- or b- keto groups.
Ktz: kutzneride synthetase.
Lgr: linear gramicidin synthetase
LC-MS: liquid chromatography-mass spectrometry.
M domain: methylation domain, can catalyze N-, O-, S-, or C- methylation.
NRP: nonribosomal peptide.
NRPS: nonribosomal peptide synthetase.
Ox domain: oxidation domain, catalyzes oxidation of thiazolines.
PCP domain: peptidyl carrier protein domain, serves as attachment point of phosphopantetheine.
PKS: polyketide synthetase. Assembly-like megaenzymes that produce polyketides.
Ppant: phosphopantetheine, serves as a covalent attachment point for monomers and growing chain.
PPi: inorganic pyrophosphate.
PPTase: phosphopantetheinyl transferase, catalyzes attachment of ppant to PCP domains.
R domain: reductase domain, catalyzes chain release.
SDR: short chain dehydrogenase
SFP: PPTase from B. subtilis.
Srf: surfactin synthetase.
StrA: first module (A-KR-PCP) of a Ces homologue in *Bacillus stratosphericus*.
TE domain: thioesterase domain, catalyzes chain release.
TLC: thin layer chromatography
Vlm: valinomycin synthetase, produces the K+ ionophore valinomycin.

Preface

This is a manuscript-based thesis consisting of two published articles and one manuscript in preparation. Part of the introduction is in preparation to be published as a review in Protein Science.

Chapter 2

Diego A. Alonzo, Nathan A. Magarvey, T. Martin Schmeing. (2015) Characterization of cereulide synthetase, a toxin-producing macromolecular machine. PlosOne; 10(6): e0128569

Chapter 3

Diego A. Alonzo*, Clarisse Chiche-Lapierre*, Jimin Wang, Michael J. Tarry, Martin Schmeing. Structural and functional studies of a module that adenylates and reduces alpha-keto acids. Manuscript in preparation.

Chapter 4

Nicolas Huguenin-Dezot*, Diego A. Alonzo*, Graham W. Heberlig, Mohan Mahesh, Duy P. Nguyen, Mark H. Dornan, Christopher N. Boddy, T. Martin Schmeing, Jason W. Chin. Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid. (2019) Nature, 565, pages 112–117

*denotes co-author

Contributions of Authors

Chapter 2

I performed all the experiments, except for LC-MS runs which were run by Thierry Ntimbane and Gaëlle Bridon at the GCRC Metabolomics Core Facility, and by Alexander Wahba in the Mass Spectroscopy Facility of the Faculty of Chemistry, both at McGill University. Nathan Magarvey provided the plasmids for CesA and CesB. T. Martin Schmeing and I designed the experiments and wrote the manuscript with input from Nathan Magarvey.

Chapter 3

I designed and sub-cloned the constructs, optimized expression conditions and purification protocols and performed all the biochemical experiments, except for cloning and assays of the CesA A10-Lys mutants, which were performed by Michael J Tarry. Together with Clarisse Chiche-Lapierre, we equally performed purification, crystal optimization and X-ray crystallography data collection. I obtained initial phases for the A-KR structure, while Jimin Wang, T. Martin Schmeing and me, determined the A-KR structure. I determined the adenylate-bound A domain structure. T. Martin Schmeing conceived the study. I wrote the manuscript with input from all authors.

Chapter 4

Nicolas Huguenin-Dezot (NHD) and Duy P. Nguyen (DPN) selected the synthetases for **2-6**. NHD characterized the incorporation of **6** and its deprotection to DAP and performed the TEV conjugation. Mohan Mahesh (MM) synthesized **2-6**. T. Martin Schmeing (TMS) and I designed, and I performed the biochemistry and crystallography with Vlm TE_{wt}. Together with NHD, we equally performed biochemistry and crystallography with Vlm TE_{DAP}. Graham W. Heberlig and Mark H. Dornan synthesized Vlm TE substrates. Jason W. Chin (JWC) supervised NHD, DPN and MM. TMS supervised DAA. Christopher N. Boddy supervised GWH and MHD. I and TMS, NHD, and JWC wrote the paper with input from all authors.

Original Contributions of Knowledge

In **chapter 2**, I expressed and purified an intact depsipeptide synthetase (Ces), performed the kinetic characterization of each A domain in the system, showed that the A domains do not require MbtH-like proteins for activity, assessed the substrate selectivity of ketoacid-selecting A domains, evaluated KR-NADPH binding and reconstituted the biosynthesis of the linear tetradepsipeptide intermediate.

In **chapter 3**, I and collaborators solved the crystal structure of an A-KR di-domain construct for the first time, showing the relative positioning of the A and KR active sites and the presence of a novel structural element that participates in domain swapping, the pseudo A_{sub} domain. I determined the high-resolution structure of the excised A domain that provides novel information on keto acid selectivity, showing that the α-ketone interacts with the protein through a carbonyl-carbonyl interaction. Finally, biochemical assays demonstrate that A-KR-PCP catalytic activity is independent of its oligomeric state and that the monomer may be the functional state of the protein.

In **chapter 4**, I solved the X-ray structure of the iterative-oligomerizing TE domain from valinomycin biosynthesis (Vlm TE), showing that its active site lid is larger and more complex than most TE structures solved to date. Through biochemical assays, I demonstrated that Vlm TE uses analogues of the tetradepsipeptidyl-*S*-PCP intermediate, catalyzing oligomerization and cyclization to generate the final product, valinomycin. The assays demonstrated that the reaction proceeds through a specific interplay of transesterification reactions between the ppant thiol and the TE active site Ser. In the same chapter, through a collaboration with Nicolas Huguenin-Dezot, we replaced the active site Ser by diamino propionic acid (DAP) through a genetic code expansion system developed by Nicolas et al. in Jason Chin's group at the LMB. Together, we formed stable tetradepsipeptidyl- and dodecadepsipeptidyl-TE complexes and solved their crystal structures, showing that a massive conformational transition of the active site lid creates a cavity that favours cyclization of the bound dodecadepsipeptide.

Acknowledgements

I was fortunate enough to work with outstanding scientists from all over the world during this dissertation. Without their eagerness, perseverance and contagious drive, this work would not exist. First, I would like to thank my supervisor, Dr. Martin Schmeing (Martin), for his continuous support during my Ph.D. His endless efforts to provide his students with the environment to do amazing science and his eagerness to constantly improve as a supervisor are outstanding. I feel fortunate to have worked with someone that empathizes with the struggles of day-to-day science, and who was always there to provide advice during the multiple roadblocks that one could face when doing structural biology. I would also like to thank my RAC committee members Dr. Albert Berghuis and Dr. Youla Tsantrizos for their advice and guidance.

Chapters 3 and 4 were team efforts, product of the hard work of several people. In chapter 3, Clarisse Chiche-Lapierre, Jimin Wang and Michael Tarry. In chapter 4, Nicolas Huguenin-Dezot, Jason Chin, Christopher Boddy, Graham Heberlig, Mohan Mahesh, Duy P. Nguyen, and Mark Dornan. I want to further thank Nicolas Huguenin-Dezot for being an extraordinary collaborator in chapter 4. It was extremely refreshing to do science while having fun and to share the thrills of scientific discovery. Thanks for always believing in the worth of our efforts. Also, special recognition goes to Dr. Jimin Wang, whose astronomical knowledge in X-ray crystallography allowed full determination of the A-KR structure. Thanks to Clarisse Chiche-Lapierre for her hard work, perseverance, organization and valuable contributions as a collaborator for chapter 3.

Data collection was performed in multiple research facilities. I want to give special recognition to all the people who keep synchrotron beam lines running smoothly and who provide amazing advice during X-ray crystallography data collection. A good part of the success of this dissertation is a consequence of their day-to-day efforts. CLS-CMCF: S. Labiuk, J. Gorin, M. Fodje, K. Janzen, D. Spasyuk and P. Grochulski; APS: Frank Murphy and Surajit Banerjee. Thanks to Dr. Alex Wahba from the Department of Chemistry and Drs. Gaëlle Bridon, Thierry Ntimbane, and Daina Avizonis in the GCRC Metabolomics Core Facility who provided support in running and analyzing MS data.

Thanks to Alexei Gorelik for his advice during these years and for translating the abstract to French.

I extend my gratitude to CONACYT, MELS-SRE, FQRNT-PBEEE, CIHR Strategic Training Program in Chemical Biology, GRASP and the Faculty of Medicine for providing financial support.

To succeed, science requires a good environment. Thanks to all current and past members of the Schmeing group for providing a great place to do science and for their support. Also, thanks for all the good times inside and outside the lab! Mike, for being a great friend, and for keeping the lab running smoothly and for the advice and contributions to the A-KR-PCP project. Janice, for

being an extraordinary colleague and a superb friend, whose tenacity and work ethic are examples for all scientists in the field (also thanks for the amino CoA used in chapter 3!). Camille, for all the new vocabulary, dinners at le Korean and for being a positive example of scientific perseverance and achievement. Frederik, for the enthusiastic conversations on science, science fiction and fiction. Clarisse, for being an amazing collaborator in Chapter 3. Max, Kris, Angelos, Asfarul and Itai for being outstanding and supportive colleagues. Flipper Kapensis for making lab fun trips even more entertaining. Several people from the 4th floor Bellini also contributed to a great research environment but it would take too long to mention them all: Juli por estar ahí en momentos alegres y en momentos difíciles, siempre con una sonrisa para seguir adelante, Dmitry Rodionov for keeping equipment functional and for his technical advice during the time we overlapped at McGill. I want to mention that Mike and Juli not only are exceptional scientists, but also amazing friends who always opened their home to celebrate Christmas when I stayed in Montreal during the Holidays.

As we all scientists know, a PhD is an emotional rollercoaster, with ups and downs of slopes of varying magnitudes. For me, in addition to the support from my supervisor and colleagues, the key to continue was to always remember the good things that happen outside of the lab (along with full-volume renditions of “The show must go on” by Queen):

Gracias a Rosa por ser una gran amiga y por crear un hogar juntos durante buena parte del doctorado. Eres extraordinaria y tu apoyo incondicional lo recordaré por siempre.

Merci Eric, pour tous les bons moments, le bonheur, les repas partagées et pour être là pendant les premieres années a Montreal. Gracias por tu afecto, por tu apoyo, risas, sabiduría y enseñanzas.

Alejandro, estaré agradecido de por vida por enseñarme que no hay palabras para describir el significado que puede tener bailar. Asimismo, gracias por ser un gran amigo. Te quiero mucho

A la comunidad tanguera de Montreal y del mundo, gracias por existir.

Thanks Ruth and Mathew for your support and for your words of wisdom. My best wishes for your future.

Itza, Islandia y Martha, fue genial estar felices los cuatro. Geniales los mezcales con naranjas (Islandia) y los dates de pizza y cine (Martha). Y a Itza, gracias por las tandas y por poner en cada verso tu corazón.

Gracias a Elena por tu amistad, tu entereza y tu sapiencia. Y por supuesto, por todas las comidas y las jocosidades que compartimos.

Gracias a Paco y Liz por “abrir una sucursal de su cariño”, como dice Mafalda, en Québec y en Riviére du Loup.

Gracias a Alejandro Freyermuth, Uriel, Best, Juan y Beto por ser amigos de los que están.

Thanks to the people of Montreal. Thanks for making life interesting and worth living. Thanks for all the joy of the good times and for all the lessons from the not-so good times. Thanks for the wine and the food, and for the beauty of nature and the possibility of living in peace.

Agradezco a la señora Francesca por la mejor comida italiana de la isla.

Our work as scientists would not be possible without the tax payers of Canada.

Gracias al pueblo de México.

Finalmente, sin mi familia esto no hubiera sido posible. Gracias a Mima, Pau, Frida y Ernesto por su cariño de toda una vida y por haber hecho el viaje durante el invierno y pasarla fantástico. Gracias a Mamolga y abuelita Nidia por su sabiduría y fuerzas enviadas a distancia. A Olgui y Vale por su amorosa hospitalidad en mis regresos a México. Por los desayunos, cenas, reuniones en la Península y mucho, pero mucho amor.

Ma, pa y Rod. Llegó el momento. Gracias por apoyarme durante todo este tiempo, por escucharme y por ser pacientes. Por cuidar de ellaaa. Gracias por extender su amor a Canadá, y por siempre darme momentos de alegría al regresar a casa en México. Un inmenso, enorme agradecimiento por su amor incondicional y por tantas risas. Esas risas que son las que nos hacen seguir. Como diría Cerati: gracias totales. Los amo.

CHAPTER 1. INTRODUCTION

1.1. Nonribosomal peptide synthetases and their products

Nonribosomal peptide synthetases (NRPSs) are multi-domain microbial enzymes that generate natural products with medical and industrial importance (Felnagle, Jackson et al. 2008). Well known examples of nonribosomal peptides (NRPs) include the antibiotics penicillin, gramicidin and daptomycin, as well as the antifungal caspofungin, the immunosuppressor ciclosporin and the emulsifier surfactin (**Fig. 1.1**).

NRPs are secondary metabolites, a group of organic compounds that include unusual sugars, terpenoids, alkaloids, flavonoids, polyketides and ribosomal peptides (Demain and Fang 2000). The production of secondary metabolites is not essential for organism survival; however, they have functions in growth, communication and reproduction of the producing organism (Demain and Fang 2000). All domains of life make secondary metabolites, but the most prolific producers of NRPs are members of the bacterial phyla *Proteobacteria*, *Actinobacteria*, *Firmicutes* and *Cyanobacteria*, and the fungal phylum *Ascomycota (Wang, Fewer et al. 2014)*. For example, among the actinobacteria, the genus *Streptomyces* accounts for production of around 50% of the known antibiotics (Demain and Fang 2000).

The natural function of NRPs and other secondary metabolites is to confer an advantage to the producing microbe in its ecological niche, for example by killing competitors, modulating symbiont interactions or by improving nutrient uptake (van der Meij, Worsley et al. 2017). Symbiosis and chemical warfare modulate the relation of the producing microbe with other microbes, plants, nematodes, insects and higher animals (Demain and Fang 2000).

Genes involved in NRP production are very often arranged in biosynthetic gene clusters whose expression is regulated through different mechanisms (van Wezel and McDowall 2011, Brakhage 2013). The production of the metabolite by their natural producer in laboratory conditions is challenging (van Wezel and McDowall 2011, Brakhage 2013), making it difficult to adjudicate a specific role to each secondary metabolite. However, some natural functions have been assigned

in a number of cases, either by co-culture of the producing organism with other microbes, by drawing conclusions from the physicochemical properties of the metabolite, or by a combination of both. For instance, the *Escherichia coli* siderophore enterobactin can scavenge iron with extremely high affinity (Raymond, Dertz et al. 2003) indicating that it has a role in providing *E. coli* with iron, whereas production of the K^+ ionophore cereulide increases fitness of *Bacillus cereus* in K^+-deprived environments (Ekman, Kruglov et al. 2012). In another example, co-cultivation of *Bacillus subtilis* surfactin producer was found to have effects on the development of co-cultured *Staphylococcus aureus (Gonzalez, Haste et al. 2011)*.

A striking and very interesting case of the natural function of NRPs is the role of actinomycin, valinomycin and antimycin in a triangle of symbiotic relationships between *Streptomyces sp.*, a fungus and leaf-cutter ants (Schoenian, Spiteller et al. 2011). Leaf cutter ants grow gardens of the fungus *Leucoagaricus gongylophorus*, whose growth is sustained by decaying vegetal matter in dedicated nest chambers. The ants feed off this fungus, making it essential to keep it clean of any other microorganism. Schoenian and coworkers found that actinomycin, valinomycin and antimycin are present in the nests or even the bodies of these ants, and that they can inhibit the growth of other microorganisms that would contaminate the fungal gardens. The authors suggest that these small molecules participate in the regulation of the microbial ecosystem in the ant's nest. Even more surprising, they found that valinomycin is present on the body of the ant, and furthermore that its concentration is different in each ant body part.

NRPs can be cyclized, methylated, formylated or even be coupled to fatty acids and polyketides, making them one of the most chemically diverse natural products (Konz and Marahiel 1999).

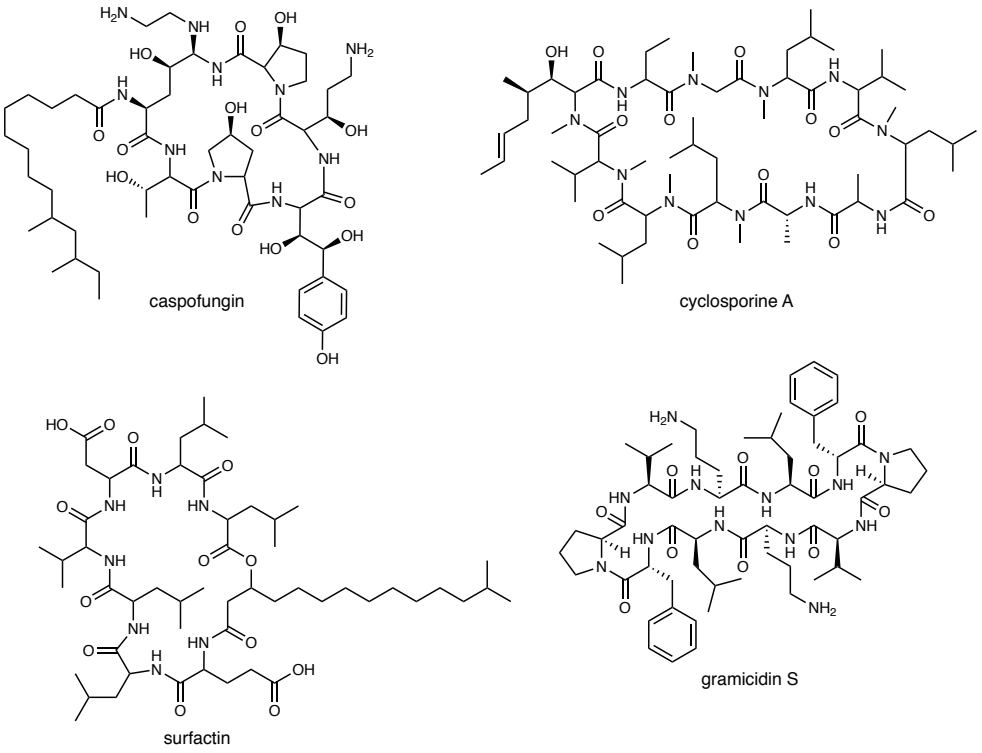


Figure 1.1. Examples of nonribosomal peptides.

Initial evidence that the ribosome was not responsible for the production of NRPs came from elegant experiments in the 1960s (Felnagle, Jackson et al. 2008). A study from 1963 (Mach, Reich et al. 1963) discovered that *Bacillus brevis* could still generate tyrocidine even when ribosomes were stalled with ribosome-binding antibiotics, suggesting involvement of an alternative macromolecular machine. In the following years, studies showed similar results for other unusual peptides. Microbial cells still made gramicidin (Tomino, Yamada et al. 1967) and polymyxin B (Daniels 1968) when treated with ribosome-targeting antibiotics or RNAses. The seminal paper that showed direct involvement of enzymes in the biosynthesis of these peptides came in 1968 (Gevers, Kleinkauf et al. 1968), using the gramicidin S system from *B. brevis* as a model. Lipmann and colleagues showed that purified protein fractions from a ribosome-free cell lysate produced gramicidin S in the presence of free amino acids and ATP. This was the first NRPS system reconstituted *in vitro*. From then until now, more research on purified NRPS systems has revealed the elegant biosynthetic mechanism of peptide formation, along with detailed structural studies that have given insight into these molecular mechanisms at an atomic level.

1.2. The biosynthetic cycle of NRPSs.

NRPSs form peptides through a thiotemplated mechanism that resembles an assembly line, based on the arrangement of their domains into modules, functional units that contain all the catalytic domains needed to add one acyl residue to the peptide chain (Schwarzer, Finking et al. 2003, Gulick 2016, Reimer, Haque et al. 2018). An elongation module is formed by an adenylation (A), a peptidyl carrier protein (PCP) and a condensation (C) domain, in the arrangement C-A-PCP (**Fig. 1.2a**). The modules at the N- and C- terminus of the assembly line are called initiation and termination modules respectively, with the minimal architectures A-PCP (initiation) and C-A-PCP-[*termination domain*] (termination). An elongation cycle (**Fig. 1.2b**) starts by the adenylation of an acyl monomer in the A domain. Thiolation occurs next, when the adenylate is attacked by the sulfhydryl group in a 4'-phosphopantheteinyl (ppant) moiety attached to the PCP domain, forming a thioester bond and therefore a covalent linkage between the monomer and the enzyme. The acyl thioester is transported to the C domain, which will catalyze peptide transfer by amide bond formation between the growing peptidyl chain from the preceding module (donor) and the recently formed acyl thioester (acceptor). Further elongation steps take place along the assembly line until a termination module releases the fully elongated chain through different mechanisms depending on the kind of C-terminal termination domain. For example, thioesterase domains (TE domains) release the chain mainly through hydrolysis or cyclization (Horsman, Hari et al. 2015), while reductase (R domains) catalyze thioester to alcohol reductive release (Manavalan, Murugapiran et al. 2010).

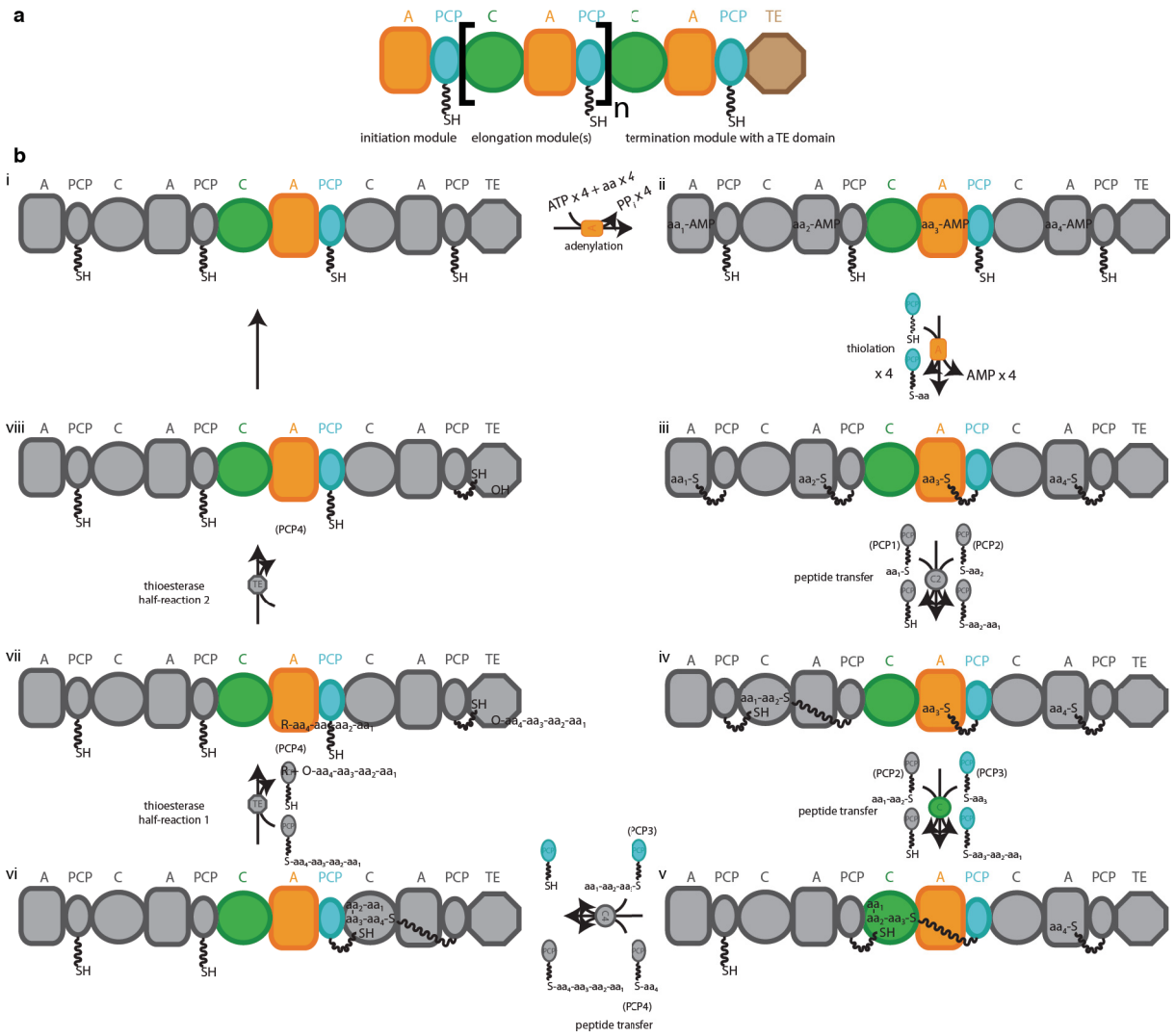


Figure 1.2. NRPSs and their biosynthetic cycle.

a, Organization of a typical NRPS. **b**, The biosynthetic cycle of NRPSs. Adapted from (Huguenin-Dezot, Alonzo et al. 2019)

The number of modules in an NRPS generally corresponds to the number of monomers in its final product, following a so-called "collinearity rule". For example, a dipeptide would be produced by a 2-module NRPS and a dodecapeptide by a 12-module NRPS. There are numerous exceptions to the collinearity rule, such as iterative (Hoyer, Mahlert et al. 2007) and module skipping mechanisms (Wenzel, Kunze et al. 2005) that can produce peptides of a different length than the one predicted from the number of modules in the biosynthetic cluster. The average size of a single module is ~1000 amino acids, and an NRPS can contain up to 18 modules in a single

polypeptide chain. Modules can also be distributed in different polypeptide chains that interact through protein-protein interactions.

NRPSs follow more exceptions than rules concerning their biosynthetic cycle. Building blocks are not restricted to the 20 proteinogenic amino acids. Instead, A domains can use ~500 different monomers, allowing NRPs to contain non-proteinogenic amino acids, fatty acids, α-hydroxy acids and many other acyl moieties (Caboche, Leclere et al. 2010). NRPSs can achieve even more chemical diversity in their products by incorporating tailoring domains in their modules (Walsh, Chen et al. 2001), including methylation (M), ketoreductase (KR), heterocyclization (Cy), formylation (F), oxidation (Ox) and reductase (R) domains. Additionally, off-assembly line enzymes can also act as tailoring domains (Chen and Walsh 2001), modifying the peptide chain during elongation or after chain release. Hybrid systems that interact with polyketide synthetases (PKSs) or fatty acid metabolism can yield mixed PKS-NRP (Muller, Garcia-Gonzalez et al. 2014) or fatty acyl-NRP products (Kraas, Helmetag et al. 2010), further increasing chemical diversity.

Each of the NRPS core reactions has been thoroughly characterized, ranging from bioinformatic studies (Skinnider, Dejong et al. 2015) to structural determinations of intact modules (Tanovic, Samel et al. 2008) (Miller, Drake et al. 2016). These advances will be discussed in the following sections.

1.2.1. Peptidyl carrier protein (PCP) domains

Peptidyl carrier proteins (PCP domains) participate in almost every reaction of the NRPS biosynthetic cycle (Weissman 2015). Following the assembly line analogy, PCP domains act as a transport arm, using a ppant moiety attached to a serine residue to move monomers from one catalytic unit to the next. To do this, the thiol group in ppant attacks high-energy intermediates at different stages of the cycle, moving the resulting thioesters to other domains to participate in downstream reactions.

The PCP domain structure is a bundle of 4 α-helices (Lohman, Ma et al. 2014). Helix number 2 bears a GGHS motif, where the Ser serves as an attachment point for ppant in a transfer reaction catalyzed by a phosphopantetheinyl transferase (PPTase) using coenzyme A as co-substrate. Most biosynthetic clusters encode a gene for the PPTase responsible of PCP priming. Some of these PPTases, like Sfp from *Bacillus subtilis* (Quadri, Weinreb et al. 1998), are highly promiscuous and can modify PCP domains from different biosynthetic clusters, a feature that has been exploited as a biochemical tool for NRPS studies.

The thiol in the ppant arm in PCP domains must be acyl-free to act as a nucleophile in NRPS reactions. However, there are several species of CoA with substitutions in the SH group in cells, mainly acetyl CoA, which can be loaded to PCP domains. These unproductive acyl-*S*-PCP species could be a potential bottleneck during nonribosomal peptide production (Schwarzer, Mootz et al. 2002). Therefore, specialized enzymes called type-II thioesterases edit mis-primed PCP domains, restoring them to their free-thiol state and allowing the biosynthetic reactions to proceed (Schwarzer, Mootz et al. 2002).

1.2.2. Adenylation and thiolation

Adenylation domains (A domains) catalyze the two half-reactions that start a NRPS biosynthetic cycle (Gulick 2009) (**Fig. 1.3**). The first half-reaction (**Fig. 1.3a**). is adenylation of a free acyl monomer using ATP•Mg^{2+} as a co-substrate, yielding PP_i and an acyl-adenylate as products. The second half-reaction, thiolation, relies on a crosstalk between the A domain and its downstream PCP domain. The sulfhydryl in the ppant arm attacks the carboxylate carbon in the high-energy adenylate, forming an acyl thioester intermediate and AMP.

A domains are 60 kDa (~550 residues) Rossmann-like fold proteins (**Fig. 1.3b**) that belong to the ANL superfamily of adenylating enzymes (ANL), which also encompasses luciferases, acyl- and aryl- CoA ligases (Gulick 2009). The A domain structure consists of a core domain of approximately 50 kDa (A_{core}, **Fig. 1.3b,** orange) and a small sub-domain of approximately 10 kDa (A_{sub}, **Fig. 1.3b,** yellow) (Conti, Stachelhaus et al. 1997, Yonus, Neumann et al. 2008). Comparison

of the first NRPS A domain structure (PDB ID: 1amu) with structural and functional studies of other ANL members revealed 10 consensus sequence motifs (A motifs) which play different roles in A domain function (Stachelhaus, Mootz et al. 1999). The A_{core} and A_{sub} domains are connected through a small linker formed by motif A8.

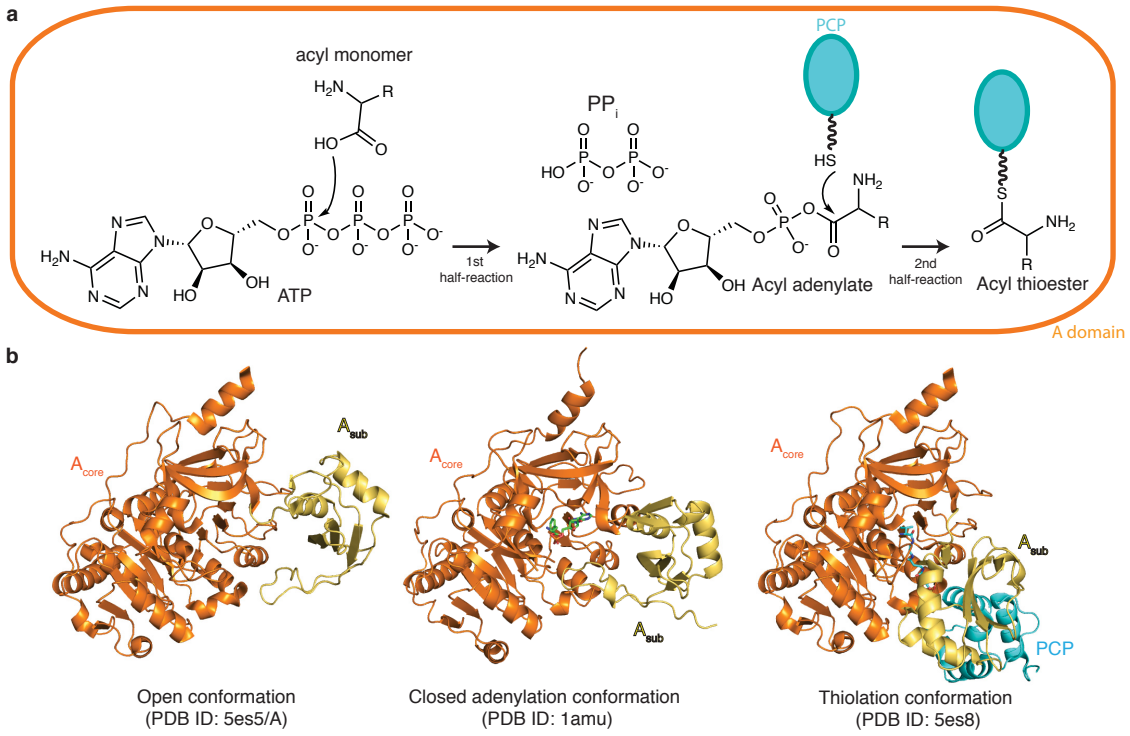


Figure 1.3. Adenylation.

a. Adenylation domain half-reactions. **b**. Adenylation domain structures.

Monomers bind to a pocket formed predominantly by residues in the A_{core} domain, with essential contributions from two residues in the A_{sub} domain. A_{core} motif A4 contains an Asp residue (Asp235 based on GrsA-Phe1 structure numbering) that is located at the entrance of the binding pocket and interacts with the α-amino group during amino acid binding (Conti, Stachelhaus et al. 1997). Asp235 is invariant in amino acid-selecting A domains but is replaced by other residues in A domains that select acyl monomers with other α-groups, such as α-hydroxy acids (Xu, Orozco et al. 2008), α-keto acids (Magarvey, Ehling-Schulz et al. 2006) or 2,3-dihydroxybenzoate (May, Kessler et al. 2002). Regardless of the nature of the α-group, all selected acyl monomers interact through their carboxyl group with an invariant lysine residue in motif A10 (A10-Lys) from the A_{sub}

domain. This residue also makes hydrogen bonds with the linking oxygen in ATP and the oxygen in the ribose ring and its mutation leads to a severe loss of adenylation activity (Hamoen, Eshuis et al. 1995).

A group of variable amino acids lining the monomer binding pocket are involved in substituent selectivity. This selectivity code has been coined as the Stachelhaus code, based on the paper that first proposed it (Stachelhaus, Mootz et al. 1999). Informatic prediction tools based on this initial analysis and on genome mining allow *in silico* monomer prediction based on sequence (Bachmann and Ravel 2009, Prieto, Garcia-Estrada et al. 2012). As more A domain structures and biochemical characterizations are revealed, the predictive power of these algorithms improves. However, these tools remain inaccurate in the prediction of non-amino acid monomers, as structural and biochemical data on these systems is more limited.

The A_{sub} domain undergoes massive conformational changes during the course of the two half-reactions (Gulick 2009) (Reimer, Haque et al. 2018) (**Fig 1.3b**). In the absence of substrates, the A_{sub} domain is positioned in an open conformation (**Fig 1.3b, left**) relative to the A_{core}, allowing the substrate and ATP to bind to their respective active sites (Conti, Franks et al. 1996) (Reimer, Aloise et al. 2016). The A_{sub} domain closes (**Fig 1.3b, middle**) after acyl monomer and ATP binding, interacting through A10-Lys with the ATP-monomer complex (Conti, Stachelhaus et al. 1997). The closed conformation is rotated by 90° relative to the substrate-free open conformation.

After adenylate formation, the A_{sub} domain undergoes another conformational change that facilitates the second half-reaction, thiolation (**Fig 1.3b, right**). The thiolation conformation in the ANL superfamily was first revealed in CoA ligase structures bound to CoA and non-hydrolyzable adenylate analogues (Gulick, Starai et al. 2003), showing that A10-Lys is rotated away from the active site by 25° and motif A8 rotates inwards compared to the adenylation conformation. Later, the structure of the standalone DltA A domain confirmed the conformational change in NRPS A domains (Yonus, Neumann et al. 2008). This study suggested that binding of the negatively

charged ATP and acyl monomer to the positively charged binding pocket caused small changes in the electrostatic interactions of active site residues, driving the adenylation reaction and subsequent conformational changes. Even though the DltA structures provided an initial experimental and theoretical framework on the crosstalk between the A domain half-reactions, information on the interactions between A-domains and the PCP domain were still missing. Later work by the Gulick and Aldrich labs (Mitchell, Shi et al. 2012) (Sundlov, Shi et al. 2012) revealed the first A-PCP di-domain structures using mechanism-based vinyl sulfonamide inhibitors that emulate the thioesterification step. After thioester formation, the A_{sub} domain opens again, releasing AMP. The resulting acyl-*S*-PCP intermediate can proceed to the condensation reaction as a donor group if it was formed on an initiation module, or as an acceptor if it was formed in an elongation module. In some systems, acyl-*S*-PCP gets transported to tailoring domains to undergo chemical modifications before proceeding to condensation.

1.2.3. Condensation

After adenylation and thiolation, the acyl-*S*-PCP intermediate is transported to the condensation (C) domain located within the same elongation module ($module_n$), where it will form a peptide bond (**Fig. 1.4a**) with the acyl/peptidyl-*S*-PCP presented by the upstream module ($module_{n-1}$). During the condensation reaction, the acyl-thioester in $module_n$ is called “acceptor substrate”, as it will “receive” the acyl/peptidyl chain from $module_{n-1}$, the “donor substrate”. The product of the condensation reaction is a chain elongated by one monomer unit, attached to the PCP in $module_n$. The condensation reaction occurs through nucleophilic attack of the α-NH_2 in the acceptor substrate against the thioester carbon of the donor substrate.

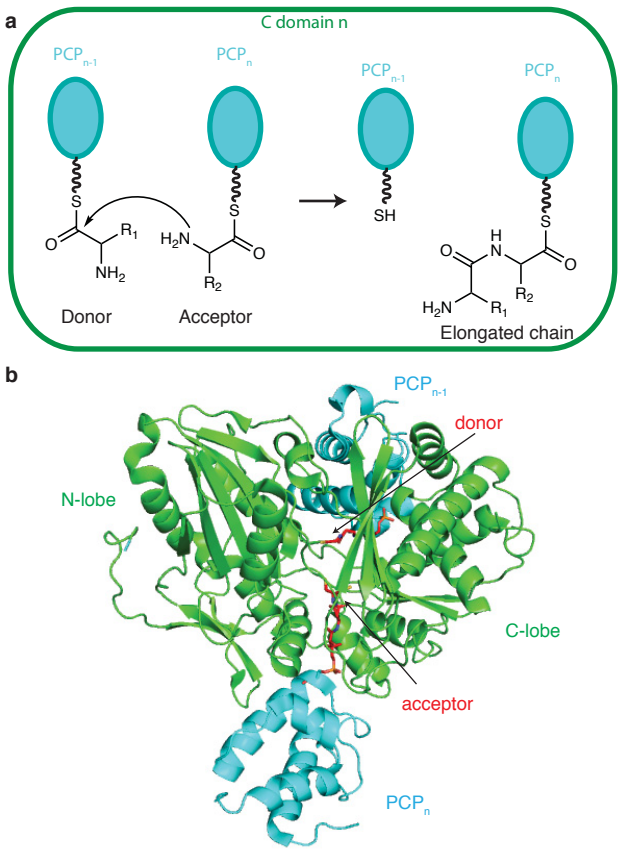


Figure 1.4. Condensation.

a. Condensation domain reaction. **b**. Condensation domain reaction model made from several structures at different reaction stages (PDB IDs: 5ejd, 4zxh, 5du0).

The first structure of a C domain revealed that they are pseudo-dimeric V-shaped proteins formed by C-terminal and N-terminal lobes (Keating, Marshall et al. 2002) (**Fig. 1.4b**). Each lobe is formed by a central β-sheet surrounded by α-helices. The active site, identifiable by the conserved motif HHxxxDG lies in a cleft than runs through the middle of both lobes. It was suggested that the second His in the motif (His157 based on numbering of first condensation domain of CDA , PDB ID: 4jn3 (Bloudoff, Rodionov et al. 2013)) played the role of a strong general base, deprotonating the α-NH_3^+ group of the acceptor substrate in preparation for nucleophilic attack. In some systems, mutation of His157 leads to loss of condensation function, whereas in others activity is retained (Bloudoff and Schmeing 2017). Along with structural evidence, it was proposed that instead of acting as a general base during the reaction, His157 participates in

acceptor substrate positioning, as in the mechanism of peptide bond formation catalyzed by the ribosome (Bloudoff, Alonzo et al. 2016).

The C domain must interact productively with the PCP domains bearing the acceptor and donor substrates in preparation for catalysis. From the V-shaped view of the C domain structure (**Fig. 1.4b**), the donor substrate enters through a tunnel located on top of the domain, while the acceptor substrate enters from the bottom. The donor substrate tunnel and PCP_{n-1} binding site were initially identified in two structures of closely related proteins, one from a PCP-epimerization domain construct (Chen, Li et al. 2016) (see 1.2.5.1. Epimerization) and another from a PCP-C_T domain construct (Zhang, Liu et al. 2016). C_T domains participate in chain termination in some fungal systems, and their structure is strikingly similar to C domains. The acceptor substrate tunnel and PCP_n binding site have been identified in several structures: the calcium dependent antibiotic C domain coupled to a substrate analogue covalent probe (Bloudoff, Alonzo et al. 2016) (see below), the holo termination module of AB3403 (Drake, Miller et al. 2016), and indirectly in the termination module of SfrAC (Tanovic, Samel et al. 2008).

Elucidating the catalytic mechanism of C domains has been hampered by the difficulty of obtaining substrate-bound C domain structures. No study has been able to obtain interpretable electron density for substrate analogues. Bloudoff et al. developed a chemical biology strategy to obtain holo-complexes of excised C domains (Bloudoff, Alonzo et al. 2016), using a probe consisting of L-Ala bound through its carboxylate group to an alkyl chain bearing a distal Br group. The rationale was that a Cys engineered by mutagenesis at the entrance of the acceptor substrate tunnel would displace the Br group, leading to a covalently linked acceptor acyl moiety, resembling the C-acyl-*S*-PCP complex. Testing different alkyl chains while evaluating C domain activity led to a structure where the acceptor residue was in a productive conformation, close to His157. The structure of holo AB3403 termination module (Drake, Miller et al. 2016) shows that the probe resembles acceptor acyl-S-PCP positioning, indicating that the chemical biology strategy could be applied to other C domains.

C domains interact dynamically with the A domain from their module (A_n). This evidence comes from outstanding efforts from several groups that solved modular structures of NRPSs, specifically the C-A-PCP-TE termination modules of SrfAC (Tanovic, Samel et al. 2008), AB3403 and EntF (Drake, Miller et al. 2016). In SrfAC, the C-A domains form a rigid "platform" through a large interaction that buries 800 A of residue surface area. It was hypothesized that this platform was the core of NRPS supramolecular structure and that it remained invariable in the course of the biosynthetic cycle. However, the AB3403 and EntF interactions showed different relative positioning of the C-A domains, indicating flexibility of the so called "catalytic platform".

Interaction of C domains with upstream A domains (A_{n-1}) is even more flexible, as shown by the A_{n-1}-PCP-C_n structure of DhbF (Tarry, Haque et al. 2017). A_{n-1} and C_n share almost no contacts, and EM data shows that the relative positioning of the domains is extremely variable. This was strong evidence that NRPSs do not have a supramolecular architecture.

1.2.4. Termination reactions

Different mechanisms can release the elongated chain attached to the last PCP domain, depending on the type of domain located at the C-terminus of the NRPS. The most common release domain is the thioesterase domain (TE domains), but some NRPSs employ other types of termination domains with diversified strategies ranging from reductive release to coupling with the solute tryptamine using specialized C domains (Bloudoff, Fage et al. 2017).

1.2.4.1 Thioesterase domains

TE domains catalyze chain release through two half-reactions using a catalytic Ser-His-Asp triad, by a similar mechanism to Ser proteases (Horsman, Hari et al. 2015) (Hedstrom 2002) (**Fig. 1.5a**). In the first half reaction, a transesterification, the His residue increases nucleophilicity of the Ser alcohol by proton abstraction. The activated Ser OH attacks the carbon thioester in peptidyl-*S*-PCP, forming a covalent acyl-*O*-TE intermediate. In the second half reaction, a nucleophile (R in **Fig. 1.5a**) attacks the acyl-*O*-TE intermediate, displacing the Ser oxygen and releasing the chain. The nature of the incoming nucleophile in the second half-reaction determines whether the released chain is linear, cyclic or will undergo oligomerization (Du and Lou 2010).

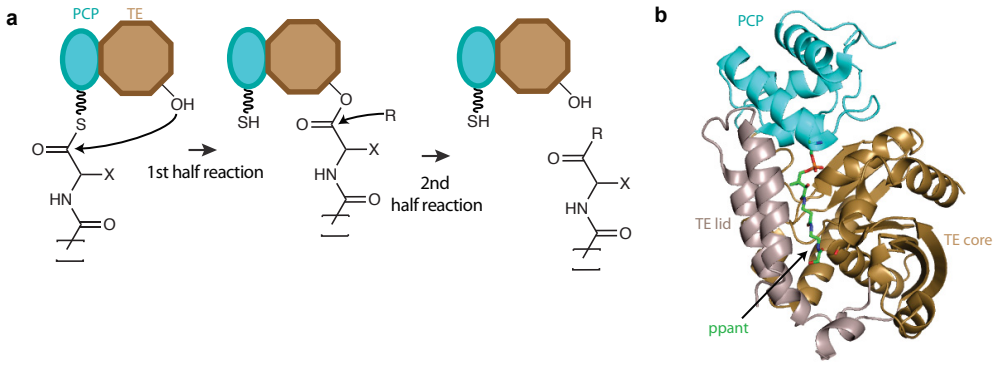


Figure 1.5. Thioesterase domain.

a. Thioesterase domain reaction. **b**. holo-PCP-TE structure (PDB ID:3tej)

For instance, hydrolysis occurs if the nucleophile is H_2O, resulting in a linear released chain (Gunsior, Breazeale et al. 2004). If the nucleophile is an NH_2 group located in an internal residue, the resulting peptide is a macrocycle (Trauger, Kohli et al. 2000). The NH_2 group can arise from the α- or β-NH_2 group in the most distal residue, resulting on a macrocycle encompassing the full sequence of the linear intermediate (Trauger, Kohli et al. 2000). A nucleophilic ε-NH_2 originating from a Lys side chain in the middle of the peptide sequence results in a branched macrocycle like in the structure of bacitracin (Konz, Klens et al. 1997).

Some TE domains catalyze oligomerization followed by cyclization, exemplified by the termination of gramicidin S biosynthesis (Hoyer, Mahlert et al. 2007), a cyclic dodecapeptide with antibiotic properties. In the proposed gramicidin S biosynthesis, a pentamodular NRPS generates the linear precursor D-Phe-L-Pro-L-Val-L-Orn-L-Leucyl-PCP-thioester (pentapeptidyl-*S*-PCP). The active site Ser in the TE domain attacks this intermediate, forming a transient pentapeptidyl-*O*-TE complex. The NRPS will synthesize another equivalent of pentapeptidyl-*S*-PCP that attacks the ester in pentapeptidyl-*O*-TE using the α-D-NH_2 group in D-Phe, forming linear decapeptidyl-*S*-PCP. The active site Ser in the TE domain will capture the decapeptidyl chain, forming decapeptidyl-*O*-TE. Then, through a still unknown mechanism, the linear decapeptidyl chain "curls" back and the distal α-D-NH_2 group in D-Phe attacks the ester group, forming a macrocycle

and releasing the peptide. The interplay of PCP and TE domains and the retro-transfer mechanism was verified using substrate analogues (Hoyer, Mahlert et al. 2007).

Several PKS and NRPS TE domain structures have been solved (Tsai, Miercke et al. 2001, Tsai, Lu et al. 2002, Frueh, Arthanari et al. 2008, Whicher, Florova et al. 2011, Horsman, Hari et al. 2015, Guntaka, Healy et al. 2017), showing that the TE fold belongs to the α/β hydrolase superfamily, consisting of a central β-sheet surrounded by α-helices (**Fig. 1.5b**). The catalytic triad has been identified in each structure, as well as an oxyanion hole that stabilizes the negatively-charged tetrahedral intermediate. A structural element called the active site lid is inserted between strands 6 and 7. This element has high sequence and structural variability within TE domains, some of them consisting of a relatively simple 54-residue two-helical structure (DEBS-TE) (Tsai, Miercke et al. 2001), while others could be consider small subdomains comprised of an α-helical bundle of 100 residues (AB3403-TE) (Drake, Miller et al. 2016). It is thought that the lid can act as a solvent shield to protect the acyl ester intermediate from hydrolysis in TEs that catalyze cyclization (Liu, Zheng et al. 2011), but its sequence, length and structure bear no correlation with the substrate length, complexity or release mechanisms (Horsman, Hari et al. 2015). Lids have been found in different positions in structures of the same TE domain (Koglin, Lohr et al. 2008), often have higher B-factors than the protein core (Tsai, Lu et al. 2002) and contain regions absent from the electron density (Samel, Wagner et al. 2006), indicating that they are mobile and dynamic structural elements. To date, the exact roles of the active site lid during TE reactions remain unknown.

Attempts to generate substrate-bound TE structures are hampered by the relative short lifetime of the acyl ester intermediates (Bruner, Weber et al. 2002, Samel, Wagner et al. 2006). A few structures of PKS TE-substrate complexes exist, which used polyketide-diphenyl-phosphonate analogues (Akey, Kittendorf et al. 2006, Argyropoulos, Bergeret et al. 2016). The most informative of these substrate-bound structures is the pikromycin TE bound to a pentaketide phosphonate (Akey, Kittendorf et al. 2006). The probe binds to the active site serine through the phosphate, emulating the tetrahedral transition state. The ketide portion is observed curling back

into the active site cavity formed in the space between the mobile lid and the region around the catalytic Ser. The authors suggest that the curled ketide conformation is enforced by a "hydrophylic barrier" formed by two well-coordinated waters at the beginning of the lid, preventing the ketide from exiting the active site, directing it towards the hydrophobic cavity and favouring the cyclization conformation. The authors also propose that cyclization is driven by conformational restraints of the ketide. It would be interesting to confirm these findings with a full-length ketide probe; however, the chemical synthesis of these molecules can be extremely challenging.

TE domains interact with the PCP domain during their biosynthetic cycle (**Fig. 1.5b**). An NMR structure of the enterobactin EntF apo PCP-TE (Frueh, Arthanari et al. 2008) indicated that the domain pair adopts multiple conformations. Eventually, the Bruner group solved the holo PCP-TE structure from EntF using the active site probe α-Cl-acetyl-amide-ppant (Liu, Zheng et al. 2011), which after loading to the PCP domain was expected to form a covalent bond upon displacement of Cl by the OH from the TE catalytic Ser. The holo structure was useful to identify a ppant tunnel formed by one of the helices of the lid and the active site canyon located in the core of the protein, which leads the ppant moiety towards the catalytic triad. Surprisingly, the PCP-ppant attachment point makes few interactions with the TE domain and most TE-ppant interactions occur through a TE loop. TE-PCP interactions occur mainly through the first β-strand and α-helix of the TE domain and two helices from the PCP domain. The helices from the active site lid also make contact with the PCP domain, suggesting that the lid may also play a role during substrate delivery. EntF TE (Shaw-Reid, Kelleher et al. 1999) and other TEs use OH groups (other than H_2O) as nucleophiles in their reactions and will be further discussed in the section **1.3. Depsipeptide synthetases**.

1.2.5 Tailoring reactions

Specialized domains catalyze chemical modifications on acyl-*S*-PCP intermediates and can exist as insertions into NRPS modules or as free-standing enzymes. Modifications include L- to D-amino acid epimerization, N-formylation, N- and S- methylation, heterocyclization, oxidation and

β-hydroxylation. Several of these reactions have been characterized, along with the structures of the domains involved.

1.2.5.1. Epimerization

Epimerization (E) domains catalyze the conversion of L-amino acids to D-amino acids and are present in modules with the minimal architecture A-PCP-E (initiation) and C-A-PCP-E (elongation) (Bloudoff and Schmeing 2017). The structure of E and C domains is almost identical, as they are phylogenetically related and share the sequence motif HHXXXDG (Samel, Czodrowski et al. 2014). However, the E domain structure is distinguished by a blocked acceptor site where 11 residues impede the access of an aminoacyl-S-PCP substrate from downstream PCP domains, restricting catalysis to donor substrates from the upstream PCP domain.

E domains in initiation modules (A-PCP-E) catalyze epimerization on an L-aminoacyl-*S*-PCP substrate, whose D- product is usually transferred to a downstream C domain with D- specificity in the donor site and L- specificity in the acceptor site ($^{D}C_{L}$). However, E domains in elongation modules (C_n-A_n-PCP_n-E_n) catalyze epimerization only when L-aminoacyl-*S*-PCP_n undergoes condensation with L-aminoacyl-*S*-PCP_{n-1} in the C_n domain (Bloudoff and Schmeing 2017), generating L-amino-L-aminoacyl-*S*-PCP_n. The condensation product acts as a donor substrate of the E_n domain, generating L-amino-D-aminoacyl-*S*-PCP_n.

1.2.5.2. Formylation

Formylation (F) domains catalyze the transfer of a formyl group from formyltetrahydrofolate to the α-amino group in an aminoacyl-*S*-PCP (Reimer, Harb et al. 2018). They are inserted at the N-terminus of initiation modules and are related to formyltransferases from other metabolic pathways (Reimer, Harb et al. 2018). The biosynthetic cycle of the F domain-containing initiation module from linear gramicidin synthetase (LgrA) has been thoroughly characterized in a spectacular paper by Dr. Reimer and co-workers (Reimer, Aloise et al. 2016). They solved different structures of the initiation module in the pre-adenylation, adenylation, thiolation and formylation stages. They discovered that massive conformational transitions occur during the cycle, all mediated by the A_{sub} domain which provides additional flexibility to the PCP domain to

transport thiolated substrates within catalytic units. The most striking movement occurs upon delivery of the aminoacyl-*S*-PCP substrate from the A to the F domain active site. The PCP domain rotates 75° and translates its center of mass across 61 Å, a movement facilitated by a 180° rotation and 21 Å translocation of the A_{sub} domain. This work was also the first time that so many conformational states were observed for a single NRPS.

1.2.5.3 Methylation

Methylation (M) domains catalyze transfer of a methyl group from S-adenosyl methionine to an amino acyl-*S*-PCP substrate, at either the α-NH_2 (Velkov, Horne et al. 2011) or γ-S (Al-Mestarihi, Villamizar et al. 2014) groups. Most M domains are inserted between motifs A8 and A9 of the A_{sub} domain, creating an interrupted A domain at an insertion point that is thought to be the location of other tailoring domains such as ketoreductase (KR) and oxidation (OX) domains (Labby, Watsula et al. 2015). The M domain from the thiocoraline pathway, that catalyzes thiol methylation, is unexpectedly inserted between A_{core} motifs A2 and A3 (Al-Mestarihi, Villamizar et al. 2014).

The structure of an A-M di-domain construct from the thiocoraline gene cluster was solved in 2018 (Mori, Pang et al. 2018). The Rossmann α/β-fold M domain is inserted between motifs A8 and A9 and connected through structured linkers. The A and M active sites are 60 Å apart and bound to substrates; a valyl-adenylate in the A domain and S-adenosyl-L-homocysteine in the M domain. Even though the structure lacks a PCP domain, the A_{sub} domain is oriented in the thiolation conformation, likely due to the presence of the adenylate in the A domain active site. The distance between catalytic pockets indicates that massive conformational rearrangements are needed to transport the acyl-*S*-PCP monomer during the cycle. Further research, specifically structures of different conformational states, might indicate the exact nature of the required conformational rearrangements, especially those involved in delivery of the acyl-*S*-PCP to the M domain.

1.2.5.4 Heterocyclization

Heterocyclization (Cy) domains are related to C domains and catalyze peptide bond formation followed by intramolecular cyclodehydration using nucleophiles in Ser, Cys or Thr sidechains, generating oxazolines, thiazolines and methyloxazolines (Bloudoff, Fage et al. 2017). Even though their structure is very similar to C domains, Cy domains contain a DxxxxD motif instead of the canonical C domain HHXXXDG motif. The structure of the Cy domain from bacillamide synthetase (BmdB) solved by Dr. Kris Bloudoff *et al.* (Bloudoff, Fage et al. 2017) revealed that the DxxxxD motif is in an extremely similar conformation to the C domain motif and suggested that it might also play a positioning role opposed to a catalytic role. Mutational analysis of the reconstituted intact NRPS system by LC-MS revealed that an Asp-Thr catalytic dyad is critical for cyclodehydration.

1.2.5.5 Oxidation

The thioazoline rings produced by Cy domains can undergo oxidation to thiazoles in a reaction catalyzed by oxidase (OX) domains, FMN-containing flavoproteins (Schneider, Shen et al. 2003). In the epothilone and bleomycin pathways, the OX domain is embedded into an interrupted A domain at the same insertion point of other tailoring domains, such as M domains (see 1.2.5.3. Methylation) (Schneider, Shen et al. 2003). In the same work, Walsh and co-workers were able to express and purify the excised OX domains from the epothilone and bleomycin pathways for the first time, confirming their identity as flavoproteins and performing a kinetic analysis using thiazoline-*S*-PCP substrate analogues.

Bacillamides A-D are thiazole-containing algicides synthesized in part by BmdB, a dimodular NRPS that contains a Cy domain responsible for thiazoline formation (see 1.2.5.4. Heterocyclization) (Bloudoff, Fage et al. 2017). However, there is no embedded OX domain in any of the two A domains that could participate in thiazoline to thiazole oxidation. Analysis of the cluster revealed that a stand-alone OX domain encoded by BmdC (Bloudoff, Fage et al. 2017) might be responsible for the oxidation reaction, but further research is needed to confirm this role and its plausible modes of interaction with the rest of the assembly line.

1.2.5.6. Hydroxylation

Stand-alone P450 enzymes can interact with NRPSs to catalyze β-hydroxylation of amino acyl-*S*-PCP intermediates. β-hydroxyl groups can be found in vancomycin-type peptides (van Wageningen, Kirkpatrick et al. 1998) among many other natural products. The activity of the P450 responsible for β-hydroxylation of three different residues in skyllamicin was evaluated *in vitro* (Uhlmann, Sussmuth et al. 2013), showing that it is selective for L-amino over D-amino acyl-*S*-PCP substrates and is stereospecific, as the β-hydroxylated products have the (2S,3S)-configuration. Additionally, tests with non-cognate PCP domains revealed that the P450 enzyme acts exclusively on cognate substrates.

1.2.5.7. Reactions involved in ester bond formation

Some NRPS reactions generate hydroxyl groups that can be used as nucleophiles in ester-bond forming downstream reactions. The resulting natural products, containing amide and ester bonds, are called depsipeptides and their biosynthesis is the main focus of the thesis. The following section is an introduction to depsipeptides, their biosynthetic machinery and the questions on the field that were addressed during the dissertation.

1.3. Depsipeptide synthetases: specialized NRPSs that incorporate hydroxy acids and catalyze ester bond formation

1.3.1 Depsipeptides: a family of useful compounds

Nonribosomal cyclic depsipeptides are NRPs with ester and amide bonds in their structures (Ballard, Yu et al. 2002), produced by specialized NRPSs called depsipeptide synthetases. They are ubiquitous natural products with vast biological activities of medical and industrial relevance; for example, daptomycin is used as an antibiotic for severe infections (Robbel and Marahiel 2010), antimycin as a piscicide in industrial fish farming (Derse and Strong 1963) and surfactin as a potential industrial tenside (Kraas, Helmetag et al. 2010). Some of them are involved in pathogenesis, like the emetic toxin cereulide (Agata, Ohta et al. 1995). In addition to the occurrence of at least one ester and one amide bond in their structures, depsipeptides may contain many other chemical groups added through the assembly line logic of NRPSs through different mechanisms of hydroxy acid incorporation, ester bond formation and tailoring reactions.

1.3.2. Chemical classification of depsipeptides.

Like other NRPs, the chemical diversity of cyclic depsipeptides is outstanding. Some of them contain modified amino acids, non-proteinogenic amino acids, fatty acids and ketide portions. Taevernier *et al*. (Taevernier, Wynendaele et al. 2017) made a thorough classification of cyclic depsipeptides based on their chemical features (**Fig. 1.6,** adapted from (Taevernier, Wynendaele et al. 2017)), generating a curated database separating depsipeptides by the type of hydroxy acid involved in ester bond formation and further classification tiers based several chemical characteristics.

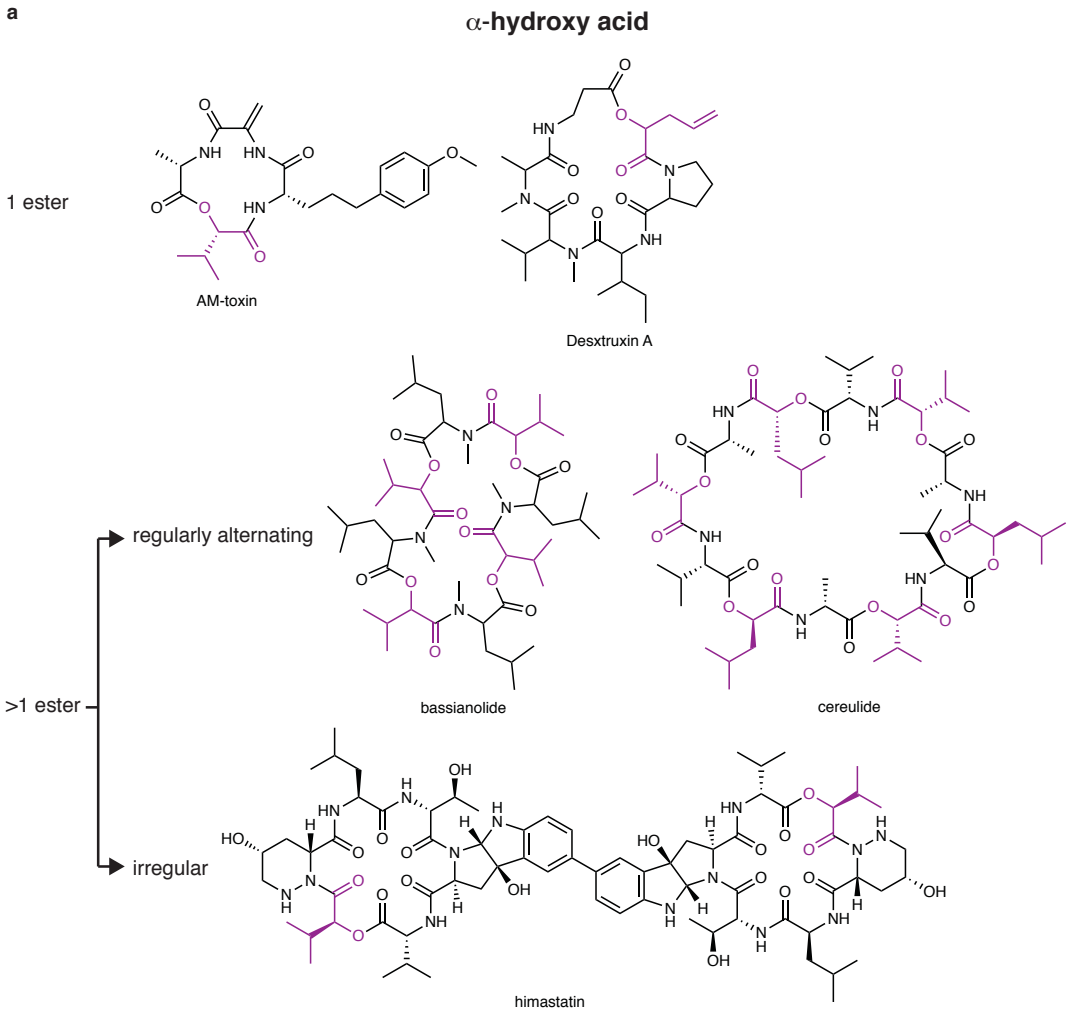


Figure 1.6a. Depsipeptides with one or more α-hydroxy acids in their structures.

Hydroxy acids are highlighted in magenta. Adapted from (Taevernier, Wynendaele et al. 2017)

Figure 1.6b. Depsipeptides with one or more β-hydroxy acids in their structures.

Hydroxy acids are highlighted in magenta. Adapted from (Taevernier, Wynendaele et al. 2017).

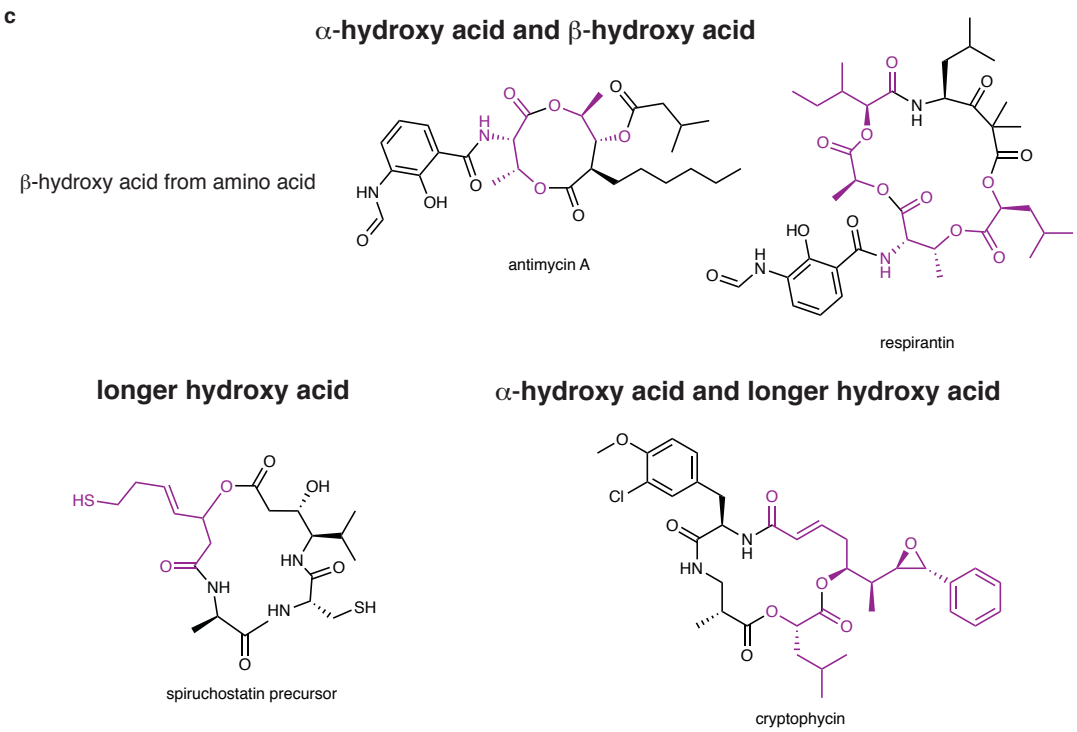


Figure 1.6c. Depsipeptides with mixed α- and β- and longer hydroxy acids in their structures.

Hydroxy acids are highlighted in magenta. Adapted from (Taevernier, Wynendaele et al. 2017).

There are several NRPS mechanisms that incorporate the different types of hydroxy acids and that catalyze varied mechanisms of ester bond formation, described in detail in the following sections. The simplest depsipeptides rely on a single specialized domain working in the context of canonical pathways to generate an ester bond (**Section 1.3.3**), whereas more complex molecules arise from NRPSs with dedicated modules and domains working iteratively to form alternating arrays of amino and hydroxy acids (**Section 1.3.5**).

1.3.3. Thioesterase domains that form ester bonds

Type I TE domains are α/β hydrolase proteins specialized in chain termination through a mechanism reminiscent of serine proteases (discussed in detail in **1.2.4.1. Thioesterase domains**). After chain elongation, the terminal PCP transfers the fully elongated chain to an active site serine in the TE domain, forming an acyl-peptide intermediate. The intermediate can then be released through nucleophilic attack (**Fig. 1.5**). The nature of the nucleophile determines whether the chain will be released by hydrolysis (H_2O), cyclization (internal group),

oligomerization (external group) or a combination of some of those mechanisms. The chemical nature of the nucleophile will also determine whether the formed bond is an amide bond (α-NH_2, ε-NH_2 $_{in}$ Lys side chain), or an ester bond (OH-R, discussed here). In the case of cyclic depsipeptides, there are several OH-containing chemical groups that are used as nucleophiles during TE-catalyzed ester bond formation.

1.3.3.1. Cyclization: nucleophile OH from Thr, Ser or Tyr sidechains

The simplest mechanism of ester formation is TE-mediated cyclization using OH groups in Thr (daptomycin) (Baltz 2008) (**Fig. 1.7**) or Tyr (fengycin) (Steller, Vollenbroich et al. 1999) residues after canonical NRPS chain elongation. In this mechanism, an elongated linear peptide containing a Ser, Thr or Tyr residue in its sequence is transferred from the last PCP domain to the active site Ser of the terminal TE domain forming a peptidyl-*O*-TE intermediate, which undergoes cyclization through ester bond formation by nucleophilic attack of the side chain OH on the peptidyl ester (Samel, Wagner et al. 2006).

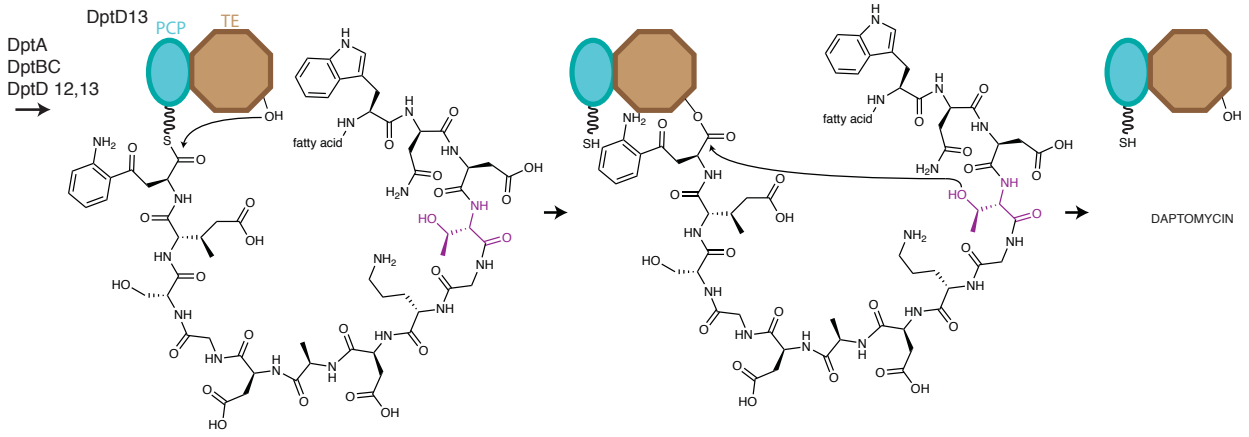


Figure 1.7. Final steps of daptomycin biosynthesis.

Only the last two domains of DptD13 are represented.

Daptomycin (**Fig. 1.6a, 1.7**) is a branched cyclic depsipeptide used as an antibiotic to treat severe skin infections (Robbel and Marahiel 2010). It is generated by 13 modules distributed in three NRPS, DptA, DptC and DptD. After elongation, the tridecapeptidyl-*S*-PCP in the last module of DptD is transferred to the DptD TE, that catalyzes the nucleophilic attack of the OH

group on Thr4 on the peptidyl-ester intermediate, releasing mature daptomycin (Grunewald, Sieber et al. 2004). Research on the excised daptomycin-TE domain (Kopp, Grunewald et al. 2006) revealed that it can catalyze the ester-forming cyclization reaction using thiophenol analogues of peptidyl-*S*-PCP substrates, showing that depsipeptides can be formed through chemoenzymatic synthesis relying only in the excised TE domain, isolated from the rest of the assembly line. The same paper showed that the TE domain has some flexibility on the position of the attacking Thr residue within the peptide, leading to the formation of new macrocycles. Accompanying data revealed the same conclusions for the analogous system that generates a depsipeptide called A54145.

Fengicyn is a decadepsipeptide with antifungal activity that is cyclized through a similar mechanism as daptomycin, using the sidechain OH in residue Tyr4 as a nucleophile (Steller, Vollenbroich et al. 1999). The structure of its cyclizing TE domain was solved (Samel, Wagner et al. 2006), showing a single active site lid conformation. An accompanying structure with PMSF covalently bound in the active site, which emulated the tetrahedral transition state, allowed the authors to unambiguously identify the oxyanion hole in the active site cavity. In spite of this information, it is still unknown what structural features determine peptide cyclization, as it is highly challenging to obtain substrate-bound structures of TE domains.

1.3.3.2. Oligomerization and cyclization: nucleophile OH from Ser or Thr in siderophore biosynthesis

Enterobactin is a bacterial siderophore with extreme affinity for Fe (III) (Raymond, Dertz et al. 2003), with a structure consisting of three-repeats of the dipeptide 2,3-dihydroxybenzoate-Ser (DHB-Ser). The ester bonds in enterobactin, formed by the side chain of Ser and the carboxylate of DHB, could be generated through either of two proposed iterative mechanisms catalyzed by the EntF TE domain (Shaw-Reid, Kelleher et al. 1999). The most likely mechanism (**Fig. 1.8**) is based on results from gramicidin S TE (**section 1.2.4.1**) and from valinomycin TE presented in this dissertation (**Chapter 4**). In this mechanism, the NRPSs EntB and EntF generate the linear intermediate DHB-Ser-*S*-PCP_{EntF} that serves as substrate for the first EntF-TE half-reaction, generating DHB-Ser-*O*-$TE_{EntF.}$ This peptidyl ester undergoes oligomerization through nucleophilic

attack by the Ser side chain OH in a newly synthesized DHB-Ser-*S*-PCP_{EntF} intermediate, generating the linear dimer $(DHB\text{-}Ser)_2$-*S*-PCP_{EntF}. An additional cycle of transfer to the EntF-TE Ser occurs, generating $(DHB\text{-}Ser)_2$-*O*-TE_{EntF}, which undergoes another cycle of oligomerization forming the trimer $(DHB\text{-}Ser)_2$-*S*-PCP_{EntF}. EntF-TE will capture the linear trimer as an ester and catalyze intramolecular cyclization using the distal OH group, forming the mature macrocycle.

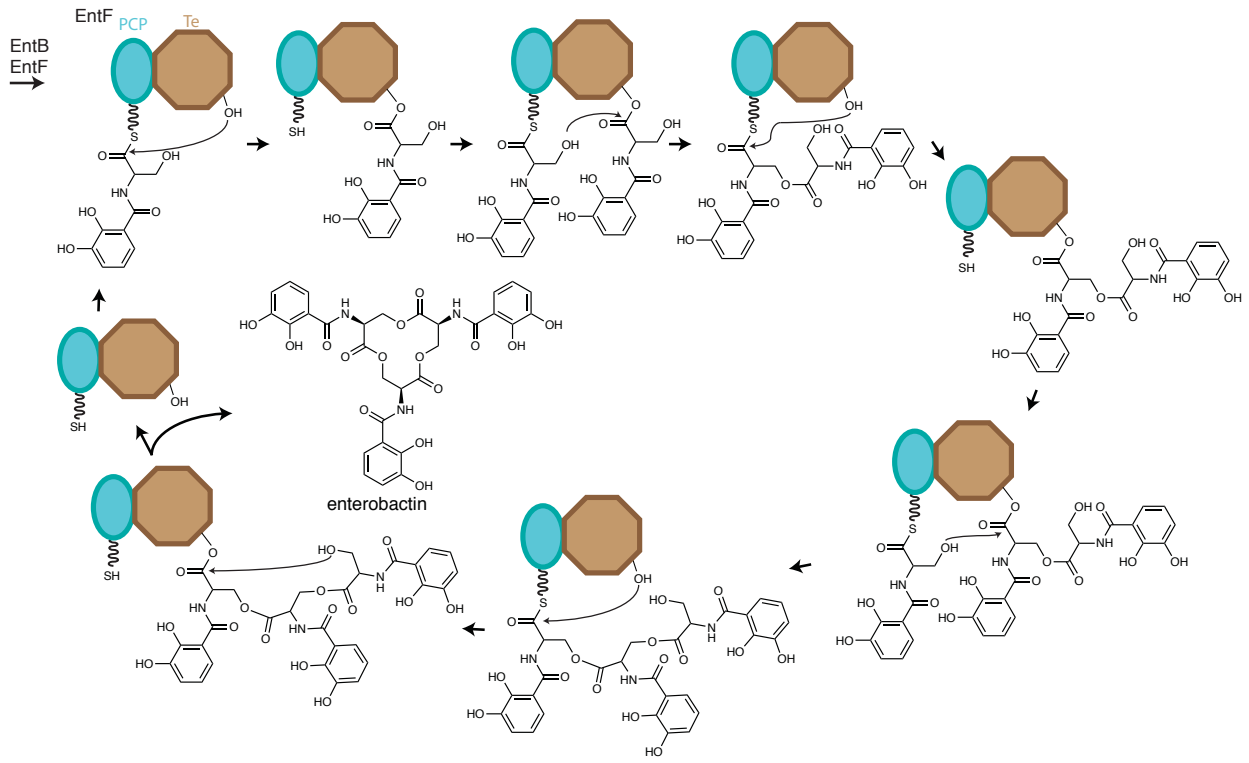


Figure 1.8. Proposed oligomerization and cyclization mechanism catalyzed by EntF TE.

The EntF-TE system has been fundamental to the structural studies of NRPS TE domains, as the first structure di-domain solved by NMR was EntF PCP-TE (Frueh, Arthanari et al. 2008), revealing the interactions between domains for the first time. Later, an EntF holo-PCP-TE structure (Liu, Zheng et al. 2011) revealed the crucial interactions of the didomain complex in the holo state and the contacts of the ppant arm with core and lid in the TE domain (see **1.2.4.1. Thioesterase domains**). However, it is still unknown what structural features determine the oligomer length and the molecular mechanism that switches the second half-reaction from oligomerization to cyclization. Related depsipeptide siderophores can use Thr side chains to form the macrocycle

(Raymond, Dertz et al. 2003) (May, Wendrich et al. 2001), likely through a similar iterative EntF TE mechanism.

1.3.3.3. Cyclization: nucleophile OH from hydroxylated fatty acids

The ester bond in some depsipeptides is formed by the β-hydroxy group in a fatty acid chain, accounting for 18% of all the molecules in the depsipeptide chemical classification database (Taevernier, Wynendaele et al. 2017). For example, surfactin is a 7-member lipodepsipeptide with tenside, antitumoral and antibiotic properties, with one of most studied biosynthetic clusters in the NRPS field (Menkhaus, Ullrich et al. 1993) (Vollenbroich, Mehta et al. 1994) (Koglin, Lohr et al. 2008) (Tseng, Bruner et al. 2002). Surfactin is generated by the NRPSs SrfAA, SrfAB and SrfAC working in conjunction with fatty acid biogenesis (Kraas, Helmetag et al. 2010). The first step in the process is the activation of 3-hydroxy myristic acid in *trans* as a CoA thioester by a fatty acid CoA ligase (Kraas, Helmetag et al. 2010). Then, the N-terminal condensation domain in SrfAA catalyzes amide bond formation between the fatty acid and the Glu substrate selected by the first module. The cycle proceeds through canonical elongation steps in SrfAB and SrfAC until linear lipopeptidyl-*O*-TE_{SrfAC} is formed. Then, SrfAC-TE catalyzes cycle formation using the OH group in the distal hydroxylated lipid (**Fig. 1.9**) (Tseng, Bruner et al. 2002) (Bruner, Weber et al. 2002).

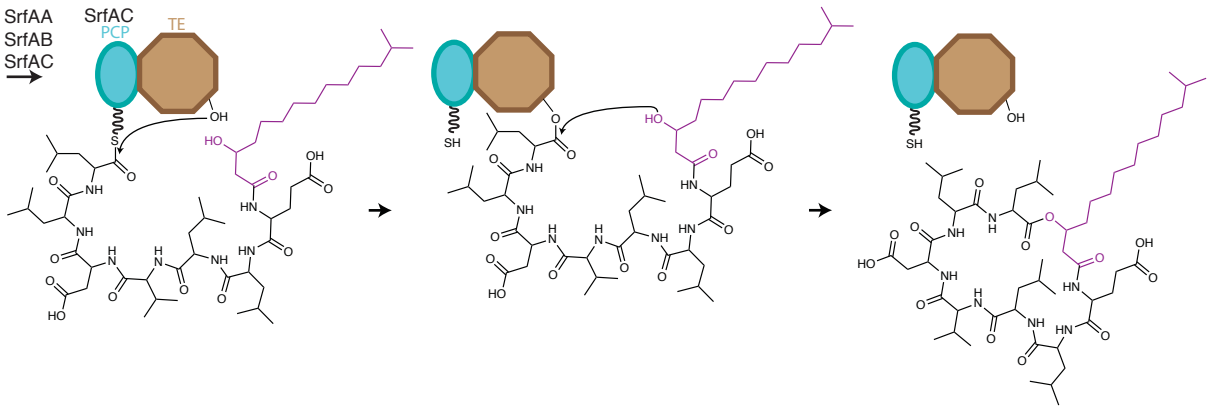


Figure 1.9. Cyclization of the surfactin linear precursor in SrfAC TE.

The structure of SrfAC-TE was solved in 2002, (Bruner, Weber et al. 2002). One of the most relevant findings of that study was that the active site lid was found in two different conformations in each asymmetric unit of the X-ray structure. This was the first structural evidence that the active site lid in TE domains could acquire different conformations and the authors suggested that it could be acting as a solvent shield during the cyclization reaction. The surfactin system was also used to solve the first full-module structure of NRPSs, leading to crucial information about interdomain interactions in termination modules (Tanovic, Samel et al. 2008).

Even though daptomycin, CDA and other antibiotics also contain fatty acids in their structures and are classified as lipodepsipeptides, the fatty acid chain is not hydroxylated and does not participate in ester bond formation.

1.3.3.4. Oligomerization and cyclization: nucleophile from hydroxylated fatty acids

Serratamolide (serrawettin w1) is a lipodepsipeptide with surfactant, antitumoral and antibacterial properties (Thies, Santiago-Schubel et al. 2014). Its structure consists of two symmetrically arranged Ser-fatty acid (Ser-FA) moieties joined through an ester bond between the hydroxylated FA and the carboxyl group in Ser (**Fig. 1.10**). This dimeric structure indicates an iterative biosynthetic pathway, like the one described for enterobactin (**Fig 1.8**). Inspection of its biosynthetic cluster reveals an NRPS with the C-A-PCP-TE domain arrangement (Li, Tanikawa et al. 2005). It is likely that the A domain selects Ser and the N-terminal C domain catalyzes amide bond formation with a fatty acid through a mechanism similar to SrfAA, forming Ser-FA-*S*-PCP. The TE domain would catalyze oligomerization to $(Ser\text{-}FA)_2$-*O*-TE followed by cyclization through nucleophilic attack of the OH group in the FA against the seryl-ester, releasing mature serratamolide.

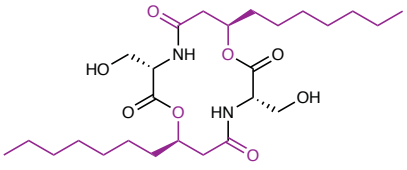


Figure 1.10. Serratamolide structure.

1.3.3.5. Cyclization: nucleophile OH from hydroxylated amino acids

Lysobactin is a cyclic depsipeptide with antibiotic properties (Hou, Robbel et al. 2011). The ester bond is formed by the OH group in β-hydroxy phenylalanine (β-OH-Phe). Hydroxylation happens after formation of Phe-*S*-PCP generating β-OH-Phe-*S*-PCP in a reaction catalyzed by free-standing hydroxylases. This mechanism is similar to that of P450 enzymes acting in *trans* (see 1.2.5.6. Hydroxylation). The C-terminal TE domain catalyzes cyclization using the β-OH group as a nucleophile (**Fig. 1.11**). The termination module has two TE domains in tandem, of which only the first participates in the cyclization reaction, while the second is a Type II TE domain that hydrolyzes mis-primed PCP domains. Furthermore, TE #2 was highly prone to proteolysis during protein expression/purification, indicative that a post-translational proteolytic event leads to a standalone TE #2 domain. In Polyoxypeptin A biosynthesis, the TE domain also uses a hydroxylated amino acid (β-OH-Leu) as nucleophile for ester bond formation and a P450 enzyme is thought to be responsible of catalyzing the hydroxylation step (Du, Wang et al. 2014).

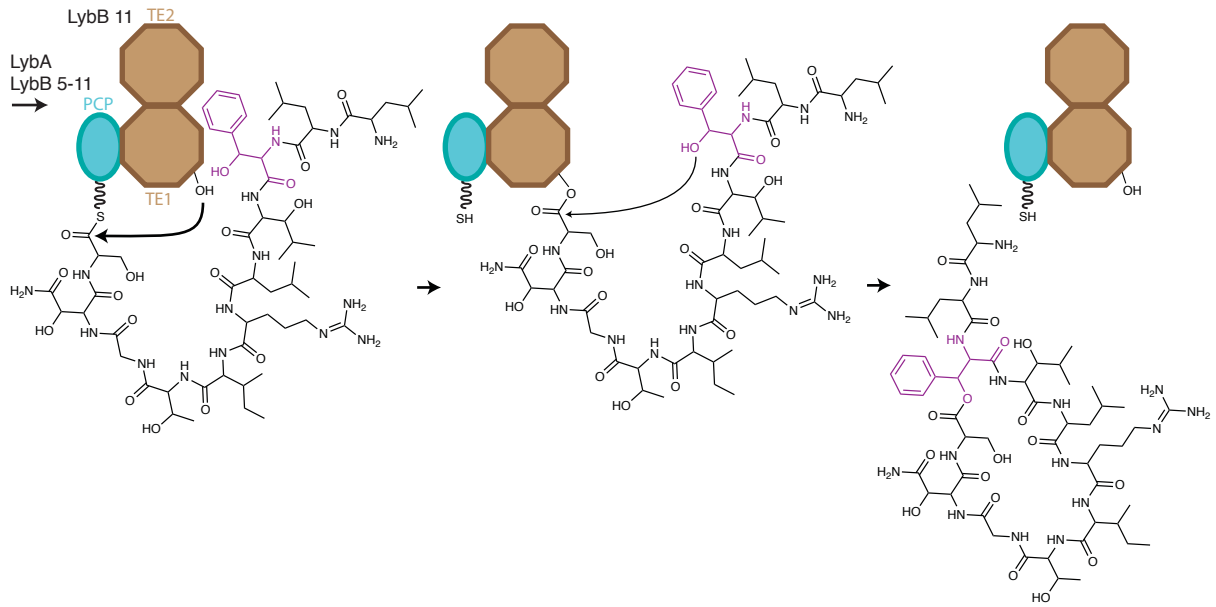


Figure 1.11. Lysobactin cyclization catalyzed by LybB TE.

1.3.3.6. Cyclization: nucleophile OH from hydroxylated polyketides

Spiruchostatins are antineoplastic agents with a mixed PKS-NRPS structure, with a biosynthetic cluster consisting of a mix of PKS and NRPS modules (Potharla, Wang et al. 2014). After

condensation of the PKS and NRPS portions and chain elongation, the mixed linear PKS-NRPS product is brought to the terminal TE domain, forming the macrolactone between the OH group in the ketide portion and the ester carbon in ketopeptidyl-*O*-TE, releasing a cyclic spiruchostatin precursor (**Fig. 1.12**). This precursor will eventually undergo disulfide bond formation, producing mature spiruchostatin. Other natural products that very likely undergo TE-mediated cyclization using the OH in mixed ketide-peptides are FK228 (romidepsin) (Narita, Kikuchi et al. 2009) and largazole (Hong and Luesch 2012).

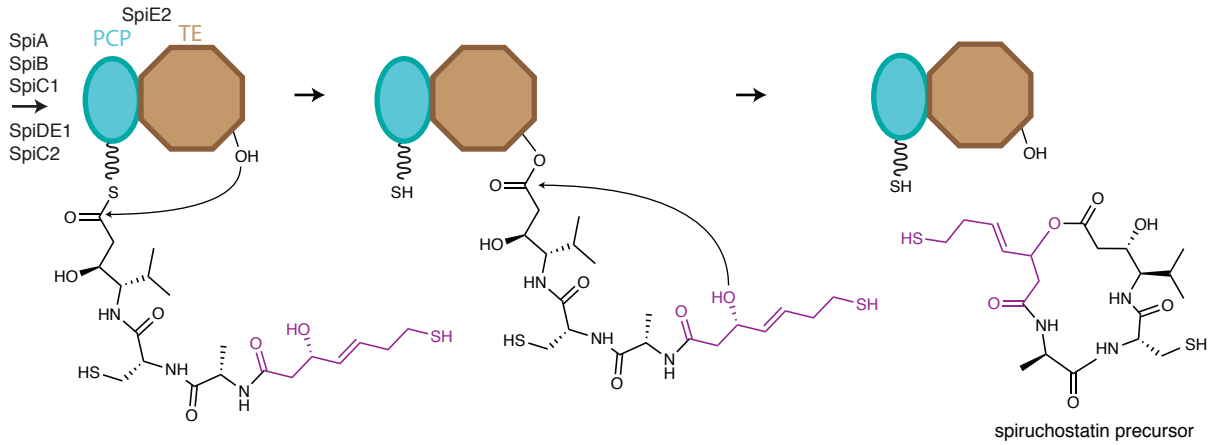


Figure 1.12. The spiruchostatin linear precursor is cyclized by SpiE2 TE.

Ester bond formation in the above examples requires β-hydroxy-monomers added to the peptide chain through the canonical NRPS elongation pathway or through mixed PKS-NRPS and fatty acid-NRPS systems. However, several depsipeptides contain α-hydroxy acids forming ester bonds in their structures whose activation requires specialized domains. The fungal and bacterial NRPS systems that generate α-hydroxy acids and use them to catalyze ester-bond forming reactions are discussed in the following section.

1.3.4. Specialized A domains that select α-hydroxy acids in fungal systems

Fungal depsipeptides like destruxin (Pedras, Irina Zaharia et al. 2002), enniatin and beauvericin (Liuzzi, Mirabelli et al. 2017) contain α-hydroxy acids in their structures (**Fig. 1.13**), selected by specialized A domains that can discriminate between α-amino and α-hydroxy groups through a yet unknown mechanism (**Fig 1.13a**). The highly-conserved Asp235 (based on GrsA-PheA

numbering) residue involved in amino acid selection is substituted by Gly in these domains (Xu, Orozco et al. 2009); however, it is unknown whether this substitution is enough to confer α-hydroxy acid selectivity. To date, there is no substrate-bound structure of a α-hydroxy-acid selecting-A domain to elucidate this question.

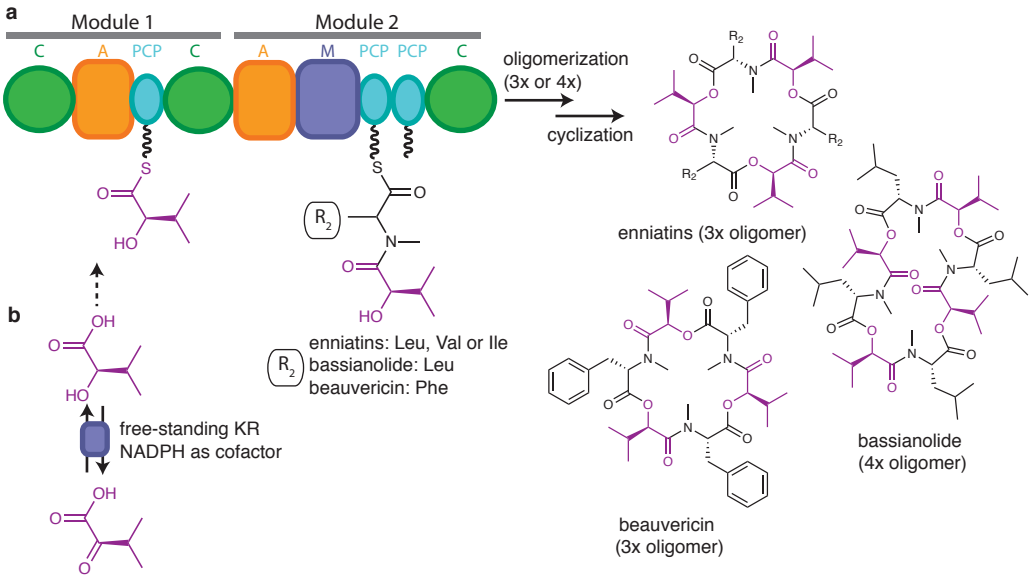


Figure 1.13. The biosynthesis of enniatin, beauvericin and bassianolide.

a, Enniatin, bassianolide and beauvericin are made by di-modular NRPSs. **b**, Free-standing ketoreductases (KR), unrelated to PKS KRs, generate α-hydroxy acids from α-keto acids.

Standalone ketoreductase (KR) domains generate the α-hydroxy acids used as A domain substrates (**Fig. 1.13b**), converting α-ketoacids in an NADPH-dependent reduction reaction (Xu, Orozco et al. 2009). The α-ketoacids α-keto isocaproic acid (α-KIC), α-keto isovaleric acid (α-KIV) and α-ketoisoleucine (α-KILE) arise from branched-chain amino acid biosynthesis pathways, while pyruvate arises from glycolysis (Kohlhaw 2003). The KR domains have been identified in some fungal depsipeptide biosynthetic clusters (Wang, Kang et al. 2012) (Xu, Orozco et al. 2009) and belong to the ApbA_C superfamily (NCBI CDD), which also encompasses ketopantoate reductases (Zhang, Jia et al. 2012). They are not related to PKS KR domains, that belong to the short-chain dehydrogenase (SDR) superfamily. Some fungal standalone KRs have been fully characterized (Lee, Gorisch et al. 1992, Zhang, Jia et al. 2012) and their deletion from the

biosynthetic cluster abolishes production of the cyclic fungal depsipeptide (Wang, Kang et al. 2012). KR deletion can be rescued by supplementing the media with α-hydroxy acids (Wang, Kang et al. 2012), indicative of the direct role of the KR in the biosynthetic pathway. It is currently unknown if these KRs interact non-covalently with the assembly line or if they produce their metabolites at high-enough concentrations to diffuse to the α-hydroxy acid-selecting A domain. After adenylation, the hydroxy acyl adenylate undergoes thiolation forming an OH-acyl-*S*-PCP intermediate that can participate in different downstream reactions depending on the location of the module and on the organization of the NRPS, leading to the different mechanisms of ester bond formation discussed in the following subsections.

1.3.4.1 Single α-hydroxy acid participating in cyclization in destruxin biosynthesis

Destruxins are insecticidal molecules produced by fungal entomopathogens, with a structure consisting of a 6-monomer ring produced by a 6-module NRPS (Pedras, Irina Zaharia et al. 2002). Cluster analysis revealed that module 1 is responsible for adding a hydroxy acid through direct selection by a specialized A domain (Wang, Kang et al. 2012). The off-line ketoreductase encoded by the dtx3 gene was identified in *in vivo* gene deletion experiments, where the destruxin-production phenotype could be rescued by addition of α-hydroxy acid (Wang, Kang et al. 2012). After chain elongation, the linear destruxin precursor remains attached to the PCP domain in module 6. This module contains a C-terminal C_T domain, which is thought to catalyze the cyclization of the peptide through a mechanism that resembles the cyclization of cyclosporine (Lawen 1996). This C_T domain is special as it catalyzes ester bond formation while most C_T domains catalyze amide bond formation. A full biochemical and structural characterization of this domain is yet to be accomplished to elucidate the determinants of depsipeptide cyclization.

We identified several fungal depsipeptides in Taevernier's database (Taevernier, Wynendaele et al. 2017) that contain a single α-hydroxy acid in their structure, but most of their biosynthetic clusters have not been identified or characterized. We hypothesize that they are generated by a similar pathway and that biochemical and biophysical characterization of the relevant domains will deliver more information of these processes.

1.3.4.2 Ester-forming C domains and iterative C domains in enniatin, beauvericin and bassianolide biosynthesis.

Enniatin, beauvericin (Liuzzi, Mirabelli et al. 2017) and bassianolide (Xu, Orozco et al. 2009) are entomopathogenic fungal toxins. They are structurally related, as they are cyclic repeats of a dipeptidol monomer formed by an α-hydroxy acid and an amino acid. Beauvericin and the enniatins consist of three dipeptidol repeats, making them cyclic trimeric esters, while bassianolide consists of four dipeptidol repeats, making it a cyclic tetrameric ester. The biosynthetic clusters for enniatin, beauvericin and bassionalide have been characterized (Jirakkakul, Punya et al. 2008, Xu, Orozco et al. 2008, Xu, Orozco et al. 2009, Liuzzi, Mirabelli et al. 2017). Surprisingly, the three different NRPS systems have the same di-modular organization: C-A-PCP-C-A-M-PCP-PCP-C. The differences between each system rely on which substrates are selected by the A domains and how many repeats of the dipeptidol are added to the final product. The first A domain adenylates and thiolates the α-hydroxy acid, which will be condensed to the amino acid from module 2 generating the dipeptidol precursor. The linear dipeptidol undergoes oligomerization followed by cyclization through a still poorly understood mechanism.

Even though Taevernier's depsipeptide chemical classification (Taevernier, Wynendaele et al. 2017) clusters enniatin, bassianolide and beauvericin with bacterial depsipeptides such as antimycin, cereulide and valinomycin, the biosynthetic mechanisms between fungal and bacterial depsipeptide synthetases are strikingly different.

1.3.5. A-KR-PCP modules select α-ketoacids and reduce them to α-hydroxy acids in bacterial systems

Several bacterial depsipeptides have one (antimycin and cryptophycin **Fig. 1.6c**), three (respirantin **Fig. 1.6c**) or several hydroxy acids (cereulide and valinomycin **Fig. 1.6a**) in their structures, and their biosynthetic clusters contain one or more specialized modules with a minimal domain arrangement A-KR-PCP (**Fig. 1.14**) (Magarvey, Beck et al. 2006, Magarvey, Ehling-Schulz et al. 2006, Jaitzig, Li et al. 2014, Liu, Zhu et al. 2016). A-KR-PCP modules adenylate an α-keto acid in the specialized A domain, thiolate the α-keto acyl adenylate and reduce the

resulting α-ketoacyl thioester in a NADPH-dependent reaction in the KR domain, which is related to PKS KR domains (Magarvey, Ehling-Schulz et al. 2006). The resulting hydroxy acyl-*S*-PCP can participate in several reactions depending on the location of the A-KR-PCP module within the NRPS: as an acceptor substrate in ester-forming condensation reactions (Magarvey, Beck et al. 2006, Ding, Rath et al. 2011), as a donor substrate in canonical condensation reactions (Magarvey, Ehling-Schulz et al. 2006, Jaitzig, Li et al. 2014) or as a nucleophile in TE-catalyzed oligomerization and cyclization reactions (Magarvey, Ehling-Schulz et al. 2006, Jaitzig, Li et al. 2014). α-keto acids αKIC, α-KIV and α-KILE originate in bacteria from branched-chain amino acid biosynthesis, while pyruvic acid originates from glycolysis (Amorim Franco and Blanchard 2017).

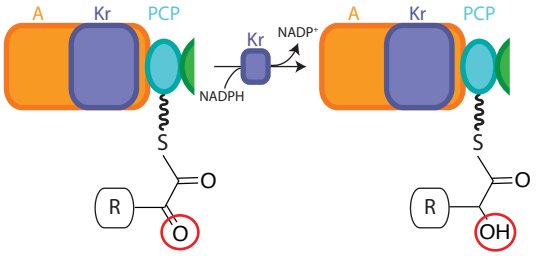


Figure 1.14. A-KR-PCP modules adenylate and reduce α-keto acids.

The genes responsible for valinomycin production were identified as early as 1990 (Perkins, Guterman et al. 1990), followed by two independent groups who sequenced and annotated the biosynthetic clusters of valinomycin (Cheng 2006) and cereulide (Ehling-Schulz, Vukov et al. 2005). These studies identified the unusual modules for the first time, but the mechanism of α-hydroxy acid incorporation remained unknown. Magarvey *et al*. sub-cloned the A-KR-PCP domains from cereulide synthetase (ces) and performed functional assays that revealed the biosynthetic pathway of A-KR-PCP modules (Magarvey, Ehling-Schulz et al. 2006). The authors showed that the A domain preferentially activates α-keto acids by testing adenylation activity against α-keto, hydroxy acids (L and D) and α-amino acids (L and D). Then, they verified that the PCP domain can form a ketoacyl thioester intermediate by evaluating the covalent incorporation of radioactive ketoacids to the module. After thiolation, the radioactive ketoacyl-loaded modules were incubated with NADPH to evaluate KR domain activity. The PCP-attached products of the

KR reaction were released with a type II TE domain and evaluated by radio-TLC against authentic α-hydroxy and α-keto acid standards, with results showing that the α-keto acids had been reduced to α-hydroxy acids.

One interesting question addressed in the 2006 study (Magarvey, Ehling-Schulz et al. 2006) regarding KR domains in A-KR-PCP modules is whether they were stereospecific, as the α-hydroxy acids present in cyclic depsipeptides such as cereulide and valinomycin have either D- or L-stereochemistry. By doing chiral chromatography of KR reaction products, Magarvey et al. discerned that the first module in CesA produced an α-D-hydroxy acid, whereas the second module produced an α-L-hydroxy acid. However, there are still several open questions on the reaction catalyzed by the KR domain.

PKS KR domains are pseudo-dimeric proteins formed by two subdomains, an N-terminal with a structural role and a C-terminal bearing the catalytic pocket (Keatinge-Clay and Stroud 2006). Both subdomains consist of a modified Rossmann fold proper of short chain dehydrogenases (SDRs), and only the catalytic subunit contains the triad Tyr-Ser-Lys needed for reduction activity (Reid, Piagentini et al. 2003, Keatinge-Clay and Stroud 2006, Keatinge-Clay 2007). KR domains in Ces and Vlm are homologous to PKS β-ketoreducing KR domains (Magarvey, Ehling-Schulz et al. 2006), which can be classified according to the stereochemistry of their product, A-type producing L-β-hydroxy acids and B-type producing D-β-hydroxy acids (Zheng, Taylor et al. 2010, Liu, Yuan et al. 2018). Regardless of the type of KR domain, NADPH binds with the 4'-*pro*-S hydrogen facing the active site Tyr and the face of the keto group presented to that hydrogen determines the stereochemistry of the reduced product (Yin, Gokhale et al. 2001). In turn, keto face positioning is dictated by interactions with the acyl carrier protein (the PCP domains of PKS systems) in a mechanism directed by contacts of ppant with one of either of two loops facing each other in the KR active site (Liu, Yuan et al. 2018). In A-type KRs, the guiding loop is characterized by a highly conserved Trp residue, while in B-type KRs it is distinguished by an LDD sequence motif (Caffrey 2003). It was still unknown whether the KR domains in depsipeptide synthetases followed the same structural determinants of substrate selectivity.

In 2006 and 2007, annotation of the cryptophycin (Magarvey, Beck et al. 2006) and kutzneride (Fujimori, Hrvatin et al. 2007) biosynthetic gene clusters revealed the presence of A-KR-PCP modules, which add hydroxy acids by the same mechanism as vlm and ces. These and more recent studies (Labby, Watsula et al. 2015) suggested that KR domains were inserted between motifs A8 and A9 of the A_{sub} domain, creating an interrupted domain arrangement as observed in the recently solved structure of a A-M construct (Mori, Pang et al. 2018). However, no structural evidence was available to support this hypothesis in A-KR-PCP modules.

Biochemical assays showing α-substituent discrimination by A-KR-PCP modules have shown that they are selective for α-keto groups (Magarvey, Ehling-Schulz et al. 2006, Fujimori, Hrvatin et al. 2007), with the keto acid-selecting Stachelhaus code suggesting that the substitution of residue Asp235 (based on 1amu) by an aliphatic residue could play an important role (Fujimori, Hrvatin et al. 2007), but there were no studies on the exact chemical consequences. McyG is a promiscuous A domain that selects unusual phenylpropanoids where the canonical Asp235 is also substituted by an aliphatic residue. Structural studies indicated that the substitution in McyG could enhance unusual hydrophobic monomer binding (Tan, Dai et al. 2015); however, the presented structure was bound to the α-amino acid L-Phe and not an unusual phenylpropanoid, giving no insight into the actual structural determinants of binding of non-amino acid monomers.

The α-hydroxy acyl thioesters produced by A-KR-PCP modules can be used in different downstream reactions, which will be discussed in the following sections.

1.3.5.1. α-hydroxy acyl thioesters as acceptor substrates in ester-bond forming C domains.

Cryptophycin and kutzneride bear a single hydroxy acid in their structures, and in agreement with the co-linearity rule, a single A-KR-PCP elongation module is present in their biosynthetic clusters (Magarvey, Beck et al. 2006, Fujimori, Hrvatin et al. 2007). Cryptophycin is a depsipeptide with tubulin-depolymerizing properties, whose synthetic analogues have anticancer properties (Magarvey, Beck et al. 2006). A hybrid PKS-NRPS system consisting of CrpA-C produces and relays

a linear intermediate to CrpD (**Fig. 1.15**), the final di-modular NRPS in the cluster with a C-A-PCP-C-A-KR-PCP-TE architecture. The first module (CrpD1: C-A-PCP) adds β-alanine to the growing PKS-peptidyl chain, while the second module (CrpD2: C-A-KR-PCP-TE) adds α-hydroxyisocaproic acid (α-HIC) through α-keto acid activation and reduction in the A-KR domains. The α-hydroxyacyl-*S*-PCP intermediate acts as an acceptor substrate in the ester-forming condensation reaction catalyzed by the CrpD2 C domain. The second ester bond in the cryptophycin structure arises from macrocyclization in a TE domain-catalyzed reaction, using the hydroxy group in the polyketide as a nucleophile.

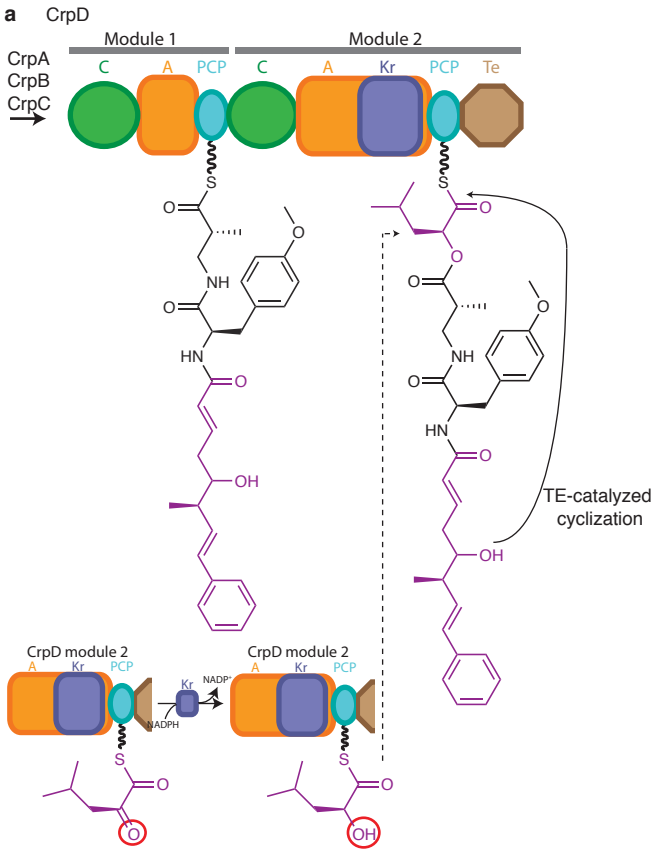


Figure 1.15. Cryptophycin uses an A-KR-PCP module for α-hydroxy acid incorporation.

Kutznerides are hexadepsipeptides with antimicrobial and antifungal properties(Broberg, Menkis et al. 2006, Fujimori, Hrvatin et al. 2007). Nine variants with different substituents throughout the core structure have been identified, where the single hydroxy acid can be either α-

hydroxyisovaleric acid or 2-Hydroxy-3,3-dimethylbutanoic acid (Fujimori, Hrvatin et al. 2007). The proposed biosynthesis of kutzneride starts with the production of 2-(1-methylcyclopropyl)-D-glycine (MecPGly) through the action of KtzB, D, A, F. Then, the module encoded by KtzE activates MecPGly. The stand-alone A-KR-PCP module KtzG generates an α-hydroxy acyl thioester, which then will be used as an acceptor substrate in the ester-forming condensation reaction in the C domain of KtzE (**Fig. 1.16**).

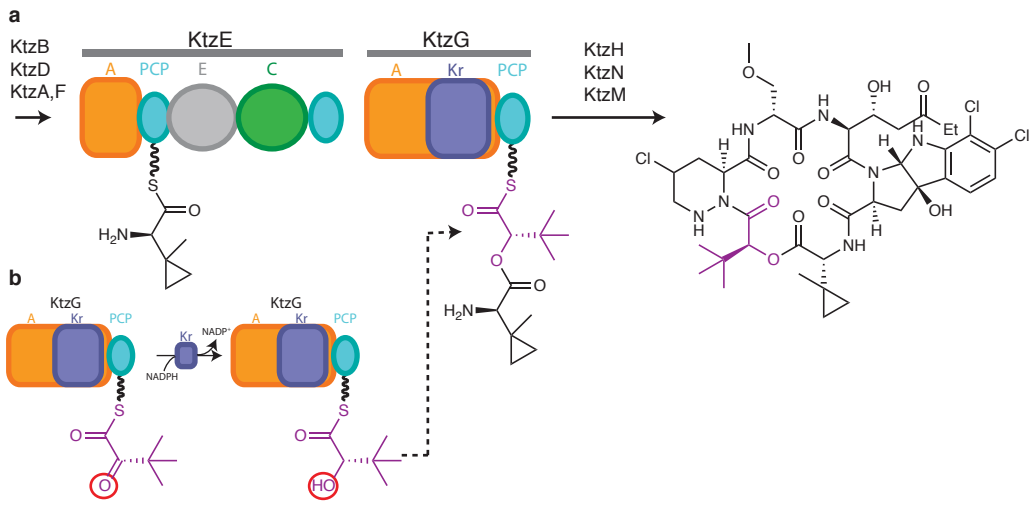


Figure 1.16. Kutzneride biosynthesis.

a. Kutznerides incorporate α-keto acids using an A-KR-PCP module (**b**).

The ester-bond forming C-domains in kutzneride and cryptophycin biosynthesis have not been fully characterized and open questions remain. Other ester-bond forming C-domains have been identified and partially characterized in non-depsipeptide systems, such as the one involved in fumonisin biosynthesis (Zaleta-Rivera, Xu et al. 2006), that catalyzes ester bond formation between a donor aconitic acid and an acceptor hydroxylated polyketide. Characterized ester bond-forming C domains bear the HHXXXD motif, indicating that the reaction mechanism is the same as amide bond-forming C domains.

1.3.5.2 α-hydroxy acyl thioesters as nucleophiles in iterative oligomerizing/cyclizing TE domains

Cereulide and valinomycin are dodecadepsipeptides formed by alternating arrays of α-amino and α-hydroxy acids (Perkins, Guterman et al. 1990, Toh, Moffitt et al. 2004, Ehling-Schulz, Vukov et

al. 2005, Matter, Hoot et al. 2009). They can also be described as trimers of a tetradepsipeptide unit. In spite of their strong resemblance to the fungal products enniatin, beauvericin and bassianolide, their biosynthesis is entirely different (**Fig. 1.17**).

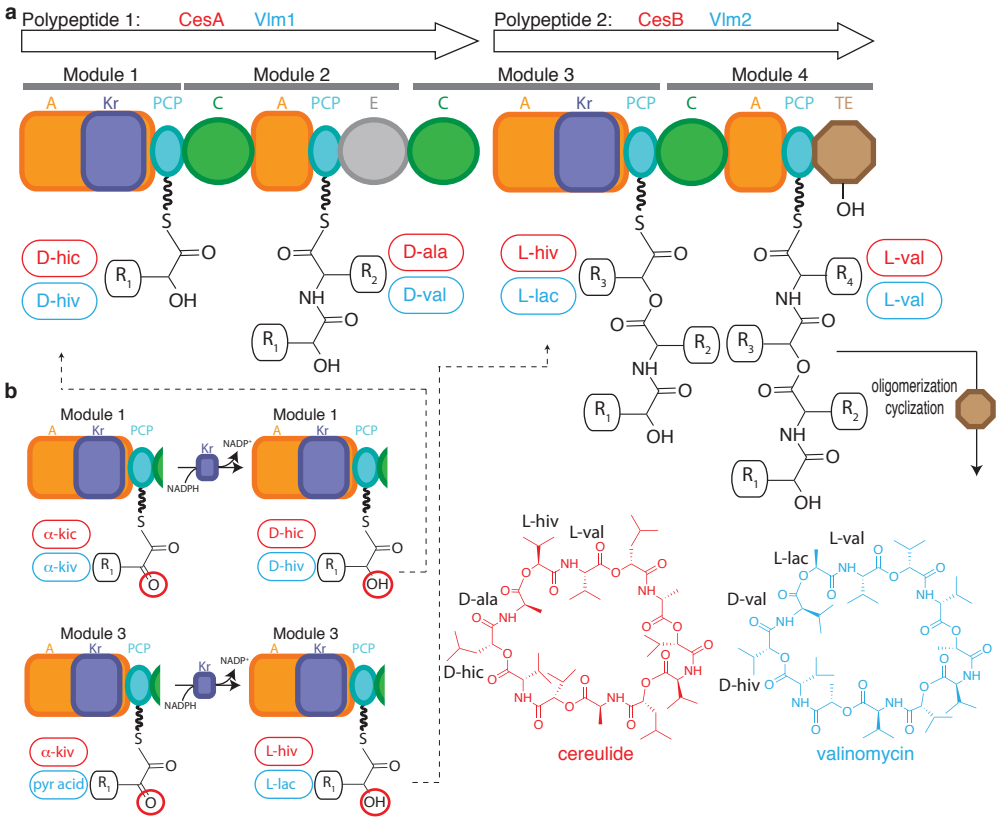


Figure 1.17. Biosynthesis of cereulide and valinomycin.

a. The biosynthesis of cereulide and valinomycin involves α-hydroxy acid incorporation through A-KR-PCP modules (**b**) and an iterative TE domain that can also catalyze cyclization.

The cereulide and valinomycin biosynthetic clusters consist of two di-modular NRPSs; CesA and CesB for cereulide, Vlm1 and Vlm2 for valinomycin (Magarvey, Ehling-Schulz et al. 2006) (**Fig. 1.17a**). CesA and Vlm1 have the architecture A-KR-PCP-C-A-PCP-E-C, while CesB and Vlm2 are constituted as A-KR-PCP-C-A-KR-PCP-TE. Initiation modules CesA1, Vlm1-1 and elongation modules CesB1 and Vlm2-1, activate and reduce α-ketoacids through the concerted action of their A-KR domains (**Fig. 1.17b**), demonstrated with biochemical assays for Ces (Magarvey, Ehling-Schulz et al. 2006). However, the fate of each generated acyl thioester is different depending on whether the module is in an initiation or elongation position. In the proposed

biosynthesis of cereulide, which can be extrapolated to that of valinomycin, the α-D-HIC-*S*-PCP intermediate in initiation module CesA1 acts as the donor substrate in the condensation reaction catalyzed by the first C domain in elongation module CesA2, using L-Ala-*S*-PCP as acceptor substrate. Based on the mechanism of E domains described on (**1.2.5.1. Epimerization**), epimerization of L-Ala to D-Ala in the D-HIC-L-Ala-*S*-PCP intermediate occurs next, forming a D-HIC-D-Ala-*S*-PCP. Elongation module CesB1 produces a α-D-HIV-*S*-PCP, which acts as the acceptor substrate in the ester-bond forming condensation reaction catalyzed by the C domain at the C-terminus of CesA2, using D-HIC-D-Ala-*S*-PCP as donor. The tridepsipeptidyl chain will then be elongated in the C domain of CesB2 with L-Val, forming a D-HIC-D-Ala-L-HIV-L-Val-*S*-PCP tetradepsipeptide intermediate. This tetradepsipeptide intermediate would then be oligomerized and cyclized by the terminal TE domain, through a proposed mechanism resembling the iterative cyclizing peptidyl (gramicidin S) (Hoyer, Mahlert et al. 2007) or depsipeptidyl (enterobactin) (Shaw-Reid, Kelleher et al. 1999) TE domains described in previous sections. However, the mechanism and structural determinants of the oligomerization and cyclization reactions remained unknown.

1.4. Objectives and overview

Ces and Vlm synthetases employ different mechanisms to generate their cyclic products. These include reactions in canonical NRPS modules, but also several steps that differ from the canonical pathway (**Fig. 1.17**). Their initiation and elongation A-KR-PCP modules (**Fig. 1.17b**) and the iterative TE domain are of particular interest. Even though some of the isolated modules of these NRPSs have been characterized (Magarvey, Ehling-Schulz et al. 2006, Fujimori, Hrvatin et al. 2007), there were several open questions on the biochemistry and structure of these enzymes, some of which have been addressed in this dissertation.

In chapter 2, I was able to express the intact components of Ces, CesA and CesB, in the heterologous host *E. coli* and to develop robust purification protocols with sufficient yields and purity for biochemical or structural characterization. Secondary objectives included performing

analysis of Michaelis-Menten kinetics of each A domain, evaluating side chain selectivity in keto acid-selecting A domains, elucidating if activity required a chromosomally-encoded MbtH-like protein, analyzing inhibition by ketoacid versions of vinylsulfonamide inhibitors and reconstituting cereulide production *in vitro*.

In chapter 3, using the expression and purification protocols from chapter 2, I and collaborators determined the di-domain A-KR structure from a homologue of the initiation module CesA1. The structure is in the adenylation conformation, revealing that the A and KR domains are 60 Å apart and that massive conformational rearrangements would be needed throughout the biosynthetic cycle. Strikingly, the structure also shows that the A domain is not interrupted, as it contains an integral A_{sub} domain connected to the KR domain through a short linker. Furthermore, an additional and unexpected subdomain located between the KR and PCP domains participates in domain swapping and may play a role during the conformational transitions of the module. The KR domain structure provided information on the stereochemistry of the ketoreduction reaction, specifically on the arrangement of the active site loops responsible for ketoacyl-ppant presentation. Another objective of chapter 3 was to answer which were the structural determinants of α-keto group selectivity. I determined the high-resolution structure of the α-keto acid-selecting A domain bound to its cognate α-ketoacyl adenylate. The structure shows that the α-keto group interacts with the active site through a carbonyl-carbonyl interaction, with no hydrogen bonds or other kind of polar contacts involved.

In chapter 4, the objective was to obtain structural information on the oligomerization and cyclization reaction of Vlm TE. Several structures of TE domains from PKS and NRPS systems have shown that the dynamic active site lid may be involved in substrate binding, ppant contacts, and cyclization (Bruner, Weber et al. 2002) (Samel, Wagner et al. 2006, Frueh, Arthanari et al. 2008); however, obtaining structures of intermediate-bound TE domains has been hampered by their inherent instability (Bruner, Weber et al. 2002, Samel, Wagner et al. 2006). In a collaborative work, we determined the structures of apo and intermediate-bound Vlm TE, using a genetic code expansion strategy and substrate analogues developed by our collaborators in the United

Kingdom and the University of Ottawa, respectively. The structures show that the lid undergoes a massive conformational change to accommodate the longest linear intermediate in a cavity that restricts its conformational flexibility, acting as an entropic trap favoring the depsipeptide cyclization conformation to release the mature product. Furthermore, the obtained results are a proof of concept for the genetic code expansion strategy, indicating that it could be employed to elucidate unanswered questions of other serine proteases.

CHAPTER 2. CHARACTERIZATION OF CEREULIDE SYNTHETASE, A TOXIN-PRODUCING MACROMOLECULAR MACHINE

Alonzo DA, Magarvey NA, Schmeing TM. (2015) Characterization of cereulide synthetase, a toxin-producing macromolecular machine. PlosOne, https://doi.org/10.1371/journal.pone.0128569

2.1. Abstract

Cereulide synthetase is a two-protein nonribosomal peptide synthetase system that produces a potent emetic toxin in virulent strains of *Bacillus cereus*. The toxin cereulide is a depsipeptide, as it consists of alternating aminoacyl and hydroxyacyl residues. The hydroxyacyl residues are derived from keto acid substrates, which cereulide synthetase selects and stereospecifically reduces with imbedded ketoreductase domains before incorporating them into the growing depsipeptide chain. We present an *in vitro* biochemical characterization of cereulide synthetase. We investigate the kinetics and side chain specificity of α-keto acid selection, evaluate the requirement of an MbtH-like protein for adenylation domain activity, assay the effectiveness of vinylsulfonamide inhibitors on ester-adding modules, perform NADPH turnover experiments and evaluate *in vitro* depsipeptide biosynthesis. This work also provides biochemical insight into depsipeptide-synthesizing nonribosomal peptide synthetases responsible for other bioactive molecules such as valinomycin, antimycin and kutzneride.

Keywords: cereulide synthetase; nonribosomal peptide synthetase; NRPS; adenylation; reduction; enzyme; depsipeptide; keto acid; *in vitro* characterization

2.2. Introduction

Cyclic depsipeptides are natural products made of diverse acyl moieties linked by amide and ester bonds (**Fig. 2.1C, D**) which have a broad range of biological and medical activities. Cyclic depsipeptides include ionophores, quorum sensing modulators, toxins and antibiotics (Stawikowski and Cudic 2007, Kjaerulff, Nielsen et al. 2013, Jaitzig, Li et al. 2014). Some examples are the anticancer agent valinomycin, the biopesticide bassianolide, the piscicide antimycin, the antihelminthic PF1022A, the anti-fungal kutzneride, and cereulide, on which this work is focused (Agata, Mori et al. 1994, Fujimori, Hrvatin et al. 2007, Xu, Orozco et al. 2009, Sandy, Rui et al. 2012, Jaitzig, Li et al. 2014, Li, Jaitzig et al. 2014).

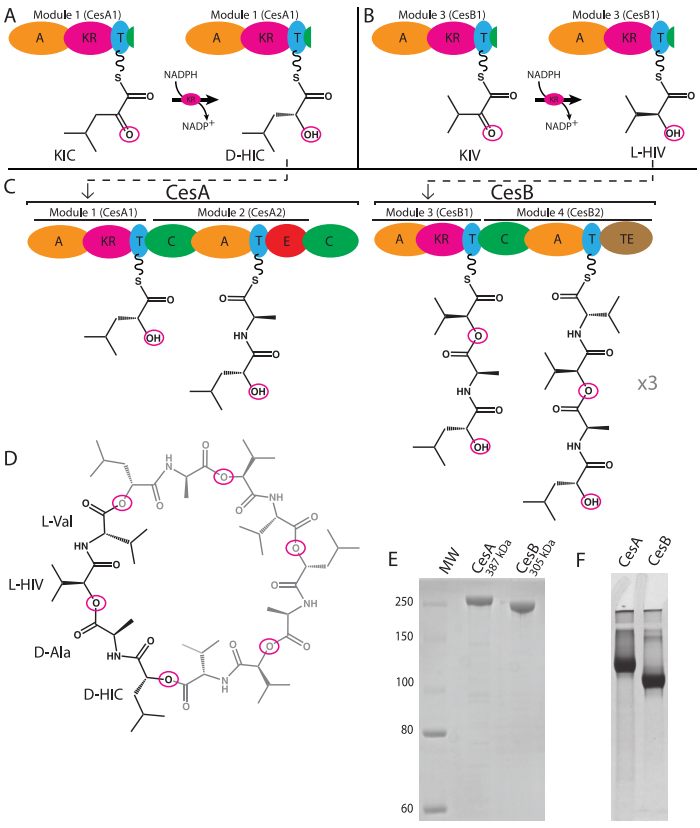


Figure 2.1. Cereulide synthetase produces the emetic toxin cereulide.

(A and B) Modules CesA1 and CesB1 contain KR domains which catalyze the reduction of bound keto acyl groups. (C) Schematic diagram of cereulide synthetase and the synthesis of *D*-HIC—*D*-Ala—*L*-HIV—

L-Val, which is trimerized to produce mature cereulide (D). (E) Denaturing and (F) native PAGE of NRPS proteins CesA and CesB. CesA and CesB migrate slightly faster than expected from their molecular masses. Domain abbreviations: A, adenylation; KR, ketoreductase; T, thiolation; C, condensation; E, epimerization; TE, thioesterase.

The cereulide toxin (**Fig. 2.1D**) is the causative agent in severe food poisoning associated with emetic strains of *Bacillus cereus*, which can lead to acute liver failure and death (Kotiranta, Lounatmaa et al. 2000, Dierick, Van Coillie et al. 2005, Stenfors Arnesen, Fagerlund et al. 2008). The main mechanism of cereulide poisoning is an emetic syndrome mediated by the serotonin 5-HT3 receptor and by stimulation of the vagus afferent nerve (Agata, Ohta et al. 1995). This toxin also exerts cytotoxicity by disrupting the mitochondrial membrane potential through its K^+ ionophoric activity (Mikkola, Saris et al. 1999). Cereulide is highly resistant to heat, extreme pH, and proteases and remains a health hazard even if the cereulide-producing bacteria are killed by thorough cooking of contaminated food (Rajkovic, Uyttendaele et al. 2008). The native role of the toxin for *B. cereus* is likely as a siderophore, as its expression increases the fitness of the microorganism in potassium deprived environments (Ekman, Kruglov et al. 2012).

Emetic *B. cereus* strains synthesize cereulide through the action of cereulide synthetase (**Fig. 2.1C**), a heterodimer of the proteins CesA and CesB (Ehling-Schulz, Fricker et al. 2006). These are non-ribosomal peptide synthetase (NRPS) proteins, modular enzymes that employ assembly-line synthetic mechanisms. Each module of an NRPS adds one monomer to the growing peptide chain. The domain arrangement of a canonical NRPS module, such as module CesB2 (**Fig. 2.1C**), consists of a condensation (C), an adenylation (A), and a thiolation domain (T). The A domain selects and adenylates an amino acid substrate, then attaches it via a thioester bond to the prosthetic phosphopantetheine arm of the T domain. The T domain then transports the bound substrate to the C domain, where it is incorporated into the growing peptide chain by amide bond formation (**Fig. 2.1C**; reviewed in (Cane, Walsh et al. 1998, Marahiel 2009, Strieker, Tanovic et al. 2010)). Because CesB2 is a termination module, it contains an extra domain not found in elongation modules, the thioesterase (TE) domain, which releases the mature nonribosomal peptide by cyclization or hydrolysis. NRPSs frequently display variations of the canonical domain

arrangement, including substitutions of canonical domains, and insertion of tailoring domains (Walsh, Chen et al. 2001, Condurso and Bruner 2012), like the epimerization (E) domain found in CesA2.

Modules CesA1 and CesA2 have a domain arrangement and mechanism exclusive to depsipeptide synthetases (Magarvey, Ehling-Schulz et al. 2006) (**Fig. 2.1A-2.1B**). Magarvey *et al.* showed that the specialized A domain in these modules selects a specific α-keto acid, which is then ligated to the T domain arm (Magarvey, Ehling-Schulz et al. 2006). The T domain next transports the tethered α-keto acid to a ketoreductase (KR) domain present in the module. This ~45 kDa domain is embedded in a mobile hinge of the A domain and catalyzes the stereospecific reduction of the α-keto acyl group to an α-hydroxy acyl group, using NADPH as a redox cofactor. The α-hydroxy-acyl-T domain is then brought to the C domain, which catalyzes ester bond formation with the upstream aminoacyl-T domain. This strategy for incorporation of α-hydroxy acids is also found in the synthesis of valinomycin, antimycin and kutzneride (Fujimori, Hrvatin et al. 2007, Sandy, Rui et al. 2012, Jaitzig, Li et al. 2014), while direct selection of α-hydroxy acids occurs in the fungal synthesis of PF0122A and bassianolide (Weckwerth, Miyamoto et al. 2000, Xu, Orozco et al. 2009).

Based on the structure of cereulide, the demonstrated substrate of the A domains and the standard NRPS synthetic logic, module CesA1 is assumed to add *D*-α-hydroxyisocaproic acid (HIC), module CesA2 to add *D*-alanine (Ala), module CesB1 to add *L*- α-hydroxyisovaleric acid (HIV), and module CesB2 to add *L*-valine (Val) to form a tetrapeptide intermediate of cereulide (*D*-HIC—*D*-Ala—*L*-HIV—*L*-Val). This tetrapeptide is presumably then passed to a serine residue in the TE domain. A second tetrapeptide is made by the upstream domains and coupled to the first, making an octapeptide, after which a third is made and coupled in the same way. The TE then cyclizes the dodecapeptide intermediate to release mature cereulide, in the manner similar to that described for the biosynthesis of cyclic peptides such as gramicidin S (Hoyer, Mahlert et al. 2007) and surfactin (Bruner, Weber et al. 2002).

To increase our understanding of the biosynthesis of cereulide, we have expressed and purified cereulide synthetase and performed a biochemical characterization. We have compared the kinetics and side chain specificity of α-keto acid selection to those of amino acid selection, investigated the requirement for an MbtH-like protein for the activity of the NRPS, assayed the effectiveness of vinylsulfonamide inhibitors on ester-adding modules, and performed *in vitro* peptide synthesis assays. This work provides insight into the functioning of all depsipeptide-synthesizing NRPSs, including valinomycin synthetase, kutzneride synthetase and the depsipeptide synthetases of the antimycin family (Li, Zvanych et al. 2013, Vanner, Li et al. 2013).

2.3 Results and discussion

2.3.1. The cereulide synthetase subunits can be expressed in *Escherichia coli* and purified to homogeneity

We developed robust expression and purification protocols of the intact NRPSs CesA and CesB, and of the excised first modules of CesA and CesB (designated CesA1 and CesB1, see **Fig. 2.1A-2.1B**), which contain the domain sequence adenylation-ketoreductase-thiolation (A-KR-T) (Magarvey, Ehling-Schulz et al. 2006). We ensured the physical integrity of the proteins by denaturing and native gel electrophoresis (**Fig. 2.1E-2.1F**), as well as by dynamic light scattering and size exclusion chromatography.

2.3.2. The apparent Michaelis constants for cognate keto acids and cognate amino acids

We performed a kinetic characterization of each of the adenylation domains in CesA and CesB to compare the keto acid–activating A domains to the amino acid–activating A domains. Two commonly used assays for adenylation are a radioactive inorganic pyrophosphate (PPi) – ATP exchange assay (Cole and Schimmel 1970) and a pyrophosphate production assay (McQuade, Shallop et al. 2009, Wilson and Aldrich 2010). The PPi–ATP exchange assay reflects both the forward and reverse rates of the adenylation reaction, as [^{32}P]ATP is generated by the reverse reaction using a product (AMP) of the forward reaction and exogenous [^{32}P]PPi. In contrast,

pyrophosphate production assays reflect only the forward rate because the signal arises from PPi produced during adenylation. It has been reported that two assays give different k_{cat} and K_m values, but that the apparent k_{cat}/K_m is always similar (Wilson and Aldrich 2010). We performed both assays with each of the purified proteins and their predicted substrates (**Table 2.1, Fig. 2.2, and Fig. 2.3**).

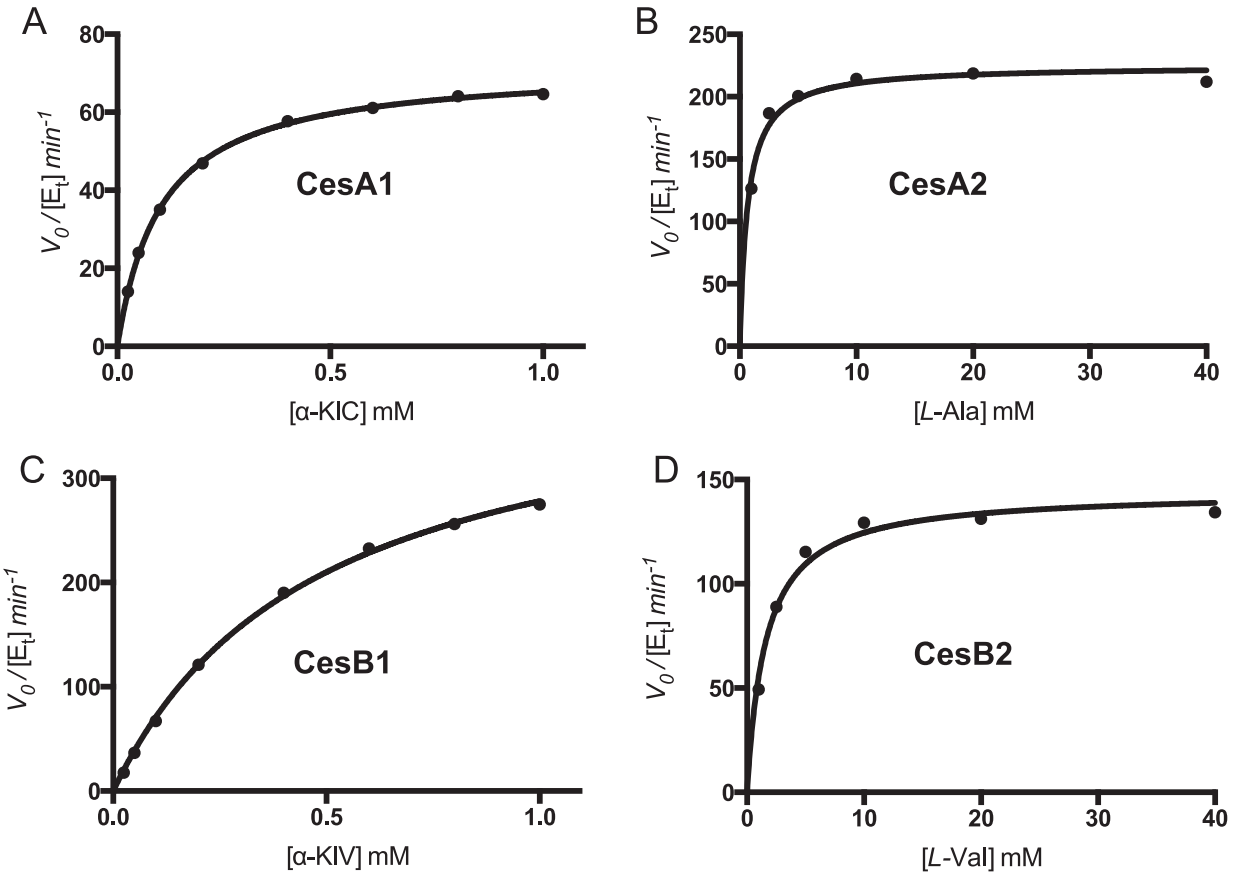


Figure 2.2. Kinetic characterization of adenylation by A domains using the ATP-PPi exchange assay. Initial velocity versus substrate concentration plots for adenylation of cognate substrates for CesA1 (A), CesA2 (B), CesB1 (C) and CesB2 (D). Curves were fit to the Michaelis-Menten equation. The kinetic parameters obtained are listed in Table 1.

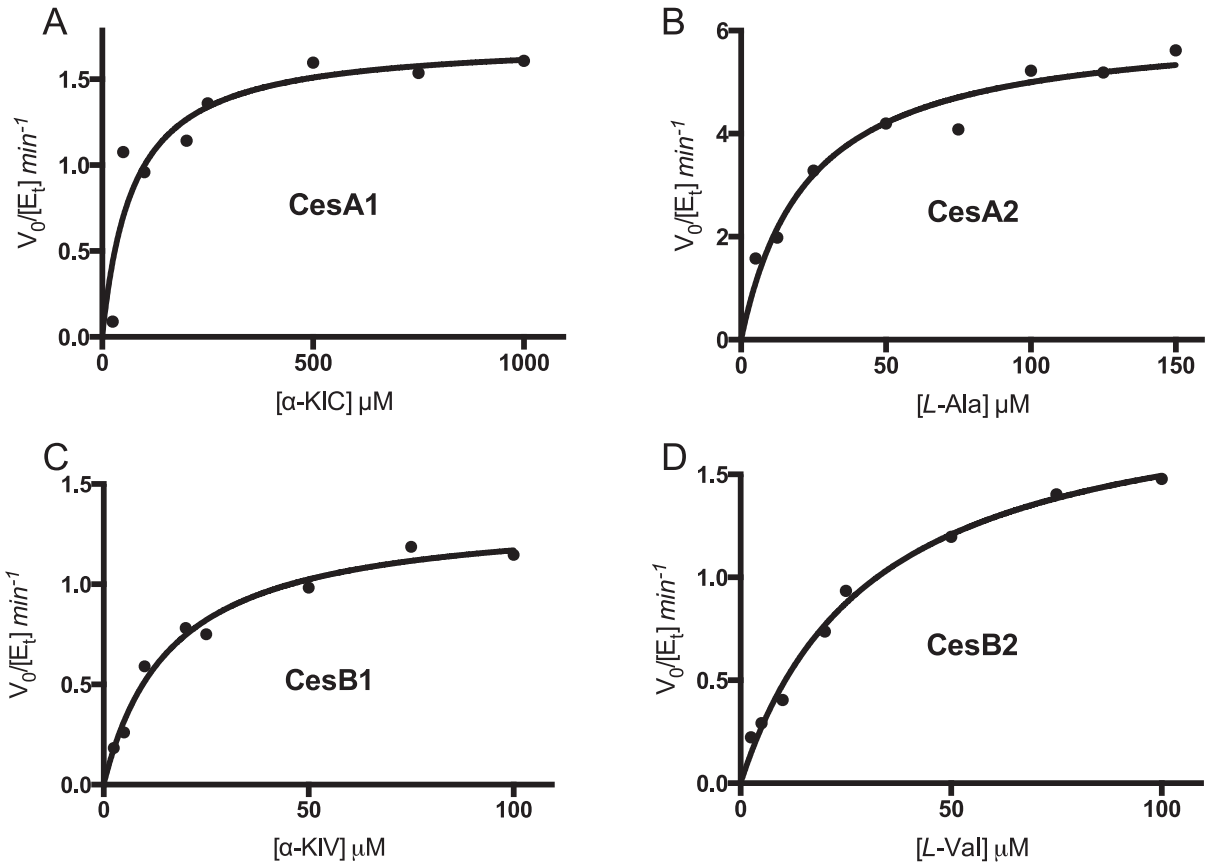


Figure 2.3. Kinetic characterization of adenylation by A domains using the pyrophosphate production assay.

Initial velocity versus substrate concentration plots for adenylation of cognate substrates for CesA1 (A), CesA2 (B), CesB1 (C) and CesB2 (D). Curves were fit to the Michaelis-Menten equation. The kinetic parameters obtained are listed in Table 1.

Assay	Module	Substrate	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m (min^{-1} μM^{-1})
ATP-PPi exchange assay	CesA1	α-KIC	103.0±2.5	71.9±0.4	0.70±0.02
	CesA2	*L*-Ala	673.5±102.6	225.0±5.2	0.33±0.05
	CesB1	α-KIV	479.3±2.5	411.1±9.4	0.86±0.02
	CesB2	*L*-Val	1592±225	144.3±4.3	0.09±0.01
Pyrophosphate-production assay	CesA1	α-KIC	22.4±2.3	3.2±0.1	0.14±0.02
	CesA2	*L*-Ala	23.0±4.9	6.2±0.4	0.27±0.06
	CesB1	α-KIV	16.5±2.6	1.4±0.1	0.08±0.01
	CesB2	*L*-Val	30.9±3.7	2.0±0.1	0.06±0.01

Table 2.1. Apparent catalytic constants of adenylation by cereulide synthetase A domains using the ATP-PPi exchange and pyrophosphate production assays.

Both assays showed that the enzymes were active in adenylation, with apparent k_{cat} and K_m values in the range reported for other adenylation enzymes with cognate substrates (Keating, Marshall et al. 2000, Wilson and Aldrich 2010, Wilson, Shi et al. 2013). Comparison of the two assays shows no trend with respect to k_{cat} or K_m. Furthermore, while the apparent k_{cat}/K_m of CesA2 with *L*-Ala and CesB2 with *L*-Val are similar across the two assays (CesA2: 0.33 $min^{-1}\mu M^{-1}$ vs 0.27 $min^{-1}\mu M^{-1}$; CesB2: 0.09 $min^{-1}\mu M^{-1}$ vs 0.06 $min^{-1}\mu M^{-1}$), they are not similar for CesA1 with α-ketoisocaproic acid (α-KIC) or CesB1 with α-ketoisovaleric acid (α-KIV) (CesA1: 0.70 $min^{-1}\mu M^{-1}$ vs 0.14 $min^{-1}\mu M^{-1}$; CesB1: 0.86 $min^{-1}\mu M^{-1}$ vs 0.08 $min^{-1}\mu M^{-1}$). There appears to be no fundamental reason that the apparent k_{cat}/K_m should be the same across two disparate assays, which report on different aspects of the reaction (forward reaction vs forward and reverse reaction). Wilson & Aldrich also saw dissimilarity in an A domain (Wilson and Aldrich 2010). We expect that as more adenylating proteins are evaluated with both methodologies, the previously observed k_{cat}/K_m similarity will be shown to be coincidental.

2.3.3. An MbtH-like protein is not required for cereulide production

We evaluated whether an MbtH-like protein is required for full adenylation activity in cereulide synthetase, as is the case with some other NRPSs (Felnagle, Barkei et al. 2010, Boll, Taubitz et al. 2011). No MbtH-like protein is encoded in the *ces* operon (Ehling-Schulz, Fricker et al. 2006), but *B. cereus* always contains an MbtH-like protein in the genome. We assayed the activity of each adenylation domain in CesA and CesB in the presence and absence of purified *B. cereus* MbtH-like protein. No enhancement of the adenylation activity was observed for any A domain in CesA or CesB (**Fig. 2.4**), and we conclude that an MbtH-like protein is not involved in cereulide biosynthesis.

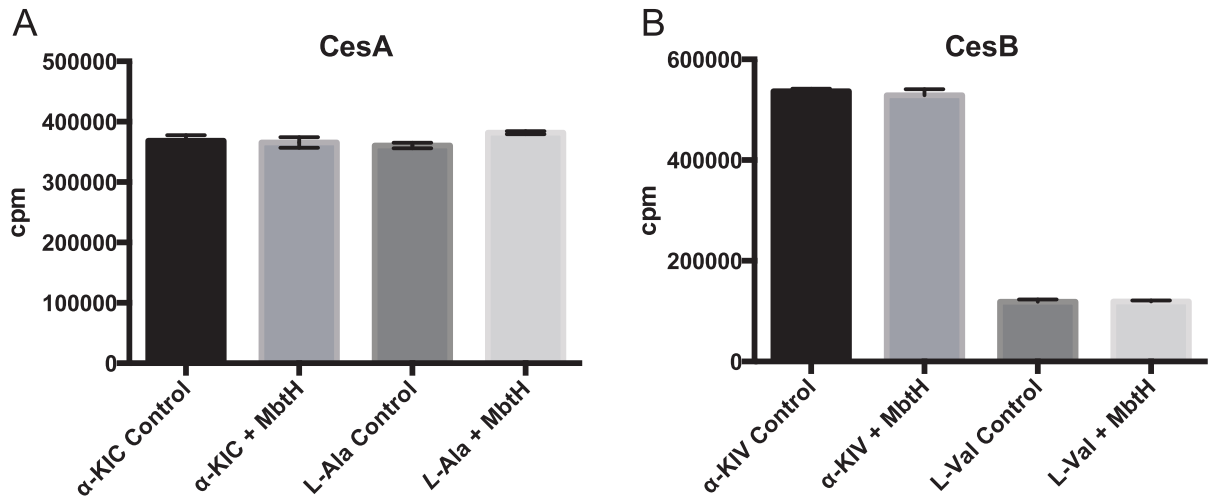


Figure 2.4. The MbtH-like protein of *B. cereus* does not affect the adenylation reactions of cereulide synthetase.

The adenylation reaction of each A domain in CesA (A) and CesB (B) with the cognate substrate was evaluated by the ATP-PPi exchange assay at a single point of substrate, with or without MbtH in the reaction, in triplicate. Data is expressed in counts per minute (cpm).

2.3.4. CesA1 and CesB1 display keto acid side chain selectivity

The specialized A domains in CesA1 (**Fig. 2.1A**) and CesB1 (**Fig. 2.1B**) have selectivity towards an α-keto over an α-amino or an α-hydroxy group in their substrates (Magarvey, Ehling-Schulz et al. 2006). However, the side chain selectivity of the α-keto acid-selecting modules had not yet been probed. We assayed the adenylation activity of CesA1 and CesB1 with α-keto acids that contain side chains present in common amino acids (**Fig. 2.5**, and **S2.1 Figure**): α-KIC, α-KIV, α-ketoisoleucine (α-KIL), pyruvic acid and oxaloacetic acid, which are the keto acid versions of leucine, valine, isoleucine, alanine and aspartic acid respectively. CesA1 shows high side chain selectivity (**Fig. 2.5A**), whereas CesB1 can activate its canonical substrate α-KIV and can also activate α-KIL (**Fig. 2.5B**).

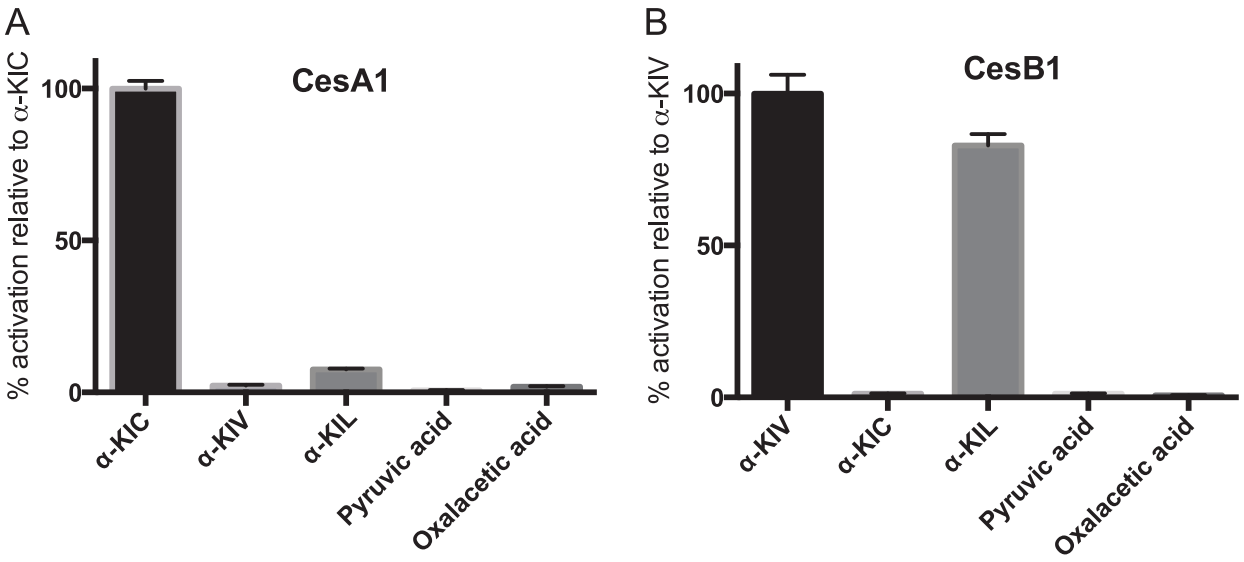


Figure 2.5. Evaluation of α-keto acid side chain selectivity of cereulide synthetase.

The ATP-PPi exchange assay was performed with CesA1 (A) and CesB1 (B) and various α-keto acids as substrates, in triplicate. See S1 Figure for a comprehensive list of the monomers, their abbreviations, and their structures. Data was normalized to the activity obtained with the cognate substrate.

That CesB1 can adenylate α-KIL was not unexpected. Pitchayawasin *et al.* reported that *B. cereus* produces homologues of cereulide which contain an α-HIC at the position usually occupied by α-HIV (**Fig. 2.1D**) (Pitchayawasin, Isobe et al. 2004), but the same data would be explained by the presence of α-hydroxyisoleucine (α-HIL), an isomer of α-HIC. In our assay, CesB1 selects α-KIL, but not α-KIC; we propose that this cereulide homologue contains an α-HIL derived from an α-KIL substrate (**Fig. 2.5B**). Whether the lower selectivity of CesB1 is advantageous for producing cereulide homologues with altered activity is of ongoing interest. Some synthetically produced homologues of cereulide show a marked decrease in toxicity towards HEp2-cells while keeping their ionophoric activity (Makarasen, Nishikawa et al. 2009), illustrating the potential utility of homologues of cereulide.

2.3.5.Depsipeptide synthetases can be inhibited by vinylsulfonamide inhibitors

The KR domain is inserted into a loop near the C-terminus of the A domain (Magarvey, Ehling-Schulz et al. 2006), but it is unclear how it is spatially accommodated into the module or what domain rearrangements must take place during α-keto acid processing. Structural studies could

address these questions, but obtaining X-ray structures of multidomain NRPS constructs is challenging, due to their flexibility and intrinsic conformational heterogeneity. Successful structural studies have typically relied on preparing conformationally homogeneous protein by either mutating the conserved serine of the T domain (Tanovic, Samel et al. 2008) or by locking the enzyme in a single conformation with mechanism based inhibitors (Liu, Zheng et al. 2011, Mitchell, Shi et al. 2012, Sundlov, Shi et al. 2012). Aminoacyl-vinylsulfonamide adenylate analogues have enabled structure determination of A-T didomain constructs, providing valuable information about conformational changes in the adenylation-thiolation cycle (Mitchell, Shi et al. 2012, Sundlov, Shi et al. 2012).

We asked whether ketoacyl-vinylsulfonamide adenylate analogues would inhibit cereulide synthetase. CesA1 was incubated with α-hydroxyisocaproic acyl-vinylsulfonamide adenylate (**Fig. 2.6A**) and its adenylation activity evaluated. The protein could be inhibited by 90%, but this required a vast molar excess of the inhibitor (**Fig. 2.6B**). Aminoacyl vinylsulfonamide adenylate compounds have been shown to have equally low inhibition efficiencies (Qiao, Wilson et al. 2007). The observed inhibition likely resulted from the established mechanism of covalent binding between the α-ketoacyl-vinylsulfonamide and the T domain's phosphopantetheine arm: Inhibition was not observed with CesA1 that was missing the phosphopantetheine modification because of mutation of the serine attachment site or lack of co-expression with the pantetheinyl transferase Sfp. Stalling of NRPSs using mechanism-based inhibitors is thus not restricted to canonical NRPSs and holds promise to facilitate crystallization of α-keto acid-selecting modules.

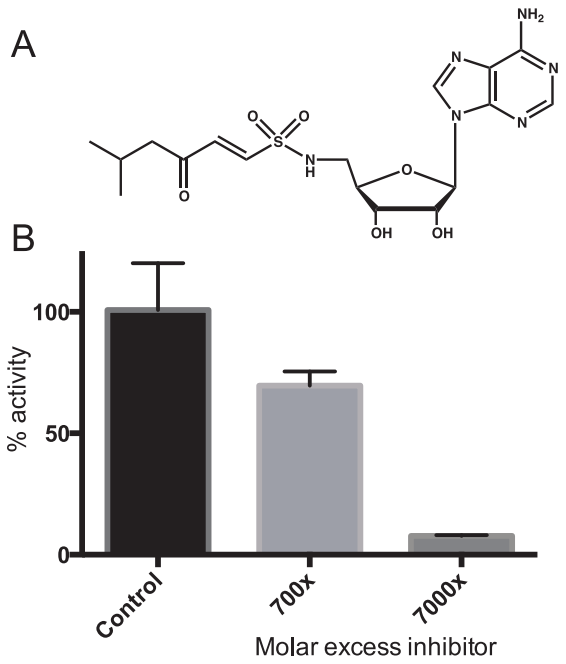


Figure 2.6. Vinylsulfonamide compounds inhibit α-keto acid adenylation.

(A) Structure of the vinylsulfonamide inhibitor for CesA1, α-hydroxyisocaproic acyl-vinylsulfonamide adenylate. The A domain should catalyze the nucleophilic attack of the pantetheine arm thiol on the vinylsulfonamide analogue, but the adenine analogue should not be released in the reaction, trapping the T domain with the A domain. (B) CesA1 is inhibited by the vinylsulfonamide inhibitor. CesA1 was incubated with DMSO or the inhibitor in DMSO. Excess inhibitor and/or DMSO was removed with a desalting column prior to ATP-PPi exchange assay using α-KIC.

2.3.6. CesA and CesB bind NADPH with a micromolar affinity

The KR domains in CesA and CesB catalyze the stereospecific reduction of the α-ketoacyl-T domain to their corresponding α-hydroxyacyl-T domain (**Fig. 2.1A-2.1B**), using NADPH as a cofactor. We assayed NADPH binding to KR domains by fluorescence emission (**Fig. 2.7**). The K_d was close to 1 µM for both proteins (**Fig. 2.7**). If the intracellular NADPH concentration in *B. cereus* is similar to that in *E. coli* (120 µM) (Bennett, Kimball et al. 2009) or *B. subtilis* (44 µM) (Soga, Ohashi et al. 2003, Maass, Sievers et al. 2011), NADPH binding would not be a bottleneck in cereulide synthesis. This K_d value is similar to those found for other ketoreductases, such as the β-ketoacyl reductase domains found in vertebrate fatty acid synthases (Dugan and Porter 1971).

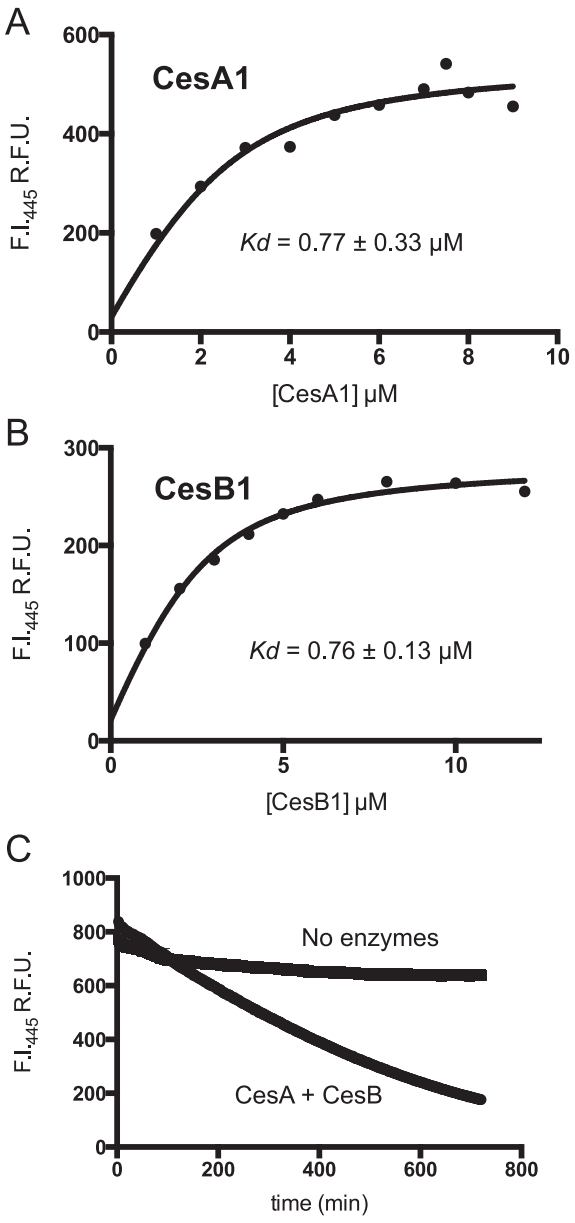


Figure 2.7. NADPH affinity and turnover in the KR domains in CesA1 and CesB1.

(A,B) Fluorescence intensity enhancement is plotted against protein concentration using a fixed concentration of NADPH (2.5 µM); λ_{exc}=340nm and λ_{emm}=445nm for CesA1 and CesB1. Curves were fitted to a Morrison equation that includes terms for enhancement of NADPH fluorescence upon binding. (C) Enzymatic NADPH consumption as reported by fluorescence intensity (λ_{exc}=340nm and λ_{emm}=445nm).

2.3.7. *In vitro* peptide synthesis assays of cereulide synthetase

We next incubated CesA and CesB with substrates and cofactors *in vitro* and assayed NADPH consumption and peptide synthesis by mass spectrometry. NADPH was consumed when both enzymes were present (**Fig. 2.7C**), and to a lesser extent by CesA alone. Peptide synthesis and control reactions were then analyzed by HPLC-MS. We could not detect *m/z* peaks corresponding to full-length cyclic cereulide. However, we did detect masses consistent with cereulide precursors dipeptide **1** and dipeptide **2**, which were verified by comparison to authentic standards (CanPeptide Inc, Pointe-Claire, QC, Canada), as well as tetrapeptide **3** (**Fig. 2.8**) and, in some samples, octapeptide **4** (**S2.5 Figure**). Thus, CesA and CesB are working in concert, but not efficiently enough to produce full cereulide against competing rates of hydrolysis (nor to produce octapeptide **4** in every *in vitro* reaction assay). The tetrapeptide **3** and octapeptide **4** precursors are likely lost from the TE domain by hydrolysis before they can be extended to a dodecapeptide and cyclized to cereulide. TE domains are known to have fairly rapid hydrolysis rates *in vitro* (Wang, Opare et al. 2009). Challenges in getting full-length cereulide production is unsurprising, as few NRPS systems that comprise two or more subunits have been reconstituted *in vitro* (Keating, Marshall et al. 2000, Gaitatzis, Kunze et al. 2001, Miller, Luo et al. 2002, Sattely, Fischbach et al. 2008) because they are highly complex systems. *In vitro* synthesis of cereulide or any similar depsipeptide is not yet reported, though a related depsipeptide, valinomycin, has been heterologously produced in *E. coli* (Jaitzig, Li et al. 2014, Li, Jaitzig et al. 2014). Detection of tetrapeptide **3** and octapeptide **4** provides hope that reaction conditions could eventually be optimized to demonstrate *in vitro* cereulide production.

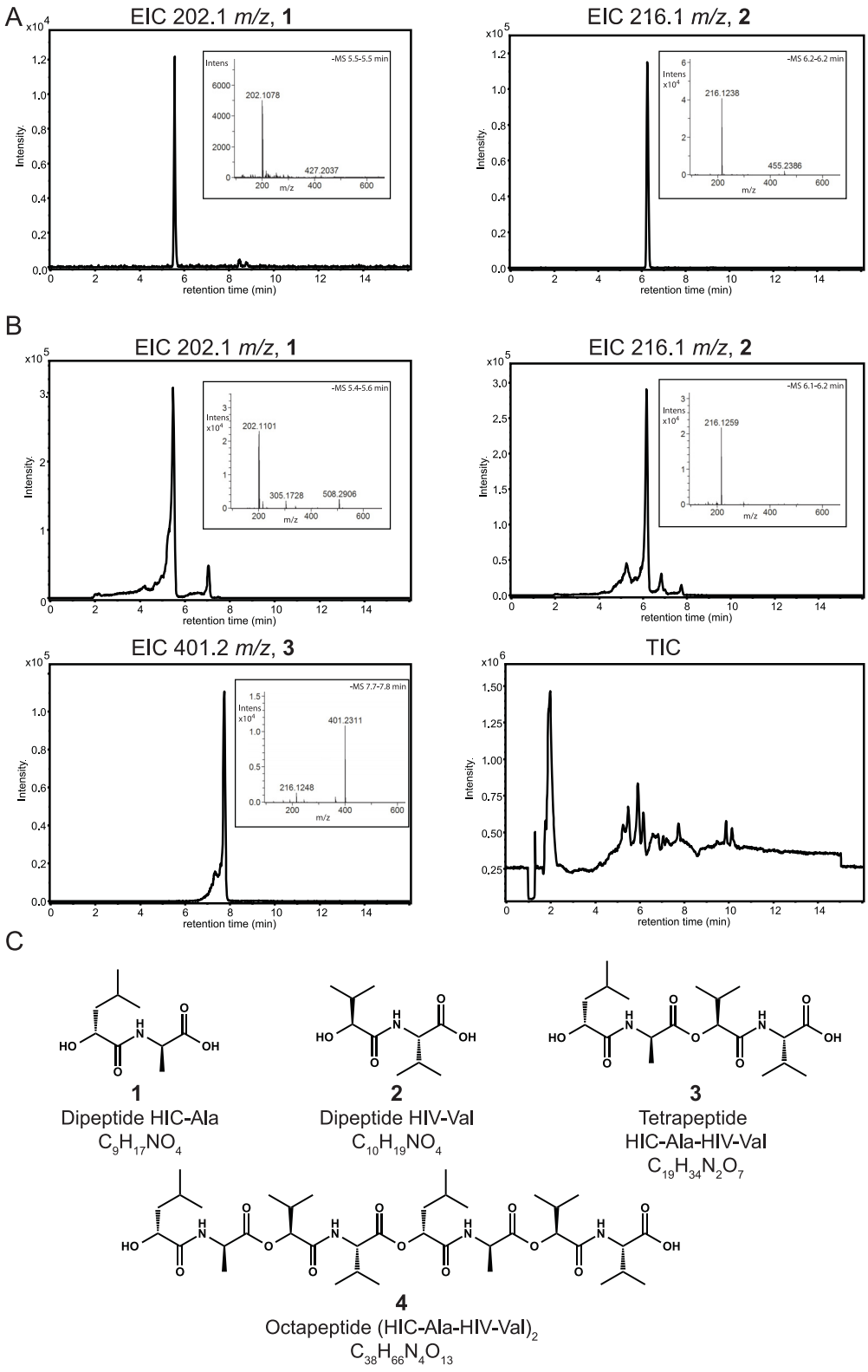


Figure 2.8. LC-MS analysis of the peptide synthesis reaction of *in vitro*-reconstituted cereulide synthetase.

Extracted ion chromatograms (EIC), from (A) synthetic standards of dipeptides **1** and **2**, and (B) from the products of a peptide synthesis reaction of CesA, CesB and substrates. EICs are extracted using exact *m/z* values (±0.05 mass units) calculated from the $[M-H]^-$ ions of the compounds shown in (C). Mass spectra corresponding to predominant peaks are shown as insets. (C) Putative chemical structures of the cereulide intermediates detected in the peptide synthesis reaction. The structures are consistent with exact masses and determined molecular formulas, as well as the elution profile of the authentic standards of dipeptides. The shown stereochemistry of the tetra- and octapeptides (**3** and **4**) are consistent with their precursor molecules (**1** and **2**) and their successor molecule (cereulide). However, these putative chemical structures are not absolutely proven here.

Sample	Compound	Retention time (min)	Measured m/z	Calculated m/z $[M-H]^-$	Error
Standards (Fig. 8A)	Dipeptide HIC-Ala **(1)**	5.5	202.1078	202.1085	3.5
	Dipeptide HIV-Val **(2)**	6.2	216.1238	216.1241	1.7
Enzymatic reaction (Fig. 8B)	Dipeptide HIC-Ala **(1)**	5.4-5.6	202.1101	202.1085	-8.2
	Dipeptide HIV-Val **(2)**	6.1-6.2	216.1259	216.1241	8.2
	Tetradepsipeptide HIC-Ala-HIV-Val **(3)**	7.7-78	401.2311	401.2293	-4.4

Table 2.2. MS analysis of the peptide synthesis reaction of *in vitro*-reconstituted cereulide synthetase. Mass spectra were extracted for the predominant peaks in Fig 8. Molecular formulae were assigned by exact mass and isotopic pattern analysis (S3 Table).

2.4. Conclusions

The components of cereulide synthetase can be expressed in *E. coli* and purified to homogeneity. Upon purification, all adenylation domains remain active and select the monomers predicted from the chemical structure of cereulide, with kinetic constants similar to those of other NRPSs. There are no substantial kinetic differences in activation of keto acids as compared to amino acids. CesA1 shows high side chain selectivity, and CesB1 can efficiently adenylate its canonical substrate α-KIV as well as α-KIL. Cereulide synthetase does not require an MbtH-like protein for its reaction cycle. Vinylsulfonamide inhibitors can inhibit α-keto acid-selecting modules, which could aid efforts to obtain crystal structures of these modules. Together, CesA and CesB were

shown to bind and consume NADPH and produce their dipeptides, and the tetrapeptide and octapeptide precursors of cereulide. The characterization of cereulide synthetase presented here and future work with cereulide synthetase could also be relevant for other NRPSs that make useful depsipeptide compounds, such as valinomycin synthetase, antimycin synthetase, kutzneride synthetase, and others yet to be identified.

2.5. Methods

Reagents

All reagents were purchased from BioShop Canada (Burlington, ON, Canada), unless specified. ATP (A7699) and α-ketoisoleucine (198978) were purchased from Sigma-Aldrich (Oakville, ON, Canada); α-ketoisovaleric acid sodium salt (522198105) and oxaloacetic acid (210056805), from MP Biomedicals (Santa Ana, CA, USA); radioactive inorganic pyrophosphate (NEX-019), from Perkin Elmer (Waltham, MA, USA). All restriction enzymes, T7 DNA ligase and Phusion DNA polymerase were purchased from New England Biolabs (Ipswich, MA, USA). All fast protein liquid chromatography media were purchased from GE Healthcare (Mississauga, ON, Canada) unless otherwise specified.

Cloning of CesA, CesB, CesA1 and CesB1

Genes for full-length CesA and CesB were cloned from *B. cereus* F4810/72 genomic DNA into plasmids pColdI-CesA-Tandem and pColdI-CesB-Tandem. These pColdI-Tandem plasmids encode an N-terminal octa-histidine tag and tobacco etch virus (TEV) protease recognition sequence, and a C-terminal TEV site and calmodulin binding peptide sequence (KRRWKKNFIAVSAANRFKKISSSGAL). The gene sequence for CesA1 (coding for amino acid residues 1-1323) and for CesB1 (amino acid residues 1-1354) were subcloned from the above plasmids into pColdI-Tandem to make pColdI-Tandem-CesA1 and pColdI-Tandem-CesB1. Mutations inserted during the cloning process were corrected by site-directed mutagenesis.

Protein over-expression and purification

CesA, CesA1, CesB and CesB1 were co-expressed with the promiscuous phosphopantetheinyl transferase Sfp, encoded in a p15A-based plasmid kindly provided by Dr. Christian Chalut and Dr. Christophe Guilhot (Chalut, Botella et al. 2006). Expression of CesA and CesA1 was in *E. coli* BL21 (DE3), and of CesB and CesB1, in *E. coli* soluBL21 (DE3) (Genlantis, San Diego, CA, USA), grown in LB medium supplemented with 50 µg/ml ampicillin and 34 µg/ml chloramphenicol. For all proteins, cultures were induced at an optical density of ~0.45 with 100 µM isopropyl β-D-1-

thiogalactopyranosid (IPTG) and further incubated for 24h or 48h at 16°C. Cells were harvested and resuspended in their buffer A. For CesA and CesA1, buffer A consisted of 50 mM Tris pH 8.0, 250 mM NaCl, 2 mM βME, 2 mM $CaCl_2$, 0.1 mM PMSF, 35% v/v glycerol, 10 mM imidazole. For CesB and CesB1, buffer A consisted of 50 mM Tris pH 8.0, 300 mM NaCl, 2 mM βME, 2 mM $CaCl_2$, 0.1 mM PMSF, 10% v/v glycerol, 10 mM imidazole. Cells were lysed by sonication, and the lysate was cleared by centrifugation at 18 000 rpm in a JA-25.50 rotor (Beckman-Coulter, Brea, CA). Supernatant was loaded onto a 5ml HiTrap-HP Ni^{2+} column pre-equilibrated with buffer A. Proteins were eluted with buffer which had the same composition but a total of 150 mM imidazole. The eluted protein was then loaded into a calmodulin sepharose 4B column pre-equilibrated with buffer A. The protein was eluted with buffer C, which consisted of 50 mM Tris pH 8.0, 300 mM NaCl, 2 mM βME, 2 mM EGTA, 0.1 mM PMSF and 35% (for CesA1 and CesA) or 10% (for CesB1 or CesB) v/v glycerol. CesB1 and CesA1 elutions were incubated with TEV protease at ratio 1:2 in dialysis against a buffer consisting of 50 mM Tris pH 8.0, 300 mM NaCl, 2 mM βME, 0.1 mM PMSF, 10% v/v glycerol before the sample was reapplied to the calmodulin sepharose 4B column and the Ni^{2+}-IMAC column. The flowthrough was loaded onto a HiPrep Sephacryl S-300 HR 26/60 column pre-equilibrated with 25 mM HEPES pH 8.0, 100 mM NaCl and 0.2 mM tris (2-carboxyethyl) phosphine (TCEP). Due to inefficient TEV cleavage, CesA samples used for most of this work were not incubated with TEV protease. Using tagless CesA did not increase the rate of NADPH turnover assays. After the calmodulin column, CesA was dialyzed overnight against 25 mM HEPES pH 8.0, 100 mM NaCl and 0.2 mM TCEP. Enzyme concentration was measured with the Bradford assay. Purified protein was flash frozen in liquid nitrogen and stored at -80°C.

The sequence encoding for the MbtH-like protein from *B. cereus* American Type Culture Collection 10987 was synthesized de novo by DNA 2.0 (Menlo Park, CA, USA) and cloned into a pet28-based vector bearing an N-terminal octa-histidine tag and TEV protease recognition sequence. The protein was overexpressed in *E. coli* BL21 (DE3) in LB media. Cultures were induced at an optical density of 0.5 with 100 µM IPTG, then further incubated at 16°C for 24h. The protein was purified as described in (Drake, Cao et al. 2007), flash frozen in liquid nitrogen and stored at -80°C.

ATP-PPi exchange assay

Reactions of volume of 100 µl consisted of 75 mM Tris pH 8.0, 10 mM $MgCl_2$, 0.2 mM TCEP, 5 mM ATP, 7% glycerol, 1 mM of unlabelled inorganic pyrophosphate, 0.5 µCi of ^{32}P labeled inorganic pyrophosphate, stated concentration of substrate and 20 - 100 nM enzyme. Reactions were incubated at room temperature for 20 minutes and then stopped with a solution of 1.6% w/v activated charcoal, 4.5% w/v inorganic pyrophosphate and 5% v/v perchloric acid. The resulting suspension was centrifuged at 14 000 rcf, the supernatant was removed, and the pellet was washed twice with 500 µl of 4.5% w/v inorganic pyrophosphate, 5% v/v perchloric acid. The pellet was resuspended in 500 µl of the wash solution and transferred to scintillation vials and 5 ml of scintillation fluid was added. Incorporated radioactivity was quantified in a Perkin Elmer scintillation counter. Total moles of produced ATP were used to calculate k_{cat} and replotted as a function of substrate concentration. Experiments were performed over multiple substrate concentrations as single experiments and analysed with non-linear regression using the Michaelis-Menten equation in the program Prism 6 (GraphPad Software, San Diego, California).

Triplicate reactions were performed with the MbtH-like protein at saturating concentrations of substrate and a 100x molar excess of purified MbtH-like protein. Absence of YBDZ, the endogenous MbtH-like protein from *E. coli,* was confirmed by denaturing polyacrylamide gel electrophoresis analysis and tryptic digest mass spectrometry analysis. Furthermore, CesA1 produced from a ΔYBDZ strain of *E. coli* (Boll, Taubitz et al. 2011), has the same kinetic profile as that from a strain not containing the ΔYBDZ mutation. Keto acid selectivity assays were performed in triplicate with keto acid concentration of 10 mM.

The pyrophosphate production assay

Pyrophosphate production assay (Wilson and Aldrich 2010) was performed with the EnzChek® Pyrophosphate Assay Kit (Invitrogen, Carlsbad, CA). Reactions of 100 µl consisted of 75 mM Tris pH 8.0, 10 mM $MgCl_2$, 0.2 mM TCEP, 5 mM ATP, 7% glycerol, 0.2 mM MESG, 1 unit ml^{-1} purine nucleoside phosphorylase, 0.03 units ml^{-1} inorganic pyrophosphatase, 150 mM hydroxylamine,

varied substrate concentration and 250 nM enzyme. Absorbance at 360 nm was recorded as a function of time in a SpectraMax M5 plate reader (Molecular Devices, Sunnyvale, CA). Slopes were calculated for the linear range of the reaction and converted and plotted as $k_{cat,}$ as a function of substrate concentration. Experiments were performed over multiple substrate concentrations as single experiments and analysed with non-linear regression using the Michaelis-Menten equation in the program Prism 6 (GraphPad Software, San Diego, California).

Vinylsulfonamide inhibition assays

α-hydroxyisocaproic acyl-vinylsulfonamide adenylate was synthesized on commission by Zamboni Chemical Solutions, Montreal, QC, Canada. CesA1 (2.8 µM) was incubated with 700x or 7000x molar excess of inhibitor in DMSO, or DMSO control, on ice for 4h, or with 100X molar excess inhibitor or DMSO at 30°C for 6h. After incubation, samples were applied to through a High Trap Desalting column to remove excess unbound inhibitor and/or DMSO. The protein concentration was re-quantified and the PPi-ATP exchange assay was performed in triplicate.

NADPH binding assays

Reactions of 100 µl volume in 96 well plates (Corning Incorporated, Corning, NY) contained 75 mM Tris pH 8.0, 0.2 mM TCEP, 7% v/v glycerol and 2.5 µM NADPH, and varied concentrations of CesA1 or CesB. After 10 minutes, fluorescence emission spectra were measured using an excitation wavelength of 340 nm and an emission wavelength of 445 nm in a SpectraMax M5 plate reader (Molecular Devices, Sunnyvale, CA). A control without NADPH was performed for each protein concentration. Fluorescence arising from the enzyme controls was subtracted from the observed fluorescence of experimental reactions. The resulting fluorescence at 445 nm was defined as $F_{mixture}$ and replotted as a function of enzyme concentration. Single experiments were performed with multiple enzyme concentration points. The resulting curve was analyzed by nonlinear regression with an equation that encompasses the Morrison equation and the molar ratio of bound NADPH expressed as a function of fluorescence enhancement (Dugan and Porter 1971).

NADPH consumption assay

Triplicate NADPH consumption reactions of 100 µl volume consisted of 75 mM Tris pH 8.0, 0.2 mM TCEP, 10 mM $MgCl_2$, 5 mM ATP, 10 mM α-KIC, 20 mM L-Ala, 10 mM α-KIV, 20 mM L-Val, and 100 µM NADPH in black polystyrene 96 well plates for fluorescence detection or 200 µM NADPH in clear polystyrene 96 well plates for absorbance detection. A mixture of CesA and CesB at equimolar concentrations (9 µM) was added and reactions were incubated for 15 minutes to achieve steady state conditions before measuring fluorescence emission at 460 nm using an excitation wavelength of 340 nm, or absorbance at 340 nm, as a function of time. Control reactions lacked keto acids and amino acids, or enzyme. The decrease in absorbance at 340 nm after completion of an independent reaction showed that the observed changes in both fluorescence and absorbance of NADPH are indeed due to enzymatic conversion to $NADP^+$.

Peptide detection by LC-MS

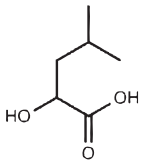
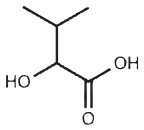
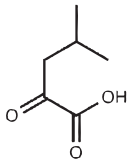
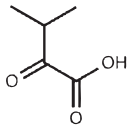
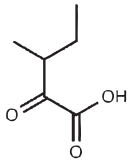
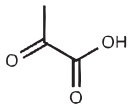
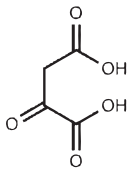
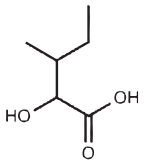
NADPH consumption assay reactions were pooled and methanol was added to a final concentration of 50% v/v. Precipitated protein was removed by centrifugation at 20 000 rpm for 15 minutes. The supernatant was recovered and centrifuged for an additional 15 minutes. LC-MS experiments were performed at the McGill Department of Chemistry Mass Spectrometry *Facility* (MSF) (Montreal, QC, Canada). Liquid chromatography was performed using a Dionex Ultimate 3000 UHPLC system with a Waters Xterra MS C8 3.5 µm, 2.1×150 mm column at 55°C, a mobile phase A of 200 mM ammonium acetate in water and a mobile phase B of 200 mM ammonium acetate in methanol. Peptides were eluted by a gradient of 90% A to 0% A over 7 minutes at a flow rate of 0.2 ml/min, followed by a further 5 minutes at 0% A. Six minutes of equilibration and one blank injections between samples were performed to ensure that there was no cross contamination. Peptide masses were analysed by an attached Bruker Maxis Impact quadrupole-time of flight (QTOF) mass spectrometer by ESI in negative and positive ionization modes with source nitrogen gas at 250 °C and 8 liters per minute. The nebulizer pressure was set at 2.0 Bar psi and capillary voltage was set a 4.5 kV. Data was collected in full scan mode (mass range: m/z 100-2000; scan time: 1 Hz). External calibration was done on each sample from an intra run

infusion at the beginning of the each analysis using Agilent ESI tune mix. Data was analyzed using the program Bruker DataAnalysis software. Exact masses were generated with the SmartFormula^{TM} tool using a 5 ppm mass window. Molecular formulae for each predominant peak were assigned by performing exact mass analysis (Table 2), isotopic pattern analysis, and rbd analysis using the SmartFormula^{TM} tool (S3 Table). The molecular formulae corresponded to the $[M-H]^-$ ions of dipeptide **1**, dipeptide **2,** and tetrapeptide **3** (Fig 8, S3 Table). Furthermore, dipeptides **1** and **2** from the enzymatic reaction elute at the same retention time as their corresponding authentic standards (Fig. 8 and Table 2). Authentic standards were synthesized on commission by CanPeptide Inc, Pointe-Claire, QC, Canada, and verified by ^{1}H NMR at the Quebec/Eastern Canada High Field NMR Facility (S7 Fig.).

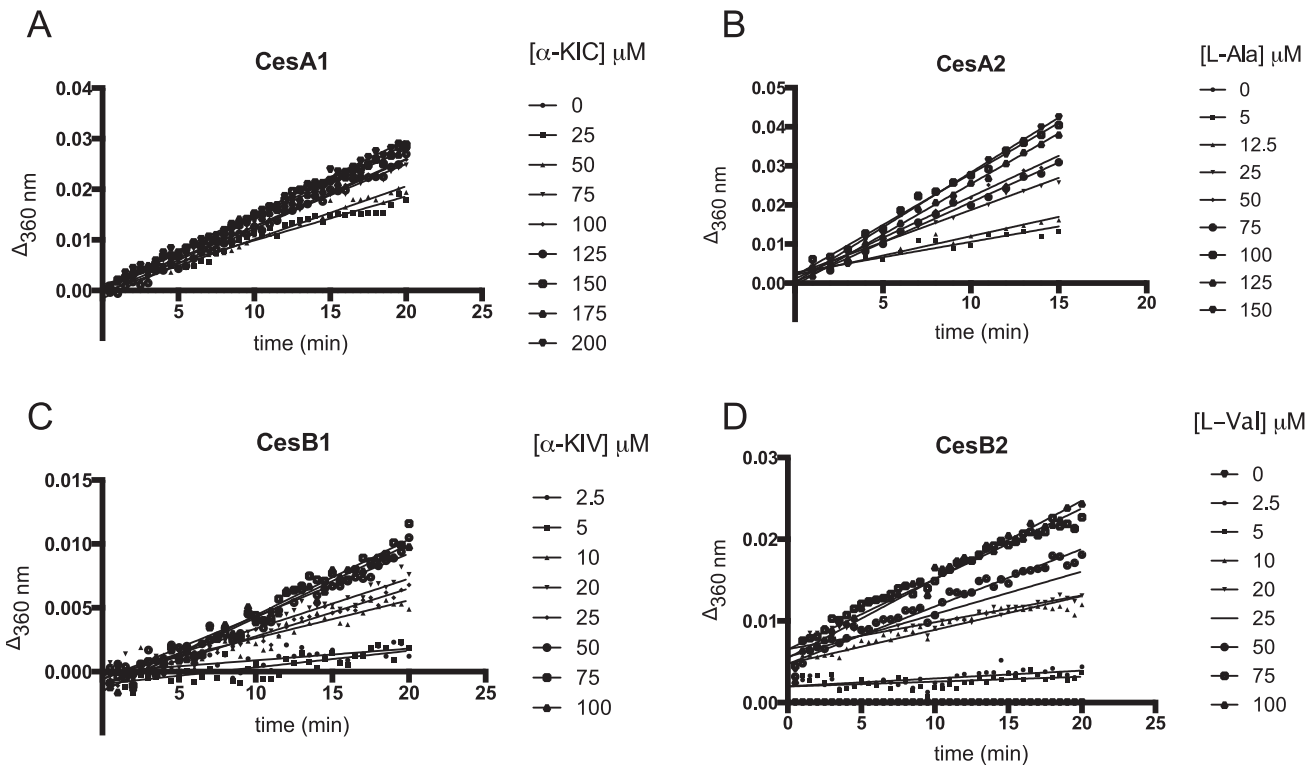
2.6. Acknowledgements

We thank Janice Reimer, Alex Wahba, Thierry Ntimbane, Gaëlle Bridon, Orval Mamer, Daina Avizonis, Chris Dejong and Chad Johnston for further help with mass spectrometry, Tara Sprules for NMR analysis, Lutz Heide for the gift of the ΔYBDZ strain, Michael Tarry, Fabien Bergeret, Kris Bloudoff and Martin Aloise for helpful discussion. We are indebted to John Colucci and Robert Zamboni (Zamboni Chem Solutions) for synthesis of the vinylsulfonamide inhibitor. We thank Carlos E. Flores Tinoco for his critical review of the first draft of the manuscript. We thank José Manuel Camacho Zaragoza for providing assistance in the subcloning and expression trials of CesA. We thank Christian Chalut and Christophe Guilhot for kindly providing the plasmid harbouring the gene for Sfp.

2.7. Supporting Information

Abbreviation	Definition	Structure
α-HIC	D-α-hydroxyisocaproic acid	
α-HIV	L-α-hydroxyisovaleric acid	
α-KIC	α-ketoisocaproic acid	
α-KIV	α-ketoisovaleric acid	
α-KIL	α-ketoisoleucine	
pyr	pyruvic acid	
oaa	oxaloacetic acid	
α-HIL	α-hydroxyisoleucine	

S2.1 Figure. Abbreviations, names and schematics of the monomers mentioned in the manuscript.

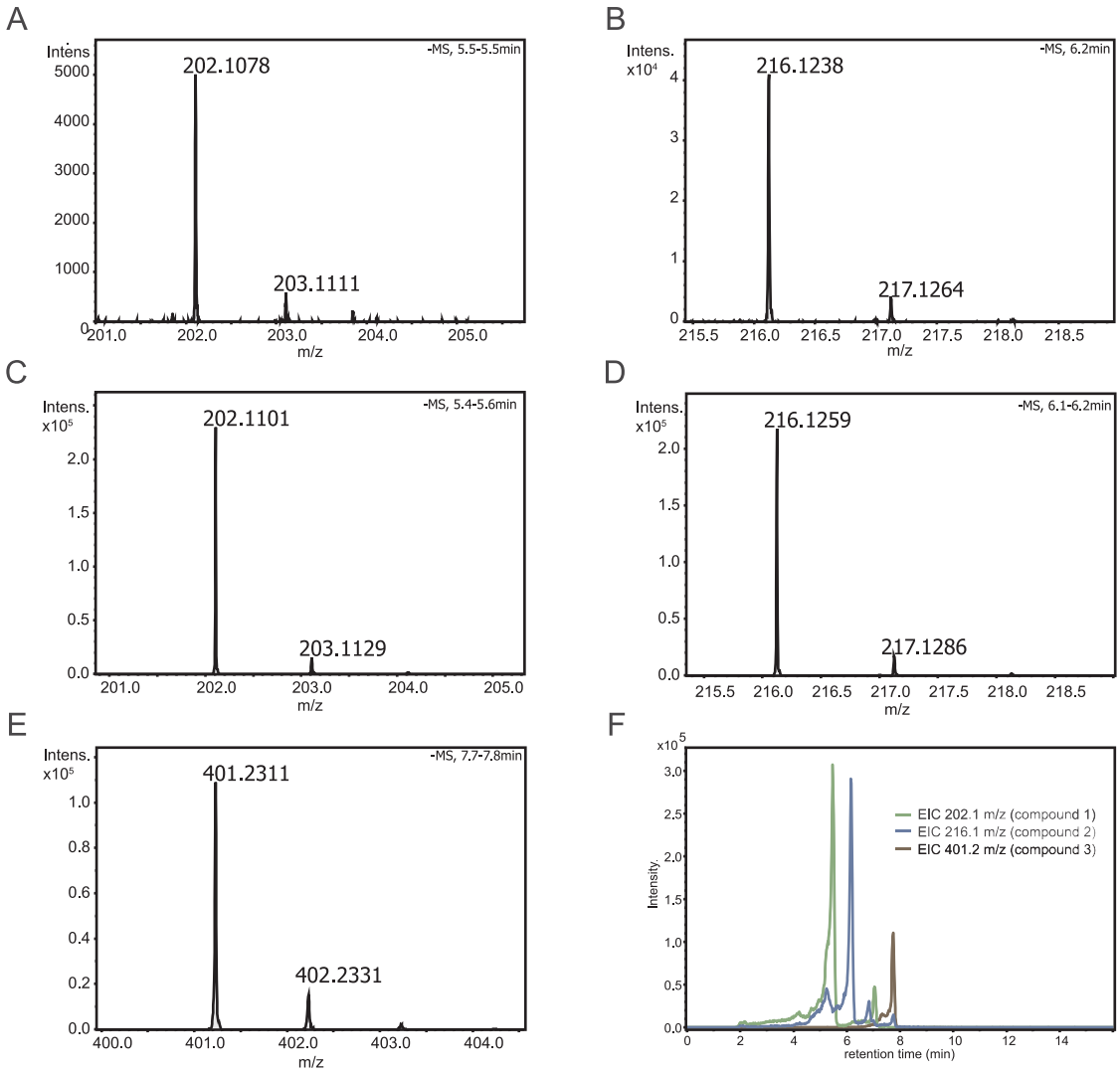


S2.2 Figure. Time course of pyrophosphate release for all A domains of cereulide synthetase with their cognate substrates using the nonradioactive pyrophosphate release assay.

Sample	Compound	Retention time (min)	Measured m/z	Ion formula $[M-H]^-$	Calculated m/z $[M-H]^-$	Error [ppm]	Error [mmu]	mSigma (#sigma)	Score	rdb
Standards (Fig. 2.8A)	Dipeptide **1**	5.5	202.1078	$C_9H_{16}NO_4$	202.1085	3.5	0.7	11.5 (1)	100	2.5
	Dipeptide **2**	6.2	216.1238	$C_{10}H_{18}NO_4$	216.1241	1.7	0.3	12.3 (1)	100	2.5
Enzymatic reaction (Fig. 2.8B)	Dipeptide **1**	5.4-5.6	202.1101	$C_9H_{16}NO_4$	202.1085	-8.2	-1.6	22.1 (1)	64.19	2.5
	Dipeptide **2**	6.1-6.2	216.1259	$C_{10}H_{18}NO_4$	216.1241	8.2	-1.8	21.3 (1)	62.31	2.5
	Tetrapeptide **3**	7.7-78	401.2311	$C_{19}H_{33}N_2O_7$	401.2293	-4.4	-1.8	39.9 (1)	62.82	4.5

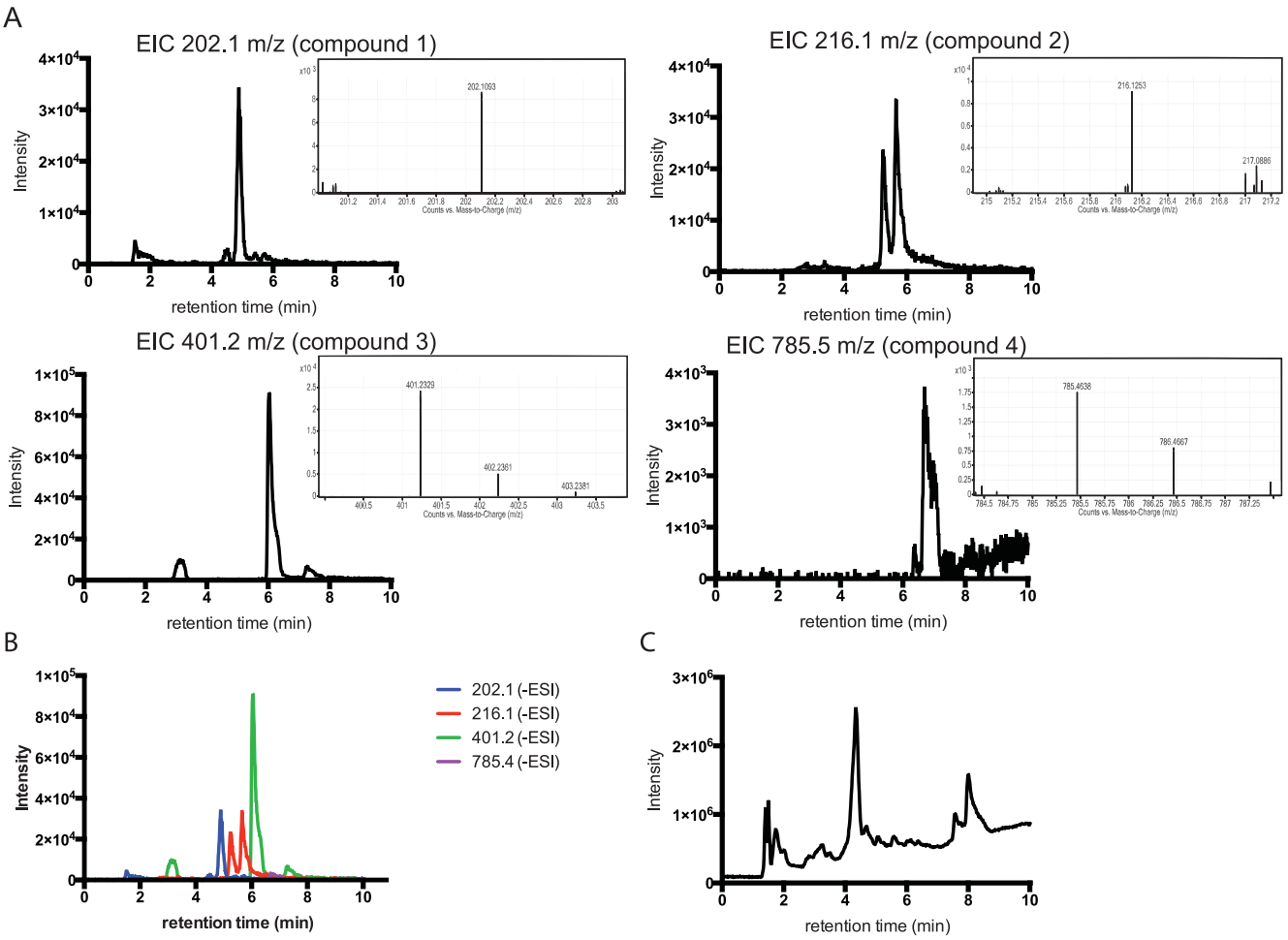
S2.3 Table. SmartFormula™ analysis of mass spectra of putative cereulide precursors.

Analysis was performed on the mass spectra shown in Figure 2.8.



S2.4 Figure. Additional LC-MS data from analysis of peptide synthesis assay shown in Figure 2.8

Mass spectra of (A) dipeptide **1** standard, (B) dipeptide **2** standard, (C) dipeptide **1**, (D) dipeptide **2** and (E) tetrapeptide **3**, showing isotopic distribution. (E) Overlay of EICs.



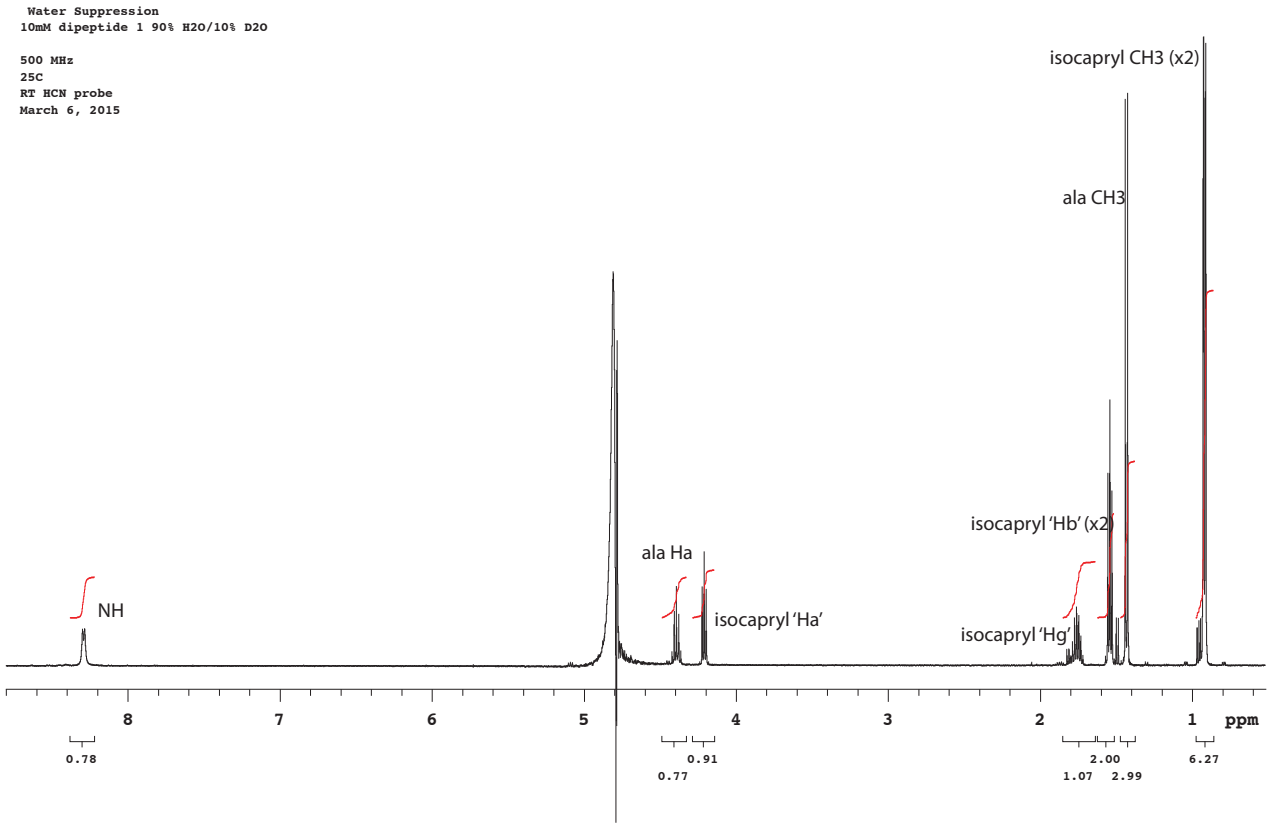
S2.5 Figure. LC-MS of a reaction of *in vitro*-reconstituted cereulide synthetase.

(A) Extracted ion chromatograms (EIC) from a reaction of CesA, CesB and substrates. EICs are extracted using the m/z values calculated from the expected molecular masses of peptides. The insets show the mass spectra. (B) Overlay of EICs. (C) Total ion chromatogram (TIC) of the same reaction.

Sample	Compound	Retention time (min)	Measured m/z	Ion formula $[M-H]^-$	Calculated m/z $[M-H]^-$	Error [ppm]	Error [mmu]
Enzymatic reaction mix (S2.4 fig)	Dipeptide **1**	4.9	202.1092	$C_9H_{16}NO_4$	202.1085	3.5	0.7
	Dipeptide **2**	6.2	216.1241	$C_{10}H_{18}NO_4$	216.1241	0	0
	Tetrapeptide **3**	5.4-5.6	401.2345	$C_{19}H_{33}N_2O_7$	401.2293	12.9	5.2
	Octapeptide **4**	6.1-6.2	785.4608	$C_{38}H_{65}N_4O_{13}$	785.4554	6.9	5.4

S2.6 Table. MS analysis of a reaction of *in vitro*-reconstituted cereulide synthetase.

Analysis was performed on the mass spectra shown in S5 Figure.



S2.7 Figure. ^{1}H-NMR of the authentic dipeptide standards.

S2.8 Supporting Information. HPLC-MS method for peptide synthesis assay shown in S2.5 Figure.

For liquid chromatography and mass spectrometry analysis of the peptide synthesis reaction of *in vitro*-reconstituted cereulide synthetase shown in S5 Figure, which showed masses consent with cereulide precursor dipeptides (1 & 2), tetrapeptide (3) and octapeptide (4), high pressure liquid chromatography and mass spectrometry equipment and media was from Agilent Technologies (Santa Clara, CA, USA), and the experiments were performed at the *Rosalind and Morris Goodman Cancer Research Centre Metabolomics Core Facility (Montreal, QC, Canada).* Liquid chromatography was performed using a 1290 Infinity ultra-performance UPLC system with a Eclipse Plus C8 1.8 µm, 2.1×100mm column at 55°C, a mobile phase A of 200 mM ammonium

acetate in water and a mobile phase B of 200 mM ammonium acetate in methanol. Peptides were eluted by a gradient of 90% A to 0% A over 5 minutes at a flow rate of 0.2 ml/min, followed by a further 5 minutes at 0% A. Six minutes of equilibration and blanks injections between samples were performed to ensure that there was no cross contamination. Peptide masses were analysed by an attached Agilent 6540 UHD Accurate-Mass Q-TOF mass spectrometer equipped with an electrospray ionization (ESI) source in negative ionization mode with source nitrogen gas at 325°C and 9 liters per minute. The nebulizer pressure was set at 45 psi and capillary voltage was set a 4.0 kV. Data were collected in full scan mode (mass range: m/z 50-2000; scan time: 1.4s; data collection: centroid and profile). Data were analyzed using the program Mass Hunter Qual. Calculated exact masses were extracted using a 10ppm window.

2.8. Connector to chapter 3

I tried to crystallize the A-KR-PCP module from CesA and the di-modular NRPS CesB without success. Therefore, I decided to explore alternative crystallization targets by extremophile orthologue screening (Giege 2013). Initially, I performed an NCBI-BLAST search of the CesA A-KR-PCP sequence, followed by manual filtering of extremophile orthologues with an identical domain organization and >90% sequence coverage. The criteria of extremophile microorganism filtering were based on the source of the isolation or growth conditions (pH, temperature, media, and salt concentration). We chose the A-KR-PCP module from a biosynthetic cluster in *Bacillus stratosphericus LAMA 585,* isolated from the Mid-Atlantic-Ridge seafloor (5,500-m depth) (Lima, Cabral et al. 2013). The coding sequence was codon-optimized for expression in *E. coli* and synthesized *de novo*. Using modified expression and purification protocols from Chapter 2, I was able to crystallize and determine the structure of the A-KR construct and of the excised A domain complexed to an α-ketoacyl adenylate, providing novel information on the mechanisms of depsipeptide biosynthesis.

CHAPTER 3. STRUCTURAL AND FUNCTIONAL STUDIES OF A MODULE THAT ADENYLATES AND REDUCES ALPHA-KETO ACIDS

Diego A. Alonzo*, Clarisse Chiche-Lapierre*, Jimin Wang, Michael J. Tarry, Martin Schmeing. Structural and functional studies of a module that adenylates and reduces alpha-keto acids. Manuscript in preparation.

*Denotes co-author

3.1. Abstract

Cyclic depsipeptides are bioactive compounds of medical and industrial importance. Examples include the emetic toxin cereulide and the antiviral agent valinomycin, which are synthesized in an assembly-like fashion by microbial enzymes called nonribosomal peptide synthetases. The hydroxyl acid moieties in cyclic depsipeptides are added by specialized modules constituted of adenylation (A), ketoreductase (KR) and peptidyl carrier protein (PCP) domains. The module A-KR-PCP catalyzes the selection/adenylation of an α-keto acid from the metabolite pool, followed by thiolation and transfer to the KR domain, which catalyzes the stereospecific reduction of the α-keto into an α-hydroxyl group using NADPH as a cofactor. The lack of structural information has limited our knowledge of these processes. Here we present the first crystal structure of an initiation A-KR-PCP module. We show that the KR domain is not inserted into the A domain, as it was previously thought, but that an integral A domain is present and is connected to the KR domain by a short linker. We also show a novel structural element connecting the KR and PCP domains. This element is reminiscent of an A_{sub} domain and can participate in *cis* or *trans* strand swapping. Finally, a high-resolution structure of the specialized A domain bound to its cognate adenylate gives insight into the selectivity of monomers containing an α-keto versus an α-amino group during depsipeptide biosynthesis.

3.2. Introduction

Nonribosomal peptide synthetases (NRPSs) are multi-domain microbial enzymes that produce a vast array of biologically active compounds, including FDA-approved drugs such as the antibiotic daptomycin, the immunosuppressant cyclosporine and the antifungal caspofungin (Felnagle, Jackson et al. 2008). NRPSs are organized in modules, functional units that contain all the catalytic domains needed to add one monomer to the growing peptide chain. The biosynthetic cycle of NRPSs resembles an assembly line and has been thoroughly reviewed by several groups (Strieker, Tanovic et al. 2010, Miller and Gulick 2016, Sussmuth and Mainz 2017, Reimer, Haque et al. 2018). It is described as a sequence started by an initiation step, followed by at least one elongation step and ended by a termination reaction. A canonical elongation step starts with the selection and adenylation of a carboxylate substrate from the metabolite pool by an adenylation (A) domain using ATP as a co-substrate. Then, the phosphopantheinyl (ppant) group attached to a Ser residue in the peptidyl carrier (PCP) domain attacks the adenylate, creating a covalent linkage between the enzyme and the substrate. The resulting acyl thioester is transferred to the condensation (C) domain, where it is added to the growing peptide chain through amide bond formation. The peptide chain is further extended in successive elongation modules until it is released via termination domains by oligomerization, cyclization, hydrolysis, reduction, or amidation. The most common termination domain is the thioesterase (TE) domain.

The products of NRPSs have a vast chemical diversity, which arises from different features of the assembly line. Different module combinations allow a theoretically infinite number of peptide sequences, as NRPSs follow a collinearity rule where the order and number of modules is directly correlated with the length and sequence of the product (Challis and Naismith 2004). Another source of chemical variation is the different substrates selected by A domains, which include at least 500 monomers in addition to the 20 proteinogenic amino acids (Caboche, Leclere et al. 2010). At the end of the cycle, different mechanisms of chain release can result in a variety of structures, for instance by oligomerization (Hoyer, Mahlert et al. 2007) or macrocycle formation (Bruner, Weber et al. 2002). Tailoring domains can add chemical groups during different stages

of the assembly process or after chain release. These modifications include N-formylation (Reimer, Aloise et al. 2016), C-, N- S- methylation (Mori, Pang et al. 2018), epimerization (Samel, Czodrowski et al. 2014) and hydroxylation (Chen and Walsh 2001). NRPSs can combine several of these mechanisms to add a massive diversity of chemical features into their peptide products, as in the case of cyclosporin, a cyclic *N*-methylated peptide with non-proteinogenic amino acids in its structure (Tedesco and Haragsim 2012).

Depsipeptides, molecules that contain ester bonds in addition to amide bonds in their structure (Taevernier, Wynendaele et al. 2017), are ubiquituous natural products made by specialized NRPSs called depsipeptide synthetases. Some medically and biologically-relevant examples include the piscicide antimycin, the K^+ ionophores cereulide (Ekman, Kruglov et al. 2012, Alonzo, Magarvey et al. 2015) and valinomycin (Jaitzig, Li et al. 2014, Huguenin-Dezot, Alonzo et al. 2019) (**Figure 3.1**), the anticancer agent cryptophycin (Ding, Rath et al. 2011), and the insecticide mycotoxins bassianolide (Xu, Orozco et al. 2009) and beauvericin (Xu, Orozco et al. 2008).

The ester bonds in depsipeptides arise from distinct mechanisms that depend on both the domain composition and the module organization of the depsipeptide synthetase involved in their biosynthesis (**Chapter 1**). Some bacterial depsipeptide synthetases generate products with alternating arrays of ester and amide bonds, which are best exemplified by cereulide and valinomycin synthetases (**Figure 3.1**). These macromolecular machines produce a linear tetradepsipeptide precursor through the alternative selection and condensation of α-hydroxy and α-amino acids (Magarvey, Ehling-Schulz et al. 2006). Three tetradepsipeptide precursors are then oligomerized in the C-terminal TE domain into a linear dodecadepsipeptide precursor, which is then cyclized to form the mature cyclic product (Huguenin-Dezot, Alonzo et al. 2019).

The α-hydroxy acids in cereulide and valinomycin arise from the interplay of unusual A domains and α-ketoreductase (KR) domains that are arranged in specialized modules with an A-KR-PCP organization (**Figure 3.1b)**. The A domains in these systems adenylate α-ketoacids instead of α-amino acids or α-hydroxy acids (Magarvey, Ehling-Schulz et al. 2006) (Fujimori, Hrvatin et al.

2007) (**Figure 3.1b)**. The resulting ketoacyl adenylate is then thiolated and transported to the KR domain, where it undergoes NADPH-mediated reduction to generate an α-hydroxy acyl thioester. The OH group in the α-hydroxy acyl thioester can then act as a nucleophile in downstream condensation, cyclization or oligomerization reactions, depending on the position of the A-KR-PCP module within the rest of the assembly line. For instance, in cereulide and valinomycin biosynthesis (**Fig. 3.1**), the α-hydroxy acid added by elongation module 3 (cesB1 or vlm2-1) forms an ester bond during formation of the linear tetradepsipeptide precursor through C domain-catalyzed condensation in module 2 (cesA2 or vlm1-2), whereas the α-hydroxy acid added by initiation module 1 (cesA1 or vlm1-1) acts as a nucleophile in the TE-catalyzed oligomerization and cyclization reactions (Huguenin-Dezot, Alonzo et al. 2019).

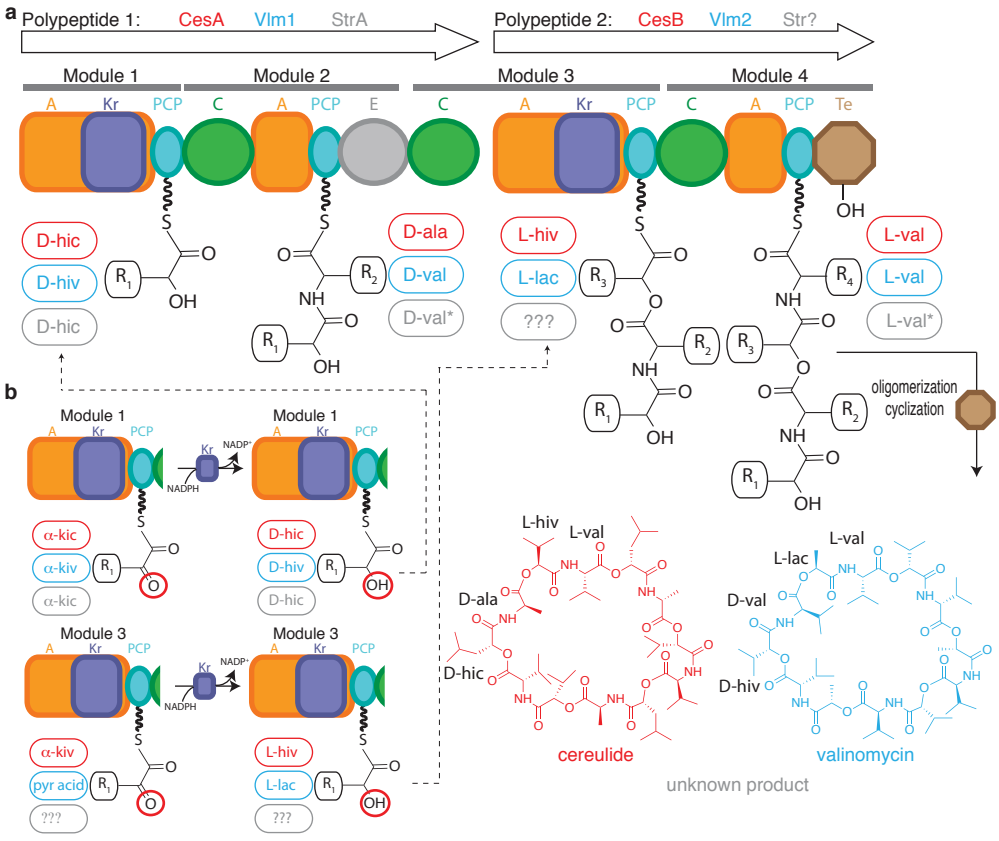


Figure 3.1. Biosynthesis of cyclic depsipeptides.

a. Cereulide and valinomycin (bottom, right) are synthesized by tetramodular NRPSs encoded in two polypeptide chains (cereulide: CesA and CesB, valinomycin: Vlm1 and Vlm2). CesA and CesB homologues from the extremophile *Bacillus stratosphericus* LAMA 585 are also encoded in two polypeptide chains

(NCBI IDs: WP_007498213 and WP_039963193.1). We called WP_007498213 StrA, due to its homology with CesA. See **Figure S3.1** for a detailed comparison of the biosynthetic gene clusters. In cereulide and valinomycin biosynthesis, Ces and Vlm catalyze the alternative addition of α-hydroxyl and α-amino acids to produce linear tetradepsipeptide intermediates, which are oligomerized and then cyclized by the terminal thioesterase domain. **b.** Modules 1 and 3 catalyze the selection and activation of α-keto acids instead of α-amino acids through specialized adenylation domains. Then, the phosphopantetheinyl arm in the PCP domain thiolates α-ketoacyl adenylates. Transfer to the KR domain follows, where the α-keto group is reduced into an α-hydroxyl group using NADPH as a cofactor. The resulting hydroxyl acid thioesters can then participate as nucleophiles in downstream condensation, oligomerization or cyclization reactions. NRPSs: nonribosomal peptide synthetases, D-HIC: D-alpha hydroxy-isocaproic acid, D-HIV: D-alpha hydroxyisovaleric acid, L-HIV: L-alpha hydroxy-isocaproic acid, L-lac: L-lactic acid, α-kic: alpha ketoisocaproic acid, α-kiv: alpha ketoisovaleric acid, pyr acid: pyruvic acid.

A-KR-PCP modules are also present in the kutzneride (Fujimori, Hrvatin et al. 2007), antimycin family (Liu, Zhu et al. 2016) and cryptophycin (Ding, Rath et al. 2011) (Magarvey, Beck et al. 2006) biosynthetic gene clusters (see **Chapter 1**). Most of these products have up to three α-hydroxy acids in their structures, and subsequently the same number of A-KR-PCP modules in the biosynthetic cluster, in agreement with the collinearity rule of NRPSs. Unlike valinomycin and cereulide synthetases, these clusters lack an iterative TE domain.

Even though the structural knowledge of NRPSs in the last decade has increased dramatically (Reimer, Haque et al. 2018) (Drake, Miller et al. 2016) (Gulick 2016) (Tanovic, Samel et al. 2008, Liu, Zheng et al. 2011, Miller, Drake et al. 2016, Tarry, Haque et al. 2017), multi domain or full-module structures possessing tailoring domains are only starting to be solved in the past three years (Mori, Pang et al. 2018) (Chen, Li et al. 2016, Reimer, Aloise et al. 2016). The architecture of A-KR-PCP modules was hypothesized to resemble the interrupted A domain configuration (Magarvey, Ehling-Schulz et al. 2006) (Labby, Watsula et al. 2015), where tailoring domains are inserted between motifs A8 and A9 of the A domain. The interrupted domain arrangement has been confirmed for methyltransferase domains in a recently published structure (Mori, Pang et

al. 2018). However, there was no direct structural and biochemical evidence to verify this domain arrangement in A-KR-PCP modules.

Another open question of depsipeptide synthetases was the selection mechanism of α-keto acids by A domains in A-KR-PCP modules. These A-domains differ substantially from amino acid-selecting A domains, as not only can they select α-keto acids and discriminate against α-amino and α-hydroxyl substituents, but they can also maintain α-keto acid side chain selectivity (Alonzo, Magarvey et al. 2015) (Fujimori, Hrvatin et al. 2007). Even though the Stachelhaus selectivity code of α-keto acid-selecting A domains reveals some differences with amino acid-selecting A domains, such as the replacement of the highly conserved residue Asp of amino acid-selecting A-domains for an aliphatic residue (Magarvey, Ehling-Schulz et al. 2006) (Magarvey, Beck et al. 2006) (Fujimori, Hrvatin et al. 2007), it was still unknown which structural determinants were used to discriminate between α-keto and α-amino groups.

The KR domains in vlm and ces are similar to β-ketoreducing PKS KR domains (Magarvey, Ehling-Schulz et al. 2006), which are classified based on their product stereochemistry as A1-type (β-L-hydroxy acids) or B1-type (β-D-hydroxy acids) (Liu, Yuan et al. 2018) (Zheng, Piasecki et al. 2013). Based on this classification, the ces and vlm KR domains belong to novel subtypes that catalyze α- instead of β-ketoreduction, with cesA1- and vlm1-1- KRs classified in the B1-subtype (α-D-resulting stereochemistry) and cesB1 and vlm2-1 classified in the A1-subtype (α-L- resulting stereochemistry). It is unknown whether the structural determinants of stereochemistry in α-ketoreducing KRs are similar to those of PKS KR domains, dictated by the face of the β-ketoacyl thioester presented to the 4'-*pro*-S hydrogen of NADPH. This orientation depends on the entry point of the ppant arm to the KR active site and is controlled by interactions with one of two loops facing each other in the catalytic pocket (Liu, Yuan et al. 2018). The orienting loops are identified by conserved sequence motifs, an LDD motif in the loop of B-type KR domains and a highly conserved Trp in the loop of A-type domains. It is unknown if similar sequence footprints or if novel structural features control stereochemistry in A-KR-PCP modules.

Here we present the di-domain A-KR and the adenylate-bound A domain structures of the initiation module StrA1 (WP_007498213) from *Bacillus stratosphericus* LAMA 585, a homologue of the initiation module CesA1. Our structures, along with accompanying biochemical, bioinformatic and biophysical analysis, show the arrangement of a KR domain within an NRPS for the first time, the presence of structural elements never before described for NRPSs and a detailed picture of the mechanism of α-group discrimination in the specialized A domain. We also propose that α-B1-type KR domains determine product stereoselectivity by controlling the access point of the ppant arm to the active site through interactions with a conserved loop as observed in β-B1-type KR domains. Furthermore, the ppant access tunnel that would correspond to an A1-type KR is blocked with Arg residue side chains. Our results can be extrapolated to other NRPSs containing A-KR-PCP modules that produce depsipeptides with medical, biological and industrial importance.

3.3. Results

3.3.1. Overall StrA1 A-KR di-domain structure

We solved the di-domain A-KR X-ray structure from the initiation module StrA1 of a cereulide synthetase homologue from *Bacillus stratosphericus* LAMA 585 (NCBI ID: WP_007498213) at 3.4 Å resolution (**Fig. 3.2, Table S3.1)**. The genome of *B. stratosphericus* LAMA 585 contains a biosynthetic cluster that resembles the valinomycin and cereulide clusters (see **Methods and Fig. S3.1**). The A-KR structure consists of four molecules in the asymmetric unit (**Fig. 3.2a**), which can be described as two copies of a domain-swapped dimer. Each crystallographic-related molecule shows unambiguous density for the A_{core}, A_{sub} and KR domains, along with novel connecting structural motifs (**Fig. 3.2,** wheat and grey**)**. Even though the crystallized protein included the PCP domain (modified with NH_2-ppant and incubated with substrates to generate ketoacyl-NH-ppant, see **Methods**), there is no interpretable density for the α-helical bundle, indicative of domain disorder or multiple conformations.

The A and KR domains are connected by a short linker (**Figure 3.2b**, linker residues 663-671 in red) and do not share any surface contacts and their active sites are 70 Å apart. The distance between active sites implies that the PCP domain must undergo conformational changes to reach each catalytic unit during the synthetic cycle, requiring a conformational transition of the A_{sub} domain (Reimer, Aloise et al. 2016) (Sundlov, Shi et al. 2012) (Drake, Miller et al. 2016) to allow the thiolation reaction to occur.

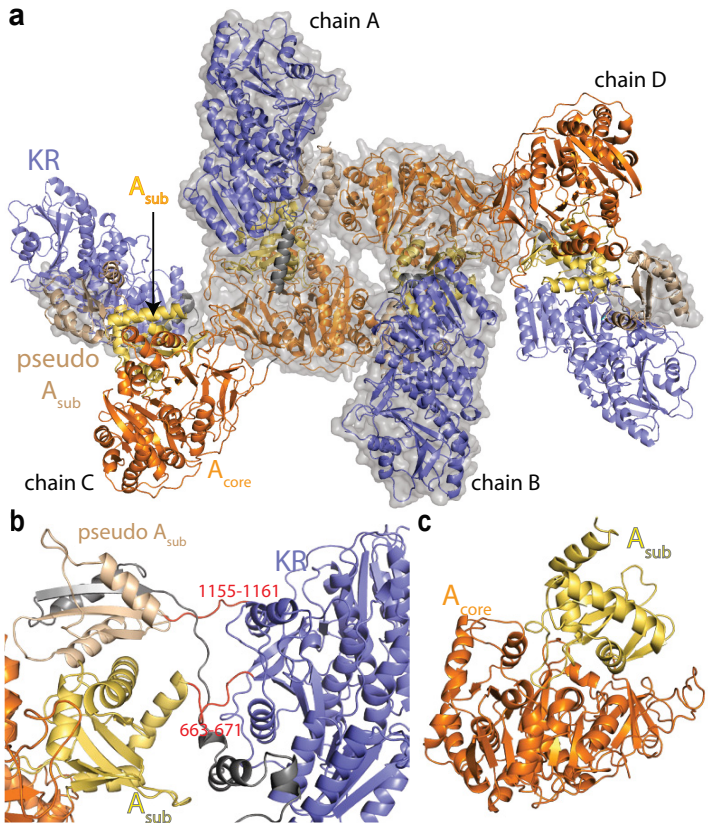


Figure 3.2. The crystal structure of StrA1 A-KR, part of an α-keto acid-processing module.

a. Cartoon representation of the asymmetric unit of Stra A-KR, color-coded by domain type. Chains A and B are depicted with a transparent surface for ease of representation. The asymmetric unit is two copies of a domain-swapped dimer, where each monomer is constituted by A_{core} (orange), A_{sub} (yellow), and KR (slate) domains. A novel structural element, the pseudo A_{sub} domain (wheat and grey) is present downstream of the KR domain. **b**. Zoomed view of the A_{sub}-KR and KR-pseudo A_{sub} linkers depicted in red. **c.** Cartoon representation of the adenylation domain of StrA1. The secondary structure elements of the A_{core} and the A_{sub} domains are in a continuous polypeptide chain. This rejects the interrupted adenylation domain hypothesis for tailoring domain insertion.

3.3.2. Structure of the adenylation domain in the StrA1 A-KR structure context

The adenylation domain has the expected Rossmann-like fold described for other A-domains (Conti, Stachelhaus et al. 1997). The A_{core} and A_{sub} domains in the A-KR structure are in the closed adenylation conformation (**Fig. 3.2c)** similar to GrsA-Phe1 and LgrA (Conti, Stachelhaus et al. 1997) (Reimer, Aloise et al. 2016) (PDB IDs: 1amu, 5es5 chain B).

The KR domain was thought to be inserted between A_{sub} motifs A8 and A9 (Labby, Watsula et al. 2015) (Magarvey, Ehling-Schulz et al. 2006). Interrupted A domains occur in other modules containing tailoring domains, as illustrated by the recently solved structure of an A-methyltransferase module (Mori, Pang et al. 2018). The uninterrupted arrangement of all the motifs of the A_{sub} domain in our A-KR structure disagrees with the interrupted A domain hypothesis. A-KR-PCP modules CesA1, CesB1, Vlm1-1, Vlm2-1, AntC and CrpD-M2 are highly likely to be uninterrupted too.

3.3.3. The StrA1 KR domain has features of type B PKS KR domains.

The strA KR domain has the expected two-subdomain (668-884 and 885-1155) Rossmann fold (Zheng, Taylor et al. 2010). The catalytic residues (Zheng, Taylor et al. 2010) tyrosine (1064), lysine (1027) and serine (1051) and the NADPH binding site were identified by structural and sequence alignments. Even though the LDD sequence motif of B-type KRs is not present in StrA1 KR, a structural alignment with the first KR domain in tylosin biosynthesis (Keatinge-Clay 2007) (PDB ID: 2z5l) revealed an IQE motif (residues 1007-1009) located at the LDD position (**Fig. 3.3**). Furthermore, the ppant entrance tunnel expected for an A-type KR domain is blocked by two Arg residues in StrA A-KR (**Fig. 3.3**), based on a structural alignment with the recently solved structure of a substrate-bound A-type KR (Liu, Yuan et al. 2018). This indicates that α-ketoreducing KR domains control product stereochemistry through a similar mechanism of β-ketoreducing KR domains.

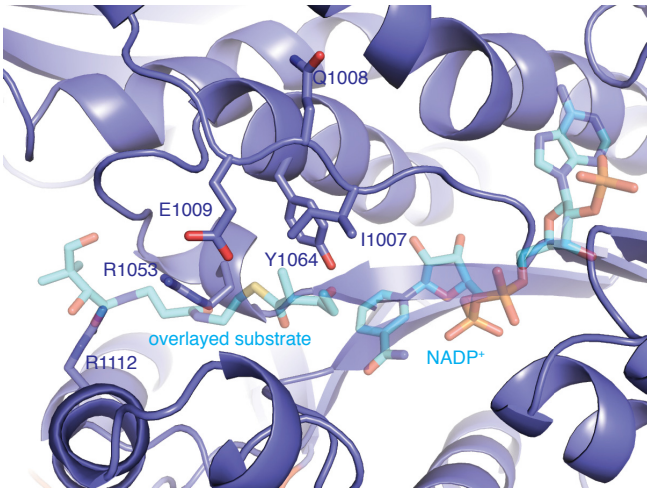


Figure 3.3. The KR domain active site.

An IQE motif (sticks, residues 1007-1009) overlays with the LDD motif characteristic of B1-type PKS KR domains. The ppant entrance tunnel distinctive of A1-type PKS domains is blocked by two Arg sidechains, R1053 and R1112. Ligands from the amphotericin KR are depicted in transparent sticks at the expected position resulting from a structural alignment of the protein chains of StrA1 KR and PDB ID 5xwv.

3.3.4. Pseudo A_{sub} domain and dimer swap

Modelling of the region connecting the KR and PCP domains (residues 1156-1227) showed a novel structural element that had not been previously described in NRPSs (**Fig. 3.4a**), a sub-domain formed by a mixed 3-stranded β-sheet surrounded by three α-helices, reminiscent of a canonical A_{sub} domain lacking the A8 motif β-sheet (**Fig. 3.4b**). Because of this similarity, we called the novel sub-domain "pseudo A_{sub} domain".

Inspection of the electron density revealed that the secondary structure elements in the pseudo A_{sub} domain are divided into two groups, each arising from different polypeptide chains. Residues downstream of the KR domain (1156-1227) constitute the first group, formed by two parallel strands of the mixed β-sheet, an α-helix preceding motif A9, the loop motif A10 and the α-helix preceding the PCP domain (**Figure 3.4a,** wheat). Unexpectedly, the second group formed by the remaining antiparallel β-strand and its preceding α-helix, are modelled by the coordinates from the N-terminus (residues 14-44) of a non-crystallographic related molecule (**Figure 3.4a,** dark

grey), creating a domain swap between two non-crystallographic symmetry related molecules. The N-terminal residues (14-44) of the non-crystallographic symmetry mate are connected to their corresponding A_{core} domain through an alpha-helical linker (residues 45-82), covering a distance of around 67 Å.

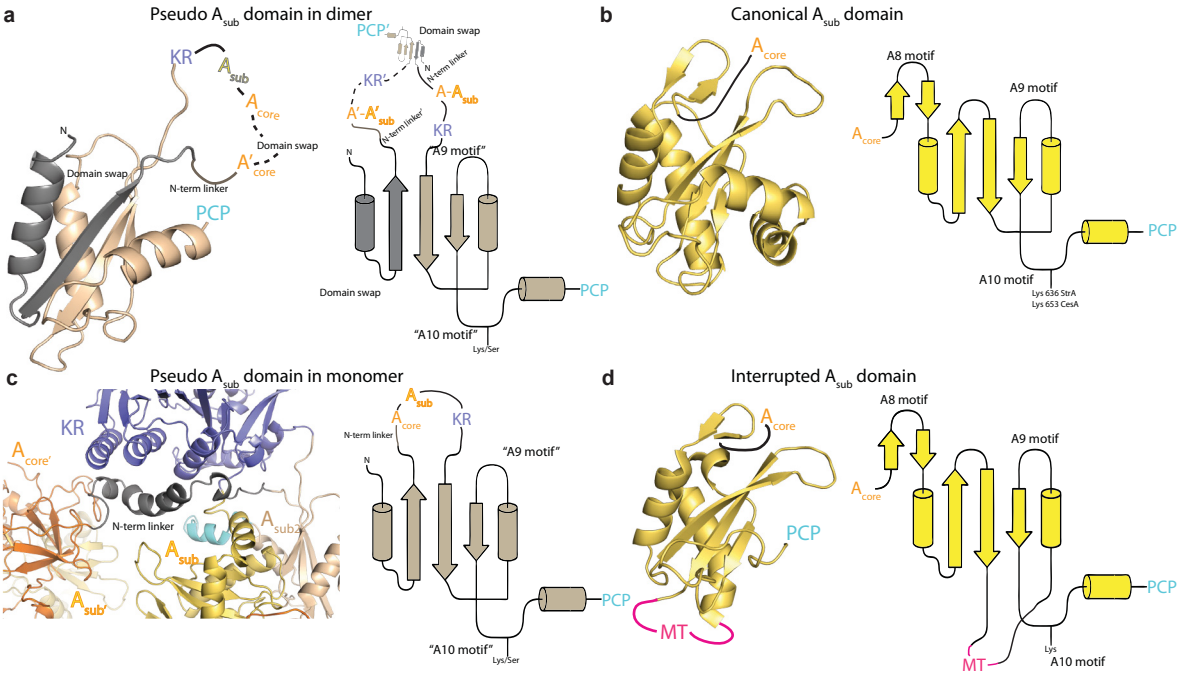


Figure 3.4. The structure of pseudo-, canonical and interrupted A_{sub} domains.

a. Cartoon (left) and topology (right) representations of the pseudo A_{sub} domain of StrA1. Secondary structure elements reminiscent of canonical A_{sub} domains (**b**) are present. A continuous polypeptide chain (colored in wheat, residues 1156-1227) contains the parallel strands of the mixed β-sheet, the α-helix preceding the A9 motif, a loop containing a pseudo A10 motif, and a C-terminal α-helix. A non-crystallographic symmetry related molecule (dark grey, residues 14-44) domain-swaps into the pseudo A_{sub} domain, complementing the tertiary structure with the N-terminal α-helix and one of the antiparallel strands of the mixed β-sheet. **b.** Cartoon (left) and topology (right) representations of the canonical A_{sub} domain from StrA1 (residues 541-667). The tertiary structure of the StrA1 A_{sub} domain is intact and resembles that of other A_{sub} domains. A small two-stranded antiparallel β-sheet that includes motif A8 connects the A_{sub} to the A_{core} domain. Most of the A_{sub} domain consists of a mixed three-stranded β-sheet surrounded by two α-helices. The A9 motif is located in a loop between the second α-helix and a strand from the parallel portion of the mixed β-sheet. A flexible loop contains the A10 motif. This motif includes the active lysine residue (636) involved in the adenylation reaction. **c.** Cartoon (left)

and topology (right) representations of a model of a monomer of StrA1 A-KR structure. StrA1 A-KR-PCP exists mainly as a monomer in solution (**Fig. S3.2**), however the structure presented in this work is a domain-swapped dimer. Modelling of the monomer shows that complementation of the A_{sub} domain can also occur in *cis*. The length of the linker (38 residues, residues 45-82) between the complementing region (residues 14-44) and the start of the A_{core} domain (residue 83) is enough to allow interaction of the complementing region and the pseudo A_{sub} in of the same molecule. **d.** Cartoon (left) and topology (right) representations of the interrupted A_{sub} domain from TioS (PDB ID: 5WMM). A methylation domain is inserted between the end of the central strand from the mixed β-sheet and the beginning of the α-helix preceding motif A9. As in canonical NRPSs, the PCP domain is located downstream of motif A10. It was hypothesized that the KR domain in StrA1 would also be located at the insertion point of methylation domains. However, the presence of an intact A_{sub} domain (**b**) in our structure rejects this hypothesis.

Even though we solved the dimer-swapped structure of the A-KR didomain construct, most of the protein is a monomer in solution, as shown by size exclusion chromatography and native-PAGE (**Fig. S3.2, a-b**). The monomer fraction never crystallized, whereas the dimer fraction crystallized in several conditions. To evaluate whether there are functional differences between monomer and dimer A-KR-PCP, we performed a characterization of the reaction cycle for each oligomeric species. We performed a selectivity assay in advance to elucidate which ketoacid is selected by StrA1 (**Fig. S3.2c, left**), revealing that it preferentially selects α-KIC, the same substrate selected by CesA1 (Alonzo, Magarvey et al. 2015).

The adenylation assay revealed that both monomeric and dimeric A-KR-PCP catalyze adenylation (**Fig. S3.2c, right**). Next, we evaluated if oligomerization influenced the phosphopantetheinylation reaction on the PCP domain **(Fig. S3.2, d-g**). Incubating apo proteins (**Fig S3.2**, **d-e**) with the promiscuous phosphopantetheinyl transferase SFP and CoA led to incorporation of a phosphopanetheinyl moiety (expected $\Delta_{m/z}$ = +342 Da), regardless of the oligomerization state (**Fig. S3.2, f-g**). We also found that both monomer and dimer catalyzed thiolation and ketoreduction through a single-turnover NADPH consumption assay (**Fig. S3.2, h-i**).

The similarities in catalytic activities between oligomeric forms, along with the monomer being more abundant in solution, indicate that the monomer is the functional form of A-KR-PCP. Given this assumption, the domain-swap elements in the pseudo A_{sub} domain must undergo self-complementation for proper protein folding in the monomer, where the N-terminal residues 1-44 must interact with the rest of the pseudo-A_{sub} elements (1156-1227). Both regions are separated by 1112 residues. To test whether the linker (residues 45-82) which joins residues 1-44 to the A_{core} is of sufficient length to allow interaction of the two regions of the pseudo-A_{sub} domain, we modelled one copy of the asymmetric unit of the A-KR-PCP structure as a hypothetical monomer (**Fig. 3.4c**). The model shows that the linker of 37 residues is long enough to connect regions 1-44 and 1156-1227 in the same molecule.

3.3.5. Role of motif 10 in pseudo A_{sub} domains

The catalytic Lys in motif A10 (A10-Lys) in A-domains interacts with ATP and the carboxyl group of the substrate during the adenylation reaction (Conti, Stachelhaus et al. 1997) and its mutation in canonical A_{sub} domains leads to abolished adenylation function (Hamoen, Eshuis et al. 1995). Sequence alignment analysis of several pseudo-A_{sub} domains from A-KR-PCP modules indicates the presence of an A10-like motif, however, A10-Lys is replaced by another residue in most of the analyzed modules, except for CesA1 (**Fig. S3.3a**). We assessed whether the exceptional pseudo-A_{sub}-A10-Lys-1218 in CesA1 could participate in the adenylation reaction or complement loss of function in a canonical-A10-Lys-653 mutant. Single mutation of canonical-A10-Lys-653 to Ala led to loss of adenylation function in CesA1 (**Fig. S3.3b**), indicating that the activity was not rescued by the presence of pseudo-A_{sub}-A10-Lys-1218. Furthermore, mutation of pseudo-A_{sub}-A10-Lys-1218 to Ala alone had no effect on the adenylation reaction, indicative that pseudo-A_{sub}-A10-Lys-1218 does not participate in keto acid adenylation and may be a remnant of a horizontal gene transfer event. Furthermore, the loss of adenylation function of the A10 motif in pseudo A_{sub} domains is strongly supported by the replacement of the Lys residue in most of the analyzed A-KR-PCP modules.

3.3.6. Plausible conformational changes during the A-KR-PCP synthetic cycle

The observed conformation of the A-KR-PCP structure is one of the multiple conformations that the module could adopt, as the protein can access several synthetic stages (**Figure 1**, **Figure 3.2**) that require massive conformational rearrangements (Gulick 2016, Reimer, Aloise et al. 2016) (Drake, Miller et al. 2016). For instance, the A-domain in the A-KR structure is in the closed conformation (**Fig. 3.2c**) observed for A-domains complexed to AMP + amino acid (Conti, Stachelhaus et al. 1997) or adenylates (Reger, Wu et al. 2008). In order to proceed to the thiolation reaction, the A_{sub} domain has to rotate several degrees to allow access of the ppant arm from the PCP domain (Gulick 2009, Reimer, Haque et al. 2018). However, the presence of the KR domain in A-KR-PCP would not allow such degrees of freedom. We propose that the pseudo A_{sub} domain plays the structural role of a canonical A_{sub} domain, translocating the PCP domain throughout the biosynthetic cycle.

3.3.7. The high-resolution structure of the StrA1 A domain-ketoacyl adenylate complex.

The A domain in the A-KR structure shows unambiguous density for most of the secondary structure elements and allowed us to evaluate its conformation and interactions in the context of the full module. However, the absence of ligands in the active site and the low resolution of the structure hampered a detailed analysis of the A-domain active site, specifically the structural determinants of α-keto against α-amino selectivity.

To address this question, we crystallized and solved the structure of an excised A domain construct of StrA1 (residues 83-649, $A_{83\text{-}649}$) at a maximum resolution of 2.3 Å, in complex with α-KIC adenylate. Initial phases were obtained by molecular replacement using the A domain from A-KR. In the initial mFo-Fc map, it was evident that a ligand was bound to the active site and after iterative manual model building and refinement, the unbiased electron density could be unambiguously identified as α-KIC adenylate (**Fig. 3.5a)**. This finding was not unexpected, as the protein was co-crystallized with α-KIC, $MgCl_2$, and ATP. After placement of the adenylate, the model refined to an R_{free} of 25.14 % (**Methods** and **Table S3.1**).

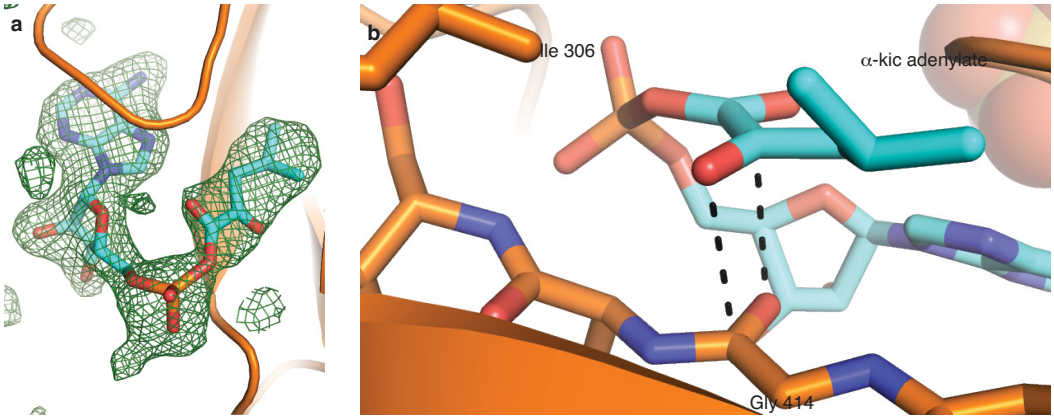


Figure 3.5. Structure of an α-keto acid-selecting A domain bound to α-kic adenylate.

a. Unbiased mFo-Fc map (green mesh, 2.5 σ) before α-kic adenylate (cyan sticks) was placed in the model. **b**. Stick representation of the A domain active site (orange) bound to α-kic adenylate (cyan). The ketone group interacts with Gly 414 through an antiparallel carbonyl-carbonyl interaction (black dashes). Ile306 is at the position of the highly conserved Asp residue of amino acid-selecting A domains.

The conserved Asp residue that coordinates the α-amino group in amino acid-selecting A domains is replaced by an aliphatic amino acid in α-keto acid selecting A-domains (Magarvey, Ehling-Schulz et al. 2006) (Fujimori, Hrvatin et al. 2007). In StrA1, this residue corresponds to Ile306 (**Fig. 3.5b, sticks**). We hypothesized that the Asp to aliphatic residue substitution in keto acid-selecting A domains would be crucial for discrimination between α-keto and α-amino groups. First, by favouring keto acid binding through the elimination of plausible electrostatic repulsion between the α-keto and the carboxyl side chain of Asp. Second, by preventing amino acid binding through the loss of the crucial interaction between α-amino groups and the carboxyl side chain in Asp. The active site in the adenylate-bound structure shows that Ile306 is in close proximity to the substrate but does not make any contact (**Fig. 3.5b**). Surprisingly, our structure shows that the crucial interaction in α-keto acid selection is between the α-keto group and the carbonyl of the peptide bond between Gly414 and Met415, through an antiparallel carbonyl-carbonyl interaction (Allen, Baalham et al. 1998) (**Fig 3.5b,** black dashed lines). This type of interaction occurs in small molecule crystals (Allen, Baalham et al. 1998) and stabilizes partially allowed Ramachandran conformations in proteins (Deane, Allen et al. 1999). The angles and

distances of the interaction match with those predicted by analysis of a dataset of small molecules (Allen, Baalham et al. 1998).

The A_{core} structure in the A-KR and A_{83-649} structures are almost identical, however, the A_{sub} domains are positioned differently. Whereas in A-KR the A_{sub} is in the adenylation conformation, we identified an open conformation in A_{83-649}, similar but not identical to the open conformation observed in one of the LgrA initiation module structures (PDB ID: 5ES5, chain A)) (Reimer, Aloise et al. 2016). The StrA A domain cannot adopt the fully open conformation observed in PDB ID: 5ES5 chain A, as the A_{sub} domain would clash with an insertion of 30 residues between motifs A5 and A6 (residues 430-460). This insertion has also been observed in fatty acyl adenylate ligases (FAALs, PDB ID: 3kxw) (Zhang, Zhou et al. 2011), where it is suggested to play a role as a steric filter to prevent binding of coenzyme-A during the adenylation reaction catalyzed by FAALs. Therefore, we propose that the range of angular motions of the A_{sub} domain might be restricted by this structural element, with currently unknown functional implications.

3.4 Discussion

Recently, knowledge of NRPS tailoring reactions has been greatly expanded by multi-domain structures with *cis*-acting domains, placed at the beginning of an initiation module (formyl-transferase) (Reimer, Aloise et al. 2016) or inserted into the A_{sub} domain (methyltransferases, **Fig. 3.4d**) (Mori, Pang et al. 2018). It was proposed that the KR domain in A-KR-PCP modules was inserted between structural motifs A8 and A9 of the A_{sub} domain (Labby, Watsula et al. 2015) (Magarvey, Ehling-Schulz et al. 2006), as in A-M-PCP or A-OX-PCP modules (Mori, Pang et al. 2018). However, our structure shows that the tailoring KR domain is not inserted in that region but located downstream of an integral A domain. This indicates that the insertion point of tailoring domains is more variable than previously thought.

In the biosynthetic cycle of LgrA, the PCP domain undergoes a massive conformational change to reach the formylation domain after Val-*S*-PCP formation, in a movement aided by a 140° rotation

of the A_{sub} domain. In A-KR-PCP, the A and KR active sites are 67 Å apart and similar movements might be needed to translocate the ketoacyl-*S*-PCP throughout the reaction; however, the A_{sub} and PCP domains are separated by the KR and pseudo A_{sub} domains. It is likely that the pseudo-A_{sub} domain plays the structural role of a canonical A_{sub} domain, swinging the PCP domain from one active site to the next. Furthermore, its structural similarity with canonical A_{sub} domains indicate that it was conserved after a probable horizontal gene transfer event that inserted a complete A domain, followed by loss of the A_{core} region. However, there are no sequence "scars" of the event in the linker region between the KR and pseudo A_{sub} domains, and the question is opened for future bioinformatic studies.

The ketoreduction reactions in the cereulide and valinomycin systems are stereoselective, as module-1 produces D-hydroxyl esters and module-3 produces L-hydroxyl esters (Magarvey, Ehling-Schulz et al. 2006). D-stereochemistry might be necessary for the oligomerization and cyclization reactions (Huguenin-Dezot, Alonzo et al. 2019), whereas L-stereochemistry might be necessary for condensation. We identified structural motifs in StrA KR that confirm its identity as a B-type α-ketoreducing domain. In spite of recent information on the interaction between PKS KR domains and carrier protein domains from a substrate-bound KR structure (Liu, Yuan et al. 2018), a multidomain PKS or NRPS holo-KR-ACP/PCP structure remains elusive.

Finally, the biosynthetic clusters for the bacterial depsipeptides antimycin, kutzneride and cryptophycin contain elongation A-KR-PCP modules (Liu, Zhu et al. 2016) (Fujimori, Hrvatin et al. 2007, Ding, Rath et al. 2011), where an ester bond-forming C domain condenses the α-hydroxy acid with the upstream growing chain. The structural information presented in this work should also apply to those systems, but the nature of the interactions with upstream and downstream domains and the structures of the ketoreduction and thiolation states remain to be elucidated.

3.5 Methods

Protein expression

The A-KR-PCP module (NCBI: WP_007498213, residues 1-1318) from a CesA homologue from *B. stratosphericus* LAMA 585 (StrA1) was originally synthesized as an *E. coli* codon-optimized sequence in a pet11a vector. Constructs A-KR-PCP_{1-1318} and A_{83-649} were sub-cloned into a pJ411-based vector with a C-terminal TEV-cleavable His_8 tag.

Both proteins were overexpressed in *E. coli* BL21 DE3 cells in LB media supplemented with kanamycin at 34 μg/L, except for some samples of A-KR-PCP_{1-1318} which were overexpressed in *E. coli* BL21 $EntD^-$ strain (kindly provided by Dr. Christian Chalut and Dr. Christophe Guilhot) (Chalut, Botella et al. 2006). Pre-inoculated 1 L cultures were grown at 37 °C under 220 r.p.m agitation until they reached an $O.D._{600}$ between 0.5 and 0.8. Then, they were cooled down on ice to 16 °C and induced with 100 μM D-1-thiogalactopyranoside (IPTG). Cultures were incubated for an additional 16 hours at 16 °C until cell harvest by centrifugation.

Protein purification

All purification steps were done at 4 °C. Wet cells were homogenized in homogenization buffer A (50 mM Tris pH 8.0, 250 mM NaCl, 2 mM $CaCl_2$, 10 % glycerol, 2 mM β-mercaptoethanol -βME-, 1 mM PMSF) and sonicated on ice for a total time of 8 minutes, with a cycle set at 10 s on, 20 s off at 50% amplitude. Lysates were clarified by centrifugation at 18 000 *g* and then applied to 2 x 5 mL Hi-Trap IMAC FF (GE Healthcare) columns connected in tandem, pre-equilibrated in homogenization buffer A.

A-KR-PCP_{1-1318} wet cells were resuspended in homogenization buffer A supplemented with DNaseI and one cOmplete (Roche) protease inhibitor tablet. After IMAC columns were washed to baseline, the protein was eluted in an isocratic step of buffer B (50 mM Tris pH 8, 500 mM NaCl, 4 mM $CaCl_2$, 150 mM imidazole, 10 % glycerol, 2 mM β-mercapto ethanol, βME). TEV protease was added to the eluted protein (1:20 w:w ratio) and dialyzed 16 h at 4 °C under mild agitation against buffer C (50 mM Tris pH 8, 10 mM NaCl, 4 mM $CaCl_2$, 10 % glycerol, 2 mM βME).

Then, the sample was clarified by centrifugation and applied to 2x5 mL Hi-Trap IMAC FF columns. Unbound, fully-cleaved protein was recovered, and it was applied to a Mono Q HR 16/10 column pre-equilibrated in buffer D (25 mM HEPES pH 8.0, 10 mM NaCl, 10% glycerol, 2 mM βME). Then, the column was washed in an isocratic step of 42% buffer E (25 mM HEPES pH 8.0, 500 mM NaCl, 10% glycerol, 2 mM βME). Protein was eluted in a gradient from 42%-70% buffer E over 400 mL. A-KR-PCP elutes over the gradient as different oligomeric species, confirmed by native-PAGE and size exclusion chromatography. A-KR dimers and monomers were pooled independently and concentrated to at least 10 mg ml^{-1} using a 30 kDa molecular weight cut off Amicon® Ultra centrifugal filter (Millipore). Purified A-KR-PCP_{1-1318} was either modified and further purified for crystallography or used directly for biochemical assays.

A_{83-649} initial purification step used the same Ni-IMAC buffers and conditions as A-KR-PCP_{1-1318}. After TEV cleavage and reapplication to the HiTrap-IMAC column, unbound fully-cleaved protein was recovered and applied to a Mono Q HR 16/10 column pre-equilibrated in buffer G (25 mM HEPES pH 8.0, 100 mM NaCl, 2 mM βME); an isocratic step of 22% buffer H (25 mM HEPES pH 8.0, 500 mM NaCl, 2 mM βME) was applied to the column and relevant fractions were concentrated using a 10 kDa molecular weight cut off Amicon® Ultra centrifugal filter (Millipore) before application to a Superdex-75 10/300 column pre-equilibrated in buffer F (25 mM HEPES pH 8.0, 100 mM NaCl, 0.2 mM tris 2-carboxyethyl phosphine -TCEP-), with A_{83-649} eluting as a single peak. Concentrated proteins were flash frozen in LiN with 20% v/v glycerol and stored at -80 ºC.

Crystallography

Crystallization of A-KR-PCP was challenging and required several iterations to obtain diffraction quality crystals. Initial crystallization conditions were identified with apo or NH_2-ppant modified A-KR-PCP dimer by sparse-matrix search using commercially available screens (Qiagen) in a vapor diffusion sitting drop setup. Even after optimization, these crystals had no edges or a gel-like appearance, diffracted poorly and were often twinned. Therefore, we re-screened for initial conditions using the sparse-matrix micro-seeding (SMMS) method (D'Arcy, Bergfors et al. 2014)

with seeds originating from two of the optimized initial hits. The number of hits increased dramatically by using SMMS compared to no-seeding screening. After iterative rounds of optimization, including streak micro-seeding, additive screens and switching to a vapor diffusion hanging-drop setup, we obtained a final condition that yielded single crystals that diffracted to a maximum resolution of 7 Å. Screening several of these crystals with different cryoprotection methods yielded datasets with maximum resolutions between 4 and 3.4 Å (**Table S3.1).**

For crystallization, purified A-KR dimer protein was incubated for 16 h at room temperature with Sfp (1:1.22 molar ratio), 3 mM NH_2-CoA (kindly provided by Janice Reimer) (Reimer, Aloise et al. 2016), 10 mM $MgCl_2$. Modification was verified by LC-MS. Then, the protein was incubated for 16 hours with 13 mM ATP and 20 mM α-KIC. The sample was analyzed by LC-ESI-MS for modification and applied two subsequent times to an S-200 16/60 column pre-equilibrated in buffer F**.** Relevant fractions were recovered, concentrated to 15.5 mg mL^{-1} and used for crystallization. Final optimized conditions consisted of a hanging drop vapor-diffusion setup of 500 μL reservoir solution (18% v/v PEG-3350, 0.314 M Ca^{2+} Acetate) equilibrating 3 x 2 μL drops. Drops were setup by sequential mixing of concentrated stock solutions of components A (6x stock), components B (3x stock) and phenol (10x stock, Hampton Additive Screen) to final concentrations of components A: 0.17 M Na^+ succinate, 0.25 % v/v PEG 2000 MME, 0.02 M HEPES pH 7.0, 0.004 M HEPES pH 8.0, 0.02 M NaCl; components B: 6% v/v PEG 3350, 0.11 M Ca Acetate; phenol: 0.01 M. Then, a 2.5 x protein stock in buffer F was added to a final concentration of 6.2 mg mL^{-1}. Crystals appeared within seconds or minutes after setting the drop and grew to their final size in 24 h.

Crystals were cryoprotected in sequential additions of increasing volumes and percentages of ethylene glycol (14 μL of 7%, 14 μL of 14%, 14 μL of 21%, 28 μL of 28% v/v) prepared as cryo-solutions containing all the components at concentrations of the pre-equilibrated crystallization drop (see above). The cover slip containing the cryo-protected drop was exposed for one hour to the environment (room temperature) to allow crystal dehydration. Then, 86 μL of drop liquid were exchanged with 86 μL of 28% v/v ethylene glycol cryo-solution, and crystals were flash

cooled under a liquid nitrogen stream. Data was collected at 100 K at the APS-NECAT 24-ID-C beamline using a Dectris PILATUS 6MF pixel array detector.

Initial crystallization conditions for the $A_{83\text{-}649}$ construct were identified by sparse matrix screening using commercial screens (Qiagen). The final optimized condition in a sitting drop vapor diffusion setup consisted of a 500 μL reservoir solution (1.75 M $(NH_4)_2SO_4$ and 22.5% v/v glycerol) equilibrating a drop of 1 μL reservoir solution, 0.2 μL of 30% w/v trimethylamine N-oxide (TMAO) and 0.8 μL protein (10 mg ml^{-1} protein, 2 mM ATP, 10 mM $MgCl_2$, 2 mM α-KIC). Crystals appeared between 24 h and 48 hours and reached their maximum size in three days. Equilibrated drop conditions were already cryoprotected, so crystals were flash cooled into liquid nitrogen. Data was collected at 100 K at the APS-NECAT 24-ID-E beamline using a Dectris EIGER 16M detector.

A-KR data was indexed and integrated to space group $P2_12_12_1$ with the program DIALS (Winter, Waterman et al. 2018) or MOSFLM (Battye, Kontogiannis et al. 2011) and scaled with the program AIMLESS. Molecular replacement (MR) failed with close homologues as search models (found by NCBI-BLAST) using the program Phaser (McCoy, Grosse-Kunstleve et al. 2007). Therefore, we generated better search models through an iterative process. The StrA1 A-KR-PCP sequence was submitted to coevolution analysis using the GREMLIN server in order to search for homologous PDB structures through HHsearch. High-scoring PDBs of A_{core} and KR domains were used as templates to build homology models of the StrA1 A-KR-PCP sequence using SWISS-MODEL (Biasini, Bienert et al. 2014). The resulting homology models were used as input for Ensembler (Adams, Afonine et al. 2011) to generate ensemble search models used by Phaser. An initial MR solution with 4 KR domains was used for iterative AUTOBUILD (Adams, Afonine et al. 2011) (Terwilliger, Grosse-Kunstleve et al. 2008) runs and subsequent placing of 2 A_{core} domains. The structure refined to a R_{free} of 42%. Structure determination was achieved by iterative steps of manual model building and domain placement in Coot (Emsley, Lohkamp et al. 2010), performed on B-factor sharpened- and NCS-averaged maps. The choice of NCS masks and groups was not trivial and required careful model and map analysis to generate meaningful NCS-averaged maps

and NCS-refinement restraints. Also, B-factor sharpening required manual iterative optimization. Refinements were performed using Refmac (Murshudov, Skubak et al. 2011) and phenix.refine (Afonine, Grosse-Kunstleve et al. 2012). The final model refined to R_{free}=28% at 3.4 Å resolution (**Table S3.1**).

Initial phases for $A_{83\text{-}649}$ were obtained using the A domain fragment from the A-KR structure as a MR search model in Phaser (McCoy, Grosse-Kunstleve et al. 2007). Initial maps showed unambiguous and unbiased electron density for a substrate bound in the active site. Iterative rounds of manual model building in Coot (Emsley, Lohkamp et al. 2010) and refinement in phenix.refine (Afonine, Grosse-Kunstleve et al. 2012) generated a near-final model used to compute an unbiased mFo-Fc map just before substrate placing (**Fig 3.4**). A final model was generated after placement of α-KIC adenylate in the active site followed by one round of refinement (**Table S3.1**).

Adenylation assays

Pyrophosphate production assays (Wilson and Aldrich 2010) (Alonzo, Magarvey et al. 2015) were performed in triplicates of 100 μL reactions containing 1 mM acyl substrate, 5 mM ATP, 75 mM TRIS pH 8.0, 0.2 mM 2-amino-6-mercapto-7-methylpurine ribonucleoside (MESG), 1 unit/ml purine nucleoside phosphorylase (PNP), 0.03 units/ml of inorganic pyrophosphatase, 0.15 M hydroxylamine. Reactions were started by the addition of protein to a final concentration of 1 μM. Time courses of the total absorbance were recorded at 360 nm. Initial rates were calculated by linear regression and plotted using the program GraphPad Prism 6.

Ppant arm incorporation and evaluation by top-down LC-MS

Proteins grown in the BL21 $EntD^-$ strain were incubated with SFP in a 1:1.2 ratio with 10 mM $MgCl_2$ and 3 mM CoA. Ppant incorporation was evaluated using the top-down LC-MS methodology described in (Huguenin-Dezot, Alonzo et al. 2019) using a deconvolution window from 135 to 165 kDa.

Single turn-over assay

Fluorescence of NADPH (λ_{exc}=340 nm, λ_{emm}=445 nm) is enhanced when bound to the active site of KR domains. Triplicates of 100 μL reactions consisting of 25 mM HEPES pH 8.0, 100 mM NaCl, 2 mM $MgCl_2$, 2.5 uM NADPH and 9 uM enzyme were set. At these conditions, all NADPH is bound to an excess of protein. Both unmodified (purified from entD$^-$ background) and ppant-modified StrA1 A-KR-PCP samples were tested. Reactions were started by addition of 20 μL of H_2O or a mix of 10 mM α-KIC + 10 mM ATP to the sample and total fluorescence intensity (λ_{exc}=340 nm, λ_{emm}=445 nm) was recorded in black polystyrene 96 well plates (Corning) in a Molecular Devices plate reader. Data was analyzed using GraphPad Prism 6.

Bioinformatics

Structural alignments were performed with Clustal Omega (Sievers, Wilm et al. 2011) or the CLC Sequence Viewer tools. Biosynthetic cluster analysis was performed with AntiSmash (Weber, Blin et al. 2015).

3.6. Supplementary Figures

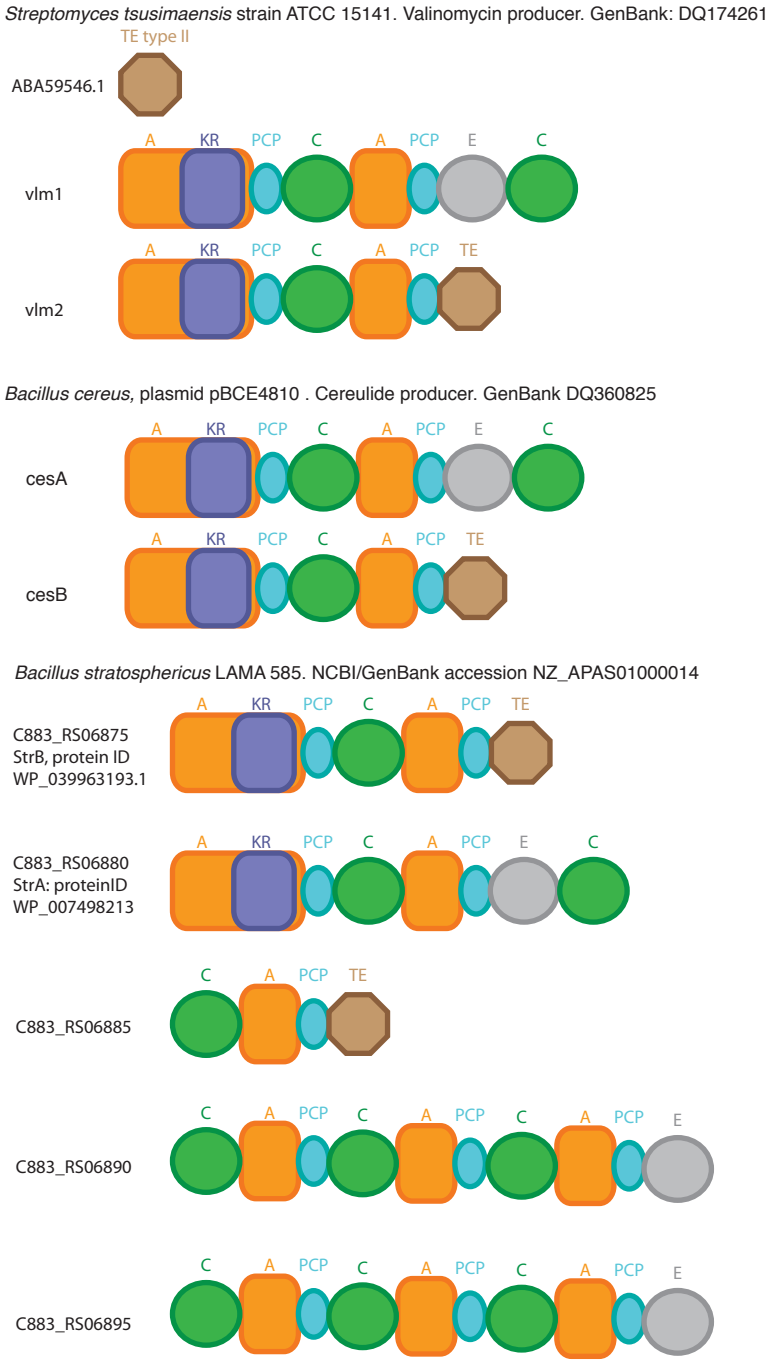


Figure S3.1. The biosynthetic gene clusters of cereulide, valinomycin and a putative cyclic depsipeptide.

AntiSmash (see **Methods**) analysis was performed on the genomic sequences from the valinomycin producer *Streptomyces tsusimaensis* (top), the cereulide producer *Bacillus cereus* (middle) and a

putative cyclic depsipeptide producer *Bacillus stratosphericus* LAMA 585 (bottom). The clusters from the three different species encode genes for two NRPSs with KR domains. Additionally, the cluster from *B. stratosphericus* LAMA 585 encodes for three additional NRPS genes, not present in the clusters from *B. cereus* and *S. tsusimaensis*.

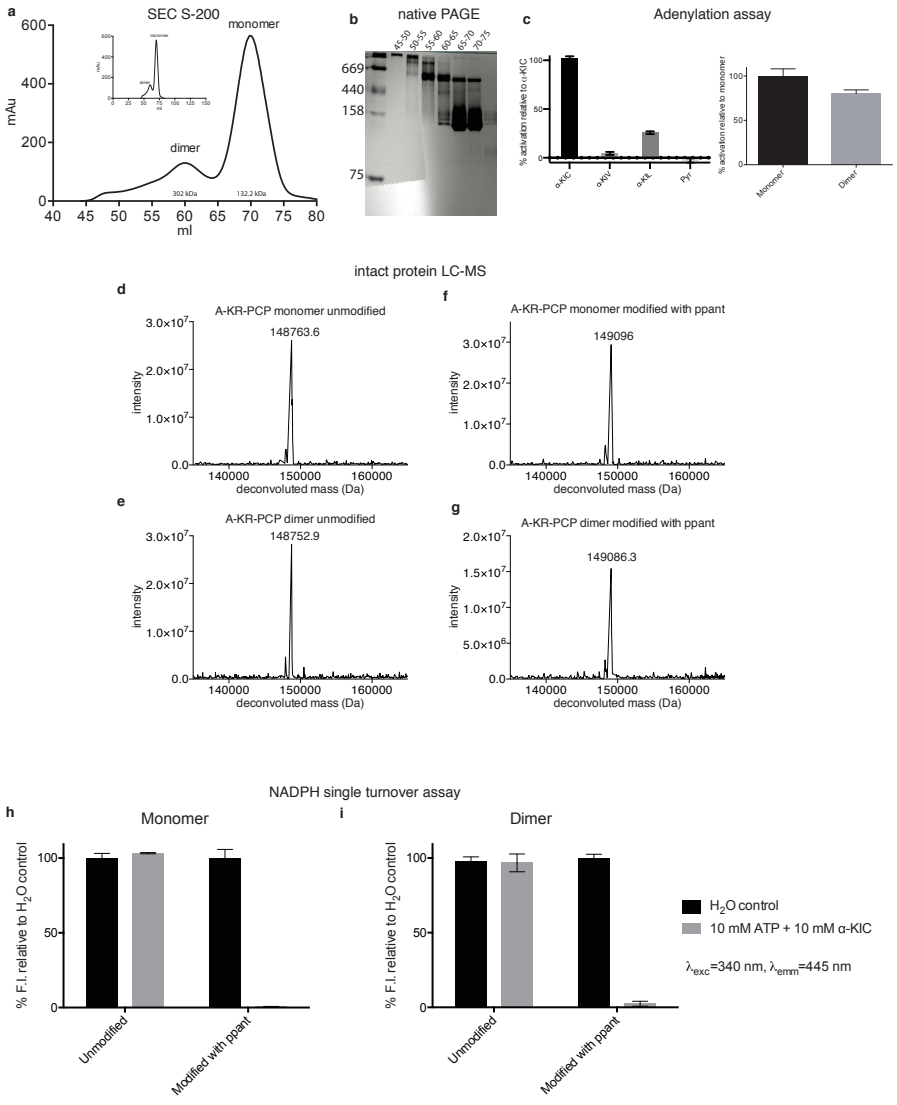


Figure S3.2. The monomer is the predominant species of StrA1 A-KR-PCP and both oligomeric species are catalytically active.

a-b. Size exclusion chromatogram (**a**) and corresponding native PAGE (**b**) from StrA1. Most of the protein elutes at the volume corresponding to the MW of the monomer. A small fraction elutes as a dimer, and an even smaller fraction elutes as a higher-order multimer. **c**. **Left**, substrate selectivity of StrA1. The

adenylation activity of StrA1 was tested in the presence of α -ketoacids with different sidechains. StrA1 selects α -ketoisocaproic acid (α-kic), and also shows partial selectivity towards α-ketoisoleucine (α-kil). α-kiv: alpha ketoisovaleric acid, α-kic: alpha ketoisocaproic acid, α-kil: alpha ketoisoleucine, pyr acid: pyruvic acid. **Right**, monomer and dimer fractions adenylate α-kic. The continuous adenylation assay was performed with purified monomer and dimer fractions in the presence of α-kic. The assays were performed in triplicate. **d-g.** Monomer (**d, e**) and dimer (**f, g**) fractions can be modified with a phosphopantetheinyl arm. Intact protein (LC-ESI-MS) of purified A-KR-PCP overexpressed in the entD$^-$ strain (left panels), and after being modified with SFP and CoA (right panels). Expected molecular masses: 148737.4 Da (unmodified), 149075.5 (modified). Observed molecular masses: 148763.6 (unmodified monomer), 149096 (modified monomer), 148752.9 (unmodified dimer), 149086.3 (modified dimer). Experiments were repeated independently three times with similar results. **h-i.** Monomer (**h**) and dimer (**i**) fractions catalyze thiolation and ketoreduction. A fluorescence-based assay was used to evaluate a single turnover of the thiolation and ketoreduction reactions. The fluorescence of NADPH (λ_{exc}=340 nm, λ_{emm}=445 nm) is enhanced when bound to the active site of KR domains. The loss of enhanced fluorescence of sub-stoichiometric amounts of NADPH indicates oxidation to NADP$^+$, which is not fluorescent. NADPH is consumed by monomer (**h**) and dimer (**i**) fractions in the presence of ATP + α-kic only when the protein is modified with ppant, confirming that the thiolation reaction is needed for ketoreduction. The assay was performed in triplicate.

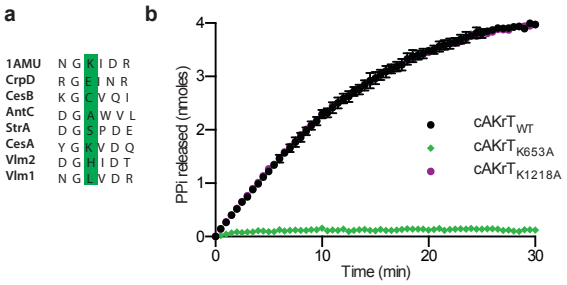


Figure S3.3. Role of the A10-like motif in the pseudo A_{sub} domains.

a. Detail of a sequence alignment of different A-KR-PCP modules. The A10-like motif present in the pseudo-A_{sub} domain is compared with the canonical A10 motif from grsA (PDB ID: 1AMU). The lysine residue in canonical A10 motifs interacts with the phosphate group of adenyl-nucleotides. The lysine residue is only present in the pseudo-A_{sub} domain of CesA. **b**, Mutating the lysine residue in the canonical A10 motif of CesA suppresses adenylation activity. If the lysine present in the pseudo-A_{sub} domain was active, activity complementation would be observed. Mutating the lysine in the pseudo A_{sub} domain does not have any effect on adenylation activity

Table 3.1. Data collection and refinement statistics.

	A-KR (cystal 5_171_b2_4)	A_{83-649} bound to adenylate (crystal 7_039_b3_4)
Data collection		
Space group	P 21 21 21	P 43 21 2
Cell dimensions		
a, *b*, *c* (Å)	142.0, 143.7, 432.31	103.0, 103.0, 172.3
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	142.0-3.4 (3.47-3.41)	88.4-2.3 (2.38-2.3)
R_{sym} or R_{merge}	0.106 (2.509)	0.152 (2.085)
$I / \sigma I$	9.2 (0.8)	11.0 (1.9)
Completeness (%)	98.4 (86.3)	99.8 (99.8)
Redundancy	6.5 (5.7)	13.2 (13.8)
Refinement		
Resolution (Å)	136.3-3.4	73.0-2.3
No. reflections	118700	41818
R_{work} / R_{free}	0.2302/0.2803	0.2161/0.2514
No. atoms	37781	4255
Protein	37772	4072
Ligand/ion	9	121
Water	0	62
B-factors	162.42	59.84
Protein	162.41	59.31
Ligand/ion	195.07	82.15
Water	na	51.19
R.m.s. deviations		
Bond lengths (Å)	0.014	0.013
Bond angles (°)	1.93	1.51

Each data set was collected from a single crystal.
Values in parentheses are for highest-resolution shell.

.

Table S3.1. Data collection and refinement statistics

3.7. Connector to chapter 4

The Ces and Vlm A-KR-PCP modules participate in chain initiation and elongation, contributing to the formation of a linear tetradepsipeptidyl-*S*-PCP intermediate that the TE domain trimerizes and cyclizes to yield mature cereulide or valinomycin (**Fig. 1.17** and **3.1**). There are several open questions about TE-catalyzed oligomerization and cyclization, specially into the mechanism of intermediate fate determination. High-resolution structures of TE domains bound to their acyl-intermediates should be powerful tools to provide insight into these processes. However, such initiatives with NPRS TE domains have been hampered by the combination of low K_m values for small-molecule substrate analogues and fast hydrolysis rates of acyl-TE intermediates (Bruner, Weber et al. 2002) (Samel, Wagner et al. 2006). In the next chapter, collaborators developed a system for genetic incorporation of the unnatural amino acid 2,3-diaminopropionic acid (DAP) that allowed us to form, crystallize and determine the structure of depsipeptidyl-*NH*-TE complexes of Vlm TE. This work provided a proof of concept for the genetic code expansion system along with crucial structural information on the reactions catalyzed by Vlm TE, especially on the participation of the active site lid during depsipeptide cyclization.

CHAPTER 4. TRAPPING BIOSYNTHETIC ACYL-ENZYME INTERMEDIATES WITH ENCODED 2,3-DIAMINOPROPIONIC ACID

Nicolas Huguenin-Dezot*, Diego A. Alonzo*, Graham W. Heberlig, Mohan Mahesh, Duy P. Nguyen, Mark H. Dornan, Christopher N. Boddy, T. Martin Schmeing,*, Jason W. Chin,*. Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid. (2019) Nature, 565, pages 112–117

* Denotes co-author

Author contributions

NHD and DPN selected the synthetases for **2**-**6**. NHD characterized the incorporation of **6** and its deprotection to DAP and performed the TEV conjugation. MM synthesized **2**-**6**. TMS and DAA designed, and DAA performed the biochemistry and crystallography with Vlm TE_{wt}. DAA and NHD performed biochemistry and crystallography with Vlm TE_{DAP}. GWH and MHD synthesized Vlm TE substrates. JWC supervised NHD, DPN and MM. TMS supervised DAA. CNB supervised GWH and MHD. TMS, DAA, NHD, and JWC wrote the paper with input from all authors.

4.1 Abstract

Many enzymes catalyse reactions that proceed through covalent acyl-enzyme (ester or thioester) intermediates (Holliday, Mitchell et al. 2009). These include serine hydrolases (Hedstrom 2002) (Long and Cravatt 2011) (that make up 1% of human genes, including serine proteases and thioesterases), cysteine proteases (including caspases), and many components of the ubiquitination machinery (Otto and Schirmeister 1997) (Swatek and Komander 2016). Their important acyl-enzyme intermediates are unstable, commonly having half-lives of minutes to hours (Yang and Drueckhammer 2001). In some cases, acyl-enzyme complexes can be stabilised using substrate analogues or active site mutations but, while these approaches can provide valuable insight (Liu, Schofield et al. 2006) (Scaglione, Akey et al. 2010) (Cappadocia and Lima 2018) (Plechanovova, Jaffray et al. 2012), they often result in complexes that are substantially non-native. Here we develop a strategy for incorporating 2,3-diaminopropionic acid (DAP) into recombinant proteins, via genetic code expansion (Chin 2014). We show that replacing catalytic cysteine or serine residues of enzymes with DAP permits their first-step reaction with native substrates, allowing efficient capture of acyl-enzyme complexes that are linked through a stable amide bond. For one of these enzymes, the thioesterase domain of valinomycin synthetase (Magarvey, Ehling-Schulz et al. 2006) (Vlm TE), we elucidate the biosynthetic pathway by which it progressively oligomerises tetradepsipeptidyl substrates to a dodecadepsipeptidyl intermediate that it cyclises to valinomycin. By trapping the first and last acyl-TE intermediates in the catalytic cycle as DAP conjugates, we provide structural insight into how conformational changes in TE domains of such nonribosomal peptide synthetases (NRPSs) control oligomerisation and cyclisation of linear substrates. The encoding of DAP will facilitate the characterisation of diverse acyl-enzyme complexes, and may be extended to capturing the native substrates of transiently-acylated proteins of unknown function.

4.2 Main text

We hypothesised that selectively replacing the sulfhydryl or hydroxyl groups in catalytic cysteine or serine residues with an amino group, making 2,3-diaminopropionic acid (DAP, **1**), would allow trapping of acyl-enzyme intermediates linked via an amide bond (**Fig. 4.1a, b**). Within a peptide, the conjugate acid of the β-amino group of DAP has a pKa of ~6.3-7.5 (c.f. lysine ~10.5) (Lan, Langlet-Bertin et al. 2010), which suggests that this amine could act as a nucleophile, and may form amide bonds with enzymes' substrates. The half-life of amides in aqueous solution is ~500 years (Radzicka and Wolfenden 1996), so the amide analogues of labile thioester and ester intermediates should be substantially stabilised, such that subsequent reactions with nucleophiles or solvent should be severely attenuated or abolished (**Fig. 4.1b**).

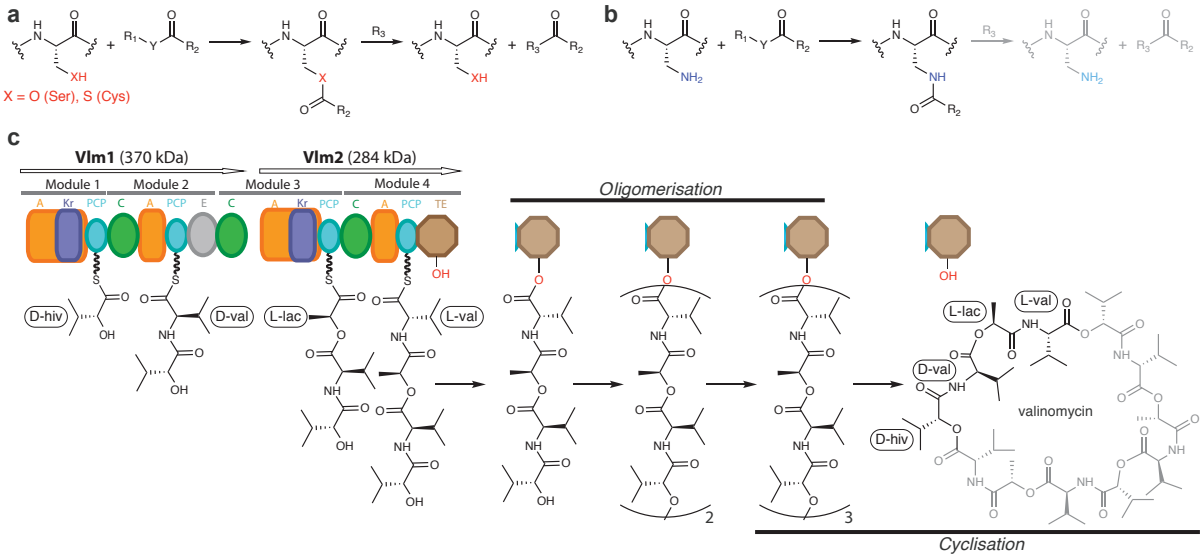


Figure 4.1: Capturing transient acyl-enzyme intermediates with DAP and the proposed biosynthesis of valinomycin.

a, Active site serine or cysteine residues react with carbonyl groups to form tetrahedral intermediates (not shown) that collapse to the acyl-enzyme intermediates by loss of R_1-YH. Attack by nucleophilic R_3 (commonly a hydroxyl, amine or thiol) releases the bound substrate fragment and regenerates the enzyme. **b**, Replacing cysteine or serine with DAP may result in a first acyl-enzyme intermediate which is resistant to cleavage. **c,** Valinomycin synthetase condenses D-α-hydroxyisovaleric acid (D-α-hiv), D-valine (D-val), L-lactic acid (L-lac), and L-valine (L-val) to form the tetradepsipeptidyl (D-hiv-D-val-L-lac-L-val) intermediate. D-α-hiv and L-lac arise from the reduction of precursor ketoacyl moieties by

ketoreductase (KR) domains. Tetradepsipeptidyl intermediates are oligomerised to a dodecadepsipeptidyl intermediate which is cyclised, by the terminal TE domain, producing valinomycin. A: adenylation; PCP: peptidyl carrier protein; C: condensation domains. See **Extended Data Fig. 4.1** for a synthetic cycle of an NRPS.

The secondary metabolite–producing megaenzymes nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) generate complex acyl-enzyme intermediates during their synthetic cycles (Reimer, Haque et al. 2018). These molecular machines use thio-templated pathways to assemble small acyl molecules into a broad array of biologically-active natural products, including antitumourals, antibiotics, antifungals, and immunosuppressants (**Extended Data Fig. 4.1**). Unravelling their molecular mechanisms has been hampered by the challenge of characterising their multiple acyl-enzyme intermediates at high resolution.

This challenge is exemplified with the thioesterase domains (Horsman, Hari et al. 2015) from NRPS pathways that oligomerise and cyclise linear peptidyl substrates (Hoyer, Mahlert et al. 2007) including the TE domain from valinomycin synthase (**Fig. 4.1c**) (Magarvey, Ehling-Schulz et al. 2006). Valinomycin synthetase, a 2-protein, 4-module NRPS, alternatively links hydroxy acids (from *in situ* reduction of α-keto acids) and amino acids into a tetradepsipeptide intermediate, which the TE domain (Vlm TE) oligomerises up to, but not beyond, the dodecadepsipeptide. Vlm TE then cyclises the dodecadepsipeptide to release valinomycin (Jaitzig, Li et al. 2014) (Magarvey, Ehling-Schulz et al. 2006) (a potassium ionophore with antimicrobial, antitumoural and cytotoxic properties) (**Fig. 4.1c**). The oligomerisations and cyclisation must be rapid enough to prevent substantial spontaneous hydrolysis to linear depsipeptides, which are useless side products.

High-resolution structures of acyl-TE intermediates in valinomycin biosynthesis could provide mechanistic insight into how TEs control substrate fate. A handful of high-resolution acyl-TE structures have been obtained, most notably with the polyketide pikromycin–forming TE and non-native substrate analogues (**Supplementary Data 4.1**) (Akey, Kittendorf et al. 2006). These have helped identify the putative oxyanion hole and demonstrated the interaction of the “lid” element of the TE domain with the substrate. Structural studies of TE domains have been

hampered by several factors, especially the hydrolysis rates of acyl-TE intermediates (Bruner, Weber et al. 2002) (Samel, Wagner et al. 2006), which are high compared to the crystallographic time scale.

Here we develop a strategy for the site-specific incorporation of DAP into recombinant proteins and demonstrate the efficient capture of acyl-enzyme intermediates for a cysteine protease and Vlm TE. We elucidate the biosynthetic pathway for converting tetradepsipeptides to valinomycin and structurally characterise deoxy-tetradepsipeptidyl-*N*-TE_{DAP} and dodecadepsipeptidyl-*N*-TE_{DAP} conjugates to provide insights into the first and last acyl-TE intermediates in the catalytic cycle of Vlm TE. Our results reveal how the fate of substrates may be determined by conformational changes in the TE domains of NRPSs that oligomerise and cyclise linear precursors.

The structural similarity of DAP (**1**) to cysteine and serine makes it challenging to discover an aminoacyl-tRNA synthetase that is selective for DAP *in vivo*. We therefore created five protected versions of DAP (**2-6**) (**Extended Data Fig. 4.2a**), for which we anticipated the successful discovery of specific aminoacyl-tRNA synthetase—$tRNA_{CUA}$ pairs would enable site-specific incorporation into proteins. The subsequent post-translational deprotection (Li, Yu et al. 2014) (Baker and Deiters 2014) would reveal DAP. **2-5** accumulated in *E. coli* at low concentrations (<10 μM) (**Extended Data Fig. 4.2b-e**) and we were unable to evolve a synthetase for these noncanonical amino acids (ncAAs) using several libraries of the orthogonal *Methanosarcina barkeri* (*Mb*) pyrrolysyl-tRNA synthetase (PylRS)/$tRNA^{Pyl}_{CUA}$ pair (Chin 2014) (**Supplementary Methods**). In contrast, **6** accumulated in *E. coli* at millimolar concentrations and we were able to evolve an *Mb* PylRS variant (named DAPRS, containing mutations Y271C, N311Q, Y349F and V366C), for the site-specific incorporation of **6** (**Extended Data Fig. 4.2f-h**). The DAPRS/$tRNA^{Pyl}_{CUA}$ pair synthesised GFP with **6** at position 150 (GFP150(**6**)) (**Fig. 4.2a**) in good yield (Virdee, Ye et al. 2010). Photo-deprotection of GFP(**6**) and subsequent incubation revealed **1** in GFP (**Fig. 4.2b, c**).

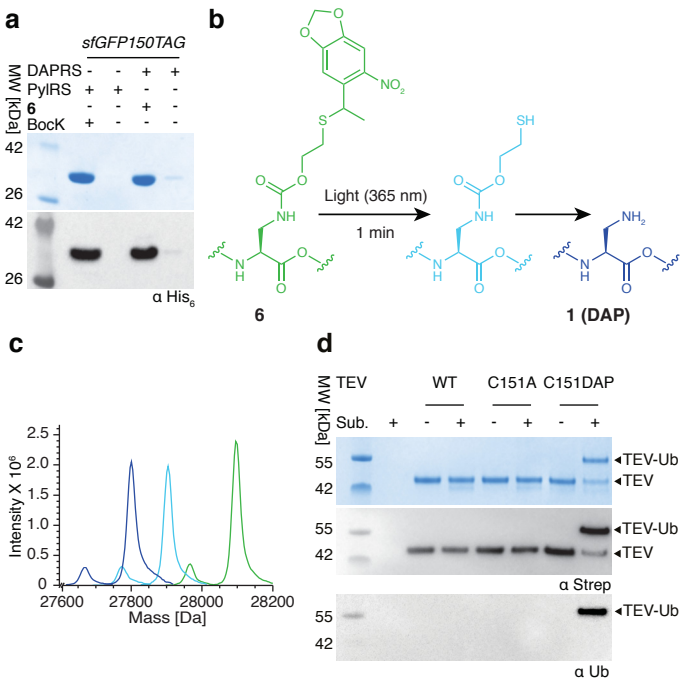


Figure 4.2: Genetically directing DAP incorporation in recombinant proteins and stably trapping the acyl-enzyme intermediate of a cysteine protease.

a, SDS-PAGE gel of GFP150(**6**) and GFP150(BocK), with protein detected by Coomassie (top) or α-His_6 antibody (bottom); the experiment was performed in two biological replicates with similar results. BocK is Nε-[(tert-butoxy)carbonyl]-l-lysine. **b**, Encoded **6** was photo-deprotected, leading to an intermediate, which spontaneously fragments to reveal DAP. **c**, Deprotection of **6** in sfGFP followed by ESI-MS analysis. Green trace: Purified GFP150(**6**): Expected: 28096.27 Da; Observed: 28097.21 Da. Light blue trace: intermediate: Expected: 27902.22 Da; Observed: 27904.14 Da. Dark blue trace: incubation (10h, 37°C) converts the intermediate to DAP **(1)**: Expected: 27798.23 Da; Observed: 27800.88 Da. Minor peaks due to loss of the N-terminal methionine are also observed. The experiment was performed in two biological replicates with similar results. **d**, TEV protease variants were incubated with Ub-tev-His. TEV(C151DAP)–Ub is the amide bond linked complex. αUb and αStrep western blots confirm the identity of the complex (TEV constructs contain a Strep tag). The experiment was performed in two biological replicates with similar results.

Cysteine proteases, including the tobacco etch virus (TEV) protease, react with substrates through a catalytic cysteine to generate an intermediate in which the protease is linked to the N-terminal portion of its substrate through a thioester (Otto and Schirmeister 1997). We replaced

the active site cysteine 151 of TEV protease with DAP, by genetically encoding **6** and deprotecting it, creating TEV(C151DAP). Incubating TEV(C151DAP) with Ub-tev-His_6, a model substrate in which the cleavage site recognised by TEV protease (tev) is flanked by ubiquitin and a hexahistidine tag (His_6), led to cleavage of the His_6 tag from ubiquitin and formation of a covalently linked TEV(C151DAP)-Ub conjugate (**Fig. 4.2d, Extended Data Fig. 4.3a-c**). Control experiments demonstrated that the stable conjugate was dependent on DAP incorporation. MS/MS demonstrated amide bond formation between DAP and Ub (**Extended Data Fig. 4.3d**). These results demonstrate that substitution of the catalytic cysteine in TEV with DAP creates a protease that performs the first step of the protease cycle, releasing the C-terminal fragment of the substrate and leaving the N-terminal fragment covalently attached to the protease through a stable amide bond that is resistant to hydrolysis.

To gain insight into TE domain function and to prepare it for use with the DAP system, we expressed and purified Vlm TE (TE_{wt}). We found that Vlm TE can use a N-acetylcysteine (SNAC) derivative of the native depsipeptide (tetradepsipeptidyl-SNAC **7**, **Extended Data Fig. 4.4**) to complete all stages of its catalytic cycle and yield valinomycin (**Fig. 4.1c**, **Fig. 4.3a, Extended Data Fig. 4.5**). Thus, **7** can mimic the natural phosphopantetheine-PCP linked substrate, consistent with prior observations with other TE domains and substrates (Tseng, Bruner et al. 2002) (Hoyer, Mahlert et al. 2007) (Samel, Wagner et al. 2006).

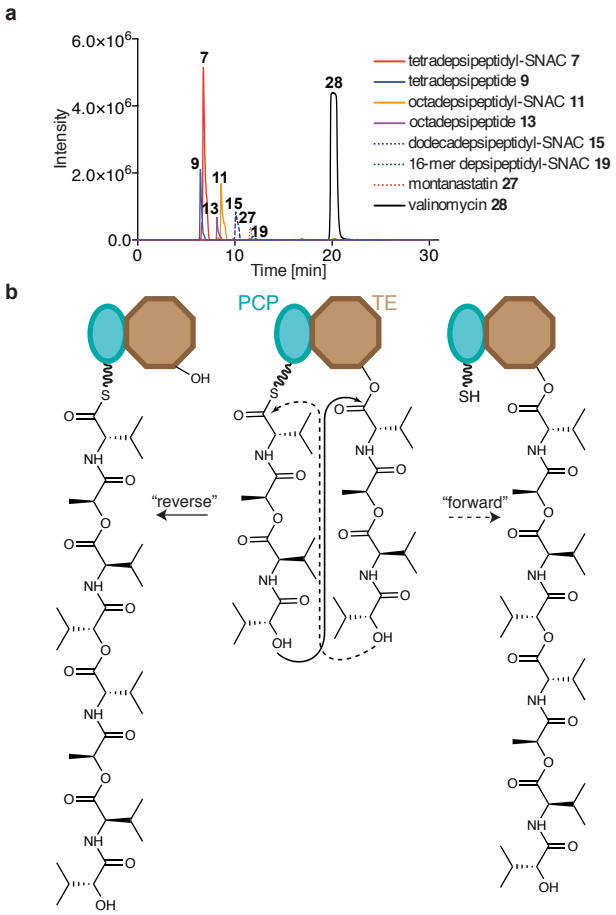


Figure 4.3: Vlm TE produces valinomycin and intermediates which delineate the oligomerisation pathway from tetradepsipeptidyl-SNAC.

a, Extracted ion chromatograms (EICs) from HR-LC-ESI-MS of reactions of tetradepsipeptidyl-SNAC **7** (1.7 mM) and Vlm TE (6.5 µM); TE_{wt} produces valinomycin as its major product. The experiment was performed two independent times with similar results. See **Supplementary Methods for Statistics and Reproducibility** for mass analysis and deviations from calculated m/z values. **b**, Two scenarios for oligomerisation(Trauger, Kohli et al. 2000, Hoyer, Mahlert et al. 2007). In the "forward transfer" scenario the distal hydroxyl group of tetradepsipeptidyl-*O*-TE attacks the thioester group in tetradepsipeptidyl-*S*-PCP, directly forming octadepsipeptidyl-*O*-TE. In the "reverse" scenario the distal hydroxyl group of tetradepsipeptidyl-*S*-PCP attacks the ester group in tetradepsipeptidyl-*O*-TE, forming octadepsipeptidyl-*S*-PCP, which would later be transferred onto the TE domain serine. Our data are consistent with the "reverse" oligomerisation scenario; see also **Extended Data Fig. 4.5**.

There are two possible pathways for oligomerisation of NRPS intermediates by TE domains, and the synthetic intermediates detected in valinomycin synthesis revealed that Vlm TE_{wt} catalyses oligomerisation via a "reverse transfer" pathway (**Fig. 4.3b, Extended Data Fig. 4.5**, **Supplementary Discussion 4.1**). This pathway is analogous to that used by the more canonical gramicidin S synthetase (Trauger, Kohli et al. 2000) (Hoyer, Mahlert et al. 2007) and we suggest that nearly all oligomerising-cyclising NRPSs (or PKSs (Zhou, Prediger et al. 2015)) will use this synthetic scheme.

We next obtained the structure of Vlm TE_{wt} (**Extended Data Fig. 4.6 and Extended Data Table 4.1**). It adopts the α/β hydrolase fold typical of type I TE domains, with a canonical Ser-His-Asp catalytic triad(Horsman, Hari et al. 2015) covered by the TE "lid". The lid is a mobile element with proposed roles that include substrate positioning and solvent exclusion(Samel, Wagner et al. 2006, Frueh, Arthanari et al. 2008, Whicher, Florova et al. 2011). The lid of Vlm TE is large, composed of an extended loop, three helices (Lα1-3, seen here as a bundle), a 5-residue helix (Lα4), a long helix (Lα5) and another short helix (Lα6) (**Extended Data Fig. 4.6a, b**). We obtained another structure of TE_{wt} that differs only in the lid. In the first, the lid is nearly completely ordered, although the B factors are markedly higher for the Lα1-4 region, which makes almost no contact with the rest of the domain (**Extended Data Fig. 4.6b**). In the second structure, Lα4-5 have similar positions to those in the first structure, whereas Lα3 is rotated 10° towards the active site and Lα1-2 are too disordered to model.

Incubating Vlm TE with depsipeptidyl-SNACs did not yield stable conjugates **(Extended Data Fig. 4.6c**), and attempts to soak TE_{wt} crystals with depsipeptidyl-SNACs failed to reveal interpretable ligand electron density in the active site or conformational changes. Others have reported similar setbacks when attempting to visualise acyl-enzyme complexes from SNAC molecules (**Supplementary Data 4.1**) (Bruner, Weber et al. 2002) (Samel, Wagner et al. 2006). We conclude that acyl-intermediates in valinomycin biosynthesis are not stable, and it is exceptionally challenging to use wild type Vlm TE to visualise biosynthetic intermediates.

We therefore produced Vlm TE in which the active site serine 2463 was replaced by DAP (TE_{DAP}) (**Extended Data Fig. 4.7a-c**) to capture stable acyl-TE conjugates (**Fig. 4.4**). To provide insight into the first acyl-TE intermediate in the catalytic cycle of Vlm TE, we captured tetradepsipeptidyl-*N*-TE_{DAP}: Incubation of TE_{DAP} with tetradepsipeptidyl-SNAC **7** led to production of a stable depsipeptidyl-TE_{DAP} intermediate at >60% yield (**Extended Data Fig. 4.7d**), and we did not observe valinomycin synthesis. However, remarkably, a small amount of octadepsipeptidyl-SNAC **11** was observed (**Extended Data Fig. 4.5f, h**). **11** is likely formed by enzyme–catalysed attack of the tetradepsipeptidyl-SNAC's hydroxyl group on the amide bond linking the tetradepsipeptide to TE_{DAP}. The attack of a hydroxyl on an amide is analogous to the first reaction employed by related serine proteases (Ekici, Paetzel et al. 2008), but it is surprising that this TE domain is capable of catalysing a more demanding chemical reaction (amide cleavage) than the reaction (ester hydrolysis) it evolved to perform. In an effort to enhance conjugate yield we optimised conditions for conjugating deoxy-tetradepsipeptidyl-SNAC **8** with TE_{DAP}, which produced (~70%) the deoxy-depsipeptidyl-TE_{DAP} conjugate (**Fig. 4.4a)**. The marginal solubility of these hydrophobic SNACs may limit the conjugation efficiency.

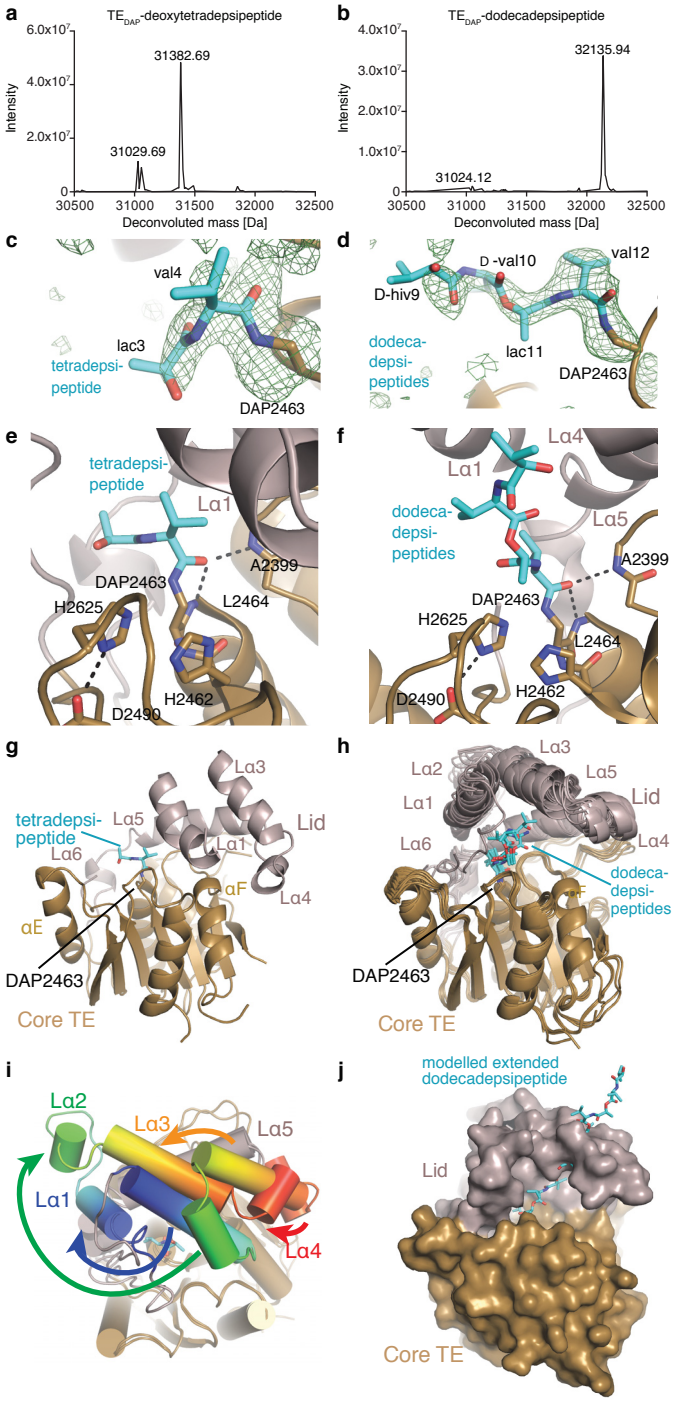


Figure 4.4: Crystal structures of complexes of TE$_{DAP}$.

Deconvoluted mass spectra of **a,** TE$_{DAP}$ incubated with deoxy-tetradepsipeptidyl-SNAC **8**: Expected masses: 31027.24 Da (unmodified) and 31383.70 Da (modified); observed: 31029.69 Da and 31382.69. **b,** TE$_{DAP}$ incubated with valinomycin: Expected masses: 31027.24 Da (unmodified) and 32139.11 Da (modified); observed: 31024.12 Da and 32135.94. The experiments were repeated independently 5 times with similar results. **c** and **d,** Unbiased mFo-DFc map (green mesh, 2.5 σ) for depsipeptide residues

of tetradepsipeptidyl-TE_{DAP} (**c,** PDB ID: 6ECD) and dodecadepsipeptidyl-TE_{DAP} (**d,** PDB ID: 6ECE, 6ECF). An amide bond links DAP (brown) and depsipeptide residues (cyan). **e** and **f**, The active site of tetradepsipeptidyl-TE_{DAP} (**e**) and dodecadepsipeptidyl-TE_{DAP} (**f**). The carbonyl oxygen of the amide formed by DAP and val4 (**e**) or val12 (**f**) is positioned close to the oxyanion hole formed by the main chain of A2399 and L2464. Catalytic triad residues H2625 and D2490 are shown in sticks. **g,** The lid of tetradepsipeptidyl-TE_{DAP} (6ECD) is in a similar position to that seen in TE_{wt} (6ECB) while **h,** all crystallographically independent molecules of the dodecadepsipeptidyl-TE_{DAP} (6ECE, 6ECF) are in a set of similar conformations, distinct from that seen in TE_{wt}. **i,** and **Supplementary Video 1 & 2** illustrate substantial conformational changes in lid helixes Lα1-Lα4 between conformations of tetradepsipeptidyl-TE_{DAP} (**g**) and dodecadepsipeptidyl-TE_{DAP} (**h**). **j,** In the dodecadepsipeptidyl-TE_{DAP} structure the lid sterically prevents the dodecadepsipeptide from extending out in a linear fashion, and favours curling back through this steric block and largely hydrophobic, non-specific interactions between lid and dodecadepsipeptide.

To determine the structure of the deoxy-tetradepsipeptidyl-*N*-TE_{DAP} conjugate, we incubated TE_{DAP} crystals with the deoxy-tetradepsipeptidyl-SNAC **8**. The resulting electron density shows somewhat weak, but unambiguous density for an amide bond between DAP2463 and *L*-val4 of the deoxy-tetradepsipeptide (**Fig. 4.4c, Extended Data Fig. 4.8a**). The carbonyl oxygen of the *L*-val4 is close to backbone amides of residues Ala2399 and Leu2464, the putative oxyanion hole (**Fig. 4.4e**) (Tseng, Bruner et al. 2002). There is also density for the next residue, *L*-lac3, but it is insufficient to reliably model *D*-val2 and *D*-hiv1 as the deoxy-tetradepsipeptide arcs out, indicating substrate flexibility. The deoxy-tetradepsipeptide does not make any interactions with the lid, which is in a conformation nearly identical to that in the first TE_{wt} (apo) structure (**Fig. 4.4g**, **Extended Data Fig. 4.6d**).

Next, we targeted insight into the last acyl-TE intermediate in the catalytic cycle. We captured dodecadepsipeptidyl-*N*-TE_{DAP} in ~65-100% yield upon incubation of valinomycin and TE_{DAP}, through a ring opening reaction analogous to the reverse of the natural cyclisation (**Fig. 4.4b**). This reaction is thermodynamically favoured by virtue of amide bond formation. Dodecadepsipeptidyl-TE_{DAP} produced crystals in similar conditions to those of TE_{wt}, but with a

different morphology and belonging to two different space groups (H3 and P1, with 2 and 6 molecules per asymmetric unit, respectively) (**Extended Data Table 4.1**). All eight crystallographically-independent molecules of dodecadepsipeptidyl-TE_{DAP} showed some density for the dodecadepsipeptide. Molecules P1_A-F and H3_A-B show strong density for 4, 3, 2, 2, 2, 2, 3 and 1 dodecadepsipeptide residues respectively (**Fig. 4.4d**, **Extended Data Fig. 4.8b-i**). Additional weaker density is present in some molecules, which could accommodate up to the full 12 residues (**Extended Data Fig. 4.8j-l**), and in others weaker density suggests multiple conformations for the distal residues, but they were not possible to definitively model. The modelled depsipeptides all follow a similar trajectory away from the active site. There is no consistent interaction between the depsipeptide beyond the *L*-val residue attached to DAP and the TE domain (**Fig. 4.8f, h)**. Rather, each depsipeptide makes different contacts with the lid. The lid forms a semi-sphere-like pocket / steric barrier made up of helices Lα1, 3, 4 and 5, and the strand N-terminal to Lα1. The lid of each crystallographically-independent molecule of dodecadepsipeptidyl-TE_{DAP} is in a similar but non-identical position, and the loops between lid helices are disordered in most molecules (**Fig. 4.4h**). This again highlights the mobility of the lid and explains why the conformation and extent of order of dodecadepsipeptides differ between molecules (**Fig. 4.4h**). The semi-sphere-like barrier occurs only because of a major rearrangement of the lid in the dodecadepsipeptidyl-TE_{DAP} structures with respect to the conformation of the lid seen in both the apo and tetradepsipeptidyl-bound structures of Vlm TE.

Comparing the position of the Vlm TE lid in the apo/tetradepsipeptide-bound structures with the position of the lid in the dodecadepsipeptide-bound structures demonstrates and emphasises its extreme mobility. To transition from one lid conformation to the other, helices Lα5-6 maintain their position, Lα3-4 rotate ~45° and translocate ~13 Å, Lα2 translocates ~25 Å, and Lα1 shortens, translocates ~13 Å, and rotates >90° in the opposite direction to Lα3-4 (**Fig. 4.4i, Supplementary Video 1, 2**). This dramatic re-arrangement means the lid helices pack together in a markedly different manner in the apo/tetradepsipeptidyl-bound structure and in dodecadepsipeptidyl-bound conformations.

The distinct lid conformations directly influence the possible location of the depsipeptide. In the apo/tetradepsipeptide-bound conformation of the lid, the C-terminus of Lα1 comes within 10 Å of Ser/DAP2463, leading the tetradepsipeptide to extend towards the TE core helix αE. In the dodecadepsipeptide-bound conformations of the lid, the loop adjacent to Lα1 blocks the location occupied by the tetradepsipeptide in the tetradepsipeptide-bound structure. Moreover, in the dodecadepsipeptide-bound structure the N-terminus of Lα1 forms part of the semi-sphere-like pocket. This pocket likely helps curl the dodecadepsipeptide back towards Ser/DAP2463 entropically controlling cyclisation as part of the oligomerisation/cyclisation pathway (**Fig. 4.4j, Extended Data Fig. 4.9** and **Supplementary Discussion 4.2**).

In summary, we have genetically encoded DAP in place of catalytic cysteine or serine residues to capture unstable thioester or ester intermediates as stable amide analogues. We have exemplified the utility of this approach for a cysteine protease and a thioesterase, and provided unique insight into intermediates in the synthesis of valinomycin: a massive lid rearrangement that is associated with the dodecadepsipeptidyl-bound Vlm TE re-orients the substrate from its position during oligomerisation and places it into a pocket that entropically controls cyclisation. Importantly, the DAP system allows formation of near-native acyl-enzyme complexes with widely-used, reaction competent, substrates (e.g. native proteins containing protease sites), substrate analogues (here SNACs), and commercially available natural products (here valinomycin, and likely other cyclic products (Tseng, Bruner et al. 2002)). We anticipate the approach will be broadly applicable and may be extended to capturing native substrates of proteins of unknown function that form acyl-protein conjugates.

Data availability

Source data for all figures are available from the corresponding authors upon reasonable request. The models and structure factors for the crystal structures are deposited in at the Protein Data Bank with accession numbers 6ECB, 6ECC, 6ECD, 6ECE and 6ECF.

Supplementary Information is linked to the online version of the paper at

www.nature.com/nature.

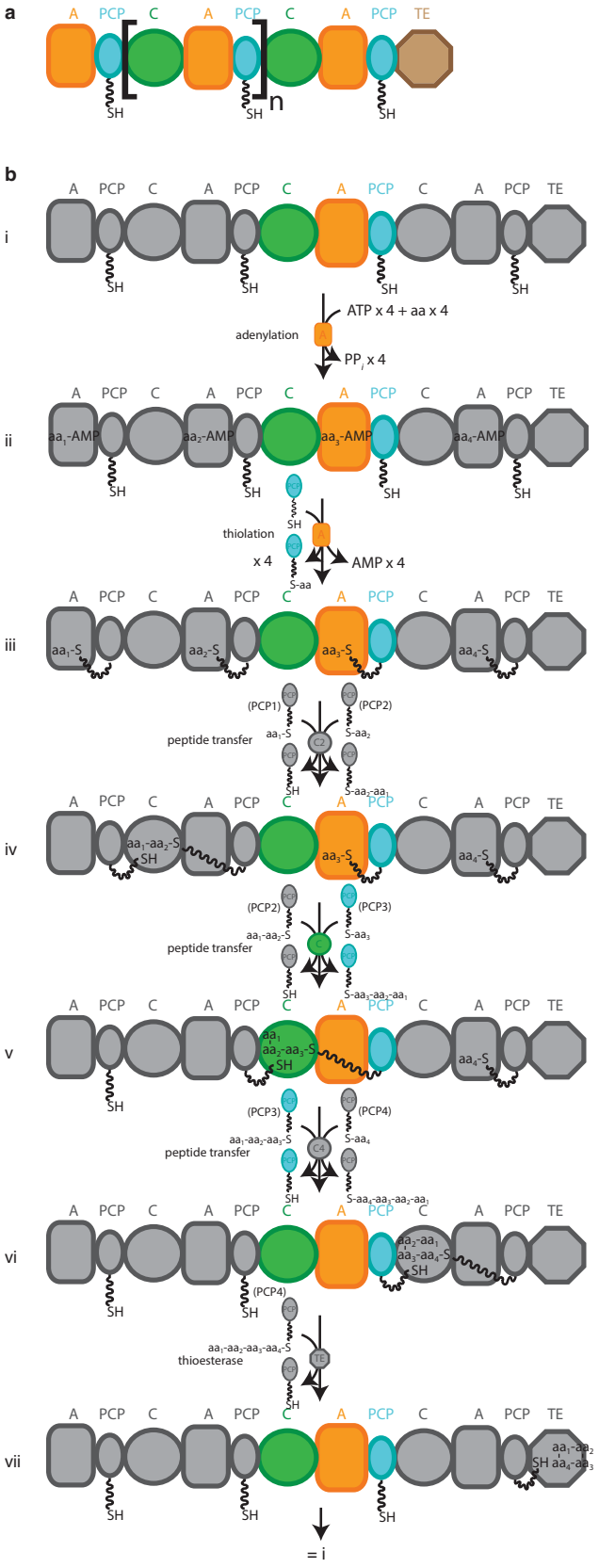
4.3. Acknowledgments

The authors thank P. Emsley and G. Murshudov for their help modelling depsipeptides, A. Wahba, J. Reimer, G. Bridon, K. Heesom (Bristol), T. Elliot for MS help, M. Tarry and N. Rogerson for editing, C. Alonso for early crystal trials, and S. Zhang help with cloning. Staff of beamlines CLS 08ID-1 (S. Labiuk, J. Gorin, M. Fodje, K. Janzen, D. Spasyuk, P. Grochulski) and APS 24-ID-C (F. Murphy) are gratefully acknowledged. This work is supported by grants to JWC from the Medical Research Council, UK (MC_U105181009 and MC_UP_A024_1008) to TMS from Canada Research Chair and NSERC (Discovery Grant 418420).

4.4. Methods

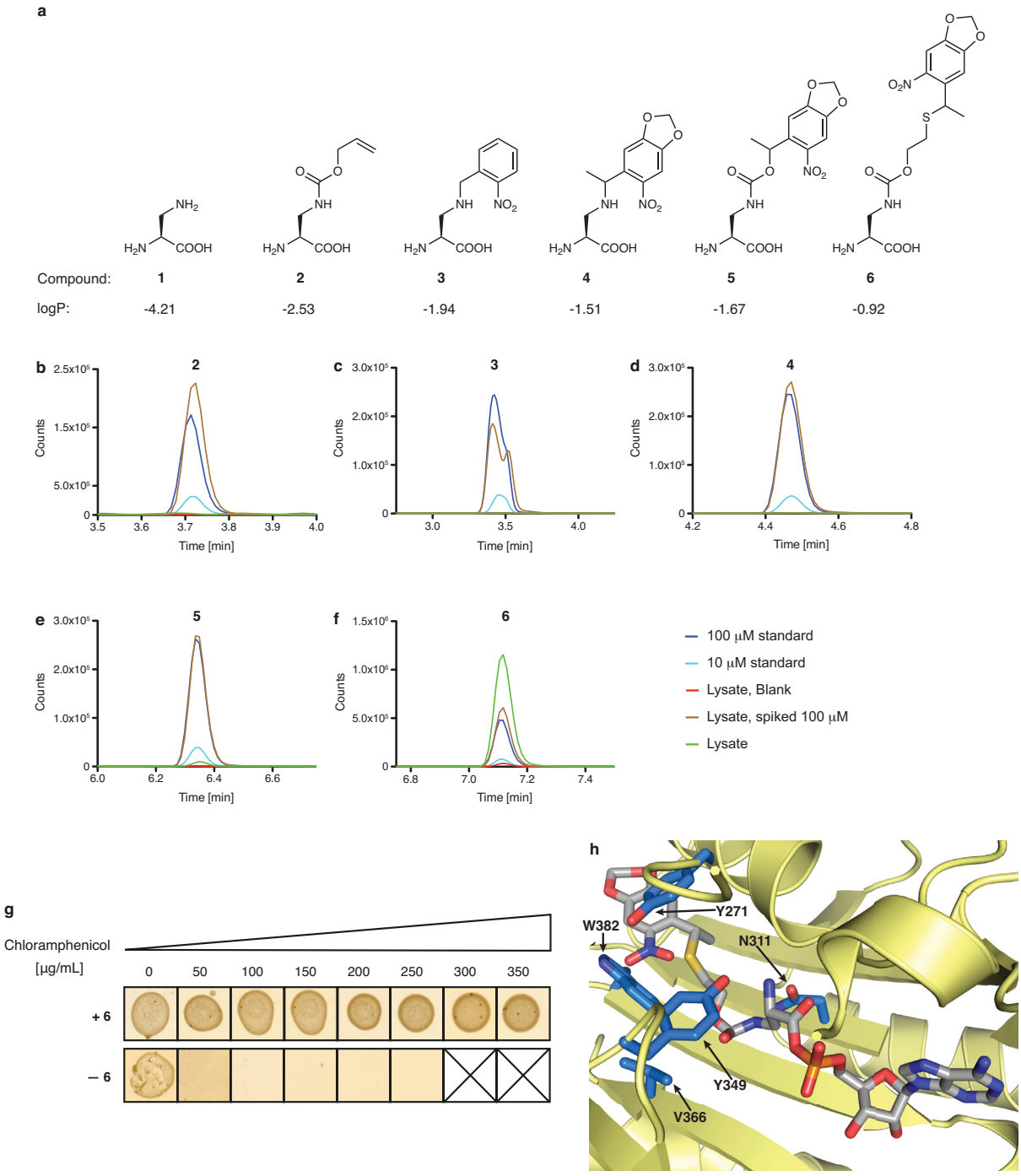
Detailed methods, including chemical syntheses, are available in the Supporting Information.

4.5. Extended Data Figures and Extended Data Tables.



Extended Data Figure 4.1. Schematic and reaction cycle of a canonical NRPS.

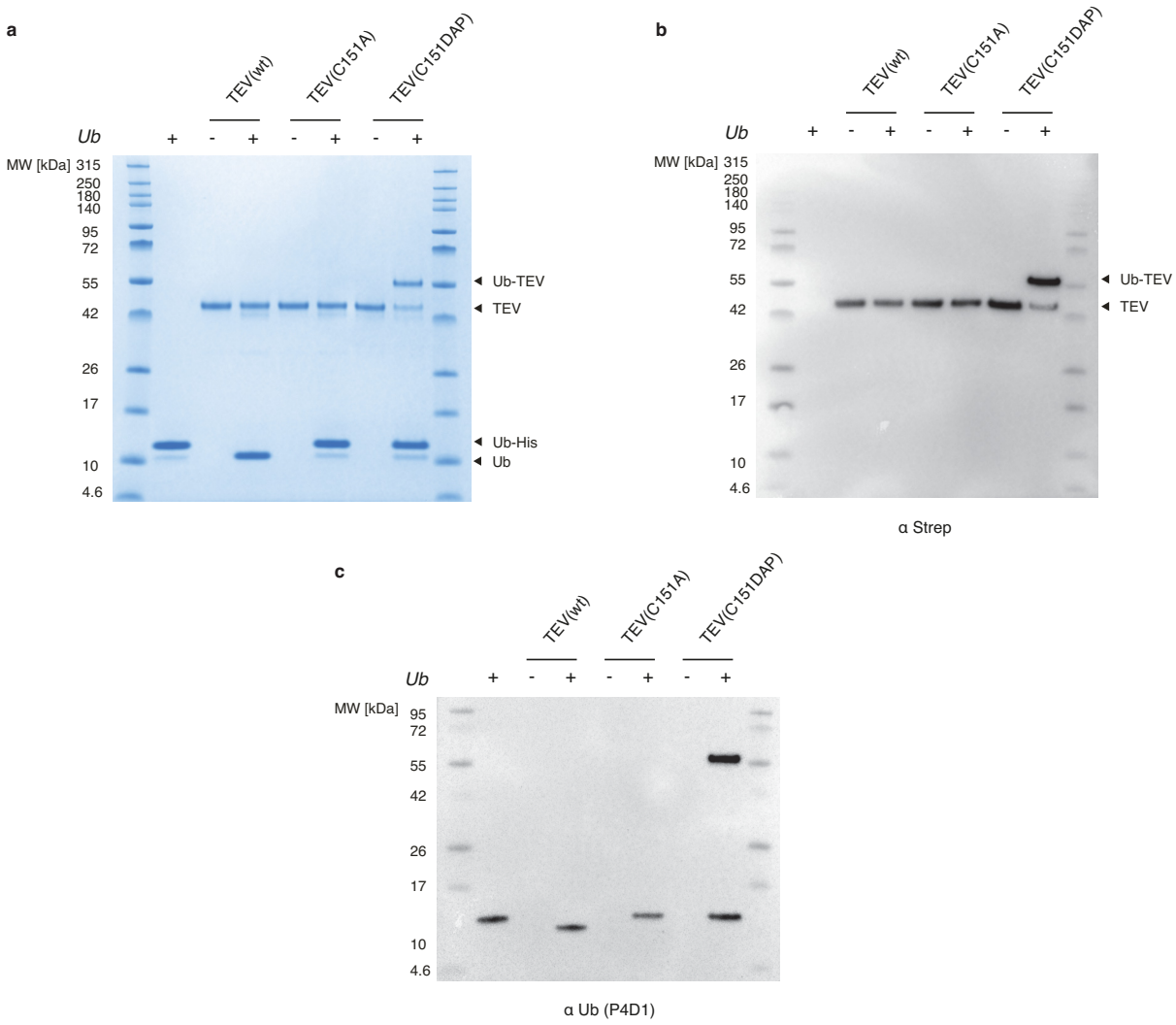
a, Schematic representation of a generic type I NRPS. **b,** Synthetic cycle of a canonical elongation module. NRPSs assemble peptides from amino acyl and other small acyl building blocks using a modular and thio-templated logic. A canonical NRPS is composed of one module for every residue in the peptide product. The initiation module contains an adenylation (A) domain, which binds cognate acyl substrate and performs adenylation and transfer of that substrate as a thioester on the phosphopantetheine arm (PPE) of a peptidyl carrier protein (PCP) domain, for transport between active sites. Each elongation module contains an A and a PCP domain, and also a condensation (C) domain, which condenses aminoacyl and peptidyl substrates bound to PCP domains, thus progressively elongating the nascent chain. Termination modules contain C, A and PCP domains, and a specialised terminating / off-loading domain responsible for the release of the peptide in its final form. The most common and most versatile terminating domain in NRPSs is the thioesterase (TE) domain. Similar TE domains terminate synthesis in polyketide and fatty acid synthases.



Extended Data Figure 4.2. Genetically directing DAP incorporation in recombinant proteins.

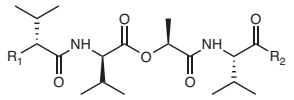
a, Structure of DAP and the protected versions investigated herein. **1**: 2,3-diaminopropionic acid (DAP). **2**: (S)-3-(((allyloxy)carbonyl)amino)-2-aminopropanoic acid. **3**: (S)-2-amino-3-((2-nitrobenzyl)amino)propanoic acid **4**: (2S)-2-amino-3-((1-(6-nitrobenzo[d][1,3]dioxol-5-yl)ethyl)amino)propanoic acid **5**: (2S)-2-amino-3-(((1-(6-nitrobenzo[d][1,3]dioxol-5-yl)ethoxy)carbonyl)amino)propanoic acid **6**: (2S)-2-amino-3-(((2-((1-(6-nitrobenzo[d][1,3]dioxol-5-

yl)ethyl)thio)ethoxy)carbonyl)amino)propanoic acid; calculated logP values are indicated (calculated using the Molinspiration molecular property calculation services at www.molinspiration.com/cgi-bin/properties). **b**-**f**, Determining the intracellular concentration of compounds **2-6** by an LC-MS assay, performed on extracts. The dark-blue trace represents a 100 μM standard for each compound. The light-blue trace represents a 10 μM standard for each compound. The red trace results from cells grown in the absence of the compound. The brown trace results from cells grown in the absence of the compound, but spiked with the compound to 100 μM. The green trace results from cells grown in the presence of 1 mM compound. The experiments were repeated in two biological replicates with similar results. **g**, Phenotyping of the DAPRS/$tRNA_{CUA}$ pair. Cells containing the DAPRS/$tRNA_{CUA}$ pair and *cat(112TAG)* were plated in the presence or absence of **6** on the indicated concentrations of chloramphenicol; the experiment was performed in two biological replicates with similar results. **h**, the side chain of **6** (grey sticks) was modelled into the active site of PylRS using a co-crystal structure of PylRS and adenylated pyrrolysine (PDB ID:2ZIM[31]). PylRS is displayed in pale yellow and amino acid positions randomised in DAPRSlib are shown in marine blue.



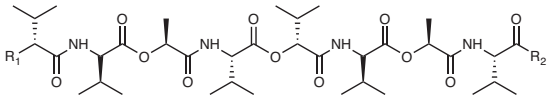
Extended Data Figure 4.3. Stably trapping the acyl-enzyme intermediate of a cysteine protease.

a, Different variants of TEV protease were reacted with Ub-tev-His. The use of TEV(wt) results in cleavage of the TEV cleavage sequence. The use of TEV(C151A) results in minimal cleavage. The presence of DAP in the active site of TEV results in the presence of an extra band in the coomassie gel, representing the isopeptide-linked TEV(C151DAP)–Ub complex. **b, c,** αStrep (**b**) and αUb (**c**) western-blots of the reactions confirm the identity of the complex. For (**a-c**), the experiment was repeated in two biological replicates with similar results. **d**, Tandem mass spectrometry following tryptic digest of the TEV(C151DAP)-Ub conjugate confirms amide bond formation at the expected position. (Top) The sequence of the branched peptide subject to fragmentation. Fragmentation of the substrate chain is predicted to lead to a series of y ions (yellow) and a series of b ions (green); the ions from this chain are labelled as 'beta'. Fragmentation of the TEV (C151DAP) derived chain is predicted to lead to a series of y ions (blue) and a series of b ions (red); the ions from this chain are labelled as 'alpha'. The MS/MS spectra with peak assignments is shown in the lower part of the panel. Ions in the alpha chain were assigned by treating DAP and the beta chain as a modification of known mass. Ions in the beta chain were manually assigned. The mass spectrometry analysis was performed once.



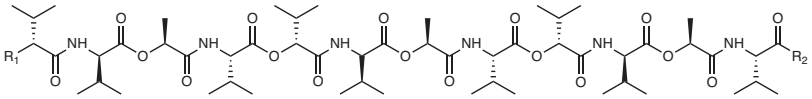
7 R_1=OH, R_2=$SCH_2CH_2NHC(O)CH_3$ **8** R_1=H, R_2=$SCH_2CH_2NHC(O)CH_3$

9 R_1=OH, R_2=OH **10** R_1=H, R_2=OH



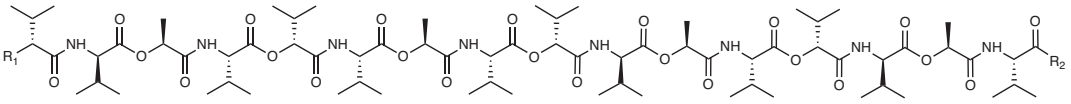
11 R_1=OH, R_2=$SCH_2CH_2NHC(O)CH_3$ **12** R_1=H, R_2=$SCH_2CH_2NHC(O)CH_3$

13 R_1=OH, R_2=OH **14**R_1=H, R_2=OH



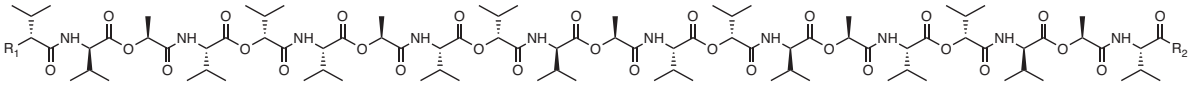
15 R_1=OH, R_2=$SCH_2CH_2NHC(O)CH_3$ **16** R_1=H, R_2=$SCH_2CH_2NHC(O)CH_3$

17 R_1=OH, R_2=OH **18** R_1=H, R_2=OH



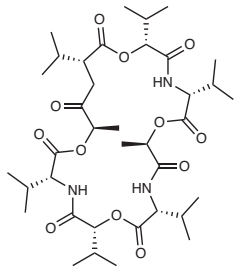
19 R_1=OH, R_2=$SCH_2CH_2NHC(O)CH_3$ **20** R_1=H, R_2=$SCH_2CH_2NHC(O)CH_3$

21R_1=OH, R_2=$SCH_2CH_2NHC(O)CH_3$ **22** R_1=H, R_2=OH

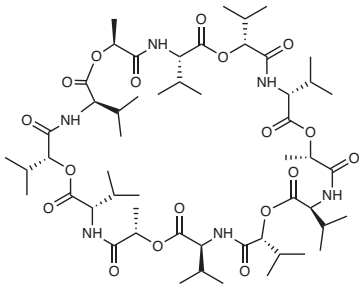


23 R_1=OH, R_2=$SCH_2CH_2NHC(O)CH_3$ **24** R_1=H, R_2=$SCH_2CH_2NHC(O)CH_3$

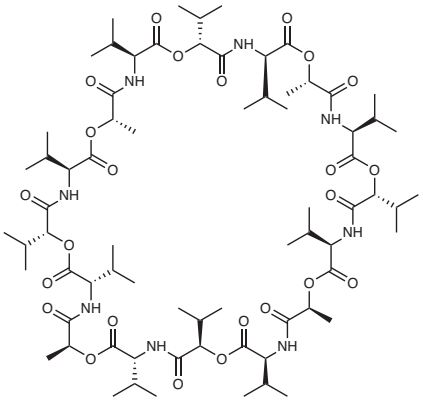
25 R_1=OH, R_2=OH **26** R_1=H, R_2=OH



montanastatin (cyclic octadepsipeptide), **27**



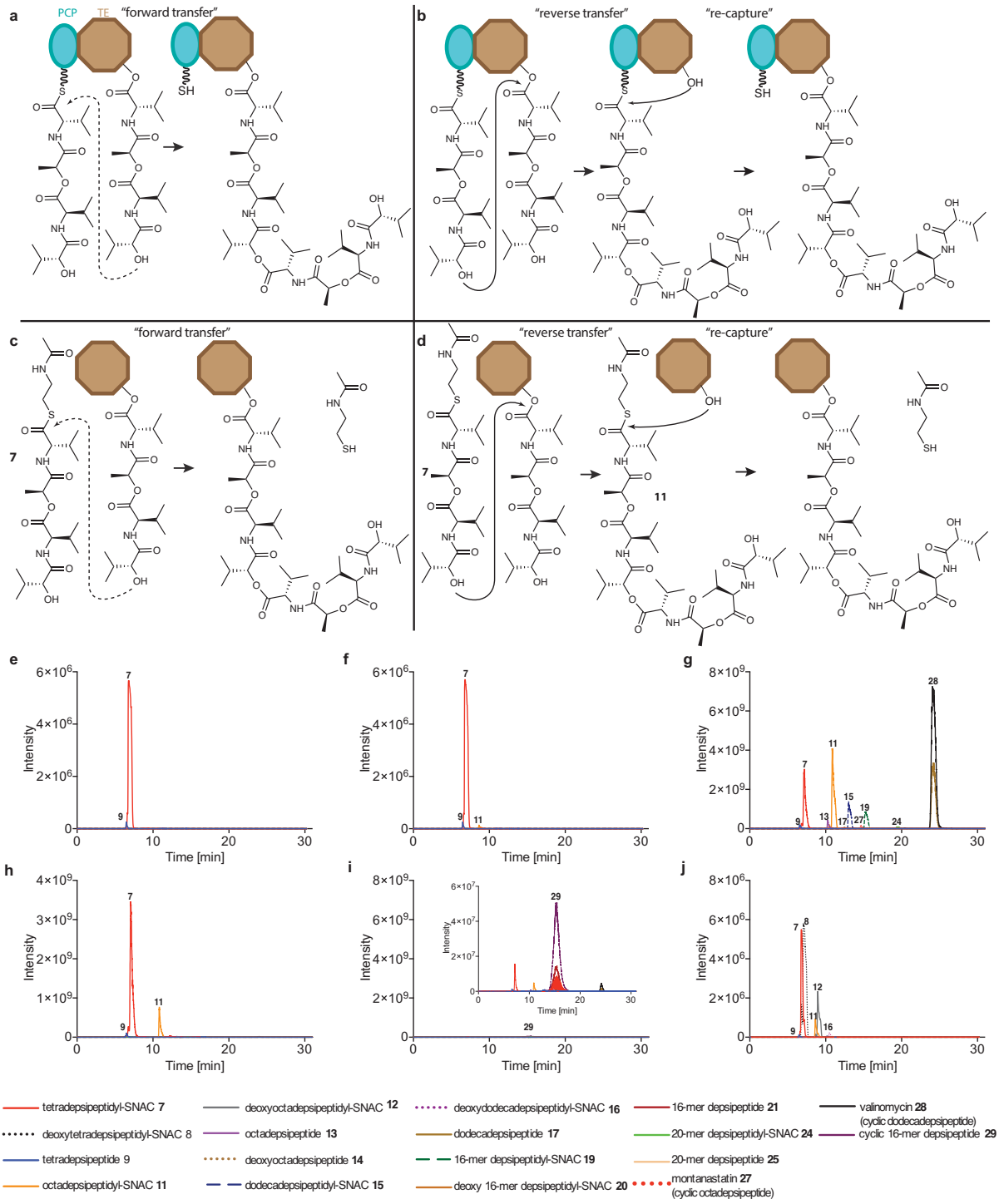
valinomycin, **28**



29

Extended Data Figure 4.4. Chemical structures of key Vlm TE substrates described in this study.

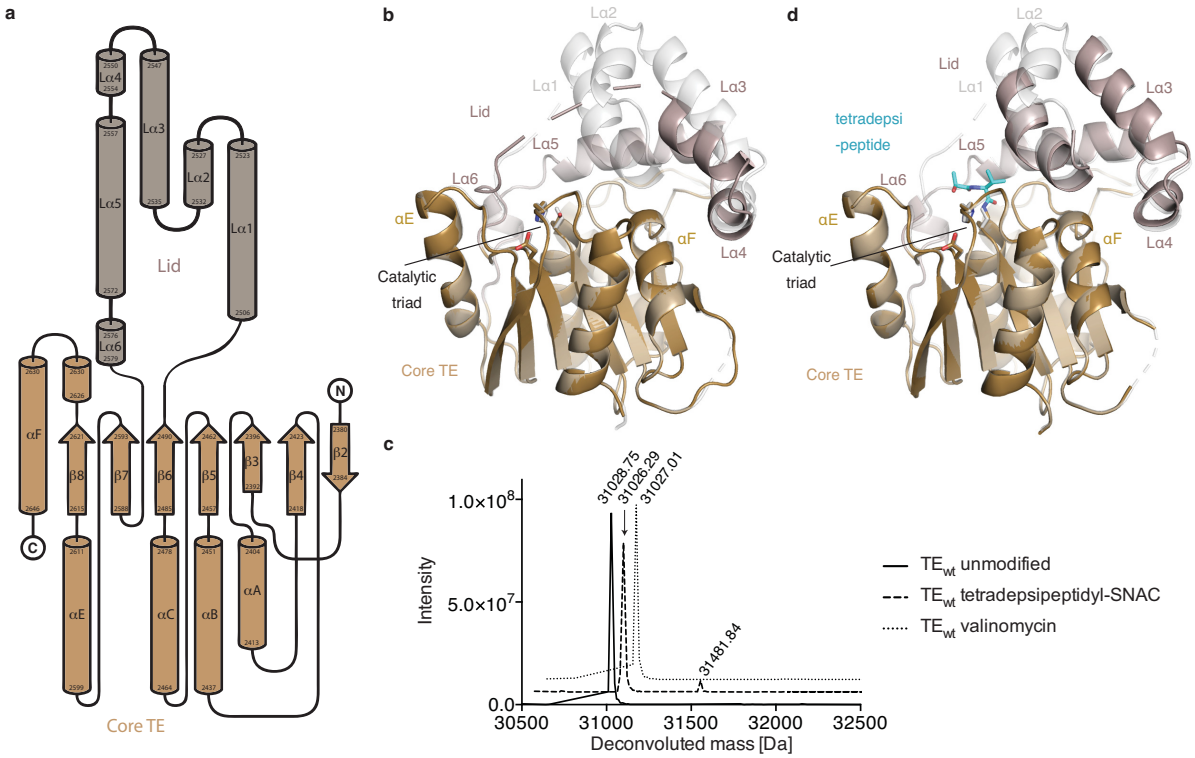
The chemical structures and the numbers used to refer to them are shown.



Extended Data Figure 4.5. The mechanism by which by Vlm TE catalyses oligomerisation.

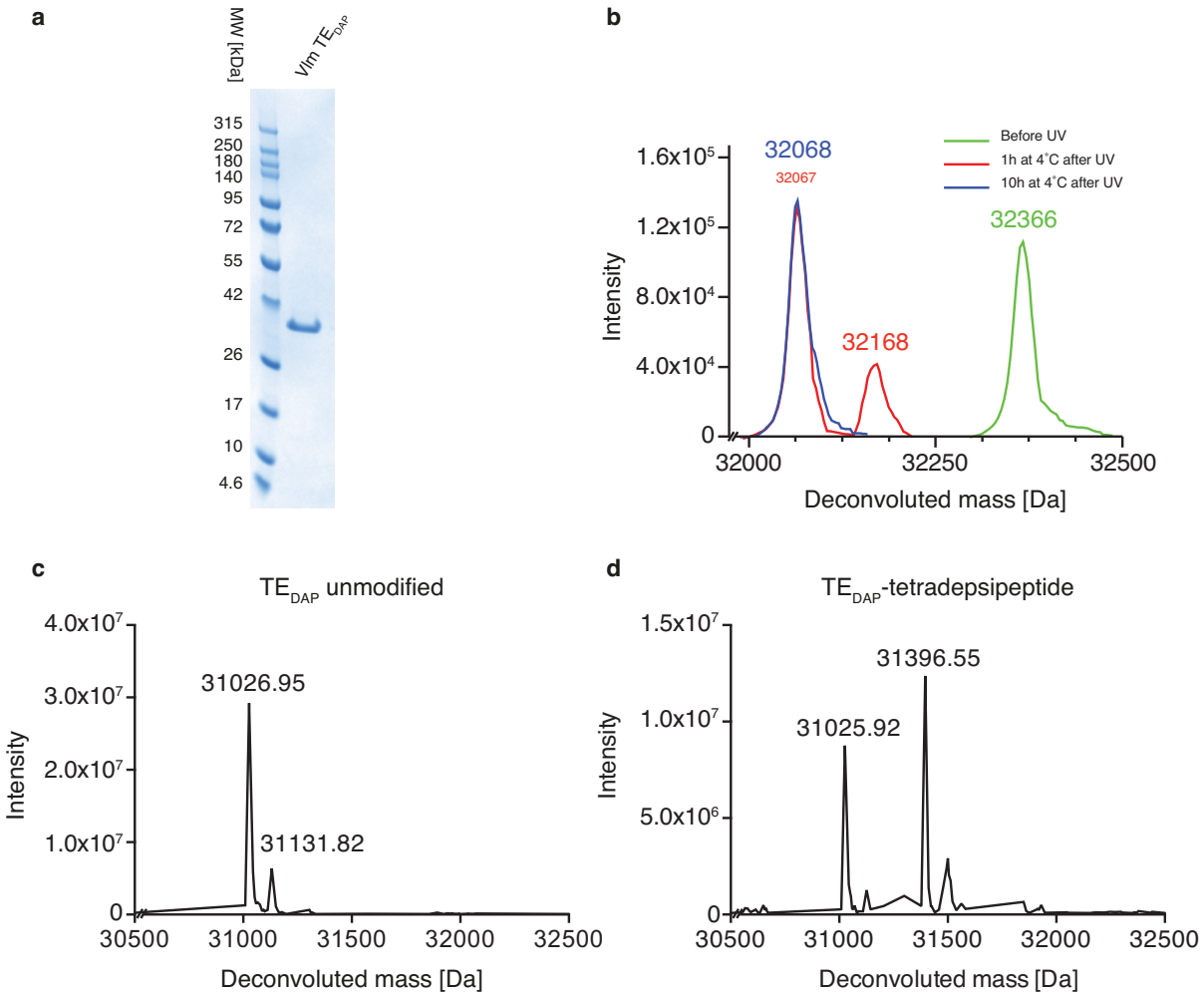
Oligomerisation could conceivably take place by two scenarios. **a**, In the first scenario the distal hydroxyl group of the tetradepsipeptidyl-*O*-TE complex attacks the thioester group in the tetradepsipeptidyl-*S*-

PCP enzyme intermediate, directly forming octadepsipeptidyl-*O*-TE as a product. **b,** In the second scenario the distal hydroxyl group of the tetradepsipeptidyl-*S*-PCP complex attacks the ester group in the tetradepsipeptidyl-*O*-TE enzyme intermediate, forming octadepsipeptidyl-*S*-PCP as a product, which would then need to be transferred onto the TE domain serine (here labelled as "recapture" by TE). **c, d:** Analogous scenarios when tetradepsipeptidyl-SNAC **7** is used as substrate instead of tetradepsipeptidyl-*S*-PCP. **e-f,** Extracted ion chromatograms (EICs) (HR-LC-ESI-MS) of a mix of tetradepsipeptidyl-SNAC **7** (1.7 mM) and buffer (**e**), or the products of a reaction between tetradepsipeptidyl-SNAC **7** (1.7 mM) and Vlm TE_{DAP} (6.5 μM) (**f**). **g-j**, EICs (LR-LC-ESI-MS) of products of reactions between tetradepsipeptidyl-SNAC **7** (1.7 mM) and Vlm TE (6.5 μM), but using a higher-volume injection into an ion trap MS instrument. **g,** The higher volume injection of the TE_{wt}-containing reaction enabled detection of a peak consistent with 20-mer depsipeptidyl- SNAC **24**. **h,** Reaction with TE_{DAP}. **i,** Small amounts of the cyclic 16-mer depsipeptide **29** elute during post-run column clean-up of experiment **g**. **j,** EICs (HR-LC-ESI-MS) of products of reactions between Vlm TE_{wt} (6.5 μM) and a mix of tetradepsipeptidyl-SNAC **7** and deoxy-tetradepsipeptidyl-SNAC **8** (1.7 mM of each). TE_{wt} produces the intermediates deoxy-octadepsipeptidyl-SNAC **12**, deoxy- dodecadepsipeptidyl-SNAC **16**, and deoxy 16-mer depsipeptidyl-SNAC **20**, which confirms the reaction pathway shown in **b.** See **Supplementary Methods for Statistics and Reproducibility** for accurate mass analysis and deviations from calculated m/z values of each compound. Experiments **e-i** were repeated independently two times with similar results. Mass spectrometry analysis of experiment **j** was performed once.



Extended Data Figure 4.6. The structures of Vlm TEwt and tetradepsipeptidyl-TEDAP and top-down LC-ESI-MS of Vlm TEwt.

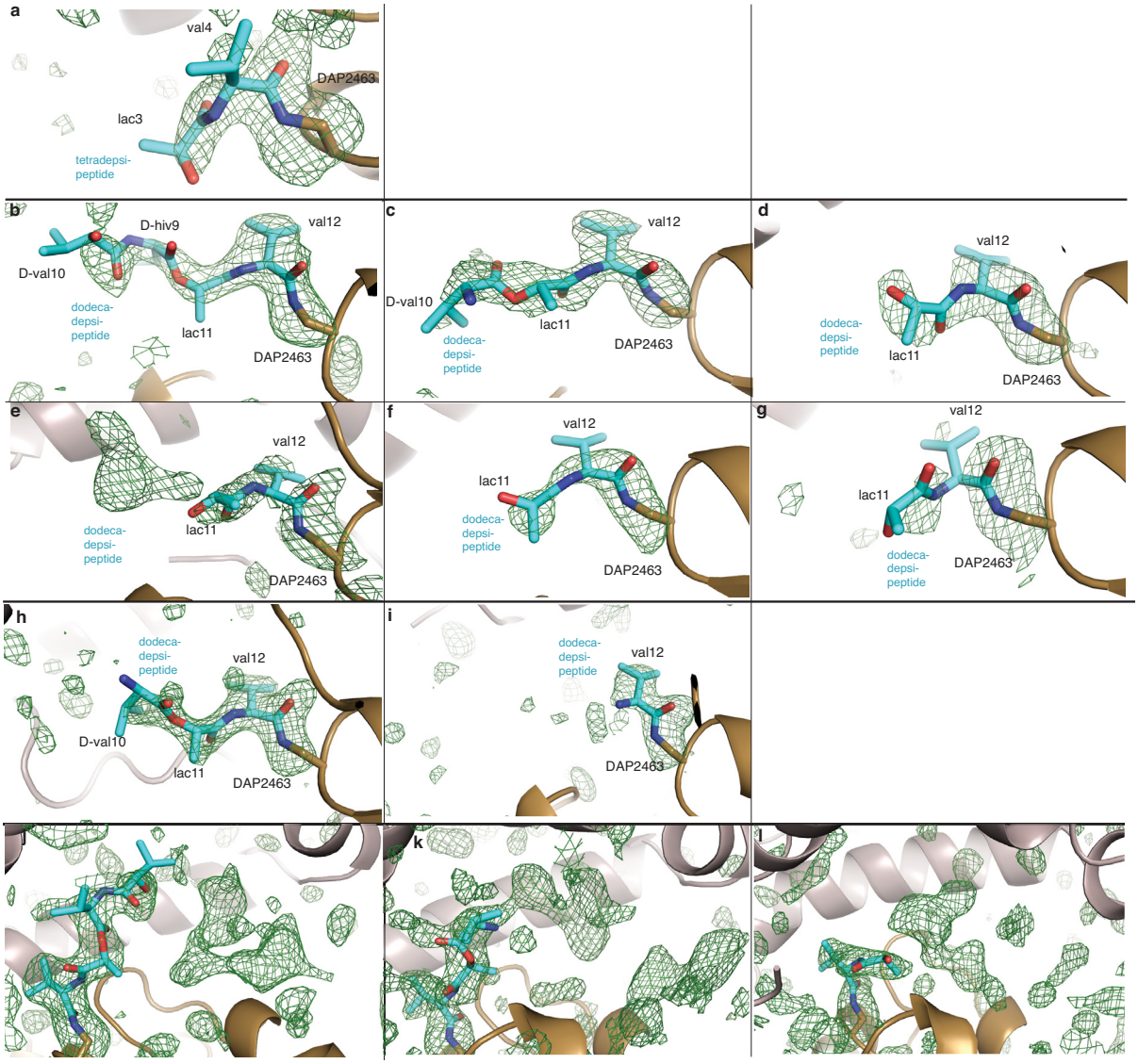
a, Secondary structure elements of Vlm TE; the naming is based on the convention for α/β hydrolase proteins. **b**, Comparison of two TE_{wt} structures (PDB ID: 6ECB and 6ECC). The active site lid of the first structure (light grey) is nearly completely ordered, while the lid of second structure (dark grey) only shows density for Lα3, Lα4 and Lα5. In the second structure, Lα3 is rotated 10°towards the active site. **c,** Deconvoluted mass spectra of TE_{wt} incubated with different substrates. Solid line, buffer control. Expected mass: 31028.22 Da; observed: 31028.75 Da. Dashed line, TE_{wt} incubated with tetradepsipeptidyl-SNAC. Expected masses: 31028.22 Da (unmodified) and 31399.44 Da (modified); observed: 31026.29 Da. Dotted line, TE_{wt} incubated with valinomycin. Expected masses: 31028.22 Da (unmodified) and 32139.86 Da (modified); observed: 31027.01 Da. Experiments were repeated independently two times with similar results. **d**, Comparison of near-identical conformations of TE_{wt} (light grey, PDB ID: 6ECB) and tetradepsipeptidyl-TE_{DAP} (tan and dark grey, PDB ID: 6ECD).



Extended Data Figure 4.7: Expression and substrate conjugation to Vlm TE containing DAP at position 2463.

a, Following expression and purification of Vlm TE$_{DAP}$, the protein was loaded on an SDS-PAGE gel and coomassie stained; the experiment was repeated in two biological replicates with similar results. **b,** The deprotection of **6** in TE$_{DAP}$-Strep was followed by ESI-MS analysis. Green trace: Purified TE$_{DAP}$-Strep containing **6** at position 2463: Expected mass: 32364.6 Da; observed: 32365.78 Da. Red trace: TE$_{DAP}$-Strep containing **6** at position 2463 following illumination to convert **6** to the intermediate: Expected mass: 32171.56 Da; observed: 32168.48 Da, and further incubation (1h, 4°C) to convert the intermediate to product (expected 32067.62, observed 32068): Blue trace: TE$_{DAP}$- Strep containing **6** at position 2463 following illumination (to convert **6** to the intermediate) and further incubation (10h, 4°C) to convert the intermediate to DAP **(1)**: Expected mass: 32067.62 Da; observed: 32067.84 Da.); the experiment was repeated in two biological replicates with similar results. **c**, Purified TE$_{DAP}$ after illumination and intermediate fragmentation: Expected mass: 31027.24 Da; observed: 31026.95 Da and 31131.82. **d,**

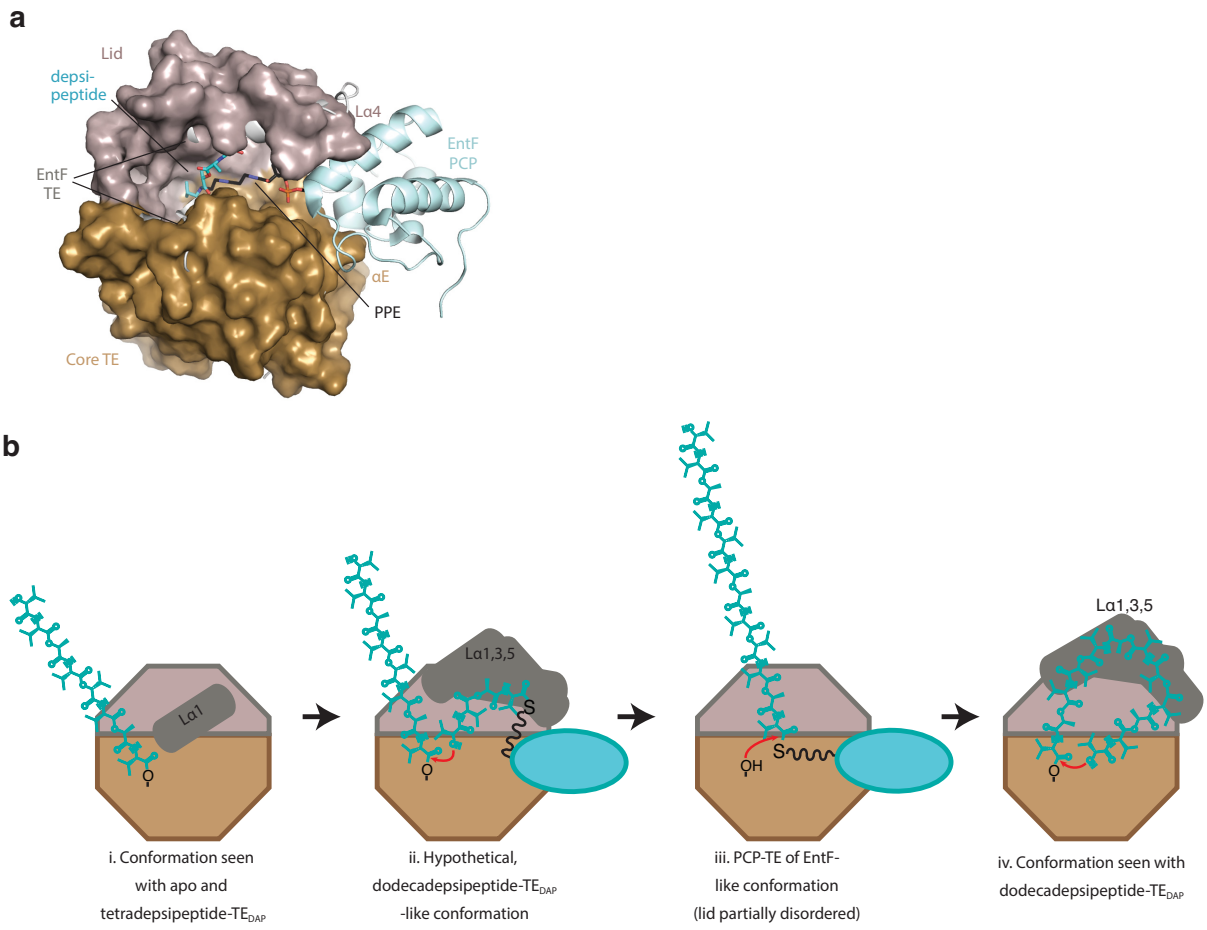
TE_{DAP} incubated with tetradepsipeptidyl-SNAC **7**: Expected masses: 31027.24 Da (unmodified) and 31398.69 (modified); observed: 31025.92 Da and 31396.55**.** Experiments (**c, d**) were repeated independently two times with similar results.



Extended Data Figure 4.8. Electron density of the active site of covalent depsipeptidyl- TE_{DAP} complexes.

Unbiased mF_o-DF_c maps (green mesh, contoured at 2.5 σ), calculated before depsipeptide residues were placed in the model. DAP (brown) and depsipeptide residues (cyan) are depicted as sticks. **a**, Tetradepsipeptidyl-TE_{DAP} (PDB ID: 6ECD). **b-g**, Dodecadepsipeptidyl-TE_{DAP} P1 space group structure (PDB ID: 6ECF), crystallographically independent molecules A to F are shown in sequential order. **h-i**: Dodecadepsipeptidyl-TE_{DAP} H3 space group (PDB ID: 6ECE), crystallographically independent molecules A

and B. **j-l**, Electron density of the active site of covalent depsipeptidyl- TE_{DAP} complexes extends beyond modelled depsipeptides. Unbiased mF_o-DF_c maps (green mesh, contoured at 2.5 σ), calculated before depsipeptide residues were placed in the model, for dodecadepsipeptidyl-TE_{DAP} P1 space group structure, crystallographically independent molecules A, B and D, in sequential order. The observed electron density that extends beyond the modelled depsipeptides (cyan sticks) could accommodate extra depsipeptide residues in different orientations. However, unambiguous modelling into this density could not be achieved.



Extended Data Figure 4.9. Modelling of PCP domain interaction with the TE domain and putative pathway.

a, Superimposition of dodecadepsipeptidyl-TE_{DAP} with the structure of EntF PCP-TE didomain (PDB ID: 3TEJ) shows the path of the PPE moiety to the active site. **b,** Hypothetical pathway for oligomerisation and cyclisation, starting from octadepsipeptidyl-TE: **i,** The position of Lα1 in observed apo/tetradepsipeptide conformation promotes an extended peptide conformation. **ii,** The tetradepsipeptidyl-PCP accepts the octadepsipeptide onto its terminal hydroxyl, perhaps using a

dodecadepsipeptide-like lid conformation which could accommodate the ~30 Å tetradepsipeptidyl-PPE bound to the PCP domain and guide it towards the active site. **iii,** The PCP domain presents the thioester for transfer back to the Ser2463. **iv,** Finally, the lid conformation observed in the dodecadepsipeptide-TE_{DAP} structures could help curl the dodecadepsipeptide back towards Ser2463 for cyclisation.

Extended Data Table 1. Data collection and refinement statistics.

	TE_{wt} structure 1 (**6ECB**)	TE_{wt} structure 2 (**6ECC**)	TE_{DAP} bound with a tetradepsipeptide (**6ECD**)	TE_{DAP} bound with dodecadepsipeptide, space group H3 (**6ECE**)	TE_{DAP} bound with dodecadepsipeptide, space group P1 (**6ECF**)
Data collection					
Space group	P 4 3 2	P 4 3 2	P 4 3 2	R 3 : H	P1
Cell dimensions					
a, *b*, *c* (Å)	151.4, 151.4, 151.4	152.2, 152.2, 152.2	152.3, 152.3, 152.3	77.6, 77.6, 235.2	77.0, 77.1,90.3
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 120	91.8, 114.9, 118.0
Resolution (Å)	151.4-1.7 (1.73-1.7)	152.2-1.8 (1.84-1.8)	107.7-1.9 (1.94-1.9)	64.59-2.0 (2.05-2.0)	78.55-2.5 (2.589-2.5)
R_{sym} or R_{merge}	0.063 (1.059)	0.08 (1.525)	0.123 (3.917)	0.079 (0.845)	0.071 (0.320)
I / sI	21.83 (1.3)	19.0 (1.1)	23.3 (1.0)	9.8 (1.5)	4.06 (1.85)
Completeness (%)	98.3 (88.3)	99.4 (94.2)	100 (100)	100 (100)	97.58 (97.10)
Redundancy	12.3 (4.5)	12.7 (6.2)	37.2 (37.4)	5.0 (4.9)	1.7 (1.7)
Refinement					
Resolution (Å)	75.7-1.7	87.87-1.8	107.7-1.9	44.24-2.0	78.55-2.5
No. reflections	64286	55853	48056	35677	53933
R_{work} / R_{free}	0.1725/0.1874	0.1757/0.1898	0.1879/0.2143	0.2091/0.2426	0.1969/0.2489
No. atoms	2188	1917	1918	3739	11761
Protein	1984	1746	1807	3645	11518
Ligand/ion	0	0	5	5	37
Water	204	171	106	89	206
B-factors	33.45	39.71	54.47	48.06	44.16
Protein	32.52	38.92	54.39	48.09	44.13
Ligand/ion	n/a	n/a	117.62	52.13	65.2
Water	42.47	47.77	52.94	46.47	41.81
R.m.s. deviations					
Bond lengths (Å)	0.009	0.019	0.018	0.005	0.005
Bond angles (°)	1.31	1.69	1.65	0.96	0.99

Each data set was collected from a single crystal.
Values in parentheses are for highest-resolution shell.

Extended Data Table 4.1. Data collection and refinement statistics for the crystal structures presented in this study.

References for Extended Data Items

Kavran, J. M. *et al.* Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 11268-11273, doi:10.1073/pnas.0704769104 (2007).

CHAPTER 5. GENERAL DISCUSSION AND FINAL STATEMENT

This dissertation provides novel structural and functional information on bacterial cyclodepsipeptide biosynthesis. Chapter 2 presents the expression, purification and characterization of the intact depsipeptide synthetase Ces, while chapter 3 analyzes the mechanisms of ketoacid adenylation and ketoreduction from a structural perspective. Chapter 4 provides structural and functional insight into the final steps of the biosynthetic pathway, the oligomerization and subsequent cyclization of the linear tetradepsipeptidyl-*S*-PCP intermediate. In this section, I discuss the implications of these findings and the open questions in the field.

In chapter 2, I showed that it is possible to express and purify the highly complex NRPS Ces system, consisting of 15 domains distributed in two di-modular NRPSs of more than 250 kDa each. I consider that one of the future endeavours in the study of Ces is to analyze the ester bond-forming C domain located at the N-terminus of CesA (CesA2-C2), as it may be expressed and purified using the conditions presented in the chapter. This domain is of special interest as its donor substrate, D-HIC-D-Ala-*S*-PCP, includes two monomers with D- stereochemistry, while its acceptor substrate is an L-hydroxy acid. Furthermore, it catalyzes ester opposed to amide bond formation. There are several questions on the function of this domain, especially on the selectivity determinants for each substrate. Ester-bond forming C domains from fumonisin (Zaleta-Rivera, Xu et al. 2006) biosynthesis or the free-standing C domain SgcC5s (Chang, Lohman et al. 2018) have been characterized, which could provide some information on CesA2-C2. Another interesting question opened by this chapter is the nature of the interaction between CesA and CesB, as they lack the communication domains commonly found on NRPSs (Hahn and Stachelhaus 2006), so they must interact through different protein-protein interfaces. The data presented in the chapter shows that they are able to form the tetradepsipeptide intermediate *in vitro*, which requires CesA-CesB to interact. A similar assay could be used for future mutational studies to elucidate the nature of the interaction.

In chapter 3, I showed that the structure of the A-KR domain is representative of the adenylation conformation of the module, with the A_{sub} domain closed into the A_{core} active site cavity, the A and KR active sites separated by 60 Å and the PCP domain absent from the electron density. The adenylation conformation is only one of the multiple conformational states that A-KR-PCP can adopt. For instance, before substrate binding, the A_{sub} domain would adopt an open conformation that could influence positioning of the KR and PCP domains in the context of a full module structure. Furthermore, the thiolation conformation would require rotation of the A_{sub} domain by several degrees, influencing the orientation of the downstream KR domain. It is also intriguing if the pseudo A_{sub} domain could play the structural role of the canonical A_{sub} domain observed in the reaction cycle of LgrA, working as an extension of the PCP domain. I expect that the thiolation and ketoreduction structures, if they are ever solved, will show massive rearrangements of both the A_{sub} and pseudo A_{sub} domains, likely acting cooperatively to drive the PCP domain through each stage of the reaction. I anticipate that synthesizing the CoA derivatives α-KIC-NH-CoA and α-HIC-NH-CoA will allow to generate holo versions of A-KR-PCP using Sfp, which could eventually lead to new crystal forms and novel structural information.

In chapter 3, I also showed that the ketone in α-ketoacids interacts with the A domain active site through a carbonyl-carbonyl antiparallel interaction, without any hydrogen bonds involved. This implies that substitution of the conserved amino acid-selective residue Asp235 for an aliphatic residue in ketoacid-selecting A-domains was likely conserved to avoid electrostatic repulsion between the ketone and the negatively charged COO^- side chain, while eliminating the charge-charge interaction with the α-NH_3^+ group in amino acids. In fungal hydroxy acid selecting-A domains Asp 235 is replaced by Gly. Extrapolating from our findings, it is very likely that the Asp 235 substitution has the same selectivity effects as in α-keto acid selecting-A domains. However, it is still unknown which favourable interaction participates in α-hydroxy acid-active site complex stabilization. The dipole in hydroxy acids is not as strong as in ketones, so a carbonyl-carbonyl interaction with the main chain amide as the one observed in our ketoacid-bound structure is unlikely to occur. Answering this question would require solving the structure of an A domain that selects hydroxy acids in the presence of substrate.

The substrate-bound TE structures in chapter 4 provide evidence that the active site lid has an important role during the reaction cycle. However, we were not able to observe all the residues of the depsipeptide due to its flexibility. We were especially interested in obtaining a structure where the distal HIV-OH was positioned for nucleophilic attack against the depsipeptidyl-TE_{DAP} amide. We anticipate that application of the DAP orthogonal system in TEs with shorter and less flexible acyl-ester intermediates may reveal the sought-after peptide cyclization conformation. These structures could also confirm that the lid in other cyclizing-TEs acts an entropic trap by restricting the number of accessible conformations of linear acyl-ester intermediates, therefore increasing the sampling of low-probability "curled" conformations and leading to cyclization. In preliminary experiments we have deleted and substituted the lid of Vlm TE by a small Gly-Gly linker and found that all catalytic activity was abolished. Experiments are ongoing, but this is a second piece of evidence that the lid plays a crucial role during the TE reaction. I also expect that a mutational analysis of the lid could indicate if there are crucial residues participating in the formation of the cavity, very likely revealing that hydrophobic residues are essential for activity.

Studying the Ces, StrA and Vlm mechanisms at an atomic level provided several details on these interesting systems that make compounds of medical, industrial and scientific interest, opening several new questions and indicating that they are amenable to structural analysis. Regarding chapter 4, I expect that the proof of concept of the DAP technology in Vlm TE will encourage its application on proteases from other biological systems, enabling structural studies of substrate-enzyme complexes that have been impossible to isolate with other methodologies.

There are increasing efforts to exploit the modularity of NRPSs to produce compounds with novel or improved activity, thoroughly summarized in (Winn, Fyans et al. 2016). I consider that the work presented in this dissertation provides valuable insights into a system that adds complex chemical features that might be challenging to add by synthetic chemistry approaches, such as ester bonds of different chirality and macrocycles. I hypothesize that in addition of using NRPSs to produce modified or novel products, WT enzymes will be increasingly used in heterologous hosts to produce already useful compounds, avoiding challenging synthetic methods.

References

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, N. Echols, J. J. Headd, L. W. Hung, S. Jain, G. J. Kapral, R. W. Grosse Kunstleve, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2011). "The Phenix software for automated determination of macromolecular structures." Methods **55**(1): 94-106.

Afonine, P. V., R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart and P. D. Adams (2012). "Towards automated crystallographic structure refinement with phenix.refine." Acta Crystallogr D Biol Crystallogr **68**(Pt 4): 352-367.

Agata, N., M. Mori, M. Ohta, S. Suwan, I. Ohtani and M. Isobe (1994). "A novel dodecadepsipeptide, cereulide, isolated from Bacillus cereus causes vacuole formation in HEp-2 cells." FEMS Microbiol Lett **121**(1): 31-34.

Agata, N., M. Ohta, M. Mori and M. Isobe (1995). "A Novel Dodecadepsipeptide, Cereulide, Is an Emetic Toxin of Bacillus-Cereus." Fems Microbiology Letters **129**(1): 17-19.

Akey, D. L., J. D. Kittendorf, J. W. Giraldes, R. A. Fecik, D. H. Sherman and J. L. Smith (2006). "Structural basis for macrolactonization by the pikromycin thioesterase." Nat Chem Biol **2**(10): 537-542.

Al-Mestarihi, A. H., G. Villamizar, J. Fernández, O. E. Zolova, F. Lombó and S. Garneau-Tsodikova (2014). "Adenylation and S-Methylation of Cysteine by the Bifunctional Enzyme TioN in Thiocoraline Biosynthesis." Journal of the American Chemical Society **136**(49): 17350-17354.

Allen, F. H., C. A. Baalham, J. P. M. Lommerse and P. R. Raithby (1998). "Carbonyl-carbonyl interactions can be competitive with hydrogen bonds." Acta Crystallographica Section B-Structural Science **54**: 320-329.

Alonzo, D. A., N. A. Magarvey and T. M. Schmeing (2015). "Characterization of cereulide synthetase, a toxin-producing macromolecular machine." PLoS One **10**(6): e0128569.

Amorim Franco, T. M. and J. S. Blanchard (2017). "Bacterial Branched-Chain Amino Acid Biosynthesis: Structures, Mechanisms, and Drugability." Biochemistry **56**(44): 5849-5865.

Argyropoulos, P., F. Bergeret, C. Pardin, J. M. Reimer, A. Pinto, C. N. Boddy and T. M. Schmeing (2016). "Towards a characterization of the structural determinants of specificity in the macrocyclizing thioesterase for deoxyerythronolide B biosynthesis." Biochim Biophys Acta **1860**(3): 486-497.

Bachmann, B. O. and J. Ravel (2009). "Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data." Methods Enzymol **458**: 181-217.

Baker, A. S. and A. Deiters (2014). "Optical control of protein function through unnatural amino acid mutagenesis and other optogenetic approaches." ACS Chem Biol **9**(7): 1398-1407.

Ballard, C. E., H. Yu and B. Wang (2002). "Recent developments in depsipeptide research." Curr Med Chem **9**(4): 471-498.

Baltz, R. H. (2008). "Biosynthesis and genetic engineering of lipopeptide antibiotics related to daptomycin." Curr Top Med Chem **8**(8): 618-638.

Battye, T. G. G., L. Kontogiannis, O. Johnson, H. R. Powell and A. G. W. Leslie (2011). "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM." Acta Crystallogr D Biol Crystallogr **67**(4): 271-281.

Bennett, B. D., E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien and J. D. Rabinowitz (2009). "Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli." Nat Chem Biol **5**(8): 593-599.

Biasini, M., S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli and T. Schwede (2014). "SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information." Nucleic Acids Res **42**(Web Server issue): W252-258.

Bloudoff, K., D. A. Alonzo and T. M. Schmeing (2016). "Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases." Cell Chem Biol **23**(3): 331-339.

Bloudoff, K., C. D. Fage, M. A. Marahiel and T. M. Schmeing (2017). "Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis." Proc Natl Acad Sci U S A **114**(1): 95-100.

Bloudoff, K., D. Rodionov and T. M. Schmeing (2013). "Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases." J Mol Biol **425**(17): 3137-3150.

Bloudoff, K. and T. M. Schmeing (2017). "Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity." Biochim Biophys Acta **1865**(11 Pt B): 1587-1604.

Boll, B., T. Taubitz and L. Heide (2011). "Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis." J Biol Chem **286**(42): 36281-36290.

Brakhage, A. A. (2013). "Regulation of fungal secondary metabolism." Nat Rev Microbiol **11**(1): 21-32.

Broberg, A., A. Menkis and R. Vasiliauskas (2006). "Kutznerides 1-4, depsipeptides from the actinomycete Kutzneria sp. 744 inhabiting mycorrhizal roots of Picea abies seedlings." J Nat Prod **69**(1): 97-102.

Bruner, S. D., T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh and M. T. Stubbs (2002). "Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE." Structure **10**(3): 301-310.

Caboche, S., V. Leclere, M. Pupin, G. Kucherov and P. Jacques (2010). "Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity." J Bacteriol **192**(19): 5143-5150.

Caffrey, P. (2003). "Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases." Chembiochem **4**(7): 654-657.

Cane, D. E., C. T. Walsh and C. Khosla (1998). "Biochemistry - Harnessing the biosynthetic code: Combinations, permutations, and mutations." Science **282**(5386): 63-68.

Cappadocia, L. and C. D. Lima (2018). "Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism." Chem Rev **118**(3): 889-918.

Challis, G. L. and J. H. Naismith (2004). "Structural aspects of non-ribosomal peptide biosynthesis." Curr Opin Struct Biol **14**(6): 748-756.

Chalut, C., L. Botella, C. de Sousa-D'Auria, C. Houssin and C. Guilhot (2006). "The nonredundant roles of two 4'-phosphopantetheinyl transferases in vital processes of Mycobacteria." Proc Natl Acad Sci U S A **103**(22): 8511-8516.

Chang, C. Y., J. R. Lohman, T. Huang, K. Michalska, L. Bigelow, J. D. Rudolf, R. Jedrzejczak, X. Yan, M. Ma, G. Babnigg, A. Joachimiak, G. N. Phillips, Jr. and B. Shen (2018). "Structural Insights into the Free-Standing Condensation Enzyme SgcC5 Catalyzing Ester-Bond Formation in the Biosynthesis of the Enediyne Antitumor Antibiotic C-1027." Biochemistry **57**(23): 3278-3288.

Chen, H. and C. T. Walsh (2001). "Coumarin formation in novobiocin biosynthesis: beta-hydroxylation of the aminoacyl enzyme tyrosyl-S-NovH by a cytochrome P450 NovI." Chem Biol **8**(4): 301-312.

Chen, W.-H., K. Li, N. S. Guntaka and S. D. Bruner (2016). "Interdomain and Intermodule Organization in Epimerization Domain Containing Nonribosomal Peptide Synthetases." ACS chemical biology **11**(8): 2293-2303.

Cheng, Y. Q. (2006). "Deciphering the biosynthetic codes for the potent anti-SARS-CoV cyclodepsipeptide valinomycin in Streptomyces tsusimaensis ATCC 15141." Chembiochem **7**(3): 471-477.

Chin, J. W. (2014). "Expanding and reprogramming the genetic code of cells and animals." Annu Rev Biochem **83**: 379-408.

Cole, F. X. and P. R. Schimmel (1970). "On Rate Law and Mechanism of Adenosine Triphosphate-Pyrosphosphate Isotope Exchange Reaction of Amino Acyl Transfer Ribonucleic Acid Synthetases." Biochemistry **9**(3): 480-489.

Condurso, H. L. and S. D. Bruner (2012). "Structure and noncanonical chemistry of nonribosomal peptide biosynthetic machinery." Nat Prod Rep **29**(10): 1099-1110.

Conti, E., N. P. Franks and P. Brick (1996). "Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes." Structure **4**(3): 287-298.

Conti, E., T. Stachelhaus, M. A. Marahiel and P. Brick (1997). "Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S." EMBO J **16**(14): 4174-4183.

D'Arcy, A., T. Bergfors, S. W. Cowan-Jacob and M. Marsh (2014). "Microseed matrix screening for optimization in protein crystallization: what have we learned?" Acta Crystallogr F Struct Biol Commun **70**(Pt 9): 1117-1126.

Daniels, M. J. (1968). "Studies of the biosynthesis of polymyxin B." Biochim Biophys Acta **156**(1): 119-127.

Deane, C. M., F. H. Allen, R. Taylor and T. L. Blundell (1999). "Carbonyl-carbonyl interactions stabilize the partially allowed Ramachandran conformations of asparagine and aspartic acid." Protein Eng **12**(12): 1025-1028.

Demain, A. L. and A. Fang (2000). "The natural functions of secondary metabolites." Adv Biochem Eng Biotechnol **69**: 1-39.

Derse, P. H. and F. M. Strong (1963). "Toxicity of Antimycin to Fish." Nature **200**: 600-601.

Dierick, K., E. Van Coillie, I. Swiecicka, G. Meyfroidt, H. Devlieger, A. Meulemans, G. Hoedemaekers, L. Fourie, M. Heyndrickx and J. Mahillon (2005). "Fatal family outbreak of Bacillus cereus-associated food poisoning." J Clin Microbiol **43**(8): 4277-4279.

Ding, Y., C. M. Rath, K. L. Bolduc, K. Hakansson and D. H. Sherman (2011). "Chemoenzymatic synthesis of cryptophycin anticancer agents by an ester bond-forming non-ribosomal peptide synthetase module." J Am Chem Soc **133**(37): 14492-14495.

Drake, E. J., J. Cao, J. Qu, M. B. Shah, R. M. Straubinger and A. M. Gulick (2007). "The 1.8 A crystal structure of PA2412, an MbtH-like protein from the pyoverdine cluster of Pseudomonas aeruginosa." J Biol Chem **282**(28): 20425-20434.

Drake, E. J., B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. L. Allen, G. Skiniotis, C. C. Aldrich and A. M. Gulick (2016). "Structures of two distinct conformations of holo-non-ribosomal peptide synthetases." Nature **529**(7585): 235-238.

Du, L. and L. Lou (2010). "PKS and NRPS release mechanisms." Nat Prod Rep **27**(2): 255-278.

Du, Y., Y. Wang, T. Huang, M. Tao, Z. Deng and S. Lin (2014). "Identification and characterization of the biosynthetic gene cluster of polyoxypeptin A, a potent apoptosis inducer." BMC Microbiol **14**: 30.

Dugan, R. E. and J. W. Porter (1971). "Mechanism of Binding of Reduced Nicotinamide Adenine Dinucleotide Phosphate to Vertebrate Fatty Acid Synthetases - Sites and Type of Bonding,

Conformation of Coenzyme, and Binding of Analogues." Journal of Biological Chemistry **246**(3): 637-&.

Ehling-Schulz, M., M. Fricker, H. Grallert, P. Rieck, M. Wagner and S. Scherer (2006). "Cereulide synthetase gene cluster from emetic Bacillus cereus: structure and location on a mega virulence plasmid related to Bacillus anthracis toxin plasmid pXO1." BMC Microbiol **6**: 20.

Ehling-Schulz, M., N. Vukov, A. Schulz, R. Shaheen, M. Andersson, E. Martlbauer and S. Scherer (2005). "Identification and partial characterization of the nonribosomal peptide synthetase gene responsible for cereulide production in emetic Bacillus cereus." Appl Environ Microbiol **71**(1): 105-113.

Ekici, O. D., M. Paetzel and R. E. Dalbey (2008). "Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration." Protein Sci **17**(12): 2023-2037.

Ekman, J. V., A. Kruglov, M. A. Andersson, R. Mikkola, M. Raulio and M. Salkinoja-Salonen (2012). "Cereulide produced by Bacillus cereus increases the fitness of the producer organism in low-potassium environments." Microbiology **158**(Pt 4): 1106-1116.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

Felnagle, E. A., J. J. Barkei, H. Park, A. M. Podevels, M. D. McMahon, D. W. Drott and M. G. Thomas (2010). "MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases." Biochemistry **49**(41): 8815-8817.

Felnagle, E. A., E. E. Jackson, Y. A. Chan, A. M. Podevels, A. D. Berti, M. D. McMahon and M. G. Thomas (2008). "Nonribosomal peptide synthetases involved in the production of medically relevant natural products." Mol Pharm **5**(2): 191-211.

Frueh, D. P., H. Arthanari, A. Koglin, D. A. Vosburg, A. E. Bennett, C. T. Walsh and G. Wagner (2008). "Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase." Nature **454**(7206): 903-906.

Fujimori, D. G., S. Hrvatin, C. S. Neumann, M. Strieker, M. A. Marahiel and C. T. Walsh (2007). "Cloning and characterization of the biosynthetic gene cluster for kutznerides." Proc Natl Acad Sci U S A **104**(42): 16498-16503.

Gaitatzis, N., B. Kunze and R. Muller (2001). "In vitro reconstitution of the myxochelin biosynthetic machinery of Stigmatella aurantiaca Sg a15: Biochemical characterization of a reductive release mechanism from nonribosomal peptide synthetases." Proc Natl Acad Sci U S A **98**(20): 11136-11141.

Gevers, W., H. Kleinkauf and F. Lipmann (1968). "The activation of amino acids for biosynthesis of gramicidin S." Proc Natl Acad Sci U S A **60**(1): 269-276.

Giege, R. (2013). "A historical perspective on protein crystallization from 1840 to the present day." FEBS J **280**(24): 6456-6497.

Gonzalez, D. J., N. M. Haste, A. Hollands, T. C. Fleming, M. Hamby, K. Pogliano, V. Nizet and P. C. Dorrestein (2011). "Microbial competition between Bacillus subtilis and Staphylococcus aureus monitored by imaging mass spectrometry." Microbiology **157**(Pt 9): 2485-2492.

Grunewald, J., S. A. Sieber, C. Mahlert, U. Linne and M. A. Marahiel (2004). "Synthesis and derivatization of daptomycin: a chemoenzymatic route to acidic lipopeptide antibiotics." J Am Chem Soc **126**(51): 17025-17031.

Gulick, A. M. (2009). "Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase." ACS Chem Biol **4**(10): 811-827.

Gulick, A. M. (2016). "Structural insight into the necessary conformational changes of modular nonribosomal peptide synthetases." Curr Opin Chem Biol **35**: 89-96.

Gulick, A. M., V. J. Starai, A. R. Horswill, K. M. Homick and J. C. Escalante-Semerena (2003). "The 1.75 A crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A." Biochemistry **42**(10): 2866-2873.

Gunsior, M., S. D. Breazeale, A. J. Lind, J. Ravel, J. W. Janc and C. A. Townsend (2004). "The biosynthetic gene cluster for a monocyclic beta-lactam antibiotic, nocardicin A." Chem Biol **11**(7): 927-938.

Guntaka, N. S., A. R. Healy, J. M. Crawford, S. B. Herzon and S. D. Bruner (2017). "Structure and Functional Analysis of ClbQ, an Unusual Intermediate-Releasing Thioesterase from the Colibactin Biosynthetic Pathway." ACS Chem Biol **12**(10): 2598-2608.

Hahn, M. and T. Stachelhaus (2006). "Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides." Proc Natl Acad Sci U S A **103**(2): 275-280.

Hamoen, L. W., H. Eshuis, J. Jongbloed, G. Venema and D. van Sinderen (1995). "A small gene, designated comS, located within the coding region of the fourth amino acid-activation domain of srfA, is required for competence development in Bacillus subtilis." Mol Microbiol **15**(1): 55-63.

Hedstrom, L. (2002). "Serine protease mechanism and specificity." Chem Rev **102**(12): 4501-4524.

Holliday, G. L., J. B. Mitchell and J. M. Thornton (2009). "Understanding the functional roles of amino acid residues in enzyme catalysis." J Mol Biol **390**(3): 560-577.

Hong, J. and H. Luesch (2012). "Largazole: from discovery to broad-spectrum therapy." Nat Prod Rep **29**(4): 449-456.

Horsman, M. E., T. P. A. Hari and C. N. Boddy (2015). "Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: Logic gate or a victim of fate?" Nat Prod Rep.

Hou, J., L. Robbel and M. A. Marahiel (2011). "Identification and characterization of the lysobactin biosynthetic gene cluster reveals mechanistic insights into an unusual termination module architecture." Chem Biol **18**(5): 655-664.

Hoyer, K. M., C. Mahlert and M. A. Marahiel (2007). "The iterative gramicidin s thioesterase catalyzes peptide ligation and cyclization." Chem Biol **14**(1): 13-22.

Huguenin-Dezot, N., D. A. Alonzo, G. W. Heberlig, M. Mahesh, D. P. Nguyen, M. H. Dornan, C. N. Boddy, T. M. Schmeing and J. W. Chin (2019). "Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid." Nature **565**(7737): 112-117.

Jaitzig, J., J. Li, R. D. Sussmuth and P. Neubauer (2014). "Reconstituted biosynthesis of the nonribosomal macrolactone antibiotic valinomycin in Escherichia coli." ACS Synth Biol **3**(7): 432-438.

Jirakkakul, J., J. Punya, S. Pongpattanakitshote, P. Paungmoung, N. Vorapreeda, A. Tachaleat, C. Klomnara, M. Tanticharoen and S. Cheevadhanarak (2008). "Identification of the nonribosomal peptide synthetase gene responsible for bassianolide synthesis in wood-decaying fungus Xylaria sp. BCC1067." Microbiology **154**(Pt 4): 995-1006.

Keating, T. A., C. G. Marshall and C. T. Walsh (2000). "Reconstitution and characterization of the Vibrio cholerae vibriobactin synthetase from VibB, VibE, VibF, and VibH." Biochemistry **39**(50): 15522-15530.

Keating, T. A., C. G. Marshall, C. T. Walsh and A. E. Keating (2002). "The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains." Nat Struct Biol **9**(7): 522-526.

Keatinge-Clay, A. T. (2007). "A tylosin ketoreductase reveals how chirality is determined in polyketides." Chem Biol **14**(8): 898-908.

Keatinge-Clay, A. T. and R. M. Stroud (2006). "The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases." Structure **14**(4): 737-748.

Kjaerulff, L., A. Nielsen, M. Mansson, L. Gram, T. O. Larsen, H. Ingmer and C. H. Gotfredsen (2013). "Identification of four new agr quorum sensing-interfering cyclodepsipeptides from a marine Photobacterium." Mar Drugs **11**(12): 5051-5062.

Koglin, A., F. Lohr, F. Bernhard, V. V. Rogov, D. P. Frueh, E. R. Strieter, M. R. Mofid, P. Guntert, G. Wagner, C. T. Walsh, M. A. Marahiel and V. Dotsch (2008). "Structural basis for the selectivity of the external thioesterase of the surfactin synthetase." Nature **454**(7206): 907-911.

Kohlhaw, G. B. (2003). "Leucine biosynthesis in fungi: entering metabolism through the back door." Microbiol Mol Biol Rev **67**(1): 1-15, table of contents.

Konz, D., A. Klens, K. Schorgendorfer and M. A. Marahiel (1997). "The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecular characterization of three multi-modular peptide synthetases." Chem Biol **4**(12): 927-937.

Konz, D. and M. A. Marahiel (1999). "How do peptide synthetases generate structural diversity?" Chem Biol **6**(2): R39-48.

Kopp, F., J. Grunewald, C. Mahlert and M. A. Marahiel (2006). "Chemoenzymatic design of acidic lipopeptide hybrids: new insights into the structure-activity relationship of daptomycin and A54145." Biochemistry **45**(35): 10474-10481.

Kotiranta, A., K. Lounatmaa and M. Haapasalo (2000). "Epidemiology and pathogenesis of Bacillus cereus infections." Microbes Infect **2**(2): 189-198.

Kraas, F. I., V. Helmetag, M. Wittmann, M. Strieker and M. A. Marahiel (2010). "Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation." Chem Biol **17**(8): 872-880.

Labby, K. J., S. G. Watsula and S. Garneau-Tsodikova (2015). "Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis." Nat Prod Rep **32**(5): 641-653.

Lan, Y., B. Langlet-Bertin, V. Abbate, L. S. Vermeer, X. Kong, K. E. Sullivan, C. Leborgne, D. Scherman, R. C. Hider, A. F. Drake, S. S. Bansal, A. Kichler and A. J. Mason (2010). "Incorporation of 2,3-diaminopropionic acid into linear cationic amphipathic peptides produces pH-sensitive vectors." Chembiochem **11**(9): 1266-1272.

Lawen, A. (1996). "Biosynthesis and mechanism of action of cyclosporins." Prog Med Chem **33**: 53-97.

Lee, C., H. Gorisch, H. Kleinkauf and R. Zocher (1992). "A highly specific D-hydroxyisovalerate dehydrogenase from the enniatin producer Fusarium sambucinum." J Biol Chem **267**(17): 11741-11744.

Li, H., T. Tanikawa, Y. Sato, Y. Nakagawa and T. Matsuyama (2005). "Serratia marcescens gene required for surfactant serrawettin W1 production encodes putative aminolipid synthetase belonging to nonribosomal peptide synthetase family." Microbiol Immunol **49**(4): 303-310.

Li, J., J. Jaitzig, F. Hillig, R. Sussmuth and P. Neubauer (2014). "Enhanced production of the nonribosomal peptide antibiotic valinomycin in Escherichia coli through small-scale high cell density fed-batch cultivation." Appl Microbiol Biotechnol **98**(2): 591-601.

Li, J., J. Yu, J. Zhao, J. Wang, S. Zheng, S. Lin, L. Chen, M. Yang, S. Jia, X. Zhang and P. R. Chen (2014). "Palladium-triggered deprotection chemistry for protein activation in living cells." Nat Chem **6**(4): 352-361.

Li, X., R. Zvanych, S. A. Vanner, W. Wang and N. A. Magarvey (2013). "Chemical variation from the neoantimycin depsipeptide assembly line." Bioorg Med Chem Lett **23**(18): 5123-5127.

Lima, A. O., A. Cabral, F. D. Andreote, A. Cavalett, M. L. Pessatti, F. Dini-Andreote and M. A. da Silva (2013). "Draft Genome Sequence of Bacillus stratosphericus LAMA 585, Isolated from the Atlantic Deep Sea." Genome Announc **1**(3).

Liu, B., C. J. Schofield and R. C. Wilmouth (2006). "Structural analyses on intermediates in serine protease catalysis." J Biol Chem **281**(33): 24024-24035.

Liu, C., M. Yuan, X. Xu, L. Wang, A. T. Keatinge-Clay, Z. Deng, S. Lin and J. Zheng (2018). "Substrate-bound structures of a ketoreductase from amphotericin modular polyketide synthase." J Struct Biol **203**(2): 135-141.

Liu, J., X. Zhu, S. J. Kim and W. Zhang (2016). "Antimycin-type depsipeptides: discovery, biosynthesis, chemical synthesis, and bioactivities." Nat Prod Rep **33**(10): 1146-1165.

Liu, Y., T. Zheng and S. D. Bruner (2011). "Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases." Chem Biol **18**(11): 1482-1488.

Liuzzi, V. C., V. Mirabelli, M. T. Cimmarusti, M. Haidukowski, J. F. Leslie, A. F. Logrieco, R. Caliandro, F. Fanelli and G. Mule (2017). "Enniatin and Beauvericin Biosynthesis in Fusarium Species: Production Profiles and Structural Determinant Prediction." Toxins (Basel) **9**(2).

Lohman, J. R., M. Ma, M. E. Cuff, L. Bigelow, J. Bearden, G. Babnigg, A. Joachimiak, G. N. Phillips and B. Shen (2014). "The crystal structure of BlmI as a model for nonribosomal peptide synthetase peptidyl carrier proteins." Proteins **82**(7): 1210-1218.

Long, J. Z. and B. F. Cravatt (2011). "The metabolic serine hydrolases and their functions in mammalian physiology and disease." Chem Rev **111**(10): 6022-6063.

Maass, S., S. Sievers, D. Zuhlke, J. Kuzinski, P. K. Sappa, J. Muntel, B. Hessling, J. Bernhardt, R. Sietmann, U. Volker, M. Hecker and D. Becher (2011). "Efficient, global-scale quantification of absolute protein amounts by integration of targeted mass spectrometry and two-dimensional gel-based proteomics." Anal Chem **83**(7): 2677-2684.

Mach, B., E. Reich and E. L. Tatum (1963). "Separation of the Biosynthesis of the Antibiotic Polypeptide Tyrocidine from Protein Biosynthesis." Proc Natl Acad Sci U S A **50**(1): 175-181.

Magarvey, N. A., Z. Q. Beck, T. Golakoti, Y. Ding, U. Huber, T. K. Hemscheidt, D. Abelson, R. E. Moore and D. H. Sherman (2006). "Biosynthetic characterization and chemoenzymatic assembly of the cryptophycins. Potent anticancer agents from cyanobionts." ACS Chem Biol **1**(12): 766-779.

Magarvey, N. A., M. Ehling-Schulz and C. T. Walsh (2006). "Characterization of the cereulide NRPS alpha-hydroxy acid specifying modules: activation of alpha-keto acids and chiral reduction on the assembly line." J Am Chem Soc **128**(33): 10698-10699.

Makarasen, A., T. Nishikawa and M. Isobe (2009). "Synthesis of Four Lysine-Linked Cereulide Analogues Showing Ionophoric Activity Towards Potassium Cations as Lead Compounds for Emetic Toxin-Detection by Immunoassays." Synthesis **2009**(13): 2184-2204.

Manavalan, B., S. Murugapiran, G. Lee and S. Choi (2010). "Molecular modeling of the reductase domain to elucidate the reaction mechanism of reduction of peptidyl thioester into its corresponding alcohol in non-ribosomal peptide synthetases." BMC Structural Biology **10**(1): 1.

Marahiel, M. A. (2009). "Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis." J Pept Sci **15**(12): 799-807.

Matter, A. M., S. B. Hoot, P. D. Anderson, S. S. Neves and Y. Q. Cheng (2009). "Valinomycin biosynthetic gene cluster in Streptomyces: conservation, ecology and evolution." PLoS One **4**(9): e7194.

May, J. J., N. Kessler, M. A. Marahiel and M. T. Stubbs (2002). "Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases." Proc Natl Acad Sci U S A **99**(19): 12120-12125.

May, J. J., T. M. Wendrich and M. A. Marahiel (2001). "The dhb operon of Bacillus subtilis encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin." J Biol Chem **276**(10): 7209-7217.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.

McQuade, T. J., A. D. Shallop, A. Sheoran, J. E. Delproposto, O. V. Tsodikov and S. Garneau-Tsodikova (2009). "A nonradioactive high-throughput assay for screening and characterization of

adenylation domains for nonribosomal peptide combinatorial biosynthesis." Anal Biochem **386**(2): 244-250.

Menkhaus, M., C. Ullrich, B. Kluge, J. Vater, D. Vollenbroich and R. M. Kamp (1993). "Structural and functional organization of the surfactin synthetase multienzyme system." J Biol Chem **268**(11): 7678-7684.

Mikkola, R., N. E. Saris, P. A. Grigoriev, M. A. Andersson and M. S. Salkinoja-Salonen (1999). "Ionophoretic properties and mitochondrial effects of cereulide: the emetic toxin of B. cereus." Eur J Biochem **263**(1): 112-117.

Miller, B. R., E. J. Drake, C. Shi, C. C. Aldrich and A. M. Gulick (2016). "Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture." J Biol Chem **291**(43): 22559-22571.

Miller, B. R. and A. M. Gulick (2016). "Structural Biology of Nonribosomal Peptide Synthetases." Methods Mol Biol **1401**: 3-29.

Miller, D. A., L. Luo, N. Hillson, T. A. Keating and C. T. Walsh (2002). "Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of Yersinia pestis." Chem Biol **9**(3): 333-344.

Mitchell, C. A., C. Shi, C. C. Aldrich and A. M. Gulick (2012). "Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains." Biochemistry **51**(15): 3252-3263.

Mori, S., A. H. Pang, T. A. Lundy, A. Garzan, O. V. Tsodikov and S. Garneau-Tsodikova (2018). "Structural basis for backbone N-methylation by an interrupted adenylation domain." Nature Chemical Biology.

Muller, S., E. Garcia-Gonzalez, A. Mainz, G. Hertlein, N. C. Heid, E. Mosker, H. van den Elst, H. S. Overkleeft, E. Genersch and R. D. Sussmuth (2014). "Paenilamicin: structure and biosynthesis of

a hybrid nonribosomal peptide/polyketide antibiotic from the bee pathogen Paenibacillus larvae." Angew Chem Int Ed Engl **53**(40): 10821-10825.

Murshudov, G. N., P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin (2011). "REFMAC5 for the refinement of macromolecular crystal structures." Acta Crystallogr D Biol Crystallogr **67**: 355-367.

Narita, K., T. Kikuchi, K. Watanabe, T. Takizawa, T. Oguchi, K. Kudo, K. Matsuhara, H. Abe, T. Yamori, M. Yoshida and T. Katoh (2009). "Total synthesis of the bicyclic depsipeptide HDAC inhibitors spiruchostatins A and B, 5''-epi-spiruchostatin B, FK228 (FR901228) and preliminary evaluation of their biological activity." Chemistry **15**(42): 11174-11186.

Otto, H. H. and T. Schirmeister (1997). "Cysteine Proteases and Their Inhibitors." Chem Rev **97**(1): 133-172.

Pedras, M. S., L. Irina Zaharia and D. E. Ward (2002). "The destruxins: synthesis, biosynthesis, biotransformation, and biological activity." Phytochemistry **59**(6): 579-596.

Perkins, J. B., S. K. Guterman, C. L. Howitt, V. E. Williams, 2nd and J. Pero (1990). "Streptomyces genes involved in biosynthesis of the peptide antibiotic valinomycin." J Bacteriol **172**(6): 3108-3116.

Pitchayawasin, S., M. Isobe, M. Kuse, T. Franz, N. Agata and M. Ohta (2004). "Molecular diversity of cereulide detected by means of nano-HPLC-ESI-Q-TOF-MS." International Journal of Mass Spectrometry **235**(2): 123-129.

Plechanovova, A., E. G. Jaffray, M. H. Tatham, J. H. Naismith and R. T. Hay (2012). "Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis." Nature **489**(7414): 115-120.

Potharla, V. Y., C. Wang and Y. Q. Cheng (2014). "Identification and characterization of the spiruchostatin biosynthetic gene cluster enable yield improvement by overexpressing a transcriptional activator." J Ind Microbiol Biotechnol **41**(9): 1457-1465.

Prieto, C., C. Garcia-Estrada, D. Lorenzana and J. F. Martin (2012). "NRPSsp: non-ribosomal peptide synthase substrate predictor." Bioinformatics **28**(3): 426-427.

Qiao, C., D. J. Wilson, E. M. Bennett and C. C. Aldrich (2007). "A mechanism-based aryl carrier protein/thiolation domain affinity probe." J Am Chem Soc **129**(20): 6350-6351.

Quadri, L. E., P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber and C. T. Walsh (1998). "Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases." Biochemistry **37**(6): 1585-1595.

Radzicka, A. and R. Wolfenden (1996). "Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases." Journal of the American Chemical Society **118**(26): 6105-6109.

Rajkovic, A., M. Uyttendaele, A. Vermeulen, M. Andjelkovic, I. Fitz-James, P. in 't Veld, Q. Denon, R. Verhe and J. Debevere (2008). "Heat resistance of Bacillus cereus emetic toxin, cereulide." Lett Appl Microbiol **46**(5): 536-541.

Raymond, K. N., E. A. Dertz and S. S. Kim (2003). "Enterobactin: an archetype for microbial iron transport." Proc Natl Acad Sci U S A **100**(7): 3584-3588.

Reger, A. S., R. Wu, D. Dunaway-Mariano and A. M. Gulick (2008). "Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase." Biochemistry **47**(31): 8016-8025.

Reid, R., M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson and R. McDaniel (2003). "A model of structure and catalysis for ketoreductase domains in modular polyketide synthases." Biochemistry **42**(1): 72-79.

Reimer, J. M., M. N. Aloise, P. M. Harrison and T. M. Schmeing (2016). "Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase." Nature **529**(7585): 239-242.

Reimer, J. M., A. S. Haque, M. J. Tarry and T. M. Schmeing (2018). "Piecing together nonribosomal peptide synthesis." Curr Opin Struct Biol **49**: 104-113.

Reimer, J. M., I. Harb, O. G. Ovchinnikova, J. Jiang, C. Whitfield and T. M. Schmeing (2018). "Structural Insight into a Novel Formyltransferase and Evolution to a Nonribosomal Peptide Synthetase Tailoring Domain." ACS Chem Biol **13**(11): 3161-3172.

Robbel, L. and M. A. Marahiel (2010). "Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery." J Biol Chem **285**(36): 27501-27508.

Samel, S. A., P. Czodrowski and L. O. Essen (2014). "Structure of the epimerization domain of tyrocidine synthetase A." Acta Crystallogr D Biol Crystallogr **70**(Pt 5): 1442-1452.

Samel, S. A., B. Wagner, M. A. Marahiel and L. O. Essen (2006). "The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide." J Mol Biol **359**(4): 876-889.

Sandy, M., Z. Rui, J. Gallagher and W. Zhang (2012). "Enzymatic synthesis of dilactone scaffold of antimycins." ACS Chem Biol **7**(12): 1956-1961.

Sattely, E. S., M. A. Fischbach and C. T. Walsh (2008). "Total biosynthesis: in vitro reconstitution of polyketide and nonribosomal peptide pathways." Nat Prod Rep **25**(4): 757-793.

Scaglione, J. B., D. L. Akey, R. Sullivan, J. D. Kittendorf, C. M. Rath, E. S. Kim, J. L. Smith and D. H. Sherman (2010). "Biochemical and structural characterization of the tautomycetin thioesterase: analysis of a stereoselective polyketide hydrolase." Angew Chem Int Ed Engl **49**(33): 5726-5730.

Schneider, T. L., B. Shen and C. T. Walsh (2003). "Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation." Biochemistry **42**(32): 9722-9730.

Schoenian, I., M. Spiteller, M. Ghaste, R. Wirth, H. Herz and D. Spiteller (2011). "Chemical basis of the synergism and antagonism in microbial communities in the nests of leaf-cutting ants." Proc Natl Acad Sci U S A **108**(5): 1955-1960.

Schwarzer, D., R. Finking and M. A. Marahiel (2003). "Nonribosomal peptides: from genes to products." Nat Prod Rep **20**(3): 275-287.

Schwarzer, D., H. D. Mootz, U. Linne and M. A. Marahiel (2002). "Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases." Proc Natl Acad Sci U S A **99**(22): 14083-14088.

Shaw-Reid, C. A., N. L. Kelleher, H. C. Losey, A. M. Gehring, C. Berg and C. T. Walsh (1999). "Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EntF catalyzes both elongation and cyclolactonization." Chem Biol **6**(6): 385-400.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson and D. G. Higgins (2011). "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega." Mol Syst Biol **7**: 539.

Skinnider, M. A., C. A. Dejong, P. N. Rees, C. W. Johnston, H. Li, A. L. Webster, M. A. Wyatt and N. A. Magarvey (2015). "Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM)." Nucleic Acids Res **43**(20): 9645-9662.

Soga, T., Y. Ohashi, Y. Ueno, H. Naraoka, M. Tomita and T. Nishioka (2003). "Quantitative metabolome analysis using capillary electrophoresis mass spectrometry." J Proteome Res **2**(5): 488-494.

Stachelhaus, T., H. D. Mootz and M. A. Marahiel (1999). "The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases." Chem Biol **6**(8): 493-505.

Stawikowski, M. and P. Cudic (2007). Depsipeptide Synthesis. Methods in Molecular Biology. G. B. Fields. Totowa, New Jersey, United States of America, Humana Press. **386:** 321-339.

Steller, S., D. Vollenbroich, F. Leenders, T. Stein, B. Conrad, J. Hofemeister, P. Jacques, P. Thonart and J. Vater (1999). "Structural and functional organization of the fengycin synthetase multienzyme system from Bacillus subtilis b213 and A1/3." Chem Biol **6**(1): 31-41.

Stenfors Arnesen, L. P., A. Fagerlund and P. E. Granum (2008). "From soil to gut: Bacillus cereus and its food poisoning toxins." FEMS Microbiol Rev **32**(4): 579-606.

Strieker, M., A. Tanovic and M. A. Marahiel (2010). "Nonribosomal peptide synthetases: structures and dynamics." Curr Opin Struct Biol **20**(2): 234-240.

Sundlov, J. A., C. Shi, D. J. Wilson, C. C. Aldrich and A. M. Gulick (2012). "Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains." Chem Biol **19**(2): 188-198.

Sussmuth, R. D. and A. Mainz (2017). "Nonribosomal Peptide Synthesis-Principles and Prospects." Angew Chem Int Ed Engl **56**(14): 3770-3821.

Swatek, K. N. and D. Komander (2016). "Ubiquitin modifications." Cell Res **26**(4): 399-422.

Taevernier, L., E. Wynendaele, B. Gevaert and B. Spiegeleer (2017). "Chemical Classification of Cyclic Depsipeptides." Curr Protein Pept Sci **18**(5): 425-452.

Tan, X. F., Y. N. Dai, K. Zhou, Y. L. Jiang, Y. M. Ren, Y. Chen and C. Z. Zhou (2015). "Structure of the adenylation-peptidyl carrier protein didomain of the Microcystis aeruginosa microcystin synthetase McyG." Acta Crystallogr D Biol Crystallogr **71**(Pt 4): 873-881.

Tanovic, A., S. A. Samel, L. O. Essen and M. A. Marahiel (2008). "Crystal structure of the termination module of a nonribosomal peptide synthetase." Science **321**(5889): 659-663.

Tarry, M. J., A. S. Haque, K. H. Bui and T. M. Schmeing (2017). "X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture." Structure **25**(5): 783-793 e784.

Tedesco, D. and L. Haragsim (2012). "Cyclosporine: a review." J Transplant **2012**: 230386.

Terwilliger, T. C., R. W. Grosse-Kunstleve, P. V. Afonine, N. W. Moriarty, P. H. Zwart, L. W. Hung, R. J. Read and P. D. Adams (2008). "Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard." Acta Crystallogr D Biol Crystallogr **64**(Pt 1): 61-69.

Thies, S., B. Santiago-Schubel, F. Kovacic, F. Rosenau, R. Hausmann and K. E. Jaeger (2014). "Heterologous production of the lipopeptide biosurfactant serrawettin W1 in Escherichia coli." J Biotechnol **181**: 27-30.

Toh, M., M. C. Moffitt, L. Henrichsen, M. Raftery, K. Barrow, J. M. Cox, C. P. Marquis and B. A. Neilan (2004). "Cereulide, the emetic toxin of Bacillus cereus, is putatively a product of nonribosomal peptide synthesis." J Appl Microbiol **97**(5): 992-1000.

Tomino, S., M. Yamada, H. Itoh and Kurahashik (1967). "Cell-free synthesis of gramicidin S." Biochemistry **6**(8): 2552-2560.

Trauger, J. W., R. M. Kohli, H. D. Mootz, M. A. Marahiel and C. T. Walsh (2000). "Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase." Nature **407**(6801): 215-218.

Tsai, S. C., H. Lu, D. E. Cane, C. Khosla and R. M. Stroud (2002). "Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases." Biochemistry **41**(42): 12598-12606.

Tsai, S. C., L. J. Miercke, J. Krucinski, R. Gokhale, J. C. Chen, P. G. Foster, D. E. Cane, C. Khosla and R. M. Stroud (2001). "Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel." Proc Natl Acad Sci U S A **98**(26): 14808-14813.

Tseng, C. C., S. D. Bruner, R. M. Kohli, M. A. Marahiel, C. T. Walsh and S. A. Sieber (2002). "Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase." Biochemistry **41**(45): 13350-13359.

Uhlmann, S., R. D. Sussmuth and M. J. Cryle (2013). "Cytochrome p450sky interacts directly with the nonribosomal peptide synthetase to generate three amino acid precursors in skyllamycin biosynthesis." ACS Chem Biol **8**(11): 2586-2596.

van der Meij, A., S. F. Worsley, M. I. Hutchings and G. P. van Wezel (2017). "Chemical ecology of antibiotic production by actinomycetes." FEMS Microbiol Rev **41**(3): 392-416.

van Wageningen, A. M., P. N. Kirkpatrick, D. H. Williams, B. R. Harris, J. K. Kershaw, N. J. Lennard, M. Jones, S. J. Jones and P. J. Solenberg (1998). "Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic." Chem Biol **5**(3): 155-162.

van Wezel, G. P. and K. J. McDowall (2011). "The regulation of the secondary metabolism of Streptomyces: new links and experimental advances." Nat Prod Rep **28**(7): 1311-1333.

Vanner, S. A., X. Li, R. Zvanych, J. Torchia, J. Sang, D. W. Andrews and N. A. Magarvey (2013). "Chemical and biosynthetic evolution of the antimycin-type depsipeptides." Mol Biosyst **9**(11): 2712-2719.

Velkov, T., J. Horne, M. J. Scanlon, B. Capuano, E. Yuriev and A. Lawen (2011). "Characterization of the N-methyltransferase activities of the multifunctional polypeptide cyclosporin synthetase." Chem Biol **18**(4): 464-475.

Virdee, S., Y. Ye, D. P. Nguyen, D. Komander and J. W. Chin (2010). "Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase." Nat Chem Biol **6**(10): 750-757.

Vollenbroich, D., N. Mehta, P. Zuber, J. Vater and R. M. Kamp (1994). "Analysis of surfactin synthetase subunits in srfA mutants of Bacillus subtilis OKB105." J Bacteriol **176**(2): 395-400.

Walsh, C. T., H. Chen, T. A. Keating, B. K. Hubbard, H. C. Losey, L. Luo, C. G. Marshall, D. A. Miller and H. M. Patel (2001). "Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines." Curr Opin Chem Biol **5**(5): 525-534.

Wang, B., Q. Kang, Y. Lu, L. Bai and C. Wang (2012). "Unveiling the biosynthetic puzzle of destruxins in Metarhizium species." Proc Natl Acad Sci U S A **109**(4): 1287-1292.

Wang, H., D. P. Fewer, L. Holm, L. Rouhiainen and K. Sivonen (2014). "Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes." Proc Natl Acad Sci U S A **111**(25): 9259-9264.

Wang, M., P. Opare and C. N. Boddy (2009). "Polyketide synthase thioesterases catalyze rapid hydrolysis of peptidyl thioesters." Bioorg Med Chem Lett **19**(5): 1413-1415.

Weber, T., K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Muller, W. Wohlleben, R. Breitling, E. Takano and M. H. Medema (2015). "antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters." Nucleic Acids Res **43**(W1): W237-243.

Weckwerth, W., K. Miyamoto, K. Iinuma, M. Krause, M. Glinski, T. Storm, G. Bonse, H. Kleinkauf and R. Zocher (2000). "Biosynthesis of PF1022A and related cyclooctadepsipeptides." J Biol Chem **275**(23): 17909-17915.

Weissman, K. J. (2015). "The structural biology of biosynthetic megaenzymes." Nat Chem Biol **11**(9): 660-670.

Wenzel, S. C., B. Kunze, G. Hofle, B. Silakowski, M. Scharfe, H. Blocker and R. Muller (2005). "Structure and biosynthesis of myxochromides S1-3 in Stigmatella aurantiaca: evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis." Chembiochem **6**(2): 375-385.

Whicher, J. R., G. Florova, P. K. Sydor, R. Singh, M. Alhamadsheh, G. L. Challis, K. A. Reynolds and J. L. Smith (2011). "Structure and function of the RedJ protein, a thioesterase from the prodiginine biosynthetic pathway in Streptomyces coelicolor." J Biol Chem **286**(25): 22558-22569.

Wilson, D. J. and C. C. Aldrich (2010). "A continuous kinetic assay for adenylation enzyme activity and inhibition." Anal Biochem **404**(1): 56-63.

Wilson, D. J., C. Shi, A. M. Teitelbaum, A. M. Gulick and C. C. Aldrich (2013). "Characterization of AusA: a dimodular nonribosomal peptide synthetase responsible for the production of aureusimine pyrazinones." Biochemistry **52**(5): 926-937.

Winn, M., J. K. Fyans, Y. Zhuo and J. Micklefield (2016). "Recent advances in engineering nonribosomal peptide assembly lines." Nat Prod Rep **33**(2): 317-347.

Winter, G., D. G. Waterman, J. M. Parkhurst, A. S. Brewster, R. J. Gildea, M. Gerstel, L. Fuentes-Montero, M. Vollmar, T. Michels-Clark, I. D. Young, N. K. Sauter and G. Evans (2018). "DIALS: implementation and evaluation of a new integration package." Acta Crystallogr D Struct Biol **74**(Pt 2): 85-97.

Xu, Y., R. Orozco, E. M. Kithsiri Wijeratne, P. Espinosa-Artiles, A. A. Leslie Gunatilaka, S. Patricia Stock and I. Molnar (2009). "Biosynthesis of the cyclooligomer depsipeptide bassianolide, an insecticidal virulence factor of Beauveria bassiana." Fungal Genet Biol **46**(5): 353-364.

Xu, Y., R. Orozco, E. M. Wijeratne, A. A. Gunatilaka, S. P. Stock and I. Molnar (2008). "Biosynthesis of the cyclooligomer depsipeptide beauvericin, a virulence factor of the entomopathogenic fungus Beauveria bassiana." Chem Biol **15**(9): 898-907.

Yang, W. and D. G. Drueckhammer (2001). "Understanding the relative acyl-transfer reactivity of oxoesters and thioesters: computational analysis of transition state delocalization effects." J Am Chem Soc **123**(44): 11004-11009.

Yin, Y., R. Gokhale, C. Khosla and D. E. Cane (2001). "Erythromycin biosynthesis. The 4-pro-S hydride of NADPH is utilized for ketoreduction by both module 5 and module 6 of the 6-deoxyerythronolide B synthase." Bioorg Med Chem Lett **11**(12): 1477-1479.

Yonus, H., P. Neumann, S. Zimmermann, J. J. May, M. A. Marahiel and M. T. Stubbs (2008). "Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains." J Biol Chem **283**(47): 32484-32491.

Zaleta-Rivera, K., C. Xu, F. Yu, R. A. Butchko, R. H. Proctor, M. E. Hidalgo-Lara, A. Raza, P. H. Dussault and L. Du (2006). "A bidomain nonribosomal peptide synthetase encoded by FUM14 catalyzes the formation of tricarballylic esters in the biosynthesis of fumonisins." Biochemistry **45**(8): 2561-2569.

Zhang, J., N. Liu, R. A. Cacho, Z. Gong, Z. Liu, W. Qin, C. Tang, Y. Tang and J. Zhou (2016). "Structural basis of nonribosomal peptide macrocyclization in fungi." Nat Chem Biol **12**(12): 1001-1003.

Zhang, T., X. Jia, Y. Zhuo, M. Liu, H. Gao, J. Liu and L. Zhang (2012). "Cloning and characterization of a novel 2-ketoisovalerate reductase from the beauvericin producer Fusarium proliferatum LF061." BMC Biotechnol **12**: 55.

Zhang, Z., R. Zhou, J. M. Sauder, P. J. Tonge, S. K. Burley and S. Swaminathan (2011). "Structural and functional studies of fatty acyl adenylate ligases from E. coli and L. pneumophila." J Mol Biol **406**(2): 313-324.

Zheng, J., S. K. Piasecki and A. T. Keatinge-Clay (2013). "Structural studies of an A2-type modular polyketide synthase ketoreductase reveal features controlling alpha-substituent stereochemistry." ACS Chem Biol **8**(9): 1964-1971.

Zheng, J., C. A. Taylor, S. K. Piasecki and A. T. Keatinge-Clay (2010). "Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase." Structure **18**(8): 913-922.

Zhou, Y., P. Prediger, L. C. Dias, A. C. Murphy and P. F. Leadlay (2015). "Macrodiolide formation by the thioesterase of a modular polyketide synthase." Angew Chem Int Ed Engl **54**(17): 5232-5235.

APPENDIX. SUPPLEMENTARY MATERIAL OF CHAPTER 4 INCLUDING METHODS

Supplementary Material
Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid

Nicolas Huguenin-Dezot[1,4], Diego A. Alonzo[2,4], Graham W. Heberlig[3,5], Mohan Mahesh[1,5], Duy P. Nguyen[1], Mark H. Dornan[3], Christopher N. Boddy[3], T. Martin Schmeing[2,*], Jason W. Chin[1,*]

Chemical synthesis can be found in https://www.nature.com/articles/s41586-018-0781-z#Sec5

Supplementary Discussion 1: The direction of the oligomerisation pathway catalysed by Vlm TE_{wt}

There are two possible pathways for oligomerisation of NRPS intermediates by TE domains such as Vlm TE, as illustrated in **Fig. 4.3b** and **Extended Data Fig. 4.5a-d**. Vlm TE_{wt} could oligomerise tetradepsipetidyl *D*-hiv–*D*-val-*L*-lac–*L*-val moieties by ester bond formation between the distal hydroxyl of *D*-hiv in tetradepsipeptidyl-O-TE and the carbonyl of *L*-val from tetradepsipeptidyl-S-PCP ("forward transfer"; **Extended Data Fig. 4.5a, c**), or else by ester bond formation between the distal hydroxyl of *D*-hiv in tetradepsipeptidyl-S-PCP and the carbonyl of *L*-val from tetradepsipeptidyl-O-TE ("reverse transfer" (**Extended Data Fig. 4.5b, d)**. The reverse transfer pathway is so called because the intermediate is passed from a more C-terminal domain of the NRPS (TE domain) to a more N-terminal domain (PCP domain), whereas typically in NRPS synthesis intermediates are progressively elongated and transferred from more N- to more C-terminal domains (**Fig. 4.1c, Extended Data Fig. 4.1)**. In addition, the octadepsipeptide would later be transferred again to the TE domain, as seen in the far right of panels and **Extended Data Fig. 4.5a-d**.

The synthetic intermediates detected in the Vlm TE_{wt}-mediated synthesis of valinomycin differentiate between two possible oligomerisation pathways[1,2]. LC-MS analysis of Vlm TE_{wt} reactions with tetradepsipeptidyl-SNAC **7** showed formation of valinomycin, as well as the octadepsipeptidyl-SNAC **11** and dodecadepsipeptidyl-SNAC **15**, (**Fig. 4.3a, Extended Data Fig. 4.5g**). Intermediates **11** and **15** are only produced in the reverse transfer oligomerisation pathway (**Extended Data Fig. 4.5b, d**) and not the forward transfer pathway, indicating Vlm TE catalyses the reverse transfer pathway. Consistent with the reverse pathway, experiments using a mixture of tetradepsipeptidyl-SNAC **7** and tetradepsipeptidyl-SNAC missing the terminal hydroxyl (deoxy-tetradepsipeptidyl-SNAC **8**), showed peaks for deoxy-octadepsipeptidyl-SNAC **12** and deoxy-dodecadepsipeptidyl-SNAC **16** (**Extended Data Fig. 4.5j**). In addition, the valinomycin synthesis assay also shows small peaks corresponding to the 16-mer depsipeptidyl-SNAC **19**, the 20-mer depsipeptidyl-SNAC **23** and the cyclic 16-mer depsipeptide **29,** (**Fig. 4.3a, Extended Data Fig. 4.5i**)

indicating that a small percentage of the time, the oligomerisation (again, by the reverse pathway) by Vlm TE domain proceeds past the dodecadepsipeptidyl intermediate, and that Vlm TE has somewhat more flexibility in final product than previously thought.

Our biochemical experiments with TE_{DAP} also supports the conclusion that synthesis occurs via the reverse transfer pathway. As discussed in the main text, incubation of tetradepsipeptidyl-SNAC with TE_{DAP} produces tetradepsipeptidyl-*N*-TE_{DAP}, where the tetradepsipeptidyl is attached to TE by a stable amine bond (**Fig. 4.4**), but all other atoms are the same as in the native transient tetradepsipeptidyl-*O*-TE_{wt} intermediate. If the synthesis occurs by the forward transfer pathway, the distal hydroxyl of *D*-hiv in tetradepsipeptidyl-*N*-TE_{DAP} should attack the carbonyl of *L*-val from the next tetradepsipeptidyl-SNAC molecule, making octadepsipeptidyl-*N*-TE_{DAP} (like **Extended Data Fig. 4.5a, c**). This should then undergo the next oligomerisation step analogously, producing dodecadepsipeptidyl-*N*-TE_{DAP}. This dodecadepsipeptidyl-*N*-TE_{DAP} intermediate should accumulate because the amide link will stall cyclisation. Thus, in the forward transfer case, incubation of tetradepsipeptidyl-SNAC with TE_{DAP} should result in dodecadepsipeptidyl-*N*-TE_{DAP} (with smaller amounts of tetradepsipeptidyl-*N*-TE_{DAP} and octadepsipeptidyl-*N*-TE_{DAP}). In contrast, if the synthesis occurs by the reverse transfer pathway, the distal hydroxyl of *D*-hiv in tetradepsipeptidyl-SNAC will attempt to attack the carbonyl of *L*-val in the amide group of tetradepsipeptidyl-*N*-TE, but will be stalled by the enhanced stability of the amide bond. Thus, in the reverse transfer case, incubation of tetradepsipeptidyl-SNAC with TE_{DAP} should result in tetradepsipeptidyl-*N*-TE_{DAP}, and negligible amounts of other acyl-TE species. Our data shows formation of tetradepsipeptidyl-*N*-TE_{DAP}, with no octadepsipeptidyl-*N*-TE_{DAP} or dodecadepsipeptidyl-*N*-TE_{DAP} detected, fully consistent with the reverse transfer pathway.

That this oligomerising-cyclising depsipeptide synthetases uses an analogous reverse pathway to the more canonical gramicidin S synthetase[1,2] and the dimerising-cyclising elaiophylin synthase[3] suggests that all oligomerising-cyclising NRPSs and PKSs will use this synthetic scheme.

Supplementary Discussion 4.2: Putative model of oligomerisation and cyclisation catalysed by Vlm TE

Several NRPS pathways feature TE domains that oligomerise and cyclise linear peptidyl or depsipeptidyl substrates. They are involved in biosynthesis of the antibiotic gramicidin S[2], the emetic toxin cereulide[4,5], the siderophores enterobactin and bacillibactin[6,7], the anticancer conglobatin[8], the DNA bis-intercalator thiocoraline[9] and valinomycin, a potassium ionophore depsipeptide with antimicrobial, antitumoural and cytotoxic properties[5,10]. These TE domains perform a challenging synthetic task: They oligomerise their (depsi)peptidyl intermediates up to, but not beyond, the number of copies found in the biologically active compound, and then change their reactivity mode to catalyse cyclisation and release of the completed product.

The structures of TE_{wt}, tetradepsipeptidyl-TE_{DAP} and dodecadepsipeptidyl-TE_{DAP} can provide insight into the oligomerisation steps, change in reactivity, and cyclisation step. While the rest of the TE domain does not alter its conformation substantially between any of our structures, there is a major difference in lid conformations of the early (apo TE_{wt}, tetradepsipeptidyl-TE_{DAP}) and late (dodecadepsipeptidyl-TE_{DAP}) stage structures (**Fig. 4.4h-i**). We do not believe the crystal packing environment unduly influences the conformation of the lid because all the structures are derived from crystallisation in very similar conditions, and because the dodecadepsipeptidyl-TE_{DAP} conformation is seen in several different crystal packing environments. However, it is likely that the presence of a particular length of depsipeptidyl-moiety does not "lock" the lid into one particular conformation. Rather, the lid will remain dynamic throughout the catalytic cycle, with the presence of the various substrates influencing which conformations predominate. Indeed, the dodecadepsipeptidyl-TE_{DAP} structures are actually a clustered of similar conformations, rather than one single conformation (**Fig. 4.4h**). When for simplicity and clarity only one dodecadepsipeptidyl-TE_{DAP} (or tetradepsipeptidyl-TE_{DAP}) structure is shown (**Fig. 4.4c-g, j, i, Extended Data Fig. 4.9, Supplementary Video 1, 2**) it is meant to be representative of the family of similar conformations, rather than "the dodecadepsipeptidyl-TE_{DAP} conformation".

As described in the main text, to transition from the early type of lid conformation (apo TE_{wt}, tetradepsipeptidyl-TE_{DAP}) to the late (dodecadepsipeptidyl-TE_{DAP}), the lid undergoes a dramatic, non-rigid body movement where some helices rotate >90° and others translocate by ~25 Å (**Fig. 4.4j, Supplementary Video 1, 2**). The mobility of the lid has been observed in other studies (often because of disorder and lack of electron density), but the conformational changes observed with TE_{DAP} are substantially more dramatic and more informative. Lid conformations vary between apo TE structures from open conformations, suggesting ease of access of the incoming substrate to the active site pocket, to closed conformation, presumably restricting access to the active sites, and lastly to a channel, restricting but not eliminating access to the active site. Examination of the few TEs with apo and holo structures show little change in lid conformation upon formation of the acyl-enzyme intermediate. For example the structure of pikromycin TE with a non-hydrolysable phosphonate based analog of the polyketide acyl chain shows the lid in a channel conformation very similar to the channel seen in the apo TE structure. Similarly, the deoxyerythronolide B TE domain shows subtle movement of the lid upon formation of a phosphonate ester. This underscores the significant changes seen in the lid conformation of the Vlm TE upon formation of the dodecodepsipeptidyl-intermediate.

In dodecadepsipeptidyl-TE_{DAP}, the lid helices form a concave pocket, which seems likely to be important during the cyclisation step in the thioesterase cycle. This pocket is mainly made up of hydrophobic residues, and it provides a steric barrier that prevents a dodecadepsipeptide attached to Ser/DAP2463 from extending out in a linear fashion (**Fig. 4.4j)**. Rather, this lid conformation favours curling back of the substrate's free end towards the acyl linkage between TE and the substrate**.** Thus, cyclisation of the dodecadepsipeptide to valinomycin may be thought of as entropically controlled by the pocket, with the dodecapeptide conformations dictated by partial confinement in the pocket and TE domain active site.

The PCP domain, a key player in the TE domain catalytic cycle, is absent from the structures we have determined. Its binding site can be inferred from the informative dead-end inhibitor trapped PCP-TE structure of EntF[11] (**Extended Data Fig. 4.9a**). The PCP domain docks at

αE of the TE, and the PPE extends the ~15 Å to position the thiol near Ser/DAP2463 (**Extended Data Fig. 4.9a, b-iii**). The position of the PPE in the EntF structure is compatible with the dodecapeptide-bound conformation of the lid, but not with the apo/tetrapeptide-bound conformation in our Vlm TE structures. (The lid in the EntF structure is partially disordered.) This EntF structure showed how the PCP and TE domains can position the thioester of the depsipeptidyl-PPE near Ser/DAP2463, but in Vlm, these domains must also be able to position the terminal hydroxyl of tetradepsipeptidyl-PPE near Ser/DAP2463 for the oligomerisation step. To do so, another ~15 Å of length of tetradepsipeptide (between terminal hydroxyl and PPE sulphur) must be accommodated in the TE domain (compare **Extended Data Fig. 4.9b-ii** and **b-iii**). The lid likely facilities this, perhaps using a pocket similar to the one we observe in the dodecapeptidyl-TE_{DAP} structures.

One can thus assemble the known structures into a hypothetical pathway for oligomerisation and cyclisation (**Extended Data Fig. 4.9b**). When in the observed apo/tetradepsipeptide-bound conformation, Lα1 of Vlm TE may inhibit any attached depsipeptide from curling around for cyclisation (**Extended Data Fig. 4.9b-i**). PCP binding could induce a TE conformation similar to those we observe for the dodecadepsipeptide-bound TE, which could accommodate the ~30 Å tetradepsipeptidyl-PPE bound to the PCP domain and guide it towards the active site (**Extended Data Fig. 4.9b-ii**). A transition to an open / largely disordered lid (as seen in EntF PCP-TE) could allow the PCP to present the thioester for transfer back to the Ser2463 (**Extended Data Fig. 4.9b-iii**). Finally, the lid conformation observed in the dodecadepsipeptide-TE_{DAP} structures, with its semi-sphere-like pocket, could help curl the dodecadepsipeptide back towards Ser2463 for cyclisation. (**Extended Data Fig. 4.9b-iv**).

The observation that there are multiple similar conformations of Vlm TE even when it is covalently bound with bona fide substrates, and the paucity of specific interactions between the lid and the rest of the TE domain make it unlikely that there is a single, fully defined conformation at any of these steps of the synthetic cycle. Formation of a pre-defined / templated conformation of the cyclisation substrate has been proposed to facilitate cyclisation in tyrocidine synthetase[1,12],

while specific interaction between the lid and the polyketide substrate was proposed to do this in pikromycin synthase, but there is no evidence for these mechanisms in Vlm TE. Indeed, specific and strong binding interactions could slow the synthetic cycle, as the tetradepsipeptide must transition back and forth between being ligated to the PCP domain and to the TE domain, and the same tetradepsipeptide must assume multiple different positions in the course of a cycle. Rather, the lid conformation likely fluctuates rapidly through the cycle, "breathing" and transiently visiting reaction-competent conformations. Interestingly, a novel inhibitor to a *Mycobacterium tuberculosis* polyketide synthase TE domain binds between the cluster of lid helices[13]. It is proposed to compete with substrate binding, but such an inhibitor could also act by preventing structural rearrangements in the lid similar to those we observe here.

Supplementary Methods

Supplementary Table 4.1: List of primers used in this study. Mutated residues are depicted in uppercase.

Primer name	Primers sequence
MbY271f	ggaaaggtctcgaccctgNNKaactatctgcgtaaactggatcgtattc
MbY271r	ggaaaggtctcagggtcggggccagcatcggacgcag
MbN311f	ggaaaggtctccatggttNNKttttgccaaatgggcagcggctgcacc
MbN311r	ggaaaggtctcaccatggtgaattcttccaggtgttctttgc
MbY349f	ggaaaggtctccatggtgNNKggcgataccctggatattatgcatgg
MbY349r	ggaaaggtctcaccatgcagctatcgcccacaatttcgaagtc
MbV366f	ggaaaggtctccagcgcgNNKgtgggtccggttagcctggatcgtg
MbV366r	ggaaaggtctcgcgctgctcagttccagatcgccatgc
MbW382f	ggaaaggtctctaaaccgNNKattggcgcgggttttggcctggaacg
MbW382r	ggaaaggtctcggtttatcaatgccccattcacgatcc
TEV_Amb_fw	ggaaaggtctcgTAGggcagtccattagtatcaactagagatgg
TEV_Amb_rev	ggaaaggtctcccCTActgcccatccttggtttgaatccaatgc
TEV_Ala_fw	ggaaaggtctcgGCTggcagtccattagtatcaactagagatgg
TEV_Ala_rev	ggaaaggtctcccAGCctgcccatccttggtttgaatccaatgc
TE_for_pNHD_fw	ttattacatatgcatcatcatcaccaccatc

TE_for_pNHD_rev	ataataactcgagttagccacgcg
Vlm2_TE_Amb_Fw	gtgtatatcggtggtcacTAGctgggtggccatat
Vlm2_TE_Amb_Rev	atatggccacccagCTAgtgaccaccgatatacac

Creation of DAPRSlib library by inverse PCR

Using the plasmid pBK-*pylS* as a template[14], the library (DAPRSlib) for amino acid **6** was generated by five consecutive rounds of inverse PCR reactions using the PrimeSTAR HS DNA Polymerase (Takara Bio) following manufacturer's guidelines. Primers randomised the codons for positions Y271, N311, Y349, V366 and W382 of the *pylS* gene to the codons for all 20 natural amino acids (all primer are listed in **Supplementary Table 4.1**). The resulting PCR products were digested with BsaI-HF and DpnI, and circularised with T4 DNA ligase. DNA was transformed into Eletrocompetent MegaX DH10B™ T1R Electrocomp™ *E.coli* cells (Invitrogen) following the manufacturer's instructions and inoculated into overnight culture with appropriate antibiotic to prepare plasmid DNA. Diversity was estimated by plating serial dilutions of the transformation rescue culture on LB-agar plates with appropriate antibiotic. A library of 10^8 transformants that was isolated covered the theoretical diversity of the library with 97% confidence.

Selections of active aaRS with DAP derivatives

Selections of synthetase mutants specific for amino acids **2**-**6** were carried out as previously reported[14] using the following libraries: DAPRSlib (Y271, N311, Y349, V366, W382), D3 (L270, Y271, L274, N311, C313), PylS fwd (A267, Y271, L274, C313, M315), Susan 1(A267, Y271, Y349, V366, W382), Susan 2 (N311, C313, V366, W382, G386), Susan 4 (A267, Y349, S364, V366, G386). Briefly, *Mb*PylRS libraries in pBK vectors were subjected to five rounds of alternating positive and negative selection. The positive selections were performed in the presence of the desired ncAA (1 mM) using a chloramphenicol acetyl transferase reporter with an amber codon at a permissive position (codon 112) and expressing the cognate tRNA. Cells that survived the positive selection on chloramphenicol (typically 50 g/mL) LB agar are predicted to use either a natural amino acid

that is constitutively present in the cell or the ncAA added to the cell. The negative selection used a barnase reporter containing amber codons and providing the cognate tRNA, in the absence of ncAA, to remove synthetase variants that use natural amino acids.

***GFP(150TAG)His6* expression and purification**

Superfolder green fluorescent protein (sfGFP) with **6** incorporated at position 150 was expressed from p15A-sfGFP150TAG in MegaX DH10B T1R cells containing pBK_DAPRS or pBK_PylRS vector. LB broth supplemented with 12.5 μg/mL tetracycline, 25 μg/mL kanamycin and 1 mM of **6** or N^{ε}-*tert*-butyloxycarbonyl-lysine (BocK) was inoculated with the transformed cells. Expression was induced with 0.2% (w/v) L-(+)-arabinose (Sigma) for 16 h at 37˚C whilst shaking at 220 rpm. Bacteria were then harvested and the protein purified by polyhistidine affinity chromatography.

His6-lipoyl-TEV-Strep expression and purification

BL21 (DE3) cells were transformed with pNHD-His6-lipoyl-TEV$_{wt}$-Strep, pNHD-His6-lipoyl-TEV$_{Ala}$-Strep (gene is a gift from Mark Allen)[15] or co-transformed with pSF-DAPRS-PylT[16] pNHD-His6-lipoyl-TEV$_{Amber}$-Strep and grown on TB-agar plates containing 25 μg/mL tetracycline and (and 50 μg/mL kanamycin for co-transformed cells) overnight at 37˚C (TB media containing 25 μg/mL tetracycline (and 50 μg/mL kanamycin for co-transformed cells) was inoculated with some transformed colonies. The cultures were diluted 1:100 into TB media containing 12.5 μg/mL tetracycline (and 25 μg/mL kanamycin and 100 μM of **6** for co-transformed cells) and incubated at 37˚C; once the OD_{600} reached 0.5-0.7, the cultures were moved to 20˚C. After 30 min of further incubation, the cultures were induced using 250 μM isopropyl β-D-1-thiogalactopyranoside (IPTG) and protein expression was carried out at 20˚C for 16 h. Cells were harvested by centrifugation and resuspended in 50 mM tris-HCl pH 7.5, 150 mM NaCl, 2 mM β-Mercaptoethanol, 1 Roche Inhibitor Cocktail tablet / 50 mL, 0.5 mg/mL lysozyme (Sigma), 50 μg/mL DNase (Sigma) and lysed by sonication. The lysate was clarified by centrifugation at 39’000 × *g* for 30 min and filtered through a 0.4 μm polyethersulfone (PES) membrane. His6-lipoyl-TEV-Strep was purified using nickel affinity chromatography (HisTrap HP column, GE Healthcare) with

a linear gradient of imidazole (0 mM to 500 mM). Fractions containing the protein were further purified by Strep-tag affinity purification using a 5 mL StrepTrap HP column (GE Healthcare). After sample loading, the column was washed with strep binding buffer (50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid [HEPES] pH 8.0, 150 mM NaCl, 1 mM ethylenediaminetetraacetic acid [EDTA], 5 mM dithiothreitol [DTT]). The protein was eluted using a linear gradient of desthiobiotin (0 mM to 1.25 mM). For His6-lipoyl-TEV$_{Amber}$-Strep, the protein was irradiated with UV light (365 nm, 35 mWcm^{-2}, 1 min) at the end of the purification.

Ub$_{tev}$ expression and purification

BL21 (DE3) cells were transformed with pNHD-Ub-tev-His6 and grown on LB-agar plates containing 25 µg/mL tetracycline overnight at 37˚C. LB media containing 25 µg/mL tetracycline was inoculated with the some colonies resulting from the transformation. The culture was diluted 1:100 into fresh LB media containing 12.5 µg/mL tetracycline; once the OD_{600} reached 0.5, the cultures were induced using 1 mM IPTG and protein expression was carried out at 37˚C for 6 h. Cells were harvested by centrifugation and resuspended 50 mM tris-HCl pH 7.5, 150 mM NaCl, 2 mM β-Mercaptoethanol, 1 Roche Inhibitor Cocktail tablet / 50 mL, 0.5 mg/mL lysozyme (Sigma), 50 µg/mL DNase (Sigma) and lysed by sonication. The lysate was clarified by centrifugation at 39'000 × *g* for 30 min and filtration through a 0.4 µm PES membrane. Ub was purified using nickel affinity chromatography (HisTrap HP column, GE Healthcare) with a linear gradient of imidazole (30 mM to 500 mM). The protein was dialysed overnight against 10 mM tris-HCl at 4˚C and Ub was further purified by ion exchange chromatography (HiTrapS 5mL column, GE Healthcare) using a NaCl gradient (0-1 M mM) in 50 mM ammonium acetate, pH 4.5. Pure fractions were pooled before overnight dialysis against 20 mM tris-HCl pH 7.4. The sample was then concentrated to ~15 mg/mL using an Amicon Ultra-15 (3 kDa MWCO) centrifugal filter device (Millipore).

Reactions of TEV with Ub$_{tev}$

15 µg of His6-lipoyl-TEV-Strep were incubated at 30˚C with 60 µg of Ub$_{tev}$ and allowed to react overnight in 150 µL of 50 mM HEPES pH 8.0, 150 mM NaCl, 1 mM EDTA, 5 mM DTT. 20 µL of the

reaction were loaded on a 4-12% NuPAGE Bis-Tris gel (Invitrogen) and allowed to run for 45 min in 2-(n-morpholino)-ethanesulfonic acid (MES) buffer. Protein was transferred on a polyvinylidene fluoride (PVSF) membrane (Roche) using 25 mM Tris pH 8.2, 192 mM glycine, 10% (v/v) methanol. Membranes were subsequently blocked for 1h in TBST buffer (25 mM Tris pH pH 7.4, 150 mM NaCl, 0.05% [v/v] Tween 20) containing 5% (w/v) milk powder at room temperature. Antibodies (Strep-Tactin-HRP conjugate (α Strep) [IBA Lifesciences] or P4D1 antibody (α Ub) [Enzo Life Sciences]) were added in 5% TBST-milk and incubated at 4˚C overnight. Secondary antibody (for α Ub antibody) was added in 5% TBST-milk and incubated at room temperature for 1 h. Blots were developed using Amersham enhanced chemiluminescence (ECL) (GE Healthcare) and a ChemiDoc XRS+ gel imaging system (Bio-Rad).

Analysis of intracellular concentration of DAP derivatives

The analysis of intracellular concentration of DAP derivatives was performed as previously described[16]. In short, DAP derivatives were added to a 5 mL solution of LB media to a final concentration of 1 mM. A control sample was also prepared with 5 mL of unsupplemented LB media. Each solution was inoculated with DH10B cells. The cultures were agitated at 220 rpm in the dark at 37˚C for 12 h. The OD_{600} of each sample was determined, and the cells from each culture were harvested. The cell pellets were washed three times with 1 mL of fresh ice-cold LB media by cycles of resuspension and centrifugation. The washed cell pellets were resuspended in a methanol:water solution (60:40). Zirconium beads (0.1 mm) were added to each suspension. The suspensions were vortexed for 12 min to lyse the cells. The lysate was centrifuged at 21000 x *g* for 30 min at 4˚C. The supernatant was carefully removed, and placed into a fresh 1.5 mL Eppendorf tube. The solutions were centrifuged again at 21000 x *g* for 2 h at 4˚C. A 100 µl aliquot of the supernatant from the resulting sample was analysed by LC-ESI-MS. A gradient of 0.5% to 95% acetonitrile in water was applied to elute the clarified lysates from a Zorbax C18 (4.6 x 150 mm) column. The concentrations were estimated using an estimate of 8×10^8 cells per 1 OD_{600} unit and a cell volume of 0.6×10^{-15} L.

Cloning, expression and purification of Vlm TE constructs

A codon-optimised construct containing *vlm2*$_{PCP4\text{-}TE}$(encoding residues 2290-2655 of Vlm2 from *Streptomyces tsusimaensis*, GenBank: ABA59548.1) was synthesised by ATUM (formerly DNA 2.0) in a *pJExpress411* vector, with an N-terminal hexahistidine tag followed by a tobacco etch virus protease (TEV) cleavage recognition sequence (*pJExpress411-vlm2-PCP4-TE*$_{wt}$). Two BamHI recognition sequences were included in *pJExpress411-vlm2-PCP4-TE*$_{wt}$, at nucleotide positions 2024-2025 and 2267-2268. Digestion with BamHI followed by ligation with T4 DNA ligase (New England Biolabs) excised the PCP4 domain sequence, yielding the plasmid *pJExpress411-vlm2-TE*$_{wt}$ which encodes residues 2368-2655 of Vlm2. To generate an expression vector for $TE_{DAP,}$ the TE_{wt} coding sequence was PCR-amplified from *pJExpress411-vlm2-TE*$_{wt}$ with primers TE_for_pNHD_fw and TE_for_pNHD_rev. The PCR product was digested with *NdeI* and *XhoI* and ligated into similarly digested pNHD plasmid using T4 DNA ligase, generating plasmid *pNHD-vlm2-TE*$_{wt}$. Next, an amber stop codon was introduced in place of the codon for serine 2463 by site-directed mutagenesis using primers Vlm2_TE_Amb_Fw and Vlm2_TE_Amb_Rev, generating *pNHD-vlm2-TE*$_{amber2463}$.

TE domains were heterologously expressed in *E. coli* BL21(DE3) cells transformed with *pJExpress411-vlm2-TE*$_{wt}$ (TE_{wt}) or co-transformed with *pNHD-Vlm2-TE*$_{amber2463}$ and pSF-DAPRS-PylT (TE_{DAP}). Cultures expressing TE_{wt} were grown in LB media supplemented with 17 mg L^{-1} of kanamycin. Those expressing TE_{DAP} were grown in TB media supplemented with 25 mg L^{-1} of kanamycin, 12.5 mg L^{-1} of tetracycline, 0.1 mM of **6** (a 100 mM stock solution of **6** was prepared in 0.4 M NaOH, added to the culture and neutralised using 5 M HCl). Cultures were incubated at 37˚C, with agitation at 220 r.p.m, until they reached an OD_{600nm} = 0.6, after which they were incubated at 16˚C for 30 min, and then expression was induced with 100 µM IPTG. Cultures were incubated for an additional 16 hours at 16˚C before harvesting by centrifugation at 5000 *g* for 20 min. Cell pellets were stored at -80˚C.

For protein purification, cell pellets of TE_{wt} were resuspended in 5 mL of buffer wt-A (50 mM TRIS pH 7.4, 150 mM NaCl, 50 mM imidazole, 2 mM β-mercaptoethanol [βME]) plus DNAseI (Bioshop)

per g of wet cells, and lysed by sonication. Lysate was clarified by centrifugation at 40 000 *g* for 20 min. Clarified lysate was applied to two 5 mL HiTrap IMAC FF (GE Healthcare Life Sciences) columns connected in series on an AKTA Prime system (GE Healthcare Life Sciences). Bound protein was eluted with buffer wt-B (buffer wt-A plus 150 mM imidazole). Fractions containing TE_{wt} (as determined by SDS-PAGE analysis) were pooled and incubated with TEV protease in a 1:100 (Te:TEV) mass/mass ratio and dialysed against buffer wt-C (50 mM TRIS pH 7.4, 10 mM NaCl, 2 mM βME) for 16 hours at 4˚C. The dialysed sample was applied to two 5 mL HiTrap IMAC FF columns connected in series, pre-equilibrated in buffer wt-A. Cleaved protein was recovered from the flow through and applied to two 5 mL HiTrap Q HP columns connected in series, pre-equilibrated in buffer Q-A (50 mM TRIS pH 7.4, 10 mM NaCl, 2 mM βME). Protein was eluted by a gradient of 0 to 100% buffer Q-B (50 mM TRIS pH 7.4, 500 mM NaCl, 2 mM βME) over 240 mL. TE_{wt}-containing fractions were concentrated in a 10 kDa molecular weight cut off Amicon® Ultra centrifugal filter (Millipore) and injected onto a Superdex S-200 16/60 PG column (GE-Healthcare) pre-equilibrated in SEC buffer (25 mM HEPES pH 7.4 or pH 8.0, 100 mM NaCl, 0.2 mM tris(2-carboxyethyl)phosphine [TCEP]). Fractions containing purified TE_{wt} were pooled, concentrated and flash frozen.

Cell resuspension, lysis, clarification and Ni-IMAC purification for TE_{DAP} were performed as described for Te_{wt}, except that prolonged exposure to light was avoided. After elution from the Ni-IMAC column, the sample was irradiated with UV light (365 nm, 35 $mWcm^{-2}$, 1 min). TEV cleavage and the subsequent IMAC column were performed as described for TE_{wt}, except that a 1:1 TE:TEV ratio was used. Anion exchange was performed as described for $Te_{wt,}$ except that 25 mM HEPES replaced TRIS as the buffer and 0.2 mM TCEP replaced βME as the reducing agent in the mobile phases. Relevant fractions were concentrated and injected onto a Superdex S-75 10/300 column pre-equilibrated in buffer T (25 mM HEPES pH 8.0, 100 mM NaCl, 0.2 mM TCEP). Fractions containing purified TE_{DAP} were pooled, concentrated, and immediately used for further experiments. The yield of purified TE_{wt} was 30-60 mg per L, the yield of purified TE_{DAP} was 0.1-0.5 mg per L.

Crystallography

Crystallisation conditions for TE_{wt} structure 1 were found in vapour diffusion crystallisation trials using commercially available screens (Qiagen) and a protein concentration of 10 mg mL^{-1}. Optimisation of an initial crystallisation hit in 24-well plates led to a final crystallisation condition where 3.2 µL of 10 mg mL^{-1} TE_{wt}, 4.0 µL 1.65 M DL-malic acid pH 9.5, and 0.8 µL of 17 % m/v IPTG were incubated against a reservoir solution of 500 µL of 1.65 M DL-malic acid pH 9.5. TE_{wt} structure 2 crystals were grown in similar conditions, where 0.5 µL of purified TE_{wt} at 22.4 mg mL^{-1} and 0.5 µL of 1.65 M DL-malic acid pH 8.1 were incubated against a reservoir solution of 500 µL of DL-malic acid pH 8.1. Crystals appeared between 24 and 48 hours and reached their maximum size in approximately one week.

Crystals of unliganded TE_{DAP} were grown in similar conditions to TE_{wt}, with a reservoir solution of 1.65 M DL-malic acid pH 8.0. In order to obtain the tetradepsipeptidyl-TE_{DAP} complex structure, TE_{DAP} crystals were incubated with deoxytetradepsipeptidyl-SNAC **8** once they achieved their maximum size. The reservoir solution was exchanged to 2.66 M DL-malic acid, pH 9.5, and 32 µL of a solution of 1 mM deoxytetradepsipeptidyl-SNAC**,** 2.66 M DL-malic acid pH 9.5, 100 mM NaCl, 25 mM HEPES pH 9.2, 10% DMSO was added to the drop. Crystals were incubated in this condition for 9 days at room temperature.

For dodecadepsipeptidyl-TE_{DAP} complex crystals, TE_{DAP} (0.1 mg mL^{-1}) was incubated in a 1.1 mg mL^{-1} suspension of valinomycin in buffer T for 16 hours at room temperature. The sample was centrifuged at 20 000 *g* and applied to a Superdex S-75 10/300 column preequilibrated in buffer T to remove excess valinomycin. Relevant fractions were pooled and complex formation was evaluated by LC-ESI-MS (see below). The sample was concentrated to 13.4 mg mL^{-1} and diffraction-quality crystals, with a different morphology from the TE_{wt} crystals, were obtained in sitting drops consisting of 1 µL of dodecadepsipeptidyl-TE_{DAP} complex plus 1 µL reservoir solution (1.30 to 1.45 M DL-malic acid pH 8.1) equilibrated against 500 µL reservoir solution. In an attempt to improve the occupancy of the ligand, a subset of these crystals were further incubated with

valinomycin, by addition of 20 µL of a solution containing 555 µM valinomycin, 2 M DL-malic acid pH 8.1, 11 mM HEPES pH 8.0, 44 mM NaCl, 0.088 mM TCEP for 24 hours.

TE_{wt} and dodecadepsipeptidyl-TE_{DAP} crystals were cryo-protected by addition of 10 µL (TE_{wt} structure 1 and dodecadepsipeptidyl-TE_{DAP}) or 20 µL (TE_{wt} structure 2) 3.6 M DL-malic acid pH 8.1 to the crystallisation drop. For dodecadepsipeptidyl-Te_{DAP} crystals that had been incubated with valinomycin, the drop solution was removed and replaced by 10 µL of 3.6 M DL-malic acid. Crystals were equilibrated for at least two minutes and then flash cooled in liquid nitrogen. Tetradepsipeptidyl-TE_{DAP} complex crystals were looped and flash cooled directly from the incubating solution. TE_{wt} data were first collected at the Centre for Structural Biology at McGill University, Montreal, Canada, on a Rigaku RUH3R generator and a R-AXIS IV++ detector. Higher resolution data for TE_{wt} and depsipeptidyl-TE_{DAP} complexes were collected at the Canadian Light Source (CLS) 08ID-1 beamline or at the Advanced Photon Source (APS) NE-CAT 24-ID-C beamline using a Pilatus detector (**Extended Data Table 4.1**).

TE_{wt} structure determination

Diffraction data from TE_{wt} structure 1 crystals were indexed and integrated in the space group P432 using iMosflm[22] or DIALS[23]. Further space group determination and scaling were performed using the programs POINTLESS and SCALA[24]. The structure was solved by molecular replacement using PHASER[25], with a SCULPTOR[26] modified version of the TE domain of srfA-C (PDB ID 2VSQ)[27] as a search model. The structure was iteratively refined and built with the programs Phenix[28] and AUTOBUILD[29]. Coot[30] was used for iterative model building. Topology diagrams were generated using TopDraw[31] based on results from PDBsum generate (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) using the TE_{wt} structure as input.

Diffraction data sets collected from dodecadepsipeptidyl-TE_{DAP} complex crystals were indexed into either P1 or H3 space groups using iMosflm[22] or DIALS[23]. Most crystals belonging to the H3 group showed evidence of twinning, and only non-twinned diffraction data were used for structure determination. Structures in both the P1 and H3 space groups were solved by molecular

replacement using PHASER[25], with the TE_{wt} structure lacking residues 2500-2647 used as a search model. The P1 structure had six molecules in the asymmetric unit whereas the H3 structure contained 2 molecules in the asymmetric unit. All depsipeptidyl-TE_{DAP} models were refined, and mF_o-F_c maps were generated before depsipeptide residues were built in the model (**Fig. 4.4c, d and Extended Data Fig. 4.8**). Depsipeptides were built from individual monomers of the PDB Chemical Component Dictionary (DPP, VAL, 2OP, DVA, VAD). Monomer libraries and restraints for the links between the monomers (namely DPP -> VAL, VAL -> 2OP, 2OP -> DVA, DVA -> VAD and VAD -> VAL) were calculated using AceDRG[32] and merged using LIBCHECK[33]. The resulting merged dictionary was then used for substrate building in Coot[30] and refinement in REFMAC5[34] and phenix.refine[28]. Final statistics are shown in **Extended Data Table 4.1**.

Depsipeptidyl-TE complex formation

TE_{DAP} or TE_{wt} at a final concentration of 0.2 mg mL^{-1} was incubated with tetradepsipeptidyl-SNAC **7** (1.7 mM), or valinomycin (50 µM) in buffer T containing 1.7% or 1% v/v DMSO for 16 hours. Reactions were concentrated in a 10 kDa molecular weight cut off Amicon® Ultra centrifugal filter (Millipore), clarified by centrifugation at 20 000 *g* and applied to a Superdex S-75 10/300 column pre-equilibrated in buffer T to remove excess depsipeptidyl-SNACs or valinomycin before final LC-ESI-MS analysis. To form the deoxytetradepsipeptidyl-TE_{DAP} complex, TE_{DAP} at a final concentration of 8.7 mg mL^{-1} was incubated with deoxytetradepsipeptidyl-SNAC **8** (2.6 mM) in 25 mM HEPES pH 8.6, 100 mM NaCl, 3.8% v/v DMSO for 40 hours. The sample was diluted in 100 mM ammonium bicarbonate pH 8.0 before final LC-ESI-MS analysis. All incubations were performed at room temperature.

LC-ESI-MS analysis of intact proteins

For experiments shown in **Fig. 4.2c**, **Extended Data Fig. 4.7b**, protein samples were subjected to a liquid chromatography (LC) system (Agilent 1200 series) followed by in-line electrospray ionisation mass spectrometry (ESI-MS) on a 6130 Quadrupole spectrometer. Using a Jupiter 5 µ C4 300A column, 150 mm x 2.00 mm (Phenomenex), proteins were run through the LC system using water with 0.1% (v/v) formic acid (solvent A) and a gradient (10% to 75% in 6 min and 75%

to 95% in 1,5 min) of acetonitrile with 0.1% (v/v) formic acid (solvent B). Proteins were detected by monitoring UV absorbance at 200 and 280 nm. Protein masses were calculated by deconvolution from the MS acquisition in positive ion mode, using the OpenLAB CDS software (Agilent Technologies).

For experiments shown in **Fig. 4.4a, b** and **Extended Data Fig. 4.6c**, **4.7c, d**, protein concentration was adjusted to 0.1 mg mL^{-1} in buffer T, and 16 uL was injected onto an Agilent PLRP-S (1000 A 5 µM, 50 x 2.1 mm ID) column pre-equilibrated in 95% mobile phase A (0.1% formic acid in water) and 5% mobile phase B (0.1% formic acid in 100% acetonitrile) on an Agilent Technologies 1260 Infinity HPLC system coupled to a Bruker Amazon Speed ETD ion trap mass spectrometer. MS data was collected with ExtremeScan mass range mode in positive ion polarity, scan range from 50 to 3000 m/z, accumulation time of 1586 µs, RF level of 96%, trap drive 69.8, PSP target Mass 922 m/z, and averaging over 5 spectra. External instrument calibration was performed using the Agilent ESI tune mix. The column compartment temperature was set up at 80˚C throughout the run. After injection, the column was washed for 5 minutes in initial HPLC conditions with the sample compartment diversion valve in the waste position. Next, a 5-minute gradient from 5% to 100% mobile phase B was performed, followed by an isocratic step of 100% mobile phase B for 8 minutes. Proteins were detected by monitoring UV absorbance at 280 nm. Data was analyzed using the Bruker DataAnalysis software (Bruker). Mass spectra were integrated from 10.5 to 13 minutes, and deconvoluted using a window between 10 000 to 40 000 m/z.

Tandem MS/MS analysis

Proteins were run on 4-12% NuPAGE Bis-Tris gel (Invitrogen) with MES buffer and briefly stained using InstantBlue (Expedeon). The bands were excised and stored in 20 mM Tris pH 7.4. Tryptic digestion and tandem MS/MS analyses were performed by Kate Heesom (Proteomics Facility, University of Bristol).

LC-ESI-MS analysis of Vlm TE reaction products

Purified TE_{wt} or TE_{DAP} at 0.2 mg mL^{-1} (6.5 µM) was incubated with tetradepsipeptidyl-SNAC **7** (1.7 mM), or a mix of tetradepsipeptidyl-SNAC **7** and deoxytetradepsipeptidyl-SNAC **8** (1.7 mM each) in buffer T. Samples were incubated at room temperature for 24 hours, and then quenched with one volume of 0.1% formic acid in acetonitrile. Next, samples were centrifuged at 20,000 *g*, flash frozen in liquid nitrogen and stored at -80˚C before HPLC analysis. For HPLC-MS analysis, frozen samples were thawed at room temperature, vortexed and clarified by centrifugation at 20,000 *g* before injection. HR-LC-ESI-MS was performed at the Mass Spectroscopy Facility (Department of Chemistry, McGill University) with an Agilent XDB-C8 (5 µm, 4.6 x 150 mm) column in a Dionex Ultimate 3000 UHPLC system coupled to a Bruker maXis impact QTOF mass spectrometer in positive ESI mode. Ion-trap LC-ESI-MS analysis was performed in an Agilent Technologies 1260 Infinity HPLC system coupled to a Bruker Amazon Speed ETD ion trap mass spectrometer in positive ESI mode. The column compartment was set up at 40˚C throughout the runs. Starting HPLC conditions were 50% mobile phase A (0.1% formic acid in H_2O), 50% mobile phase B (0.1% formic acid in acetonitrile). After injection (1µL for HR-LC-ESI-MS and 5 µL for ion-trap LC-ESI-MS), a gradient from 50% to 98% mobile phase B in 5 minutes was performed, followed by an isocratic step of 98% mobile phase B, run for 20 minutes. For HR-LC-ESI-MS, internal calibration was performed with an intra-run infusion at the beginning of the first analysis using Na^+ formate, and the resulting calibration was used as an external calibration for subsequent analysis. Ion trap external calibration was performed using the Agilent ESI tune mix. Data was analysed using the Bruker DataAnalysis software and the SmartFormula tool (Bruker).

Statistics and Reproducibility

Supplementary Table 4.2: SmartFormula analysis of the reactions shown Fig. 4.3a (a) and Extended Data Fig. 4.5j (b). Experiment on **Fig. 4.3a** was replicated twice with similar results. HR-MS analysis of experiment on **Extended Data Fig. 4.5j** was done once.

a. TE_{wt}+tetradepsipeptidyl-SNAC **7**.

Molecule	Adduct	Meas. m/z	#	Ion Formula	m/z	err [ppm]	Mean err [ppm]	rdb	N-Rule	eØ Conf	mSigma	Std I
tetradepsipeptidyl-SNAC **7**	H^+	490.2593	1	C22H40N3O7S	490.2581	-2.4	-3.4	4.5	ok	even	32.2	66.4
	$[NH_4]^+$	507.2878	1	C22H43N4O7S	507.2847	-6.1	1681	3.5	ok	even	50.2	108.7
	Na^+	512.2414	1	C22H39N3NaO7S	512.2401	-2.6	-4.4	4.5	ok	even	33.7	67.6
tetradepsipeptide **9**	H^+	389.2296	1	C18H33N2O7	389.2282	-3.5	-4.4	3.5	ok	even	32.5	59.9
	$[NH_4]^+$	406.2571	1	C18H36N3O7	406.2548	-5.7	1018	2.5	ok	even	24	48.7
	Na^+	411.2120	1	C18H32N2NaO7	411.2102	-4.6	-5.2	3.5	ok	even	21.6	38.1
octadepsipeptidyl-SNAC **11**	H^+	860.4695	1	C40H70N5O13S	860.4685	-1.1	-2.8	8.5	ok	even	35.4	55.8
	$[NH_4]^+$	877.4972	1	C40H73N6O13S	877.4951	-2.4	-0.7	7.5	ok	even	29.5	44
	Na^+	882.4508	1	C40H69N5NaO13S	882.4505	-0.4	-2.1	8.5	ok	even	32.9	52.4
octadepsipeptide **13**	H^+	759.4409	1	C36H63N4O13	759.4386	-3	-3.8	7.5	ok	even	30.4	44.7
	$[NH_4]^+$	776.4686	1	C36H66N5O13	776.4652	-4.5	-2.3	6.5	ok	even	19.4	26.7
	Na^+	781.4227	1	C36H62N4NaO13	781.4206	-2.7	-3.8	7.5	ok	even	28.1	42.1
dodecadepsipeptidyl-SNAC **15**	H^+	1230.6842	1	C58H100N7O19S	1230.6789	-4.3	-5.8	12.5	ok	even	38.3	44
	$[NH_4]^+$	1247.7096	1	C58H103N8O19S	1247.7055	-3.3	-4.4	11.5	ok	even	34.5	40.2
	Na^+	1252.6680	1	C58H99N7NaO19S	1252.6609	-5.7	-6.7	12.5	ok	even	44.5	49.8
16-mer depsipeptidyl-SNAC **19**	H^+	1600.9055	1	C76H130N9O25S	1600.8893	-10.1	799.8	16.5	ok	even	89.4	107.7
	$[NH_4]^+$	1617.9319	1	C76H133N10O25S	1617.9159	-9.9	790.7	15.5	ok	even	91.4	110.2
	Na^+	1622.8891	1	C76H129N9NaO25S	1622.8713	-11	788.5	16.5	ok	even	92	110.5
montanastatin **27**	H^+	741.4309	1	C36H61N4O12	741.4281	-3.8	644.9	8.5	ok	even	22.2	36.8
	$[NH_4]^+$	758.4574	1	C36H64N5O12	758.4546	-3.7	628.9	7.5	ok	even	25.7	41.9
	Na^+	763.4120	1	C36H60N4NaO12	763.4100	-2.6	-5.4	8.5	ok	even	29.3	43.7
valinomycin **28**	$[NH_4]^+$	1128.6673	1	C54H94N7O18	1128.6650	-2.1	0.7	11.5	ok	even	78.2	72.3

b. TE_{wt}+tetradepsipeptidyl-SNAC **7**+ deoxytetradepsipeptidyl-SNAC **8**

Molecule	Adduct	Meas. m/z	#	Ion Formula	m/z	err [ppm]	Mean err [ppm]	rdb	N-Rule	eØ Conf	mSigma	Std I
tetradepsipeptidyl-SNAC **7**	H^+	490.2607	1	C22H40N3O7S	490.2581	-5.2	-10.1	4.5	ok	even	17	35.1
	$[NH_4]^+$	507.2884	1	C22H43N4O7S	507.2847	-7.3	15.7	3.5	ok	even	25.7	38.4
	Na^+	512.2408	1	C22H39N3NaO7S	512.2401	-1.3	-3.5	4.5	ok	even	33.1	67
deoxytetradepsipeptidyl-SNAC **8**	H^+	474.2649	1	C22H40N3O6S	474.2632	-3.5	-7.9	4.5	ok	even	34.5	49.2
	$[NH_4]^+$	496.2462	1	C22H39N3NaO6S	496.2452	-2	-7	4.5	ok	even	29.9	61.8
tetradepsipeptide **9**	H^+	389.2307	1	C18H33N2O7	389.2282	-6.2	5.6	3.5	ok	even	7.6	12.1
	$[NH_4]^+$	406.2565	1	C18H36N3O7	406.2548	-4.1	19	2.5	ok	even	97	158.9
	Na^+	411.2126	1	C18H32N2NaO7	411.2102	-5.9	3.2	3.5	ok	even	7.9	12.8
deoxyoctadepsipeptide **10**	H^+	743.4482	1	C36H63N4O12	743.4437	-6	-7.5	7.5	ok	even	34.4	55.6
	$[NH_4]^+$	760.4719	1	C36H66N5O12	760.4702	-2.2	627.4	6.5	ok	even	39.4	52.3
	Na^+	765.4321	1	C36H62N4NaO12	765.4256	-6.6	622.1	7.5	ok	even	39.7	51.7
octadepsipeptidyl-SNAC **11**	H^+	860.4731	1	C40H70N5O13S	860.4685	-5.4	-21.3	8.5	ok	even	128.9	140.4
	$[NH_4]^+$	877.5009	1	C40H73N6O13S	877.4951	-6.6	-7	7.5	ok	even	26.4	37.9
	Na^+	882.4517	1	C40H69N5NaO13S	882.4505	-1.4	-4.1	8.5	ok	even	26.8	42
deoxyoctadepsipeptidyl-SNAC **12**	H^+	844.4781	1	C40H70N5O12S	844.4736	-5.2	-9.3	8.5	ok	even	28.4	46.5
	$[NH_4]^+$	861.5006	1	C40H73N6O12S	861.5002	-0.5	-7.1	7.5	ok	even	27.1	41.4
	Na^+	866.4591	1	C40H69N5NaO12S	866.4556	-4	-9.3	8.5	ok	even	30.1	48.4
octadepsipeptide **13**	H^+	759.4466	1	C36H63N4O13	759.4386	-10.5	-25.8	7.5	ok	even	354.4	364.2
	$[NH_4]^+$	776.4740	1	C36H66N5O13	776.4652	-11.4	0.1	6.5	ok	even	108.7	156.6
	Na^+	781.4250	1	C36H62N4NaO13	781.4206	-5.6	-0.7	7.5	ok	even	123.1	169.9
dodecadepsipeptidyl-SNAC **15**	H^+	1230.6992	1	C58H100N7O19S	1230.6789	-16.5	-28.4	12.5	ok	even	481.7	518
	$[NH_4]^+$	1247.7206	1	C58H103N8O19S	1247.7055	-12.1	719.1	11.5	ok	even	58.3	78.3
	Na^+	1252.6585	1	C58H99N7NaO19S	1252.6609	1.9	727.4	12.5	ok	even	51.6	64.5
deoxydodecadepsipeptidyl-SNAC **16**	H^+	1214.6971	1	C58H100N7O18S	1214.6840	-10.7	-19.9	12.5	ok	even	11.1	11.5
	$[NH_4]^+$	1231.7221	1	C58H103N8O18S	1231.7106	-9.4	-17.9	11.5	ok	even	29.4	34.6
	Na^+	1236.6816	1	C58H99N7NaO18S	1236.6660	-12.6	-13.1	12.5	ok	even	14.3	14.8
16-mer depsipeptidyl-SNAC **19**	$[NH_4]^+$	1617.9346	1	C76H133N10O25S	1617.9159	-11.6	683.6	15.5	ok	even	465.4	437.1
	Na^+	1622.8725	1	C76H129N9NaO25S	1622.8713	-0.8	994.2	16.5	ok	even	216	286.2
deoxy 16-mer depsipeptidyl-SNAC **20**	H^+	1584.9076	1	C76H130N9O24S	1584.8944	-8.3	794.8	16.5	ok	even	86	100
	$[NH_4]^+$	1601.9337	1	C76H133N10O24S	1601.9209	-8	288.2	15.5	ok	even	14.2	13.1
	Na^+	1606.8919	1	C76H129N9NaO24S	1606.8763	-9.7	554.5	16.5	ok	even	29	31

valinomycin **28**	$[NH_4]^+$	1128.6752	1	C54H94N7O18	1128.6650	-9	445.7	11.5	ok	even	24	29.7

Full Acknowledgments to Synchrotron Sources

CLS: Research described in this paper was performed using beamline 08ID-1 at the Canadian Light Source, which is supported by the Canada Foundation for Innovation, Natural Sciences and Engineering Research Council of Canada, the University of Saskatchewan, the Government of Saskatchewan, Western Economic Diversification Canada, the National Research Council Canada, and the Canadian Institutes of Health Research.

APS: This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P41 GM103403). The Pilatus 6M detector on 24-ID-C beam line is funded by a NIH-ORIP HEI grant (S10 RR029205). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

Supplementary data 1

Product, NRPS or PKS system, release mechanism	PDB	Active site ligands, covalent/non-covalent	Lid elements (based on VIm TE_{wt})	Lid movement	Commercialy available TE product?
Aflatoxin pksA, PKS, Claisen-cyclization	3ILS [35]	No.	Lα1-loops-Lα5	Only one conformation, lid has high B-factors.	Norsolorinate anthrone is not available.
Curacin, PKS, alkene formation	3QIT [36]	No. Acyl-ester intermediate modelled in the apo structure. [36]	Different fold from other TEs in the list or VIm TE. Fold includes a two-strand β-sheet.	Perpetual open state [36].	No
Deoxyerythronolide B, PKS, cyclization	1KEZ [37], 1MO2 [38], 5D3K [39], 5D3Z[39]	Allylphosphonate, covalent (5D3Z). Complex diphenyl alkylphosphonates could not be visualized. [39]	Lα-1-Lα5, N-terminal helices.	pH-mediated kink of 65° of a single helix.	Erythromycin (final product of the biosynthetic pathway).
Enterobactin, NRPS, cyclization	2ROQ [40], 3TEJ [11]	No	Lα-1- Lα1-Lα5	Breathing/opening motions, evaluated by reduced intensities in NMR signals and rapid solvent exchange of amide protons [40].	Yes
Fengycin, NRPS, cyclization	2CB9 [41], 2CBG [41]	PMSF, covalent (2CBG). SNAC-, coenzyme A- or thiophenyl–peptides could not be visualized. [41]	Largely disordered, loop-Lα5.	Not observed.	Yes
Pikromycin, PKS, cyclization	1MN6 [38], 1MNA [38], 1MNQ [38], 2H7X [42], 2H7Y [42], 2HFJ [43], 2HFK [43]	Triketide phosphonates, covalent (2H7X and 2H7Y). Pentaketide phosphonate, covalent (2HFJ). Product 10-deoxymethynolide in Ser-Ala mutant, non-covalent (2HFK). Modelled full-length ester intermediates [43].	Bent Lα1-Lα5	Similar to deoxyerythronolide B.	Yes
Prodiginine, NRPS/PKS hybrid type II TE, hydrolysis.	3QMV [44], 3QMW [44]	Polyethylene glycol, non-covalent (3QMW).	Lα1-Lα5	Open, closed and six in-between conformations.. Conformational changes in loop and "flap".	Yes, long-chain acyl-coenzyme A.
Surfactin, NRPS, cyclization.	1JMK [45], 2VSQ [27], 2K2Q (Type II) [46]	Heptapeptidyl-carboxylate, non-covalent [45]. Dipeptidyl boronate inhibitors, covalent [47].	Lα1-Lα2-Lα5	Open/close.	Yes.
Tautomycetin, PKS, hydrolysis	3LCR [48]	Formate, non-covalent.	Lα1-Lα5. Also two helices from the N-terminus, similar to deoxyerythronolide B	Dimerization is mediated by the N-terminal helices of the construct, possibly restricting movement.	No.
Unknown, NRPS, unknown	4ZXH [49], 4ZXI [49]	No	Lα1-Lα2-Lα3-Lα5	Not observed.	Unknown product.
Mycolic acids, FAS/PKS hybrid, acyl-transfer to polyols	5V3X [13]	Structure-based inhibitor, non-covalent (5V3X). Homology model with trehalose bound to a putative polyol-binding site, non-covalent [50].	Lα1-Lα2-Lα3-Lα5	Not observed.	No. Analogues could be synthesized by the pks13-TE_{wt} from commercially available palmitoyl Co-A and trehalose.
Colibactin, NRPS/PKS hybrid type II TE, hydrolysis and cyclization.	5UGZ [51]	No. No interpretable density for precolibactin derivatives [51].	Lα1-Lα5	Opened/closed loops.	Unknown.

Appendix References :

1 Trauger, J. W., Kohli, R. M., Mootz, H. D., Marahiel, M. A. & Walsh, C. T. Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* **407**, 215-218, doi:10.1038/35025116 (2000).

2 Hoyer, K. M., Mahlert, C. & Marahiel, M. A. The iterative gramicidin s thioesterase catalyzes peptide ligation and cyclization. *Chemistry & biology* **14**, 13-22, doi:10.1016/j.chembiol.2006.10.011 (2007).

3 Zhou, Y., Prediger, P., Dias, L. C., Murphy, A. C. & Leadlay, P. F. Macrodiolide formation by the thioesterase of a modular polyketide synthase. *Angew Chem Int Ed Engl* **54**, 5232-5235, doi:10.1002/anie.201500401 (2015).

4 Alonzo, D. A., Magarvey, N. A. & Schmeing, T. M. Characterization of cereulide synthetase, a toxin-producing macromolecular machine. *PloS one* **10**, e0128569, doi:10.1371/journal.pone.0128569 (2015).

5 Magarvey, N. A., Ehling-Schulz, M. & Walsh, C. T. Characterization of the cereulide NRPS alpha-hydroxy acid specifying modules: activation of alpha-keto acids and chiral reduction on the assembly line. *Journal of the American Chemical Society* **128**, 10698-10699, doi:10.1021/ja0640187 (2006).

6 Shaw-Reid, C. A. *et al.* Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EntF catalyzes both elongation and cyclolactonization. *Chemistry & biology* **6**, 385-400, doi:10.1016/S1074-5521(99)80050-7 (1999).

7 May, J. J., Wendrich, T. M. & Marahiel, M. A. The dhb operon of Bacillus subtilis encodes the biosynthetic template for the catecholic siderophore 2,3-dihydroxybenzoate-glycine-threonine trimeric ester bacillibactin. *J Biol Chem* **276**, 7209-7217, doi:10.1074/jbc.M009140200 (2001).

8 Zhou, Y. *et al.* Iterative Mechanism of Macrodiolide Formation in the Anticancer Compound Conglobatin. *Chemistry & biology* **22**, 745-754, doi:10.1016/j.chembiol.2015.05.010 (2015).

9 Robbel, L., Hoyer, K. M. & Marahiel, M. A. TioS T-TE--a prototypical thioesterase responsible for cyclodimerization of the quinoline- and quinoxaline-type class of chromodepsipeptides. *FEBS J* **276**, 1641-1653, doi:10.1111/j.1742-4658.2009.06897.x (2009).

10 Jaitzig, J., Li, J., Sussmuth, R. D. & Neubauer, P. Reconstituted biosynthesis of the nonribosomal macrolactone antibiotic valinomycin in Escherichia coli. *ACS Synth Biol* **3**, 432-438, doi:10.1021/sb400082j (2014).

11 Liu, Y., Zheng, T. & Bruner, S. D. Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chemistry & biology* **18**, 1482-1488, doi:10.1016/j.chembiol.2011.09.018 (2011).

12 Trauger, J. W., Kohli, R. M. & Walsh, C. T. Cyclization of backbone-substituted peptides catalyzed by the thioesterase domain from the tyrocidine nonribosomal peptide synthetase. *Biochemistry* **40**, 7092-7098 (2001).

13 Aggarwal, A. *et al.* Development of a Novel Lead that Targets M. tuberculosis Polyketide Synthase 13. *Cell* **170**, 249-259 e225, doi:10.1016/j.cell.2017.06.025 (2017).

14 Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat Chem Biol* **4**, 232-234, doi:10.1038/nchembio.73 (2008).

15 Rogerson, D. T. *et al.* Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat Chem Biol* **11**, 496-503, doi:10.1038/nchembio.1823 (2015).

16 Zhang, M. S. *et al.* Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. *Nat Methods* **14**, 729-736, doi:10.1038/nmeth.4302 (2017).

17 McGall, G. H. *et al.* The efficiency of light-directed synthesis of DNA arrays on glass substrates. *Journal of the American Chemical Society* **119**, 5081-5090, doi:DOI 10.1021/ja964427a (1997).

18 Nguyen, D. P. *et al.* Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *Journal of the American Chemical Society* **136**, 2240-2243, doi:10.1021/ja412191m (2014).

19 Pendrak, I., Wittrock, R. & Kingsbury, W. D. Synthesis and Anti-Hsv Activity of Methylenedioxy Mappicine Ketone Analogs. *J Org Chem* **60**, 2912-2915, doi:DOI 10.1021/jo00114a050 (1995).

20 Mayer, S. C., Ramanjulu, J., Vera, M. D., Pfizenmayer, A. J. & Joullie, M. M. Synthesis of New Didemnin B Analogs for Investigations of Structure/Biological Activity Relationships. *J Org Chem* **59**, 5192-5205, doi:10.1021/jo00097a022 (1994).

21 Faure, S. *et al.* Asymmetric intramolecular [2+2] photocycloadditions: alpha- and beta-hydroxy acids as chiral tether groups. *J Org Chem* **67**, 1061-1070, doi:10.1021/jo001631e (2002).

22 Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica. Section D, Biological crystallography* **67**, 271-281, doi:10.1107/S0907444910048675 (2011).

23 Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta crystallographica. Section D, Structural biology* **74**, 85-97, doi:10.1107/S2059798317017235 (2018).

24 Evans, P. Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography* **62**, 72-82, doi:10.1107/S0907444905036693 (2006).

25 McCoy, A. J. *et al.* Phaser crystallographic software. *Journal of applied crystallography* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).

26 Bunkoczi, G. & Read, R. J. Improvement of molecular-replacement models with Sculptor. *Acta crystallographica. Section D, Biological crystallography* **67**, 303-312, doi:10.1107/S0907444910051218 (2011).

27 Tanovic, A., Samel, S. A., Essen, L. O. & Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659-663, doi:10.1126/science.1159850 (2008).

28 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

29 Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta crystallographica. Section D, Biological crystallography* **64**, 61-69, doi:10.1107/S090744490705024X (2008).

30 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

31 Bond, C. S. TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* **19**, 311-312 (2003).

32 Long, F. *et al.* AceDRG: a stereochemical description generator for ligands. *Acta crystallographica. Section D, Structural biology* **73**, 112-122, doi:10.1107/S2059798317000067 (2017).

33 Vagin, A. A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta crystallographica. Section D, Biological crystallography* **60**, 2184-2195, doi:10.1107/S0907444904023510 (2004).

34 Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **67**, 355-367, doi:10.1107/S0907444911001314 (2011).

35 Korman, T. P. *et al.* Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 6246-6251, doi:10.1073/pnas.0913531107 (2010).

36 Gehret, J. J. *et al.* Terminal alkene formation by the thioesterase of curacin A biosynthesis: structure of a decarboxylating thioesterase. *J Biol Chem* **286**, 14445-14454, doi:10.1074/jbc.M110.214635 (2011).

37 Tsai, S. C. *et al.* Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 14808-14813, doi:10.1073/pnas.011399198 (2001).

38 Tsai, S. C., Lu, H., Cane, D. E., Khosla, C. & Stroud, R. M. Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry* **41**, 12598-12606 (2002).

39 Argyropoulos, P. *et al.* Towards a characterization of the structural determinants of specificity in the macrocyclizing thioesterase for deoxyerythronolide B biosynthesis. *Biochim Biophys Acta* **1860**, 486-497, doi:10.1016/j.bbagen.2015.11.007 (2016).

40 Frueh, D. P. *et al.* Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* **454**, 903-906, doi:10.1038/nature07162 (2008).

41 Samel, S. A., Wagner, B., Marahiel, M. A. & Essen, L. O. The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide. *Journal of molecular biology* **359**, 876-889, doi:10.1016/j.jmb.2006.03.062 (2006).

42 Giraldes, J. W. *et al.* Structural and mechanistic insights into polyketide macrolactonization from polyketide-based affinity labels. *Nat Chem Biol* **2**, 531-536, doi:10.1038/nchembio822 (2006).

43 Akey, D. L. *et al.* Structural basis for macrolactonization by the pikromycin thioesterase. *Nat Chem Biol* **2**, 537-542, doi:10.1038/nchembio824 (2006).

44 Whicher, J. R. *et al.* Structure and function of the RedJ protein, a thioesterase from the prodiginine biosynthetic pathway in Streptomyces coelicolor. *J Biol Chem* **286**, 22558-22569, doi:10.1074/jbc.M110.213512 (2011).

45 Bruner, S. D. *et al.* Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* **10**, 301-310, doi:10.1016/s0969-2126(02)00716-5 (2002).

46 Koglin, A. *et al.* Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature* **454**, 907-911, doi:10.1038/nature07161 (2008).

47 Tseng, C. C. *et al.* Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase. *Biochemistry* **41**, 13350-13359 (2002).

48 Scaglione, J. B. *et al.* Biochemical and structural characterization of the tautomycetin thioesterase: analysis of a stereoselective polyketide hydrolase. *Angew Chem Int Ed Engl* **49**, 5726-5730, doi:10.1002/anie.201000032 (2010).

49 Drake, E. J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235-238, doi:10.1038/nature16163 (2016).

50 Gavalda, S. *et al.* The polyketide synthase Pks13 catalyzes a novel mechanism of lipid transfer in mycobacteria. *Chemistry & biology* **21**, 1660-1669, doi:10.1016/j.chembiol.2014.10.011 (2014).

51 Guntaka, N. S., Healy, A. R., Crawford, J. M., Herzon, S. B. & Bruner, S. D. Structure and Functional Analysis of ClbQ, an Unusual Intermediate-Releasing Thioesterase from the Colibactin Biosynthetic Pathway. *ACS Chem Biol* **12**, 2598-2608, doi:10.1021/acschembio.7b00479 (2017).