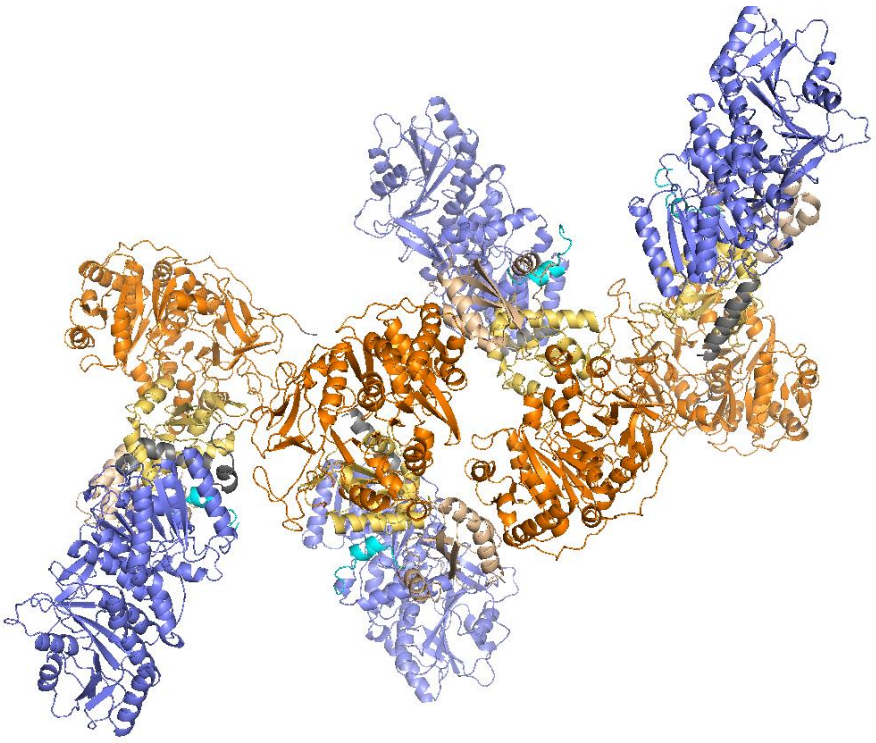
Structural Insight into a Ketoreductase Domain-Containing Nonribosomal Peptide Synthetase Initiation Module from *Bacillus stratosphericus*

Clarisse Chiche-Lapierre
Department of Biochemistry
McGill University, Montréal
August 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science.



“We delight in the beauty of the butterfly, but rarely admit the changes it has gone through to achieve that beauty.”

-Maya Angelou

Abstract

The fascinating microscopic world of bacteria and fungi is rich in small bioactive compounds produced by the organisms in their constant war of survival. Among this treasure trove of compounds were found many of the pharmaceutical compounds invaluable to humankind. In order to better understand their synthesis and work towards the bioengineering of novel natural products, more knowledge and insight must be assembled on the elegant and powerful molecular machines producing them.

One such class of compounds is the nonribosomal peptides (NRPs) produced by massive multimodular enzymes called nonribosomal peptide synthetases (NRPSs). These machines consist of a large number of catalytic domains operating in modules, where each module is responsible for the addition of one amino acid building block in the growing peptide chain of the final product. The adenylation domains (A domains) are responsible for selecting one specific residue and loading it onto the prosthetic arm of the peptidyl carrier protein domains (PCP domains) shuttling the growing chain along the NRPS assembly line. One such NRPS is found in *Bacillus stratosphericus* and is responsible for the production of a compound similar to valinomycin and cereulide, two very powerful toxins. It consists of a number of domains, organized in modules, on two different protein chains. The initial module contains an atypical domain, the ketoreductase domain (Kr domain). To shed light on how the structural elements of the synthetase work together, we embarked on solving the protein structure of parts of this initiation module, the A-Kr-PCP, through X-ray crystallography.

The crystallography of A-Kr-PCP proved to be very challenging, but three factors made it possible for us to solve: controlling the presence of the prosthetic arm attached to the PCP domain and what substrate is bound to it; the use of advanced techniques of cryoprotection and dehydration of the crystals, and finally, the process of iterative omit-map sharpening combined with four-fold non-crystallographic symmetry averaging. This work resulted in two individual protein structures: the

StsA_A-KrT initiation module solved to a resolution of 3.4-3.9Å and the StsA A domain solved to a resolution of 2.3 Å. The structure of StsA_AKrT contains three distinctly novel features. It contains the first structure of an NRPS Kr domain. It is also the first structure of a Kr-containing NRPS module, allowing us to visualize how the Kr domain is embedded into the synthetase. Finally, the structure contains the first structure of a ketoacid-selecting A domain. Contrary to expectations, the A domain is uninterrupted by the Kr domain, and the two domains show very little surface contact. The structure also revealed a surprising feature: a novel pseudo A_{sub} (pA_{sub}) domain, interrupted by the A-Kr di-domain, dividing it into two parts located over 1000 residues apart in the primary sequence, brought together through a 39-amino-acid long strand swap between dimeric partners, seemingly inducing dimerization. The dimer:monomer ratio can be controlled to some extent by controlling the length of the linker region involved in the domain swap. The pA_{sub} domain has the same structure as the canonical A_{sub} domain but holds no catalytic function. The module is able to fold properly and adenylate its substrate without this curious subdomain.

The keto-acid-selecting A domain has the same overall structure as amino-acid-selecting ones, but some of the highly conserved motifs and substrate binding residues differ. The α-keto substrate seems to be accommodated by the mutation in StsA of a proline to a methionine residue, whose backbone carbonyl may rotate 40° to interact with the α-carbonyl group of the substrate through a reciprocal carbonyl-carbonyl interaction. We showed that the reverse mutation allowed for an amino acid selecting A domain from linear gramicidin synthetase A to select and adenylate the α-keto version of its cognate substrate. The Kr domain has the same overall structure as those found in other types of synthesis systems, and seem to function through a similar mechanism. Finally, we propose a design for inhibitors that would trap the Kr domain in the substrate binding and ketoreduction state.

Résumé

Les bactéries, champignons et autres organismes microscopiques vivent au sein d'un tout autre monde invisible à nos yeux. Cet environnement fascinant regorge de composés biologiques produits par ces micro-organismes pour assurer leur survie dans ce milieu compétitif et hostile. Parmi cette multitude de produits naturels se trouvent de nombreux composants aux propriétés médicales étonnantes et essentielles au bon fonctionnement de notre société actuelle. Cependant, nous devons rassembler plus de savoir fondamental sur le fonctionnement des machines moléculaires qui les produisent si nous voulons pouvoir appliquer leur logique de synthèse à la production de nouveaux composés biologiques tels que des antibiotiques, des immunosuppresseurs ou des anti-tumoraux.

Dans cette étude, nous apportons une attention particulière à un certain type de produits naturels appelés peptides non-ribosomiques (PNR), produits par d'impressionnantes machines de synthèse appelées enzymes de synthèse des peptides non-ribosomiques (ESPNR) que l'on trouve chez les micro-organismes. Les ESPNR sont composées de nombreux modules, eux-mêmes divisés en plusieurs domaines, où chaque module est chargé d'incorporer un acide aminé à la chaine peptidique naissante. Le domaine d'adénylation (domaine A) sélectionne un substrat et l'active par adénylation avant de le transférer au bras prothétique de la protéine porteuse de peptide (domaine PPP), qui se chargera de transporter le peptide le long de la chaine d'assemblage où de nouveaux acides aminés s'ajoutent à la chaine peptidique grandissante. L'une de ces ESPNR produite par la bactérie *Bacillus Stratosphericus* synthétise un produit naturel peu étudié, similaire à la cereulide ou à la valinomycine. Cet ESPNR, appelée stratospherulide synthétase (Sts), est composée de quatre modules qui s'étendent sur deux chaines protéiques. Le premier module contient un domaine atypique intéressant: le domaine réducteur de cétone (domaine RC). Afin de comprendre le fonctionnement d'un module d'ESPNR contenant un domaine RC, nous nous proposons d'élucider la structure du module d'initiation A-RC-PPP de la stratospherulide synthétase par cristallographie à rayon-X.

La cristallographie du module A-RC-PPP s'avéra être extrêmement difficile. Cependant, trois facteurs clefs ont rendu possible l'élucidation de la structure : le contrôle du bras prothétique du

domaine PCP et le chargement du substrat; le développement de techniques de cryo-protection et déshydratation avancées; enfin, l'utilisation d'un processus itératif d'ombrage de la carte en faisant une moyenne des partenaires de symétrie non-cristallographique.

De part cette étude ont été produites deux structures : la structure A-RC-PPP du module d'initiation de la stratospherulide synthétase à une résolution de 3,4-3,9Å, ainsi que la structure du domaine A du module d'initiation à une résolution de 2,3Å. La structure A-RC-PPP nous permet de visualiser pour la première fois trois composants intéressants : il s'agit de la première structure d'un module d'ESPNR contenant un domaine RC, permettant de visualiser l'accommodation de ce domaine atypique au sein du complexe protéique; cette structure nous permet également de visualiser pour la première fois la structure d'un domaine A capable de sélectionner des acides cétonés ainsi que la structure d'un domaine CR d'ESPNR. Notre structure révèle aussi la présence d'un domaine inattendu : le domaine pseudo-A_{sub}, interrompu par le bi-domaine A-CR qui le divise en deux parties séparées par 1112 acides aminées en termes de séquence, rassemblé par un échange de chaine longue de 39 acides aminées entre partenaires dimériques, ce qui semble induire la dimérisation de l'ESPNR. Le domaine pseudo-A_{sub} présente la même structure que le domaine A_{sub}, mais il ne semble pas avoir de fonction catalytique et le module est capable d'assurer la fonction d'adénylation même si on l'excise.

Le domaine A, capable de sélectionner un acide cétoné, présente la même structure générale que les domaines A plus communs qui sélectionnent spécifiquement les acides aminés, à l'exception de certain motifs et résidus important pour la sélection du substrat : le site actif du domaine A semble pouvoir s'accommoder à un acide cétoné de par la mutation d'une proline à une méthionine. Cette mutation semble altérer la position de la charpente protéique afin d'interagir de manière positive avec le substrat par interaction carbonyle-carbonyle réciproque. De plus, la mutation inverse a permis au premier domaine A de la protéine de synthèse de la gramicidine linéaire A, spécifique pour les acides aminés, de sélectionner la version cétonée de son substrat naturel. Enfin, nous proposons la synthèse de deux inhibiteurs capables de capturer le complexe protéique d'ESPNR en phase de liaison des substrats du domaine CR, ainsi qu'en phase de réduction du groupe cétoné.

Acknowledgements

To my supervisor Dr. Thomas Martin Schmeing, I thank you for welcoming me into your lab, for your enthusiasm and your respect for ethics. I can guess that having me as a student was not easy, so I thank you for supporting me through sickness and personal tragedies, for showing me what I could be and for helping me grow. To Dr. Diego Arthuro Alonzo, who had a vision and never let any amount of "good luck" tear his spirit down. I thank you for your creativity, sensitivity, kindness, focus, and relentlessness. Your example has shaped me as a scientist and as a person, for that I am eternally grateful. To Dr. Michael Tarry, I thank you for teaching me that success never comes without failure, and not to take life too seriously. To Claudia Alonso, I thank you for having been the best student I could have hoped for, and for being there when I couldn't. I hope that you will get the opportunity that your talents deserve, and that you find success, whatever that may be. To Dr. Frederik Teilfeldt Hansen, I thank you for supporting me through the darkest and most challenging times of my life. There are no words to express the admiration and respect that I have for you. I thank you for your precious advice and scientific insight, for all the overnights you set for me, all the cultures you harvested, and all the things you did for me when I couldn't, for your support through all the late nights and long weekends, and last but not least, for editing this thesis. To Dr. Janice Reimer, I thank you for all the aminoacyl-coA, but mostly I want to thank you for your strength, your presence and your courage, you are an inspiration to any young girl aspiring to become a scientist. To Dr Kristjan Bloudoff, Ingrid Harb, Camille Fortinez, Maximillian Eivaskhani, Itai Sharon, Angelos Pistofidis, Jessie Jiang, Sean Jiang, Sarah Roffeis, Yael Ripstein, David Dai, Connor Harrigan and to all the past and current members of the Schmeing lab, I thank you for the helpful scientific discussion and advice, and I thank you for your hard work and dedication. To Jessica del Castillo, thank you for turning life into a musical; you are all that is best about humanity. To Yael Ripstein, your overflowing energy resonates with positivity and brings out the best in others, I thank you for reminding me why I am here. Your spirit is one of a kind, don't let the world drag you down. I thank Drs. Alba Guarné and James Gleason for serving on my Research Advisory Committee and contributing to my project through constructive and helpful advice. I thank John Colucci

and the chemists of Zamboni Chemical Solutions for their wonderful work on the adenosine vinylsulfonamide inhibitor and precious advice. I thank Dr. Alexei Gorelik for his invaluable help and dedication. I thank the CMCF team at the Canadian Light Source for assisting with data collection.

I am indebted to the generous sources of funding that have supported this work: The Rose Mamelak Johnstone award, the Graduate Excellence Fellowship for Merit and the GRASP Graduate Student Award.

To Dr. Valérie Lapierre, I thank you for being a fierce mama bear. Your successes have shown me what I could do, and your setbacks have shown the face of courage, true strength and true love. You are my hero. To Dr. Val Chiche-Lapierre, I thank you for shaping me into the person I am today, for being one of the few people on this earth who can understand me. I am so proud to have you as a brother. To Dr. Dan Chiche, I thank you for bringing me to my new life, and for our future scientific collaborations. To Jesse Tiberghien-Lapierre, Liam Chiche and Kaël Chiche, thank you for being kids and knowing what is important in life. To Dr. Serge Poiraudeau, I thank you for believing in me. To Dr. Alain Alterescu; as a fighter never fights without his coach, I would not be presenting this work today without your help and support; thank you for being in my corner. To Tri Nguyen, I thank you for giving me my life back, twice, and for giving me a family. I thank Sandra Claro, Jenner Junior, Leslie Castro, Kim Lachance, Samuel Chan-Taw, Dr Achyuth Chandra, Dr. Nida Latif, Guillaume Allaux, Naomi Loder, Sophie Kassel, Jess Goldman, Kaity McGee, Julien Morizio, Clara Petit, Clair Lyle, Anne Dorward, Dr Deanna Fong, Simon Mathieu, Ala Eddine, Ali Beydoun, Youcef Benabdallah, John Bossi, Terrie Schauer and all of Carlson Gracie Montréal Team, with a special thanks to my Jiu-Jitsu wife Murtaza Kazi.

Contributions

Dr. Thomas Martin Schmeing, Dr. Diego A. Alonzo and I designed the study. Dr. Diego A. Alonzo, Dr. Michael Tarry, Claudia Alonso and I performed the biochemical experiments. Dr. Diego A. Alonzo and I performed the crystallization, structure determination and refinement. Dr. Jimin Wang performed the structure determination and refinement of the A-Kr-PCP structure. Dr Frederik Teilfeldt Hansen edited this thesis.

Contributions to data acquisition by chapter:

Chapter 1: Chapter 1 was written by me and edited by Frederik Teilfeldt Hansen.

Chapter 2: Chapter 2 was written by me and edited by Frederik Teilfeldt Hansen. I performed the crystallization of A-Kr; I performed the crystallization of A-Kr-PCP together with Diego Alonzo; I performed the structure elucidation and analysis together with Jimin Wang, Thomas Martin Schmeing and Diego Alonzo. The iterative omit-map sharpening and non-crystallographic symmetry averaging processes were conceptualized and performed by Jimin Wang. Claudia Alonzo prepared the enzymes for synthesis of aminoacyl-coenzyme A, and assisted with cloning, expression, purification and crystallization of the enzymes mentioned in this chapter.

Chapter 3: Chapter 3 was written by me and edited by Frederik Teilfeldt Hansen. I performed to crystallization of Sts A domain together with Diego Alonzo. The adenylation assays for side chain specificity of the A domain were performed by Diego Alonzo. I performed the cloning and biochemical assays for the study of specificity of the alpha-substituent and the crystallization of LgrA_FA_P507M. Michael Tarry performed cloning and biochemical assays for the linker deletion constructs. I performed the cloning and biochemical assays for the pA_{sub} deletion constructs together with Diego Alonzo. I designed the Br-NADPH inhibitor and the C^1-NADPH inhibitor. Claudia Alonzo prepared the enzymes for synthesis of Br-NADPH inhibitor and assisted with cloning, expression, purification and crystallization of the enzymes mentioned in this chapter.

Chapter 4: Chapter 4 was written by me and edited by Frederik Teilfeldt Hansen.

Table of Figures

Table of Tables

List of Abbreviations

A domain	Adenylation domain	NIS	Nonribosomal peptide synthetase independent siderophore
ACP domain	Acyl carrier protein domain	N-lobe	N-terminal lobe
AMP	Adenosine monophosphate	NMAT	Nicotinamide monoclucleotide adenyltransferase
AMR	Antimicrobial resistance	NMN	Nicotinamide mononucleotide
AT domain	Acyl transferase domain	NMNAT	Nicotinamide mononucleotide adenyltransferase
ATP	Adenosine triphosphate	NPRT	Nicotinamide phosphoribosyl transferase
AVSI	Adenosine vinylsufonamide inhibitor	NRP	Nonribosomal peptide
Br-NaMN	4-bromo-nicotinic acid mononucleotide	NRPS	Nonribosomal peptide synthetase
C domain	Condensation domain	Ox domain	Oxidization domain
CBP	Calmodulin binding protein	PAGE	Polyacrylamide gel electrophoresis

CDC	Center for disease control	PanK	Pantothenate kinase
CIHR	Canadian institute of health and research	pAsub domain	Pseudo-adenylation subdomain
C-lobe	C-terminal lobe	PCP domain	Peptidyl carrier protein domain
CoA	Coenzyme A	PDB	Protein databank
Ct domain	Fungal terminal condensation domain	Pep/PK	Phosphoenolpyruvate/pyruvate kinase
Cy domain	Heterocylization domain	PK	Polyketide
dATP	Deoxyadenosine triphosphate	PKS	Polyketide synthase
DH domain	Dehydratase domain	PMSF	Phenylmethanesulfonyl fluoride
DPCK	Dephosphocoenzyme A kinase	PNP	Purine nucleoside phosphorylase
dTTP	Deoxythymidine triphosphate	PPAT	Phosphopantetheine adenylyl transferase
E domain	Epimerization domain	PPC-DC	Phosphopantothenoylcystein decarboxylase
EDTA	Ethylenediaminetetraacetic acid	R	Reductase domain
EM	Electron microscopy	SAM	S-adenylmethionine
ER domain	Enoyl reductase domain	SDS	Sodium dodecyl sulfate
GDH	Glucose dehydrogenase	SEC	Size exclusion chromatography
GT	Glycosyltransferase	TB	Terrific broth
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	TE domain	Thioesterase domain
HPLC	High pressure liquid chromatography	TECEP	Tris-(2-carboxyethyl)-phosphine
IPTG	Isopropyl-β-D-1-thiogalactopyranoside	TEV	Tobacco etch virus protease
KR domain	Ketoreductase domain	Vitamin B5	Pantothenic acid
KS domain	Ketosynthase domain	WHO	World Health Organization
LB	Lysogeny broth	α-HA	α-hydroxy acid
LIC	Ligation independent cloning	α-HIC	D-α-hydroxyl-isocaproic acid
MT domain	Methyltransferase domain	α-KIC	α-ketoisocaproic acid
NADPH NADP+	Nicotinamide adenine dinucleotide phosphate	α-KIL	α-ketoisoleucin
NADPH NADP+	Nicotinamide adenine dinucleotide phosphate	α-KIV	α-ketoisovaleric acid
NADR	NAD-repressor enzyme	B-Me	β-mercaptoethanol
NCS	Non-crystallographic symmetry		

Contents

Chapter I – Introduction

In the simplest terms, living things are made of cells, which in turn consist of lipids, proteins, sugars and water. Lipids provide us with energy, gives the cell its globular structure and protect it by forming a membrane around it, much like the skin to a human body. The sugars also provide us with energy, are involved in many cellular processes and most famously are the guardians of all the information needed for the formation, development and survival of any living organism, in the form of DNA. The water is simply the media in which all these components are dwelling, much like the air is to us, animals and plants. If lipids are the external structure, and DNA is the memory, proteins are the hardware that allows this impressive machinery that is the living organism to function. Some proteins are static and fulfil a structural role. Others function as catalysts for various cellular processes; this specific sub-type of proteins are called enzymes. Anything that happens in the cell does so because of proteins. They allow us to breathe, eat, move, think, perceive our environment and respond to it in an appropriate manner. The information encoded in the DNA is translated into proteins, which in turns perform a function that ultimately makes us who we are. The incredible diversity of life we celebrate on earth is only possible because of proteins. They are akin to a pearl necklace, where every pearl is an amino acid. In all cells, an enzyme called the ribosome, assembles all proteins and most of the shorter peptides. Within microorganisms, such as bacteria or fungi, can be found another type of peptide-making enzyme, called nonribosomal peptide synthetases (NRPSs). NRPSs fulfil a very specific role within the microorganism that gives it a survival advantage in its environment. It could be a piece of molecular weaponry that allows the organism to defend itself against other bacteria or fungi and compete for the media they all want to colonize, or it could give them an advantage to find, harvest and process nutrients. If the products of these NRPSs have many functions for the organism that produces it, it has even more application for us humans, in terms of medicine, chemistry or agriculture. For example, the famous antibiotic penicillin, which is estimated to have saved 200 million lives since its discovery by Alexander Fleming in 1928, is produced by an NRPS. In this introduction, we will examine how the NRPS machinery and its components function, and how their potential could be harnessed to help up tackle some of the biggest threats that humanity is facing.

1.1 Nonribosomal Peptide Synthetases

1.1.1 Prevalence of Nonribosomal Peptide Synthetases

Microorganisms often produce various small biologically active peptides that provides them with an evolutionary advantage to multiply and thrive in their environment [1]. These natural products might provide the organism with a form of defense against competing organisms, give them an advantage in colonizing a certain media, or infect a host [1, 2]. These peptides are not produced by the ribosome as most proteins are, but by huge enzymes produced by the ribosome, called nonribosomal peptides synthetases (NRPS) [1]. NRPS are fairly widespread in our micro-environment, with over three thousand gene clusters characterized in more than a thousand organisms [3]. The NRPS family counts some of the largest enzymes characterized to date [3]. Their size can vary from a few hundred kilodaltons to up to several megadaltons, such as the NRPS plu2670, which contains 15 modules and 46 domains on a single chain of 16,367 amino acids [3]. NRPSs have been a subject of growing interest among the scientific community for their ability to produce various natural compounds with diversified and valuable applications (such as antibacterial, antifungal, anti-tumoural or immunosuppressant) with close to a thousand articles published in the last two decades and over 50 protein structures deposited in the protein data bank (PDB) [1, 4].

1.1.2 Structural and Functional Diversity of Nonribosomal Peptides

The diversity of proteins and enzymes giving life on Earth is produced by the ribosome from 20 different amino acids building blocks. In contrast, NRPSs can use over 500 different substrates, including but not limited to, canonical and non-canonical amino acids, L- and D-amino acids, ketoacids, fatty acids of varying lengths, heterocycles or sugars, resulting in a vastly more diverse pool of nonribosomal peptides (NRPs) [5]. In addition, some NRPS have tailoring domains such as ketoreductases, methyltransferases or formyltransferases, that can act in *cis* or in *trans* to modify the substrate and add another layer of diversity [2, 5]. In Figure 1 are shown some examples highlighting the broad functional, structural and chemical diversity of NRPs.

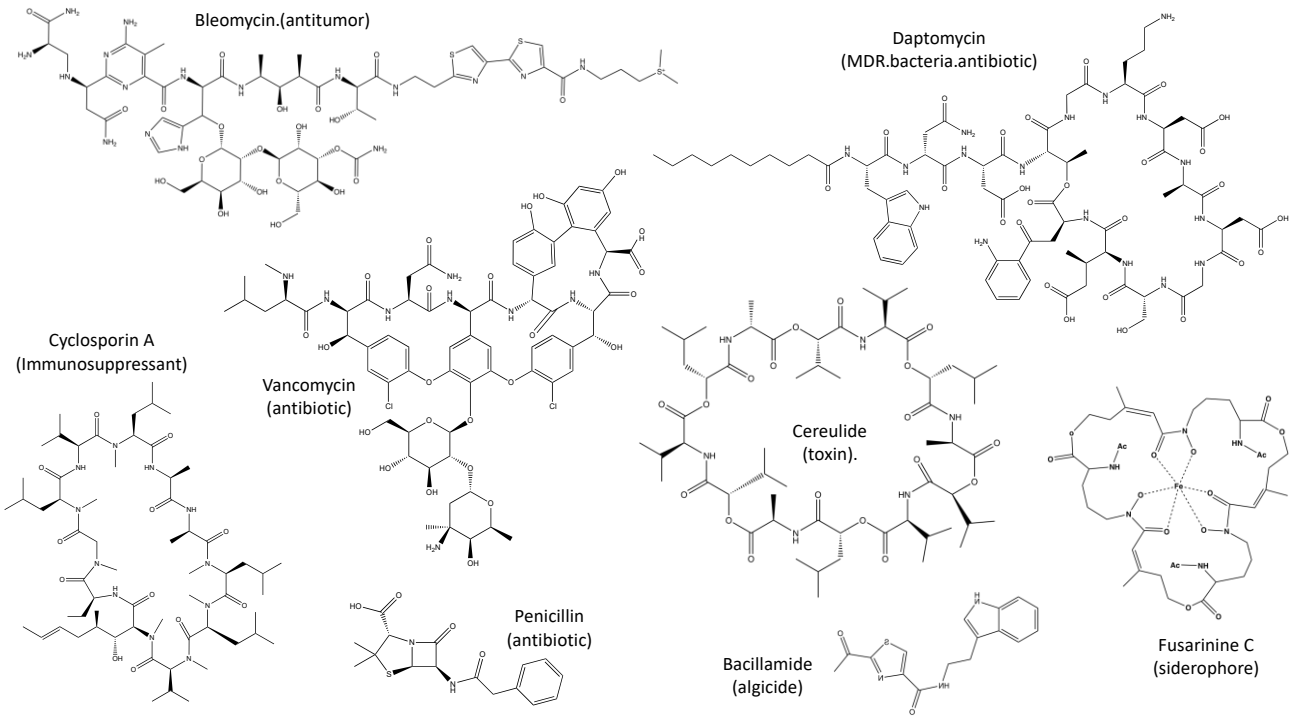


Figure 1: Structural and functional diversity of NRPs. The unusual structure of the common antibiotic penicillin includes a β-lactam ring and the non-canonical amino acid δ-(L-α-amino adipic acid) [1]. The antibiotic daptomycin used in hospitals to fight multi-resistant bacteria is a cyclic compound, containing various non-canonical amino acid and a fatty acid residue [5]. The antibiotic vancomycin is built from a hepta-NRP including non-canonical amino acids hydroxychlorotyrosine, 4-hydroxyphenylglycine and 3,5-dihydroxyphenylglycine, which is later glycosylated in *trans* by tailoring enzymes and condensed into the tri-cyclic biologically active final compound [1]. The emetic compound cereulide is a cyclic NRP which includes L- and D-amino acids as well as L- and D-hydroxy acids using a *cis* ketoreductase tailoring domain, resulting in alternating peptide bonds and ester bonds [6]. The immunosuppressant cyclosporin A is a cyclic compound containing both L- and D-amino acid, including the non-canonical amino acids (4R)-4-[(E)-2-butenyl]-4-methyl-L-threonine and α-amino-buteric acid [5]. The small algicide bacillamide contains a thiazoline ring formed by cyclodehydration of a serine residue [7]. The fungal iron-chelating siderophore fusarinine C contains the non-canonical amino acid ornithine, a very common building block for the synthesis of siderophores found in fungi [8, 9]. The anti-tumor compound bleomycin, used to treat several types of cancer including pediatric lymphoma, contains an imidazole ring, two thiazoline

rings, two sugar adducts, several non-canonical amino acids and various modification such as methylation and acetylation [5].

1.1.3 The Domains of Nonribosomal Peptide Synthetases

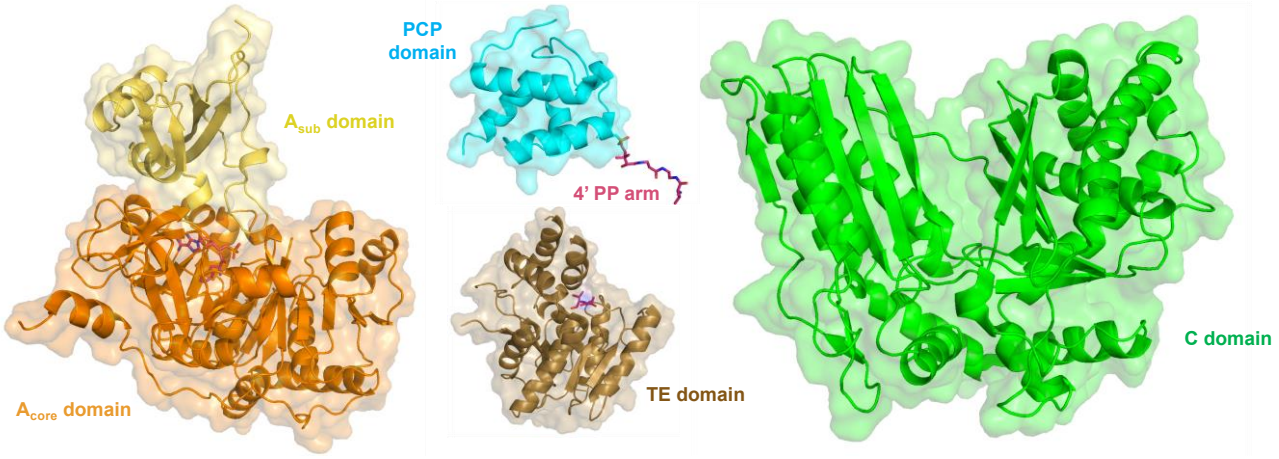


Figure 2: NRPS domains structural overview. From left to right, the first A domain and the first PCP domain of linear gramicidin synthetase, the PCP domain is modified with aminoacyl-4' ppant arm [10], the TE domain of valinomycin synthetase [11] and the C domain VibH [12].

The NRPS assembly line is composed of modules, where each module is responsible for the addition of one residue to the growing NRP chain. Each module can be divided into domains and each domain fulfils a specific function to allow initiation, elongation and termination of the NRP synthesis [2]. The minimal canonical NRPS assembly line contains four key domains: the adenylation domain (A domain), the peptidyl-carrier protein domain (PCP domain), the condensation domain (C domain) and the thioesterase domains (TE domain) [2]. Considering the case of a minimalistic di-module NRPS, the domains are arranged like the wagons of a train in the following order: A-PCP-C-A-PCP-TE. The A domain selects the residue it is specific for out of the pool of substrates available in the cell, catalyzes its activation by adenylation using ATP and subsequently loads the substrate onto the PCP domain by thiolation [13]. The PCP domain is charged with carrying the substrate from domain to domain, from modules to modules and present it to the various active sites via its prosthetic 4' ppant arm [14, 15]. The role of the C domain is to link the two substrates presented by the directly upstream and

downstream PCP domains through an enzymatic condensation reaction [16]. The TE domain is found at the very end of the assembly line; it is responsible for the termination of the synthesis and for releasing the final NRP product by hydrolysis or macrocyclization [17]. In some systems, the TE also catalyzes oligomerization of several product of the assembly line before releasing the substrate [3, 11]. Some systems include additional domains called "tailoring domains", which are either part of the NRPS assembly line, or expressed independently as part of the NRPS gene cluster. They catalyze additional reactions to bring another layer of diversity to the NRP chemical space [18]. An overview of the domains of NRPSs are shown in Figure 2 and Table 1.

Table 1: NRPS domains overview of function and reaction.

Domain	Function	Reaction
A domain	Substrate selection	--
	Substrate adenylation	residue-COOH + ATP → residue-CO-AMP (adenylated residue) + PP_i
	Substrate loading onto the 4' ppant arm of the PCP domain (thiolation)	Residue-CO-AMP + SH-4' ppant-PCP → residue-CO-S-4'ppant-PCP + AMP
PCP	Substrate channeling and presentation	--
C domain	Catalyzing the bond formation between two substrates presented by the PCP domain	$residue_1$-CO-S-4'ppant-PCP_1 + NH_2-$residue_2$-CO-S-4'ppant-PCP_2 → SH-4'ppant-PCP_1 + $residue_1$-CO-NH-$residue_2$-CO-S-4'ppant-PCP_2
TE domain	Termination of synthesis by hydrolysis	PCP-4'ppant-S-CO-NRP + H_2O → NRP-COOH + PCP-4'ppant-SH
	Oligomerization of pre-peptide produced by the NRPS assembly line	PCP-4'ppant-S-CO-NRP_2-NH_2 + TE-O-CO-NRP_1 → PCP-4'ppant-S-CO-NRP_2-NH-CO-NRP_1 + TE-OH
	Macrocyclization	TE-O-CO-NRP-NH_2 → TE-OH + cyclo(-CO- NRP-NH-)

1.1.4 The Adenylation Domain

The A domain is responsible for selecting a specific substrate, activating it by adenylation, and loading the substrate onto the 4' ppant arm of the PCP domain by thiolation. Adenylation enzymes belong to a broad superfamily of proteins involved in various cellular process such as ribosomal and nonribosomal peptide synthesis, enzyme regulation and metabolism [19]. Their role is to adenylate a substrate, forming a high energy bond susceptible to nucleophilic attack to produce esters, thioesters or amides which are important building blocks for the synthesis of natural products or secondary metabolites [19]. Even if adenylation enzymes all perform the same reaction, they vary greatly in terms of primary sequence, and can be divided into different classes according to their tertiary structures [19]. NRPS A domains are class I adenylating enzymes, characterized by the presence of two sub-domains linked by a hinge region. Class I also includes oxidoreductases and acyl/aryl-CoA synthetases [20]. tRNA synthetases are class II adenylating enzymes, and bear little structural homology with the NRPS A domain, despite fulfilling the same function in peptide synthesis [19]. The NRPS-independent siderophore (NIS) synthetases make up the class III of adenylating enzymes [19].

The NRPS A domain is approximately 60kDa, consists of a 48kDa A_{core} and a 12kDa A_{sub} subdomains, structurally very similar to the luciferase and the acyl-CoA ligase [13]. The core motifs of A domains are highly conserved in sequence among NRPSs, which allow their identification from a newly sequenced genome. About 25 NRPS A domain structures have been deposited in the PDB to date, but the phenylalanine-activating A-domain of linear gramicidin S synthetase from *Aneurinibacillus migulanus* (PDB:1AMU) [13] was the first. It is well characterized and often used as a reference. The A_{core} conserved structure consists of two β-sheets flanked on either side and separated by α-helices, forming a five-layered alternating $\alpha\beta\alpha\beta\alpha$ pattern (one 6-membered β-sheet, one 8-membered β-sheet and 12 α-helices) with an additional lobe composed of a distorted β-barrel (6 β-stands and 2 helixes) [13]. The structure and conserved motifs of the A domain is shown in Figure 3.

A

Motif	Consensus Sequence
A1	LTYxEL
A2	LKAGxAYVPID
A3	LAYxxYTSTGTTGxPKG
A4	FDxS
A5	NxYGPTE
A6	GELxIxGxGLARGYW
A7	YKTGDQ
A8	GRxDxQVKIRGxRVELEEVE
A9	LPXYMIP
A10	NGKIDR

B

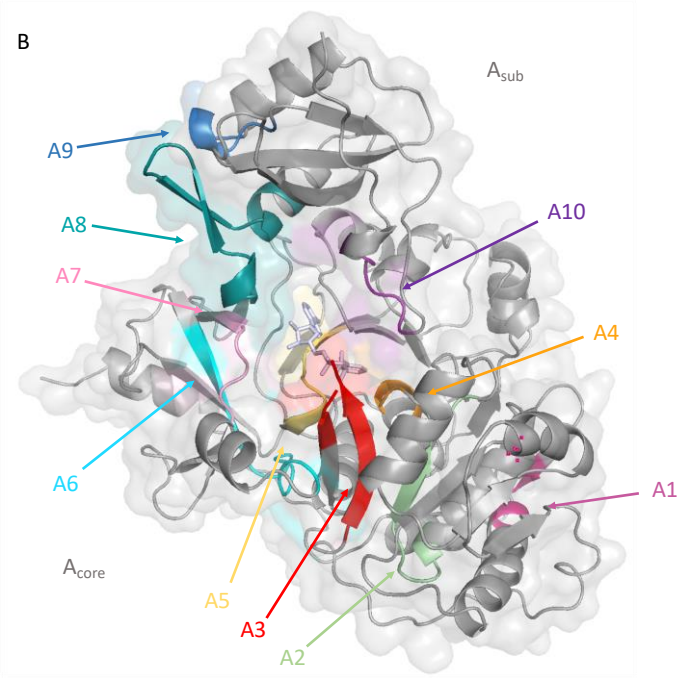


Figure 3: The conserved motifs of the A domain (a) A domain motifs and consensus sequence. (b) Representation of the motifs on 1AMU, phenylalanine and AMP shown in grey [13, 21]. Motifs A3, A4, A5, A7 and A8 are involved in substrate binding together with motif A10. A8 is also the hinge region between the A_{core} and the A_{sub}, involved in dynamic rearrangements of the A_{sub} domain and overall conformation throughout the cycle; A10 is the catalytic loop of the A_{sub} containing the catalytic lysine.

Stachelhaus et al. identified 10 residues spread out across motifs A3 to A5 and A10, which enables the prediction of the substrate specificity of the A domain [13, 21]. As shown in Figure 4, Asp235 in motif A4 coordinates the α-amino group and Lys517 in A10 coordinates the carboxyl group of the residue, together handling the backbone moiety of the residue. As for the side chain of the selected residue, it is surrounded by residues Ala236, Ile330 and Cys331 on one side, and residues Ala322, Ala301, Ile299 and Thr278 on the other side, with Trp239 in between at the bottom of the cleft. The composition of these 10 residues makes up specific motifs, called Stachelhaus codons [21, 22]. By creating phylogenetic trees using only these 10 amino acids from a large number of A domains, Stachelhaus observed that the A domains clustered into sub-groups presenting the same residue

specificity. In other words, the composition of these 10 specific amino acids can predict the side chain specificity of the A domain. In this way was established the Stachelhaus code; it is analogous to the genetic code in ribosomal synthesis, assigning a specific residue to each Stachelhaus codon. This code is based upon the assumption of the structural homology model: 3D alignment of a large number of A domain shows that the position of the backbone is highly conserved, with paired α-carbon being no more than 3Å r.m.s.d. apart between A domain models. As the side chain specificity is not conferred by the backbone of the catalytic pocket residues, the side chains of the catalytic pocket residues must be responsible for it [22].

The active site of the A domain along with its catalytic residues is shown in Figure 4.

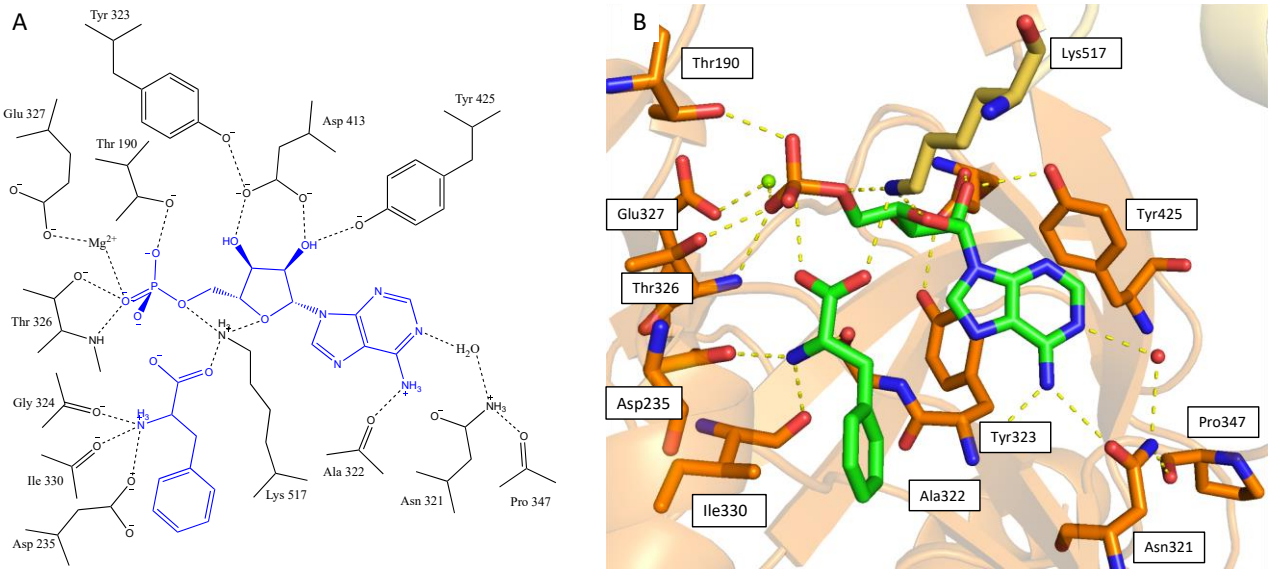


Figure 4: The A domain active site: substrate binding and catalytic residues (a) schematics representation of the hydrogen bonding network and interactions between substrates and active site residues. (b) The active site of the phenylalanine-activating A-domain of linear gramicidin S synthetase from *Aneurinibacillus migulanus*, in complex with substrate L-Phenylalanine, Mg^{2+} and AMP (PDB:1AMU).

The A domain as part of a canonical NRPS module has been observed in three different conformations. The "open" conformation that allow for substrate binding, the "closed" conformation

that allows for the adenylation reaction to occur after substrate binding, and the "thiolation" conformation that allows access to the active site for the 4' ppant arm of the PCP domain and the thiolation reaction to occur [10, 23]. As discussed later, other conformations have also been observed in non-canonical NRPS modules to accommodate for the presence of additional domains [10].

The adenylation cycle starts in the "open" conformation. The substrate and ATP bind the active site. The A_{sub} domain then rotates 30° to achieve the "closed" conformation [24]. Conserved active site residues coordinate the carboxylic moiety of the substrate and the α-phosphate of ATP, stabilizing them in an activated conformation to allow for the nucleophilic substitution reaction to occur: the carboxylic oxygen of the substrate attacks the phosphorus of the α-phosphate of ATP, releasing the pyrophosphate which acts as a leaving group [25]. The mechanism of adenylation is shown on Figure 5.

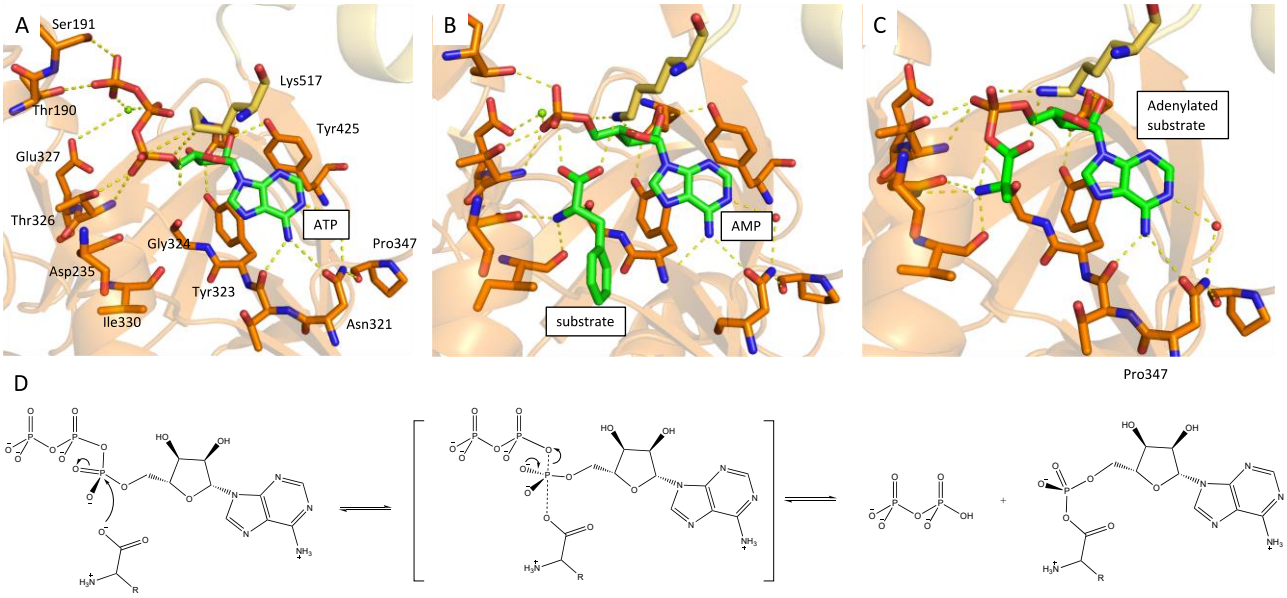


Figure 5: Mechanism of enzymatic adenylation catalyzed by the NRPS A domain. (a) Active site of the D-alanine-activating A domain from *Bacillus cereus* D-alanyl carrier protein ligase (DltA) in complex with ATP and Mg^{2+} (PDB:3FCC) [26]. (b) The active site of the phenylalanine-activating A-domain of linear gramicidin S synthetase from *Aneurinibacillus migulanus* in complex with substrate L-phenylalanine, Mg^{2+} and AMP (PDB:1AMU) [13]. (c) Active site of the D-alanine-activating A domain from *Bacillus cereus*

D-alanyl carrier protein ligase (DltA) in complex with the adenylated D-alanine-AMP product of the adenylation reaction (PDB:3DHV) [27]. (d) Mechanism of enzymatic adenylation by nucleophilic substitution.

To ease comparison, the amino acids are identified with a nomenclature based on 1AMU. As shown in Figure 5, Asp235, Gly324, Ile330 coordinate the substrate by engaging in hydrogen bonding with the amino group, while Thr190, Thr326 and Glu327 (by the intermediate of a Mg^{2+} ion) coordinate the α-phosphate of ATP [13]. In addition, the catalytic Lys517 from the A_{sub} domain engages in hydrogen bonding with the ribose oxygen, the α-phosphate oxygen and the carbonyl group of the substrate simultaneously, as shown in Figure 5. The joint action of these residues brings the active oxygen of the carboxylate moiety of the substrate to an appropriate distance of 3Å to the α-phosphate for the nucleophilic substitution to occur, as shown in Figure 5. Following the mechanism for nucleophilic substitution, the active oxygen of the carboxylic moiety acts as a nucleophile and attacks the non-hindered side of the α-phosphate opposite from the pyrophosphate, which acts as a leaving group, in an in-line mechanism [28]. The resulting pentavalent phosphorus intermediate has the conformation of a triangular-based bi-pyramid, with the phosphorus at the center of the base and the apical position occupied by the nucleophile and the leaving group. The pentavalent phosphorus intermediate is stabilized by Thr326 backbone amino group, Lys517 and Glu327 through a Mg^{2+} ion pulling electron density away from the oxygen substituents, while the electron density is delocalized throughout the adenosyl moiety of ATP stabilized by Asp413, Tyr425, Ala322 and Asn321 through a water molecule, pulling electron density away from the phosphorus [29]. The hydroxyl group of Thr326 may act as a hydrogen donor and stabilize the alcohol intermediate. The penta-covalent phosphorus intermediate relaxes back into tetrahedral form with an inversion of configuration, resulting in the release of the pyrophosphate and formation of the adenylated substrate.

Following adenylation, the A_{sub} domain rotates 140° to allow for the PCP domain and its 4' ppant arm to access the A domain's active site, in what we call the "thiolation" conformation [23]. The mechanism of thiolation is shown in Figure 6.

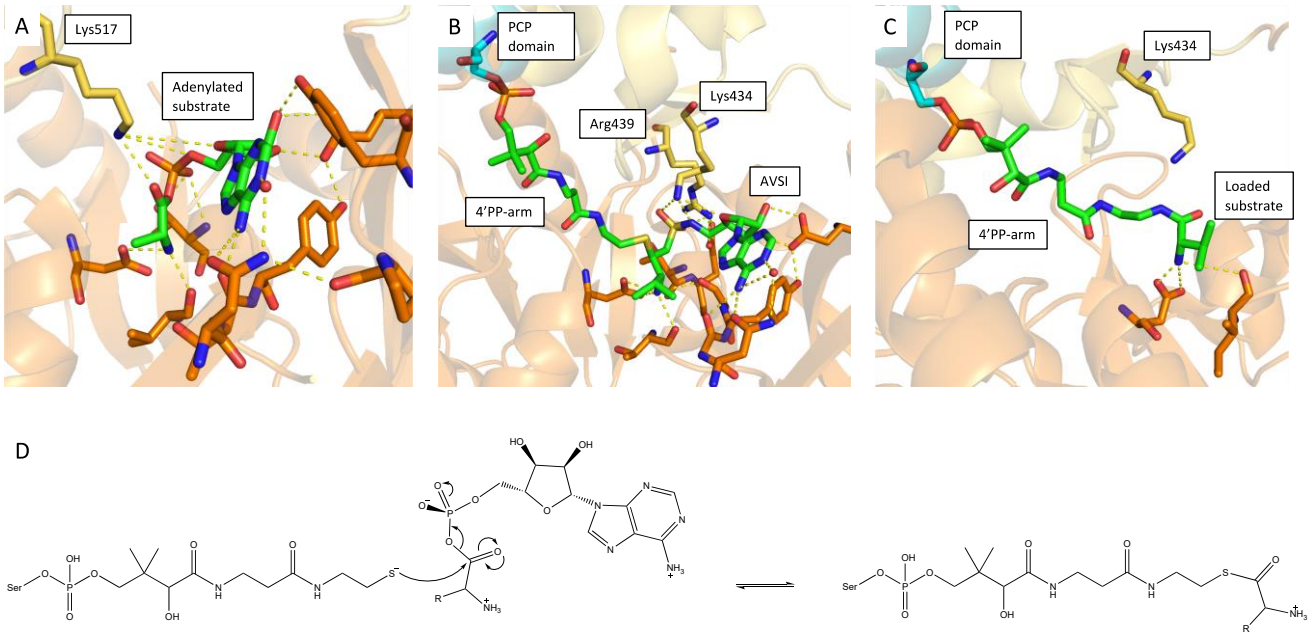


Figure 6: Mechanism of enzymatic thiolation and substrate loading onto the 4' ppant arm of the PCP domain, catalyzed by the NRPS A domain. (a) Active site of the D-alanine-activating A domain from *Bacillus cereus* D-alanyl carrier protein ligase (DltA) in complex with the adenylated D-alanine-AMP product of the adenylation reaction (PDB:3DHV) [27]. (b) Active site of the (R,R)-2,3-butanediol-activating A domain from PA1221, a di-domain NRPS from *Pseudomonas aeruginosa*, covalently linked to a vinyl-sulfonamide inhibitor (PDB:4DG9) [14]. (c) Active site of the L-valine-activating A domain from linear gramicidin synthetase with L-valine substrate covalently linked to the 4' ppant arm of the PCP domain (PDB:5ES8) [10]. (d) Mechanism of thiolation by nucleophilic attack of the activated carbonyl carbon of the adenylated substrate by the terminal thiol group of the 4' ppant arm of the PCP domain.

The structure of PA1221 A-PCP di-domain from *Pseudomonas aeruginosa* covalently linked to a vinyl-sulfonamide inhibitor [14], mimicking the thiolation reaction intermediate, as well as the structure of the initiation module of linear gramicidin A synthetase (including the A-PCP di-domain) with substrate covalently linked to the 4' ppant arm [10], as shown in Figure 6, allowed for some insight into the thiolation reaction.

Key binding residues from the A_{core} of the A domain's active site adopt the same position and orientation in the thiolation state as they do in the substrate-binding state [10]. This extends to the ATP binding state, as shown in Figure 5, suggesting that the key binding residues of the active site from the A_{core} interact with the adenylated substrate in the same way and perform essentially the same role throughout the cycle, including during thiolation [30-32].

When rearranging into the thiolation conformation to allow the PCP domain to access the active site, the A_{sub} domain is essentially turning back by rotating 140° and presents its opposite face to the active site. As such, the Lys517 from the A_{sub} no longer has access to the active site, and residues from the "back" face of the A_{sub} domain may interact with the substrate. In the adenylation conformation, the residue Glu327 from the A_{core} is coordinating the α-phosphate of ATP via a Mg^{2+} ion, as shown in Figure 5. In the thiolation conformation, the Arg439 from the A_{sub} is brought to the same space the Mg^{2+} ion used to occupy [13], as shown in Figure 6. The Arg439 interacts with the substrate directly to coordinate the α-phosphate of the adenylate moiety, negating the purpose of Glu327 which is no longer involved in substrate binding. The Arg439 also binds the backbone carbonyl of Ser191 (adjacent to Stachelhaus residue Thr190), which helps the docking of the PCP domain into the A domain's active site. In the thiolation conformation, Lys434 of the A_{sub} takes the place of Lys517 and binds the adenylated substrate in a similar way [33]. Lys434 coordinates the adenylated substrate and, together with the other active site residues, pulls electron density away from the phosphate to promote the acceptance of a lone pair of electrons, making it a better leaving group and facilitating the nucleophilic attack of the substrate by the thiol end of the 4' ppant arm, as shown in Figure 6.

1.1.5 The Condensation Domain

The C domain is a ~55kDa heart-shaped domain whose function is to catalyze the formation of a peptide bond, or other types of bonds, between the substrates presented by the PCP domains located immediately upstream and downstream of the C domain [12]. The catalytic motif HHxxxDG (where x stands for any amino acid) of the C domain is essential for activity and is highly conserved across species [16].

The overall structure of the C domain can be separated into two lobes; the N-terminal lobe (N lobe) is composed of 7 β-strands and 9 α-helices and the C-terminal lobe (C-lobe) is composed of 6 β-strands followed by 8 β-strands and 9 α-helices [12, 34]. The two structurally related lobes are arranged in a pseudo dimeric fashion with a β-sheet core enclosed in an envelope of α-helices, forming a heart-shaped domain with the 30Å-long active site along the interface of the two lobes. The "latch" region (VibH residues 338-360, PDB:1L5A) is a β-strand and a loop donated from the C-lobe to the N-lobe by a pseudo-strand-swap, believed to be a remnant of an oligomerization element [12]. Samel et al. proposed that it is able to undergo a conformational change that will expose the active site, allowing the substrate and product to access or exit it, hence its name, although there is no solid evidence for it at this time. The "floor loop" (VibH residues 262-276) lines the catalytic tunnel of the domain [12]. The structure of the C domain has been solved in the "open" and "closed" conformations. In the open confirmation, both halves of the heart are moved away from one another with respect to the closed conformation [34]. The donor PCP domain docks at the donor site of the C domain, located at the ridge between the two lobes, on the "top end" of the heart. The acceptor PCP domain docks at the acceptor site on the opposite side of the C domain, at the "sharp end" of the heart. The two prosthetic arms extends towards each other into the tunnel formed by the interface between the lobes, to reach the second catalytic histidine; this key residue is located roughly 15Å away from the surface where the two PCP domains have docked, which is precisely the length of the 4' ppant arm [34].

Other domains, most of which are structurally related, may replace, or work in conjunction with the C domain:

The epimerization (E) domain can change the chirality of the α-carbon of the substrate. The main function of the E domain is to produce D-amino acids from L-amino acids, conferring various biological functions to the NRP including protecting against proteases [12]. The E domain is often found at the end of a module. The presence of an E domain is not essential to the incorporation of D-amino acids to the NRP chain: in some systems, such as in DltA discussed above, the A domain specifically selects a D-residues [27]. In other systems, the C domain is replaced by a C/E domain that performs both condensation and epimerization [34].

The heterocyclization (Cy) domain catalyzes the condensation reaction, followed by the cyclodehydration of cysteine, serine or threonine into five membered thiazoline, oxazoline or methyloxazoline rings, such as the ones found in bleomycin or bacillamide shown in Figure 1 [5, 7]. In the Cy domain, the HHxxxD motif is replaced by a DxxxxD motif [7]. Cy domains have been reported to function in tandem in systems introducing (methyl)oxazoline rings, including some siderophore synthetases [35].

The fungal NRPS terminal condensation (Ct) domain replaces the TE domain in most filamentous fungal NRPSs and performs macrocyclization to release the final NRP product [36].

A specific sub-type of C domain, called β-lactam-forming C domain, can perform a condensation reaction followed by the formation of a β-lactam ring [37]. The donor substrate must contain a β-hydroxyl group, which is converted into an α-β-alkene through elimination, providing an electrophilic site for the α-amino group of the acceptor substrate to attack and form a β-lactam ring [34].

X-domains are a class of NPRS domains, structurally related to the C domain, required for crosslinking amino acid side chains and form multiple macrocyclic NRPs [38], as seen in vancomycin shown in Figure 1. The X domains functions by recruiting an oxygenase to catalyze the cross-linking reaction on the NRPS-bound NRP [38].

1.1.6 The Peptidyl Carrier Protein Domain

The peptidyl carrier protein (PCP) domain is a small (80-100 residues) domain responsible for carrying the substrate from domain to domain and presenting it to their active sites using its 4' phosphopantetheine (4' ppant) arm [39]. The overall structure of the PCP domain consists of a four-helix bundle and there are currently over 25 NRPS structures containing PCP domains deposited in the PDB. It is modified post-translationally by a phosphopantetheinyl transferase, such as the promiscuous Sfp protein from *Bacillus subtilis*, which covalently binds the phosphopantetheine moiety of coenzyme A to the serine of the conserved GGxS motif found in the second helix of the PCP domain [5] [40]. A PCP domain with no 4' ppant arm tethered to the serine is called apo PCP, and a PCP domain modified with the 4' ppant arm is called holo PCP. The substrate is loaded onto the 4' ppant arm by thiolation, catalyzed by the A domain: the thiol group at the end of the 4' ppant arm acts as a nucleophile to attack

the activated carbonyl group of the substrate, as previously discussed. The PCP domain picks up the monomer substrate from the A domains active site, transports the substrate along the assembly line and presents the substrate to the next active site, that is to say, the C domain for elongation, a tailoring domain for modification or the TE domain for termination [40].

1.1.7 The Thioesterase Domain

The thioesterase (TE) domain is a ~30kDa domain, located at the end of the assembly line, responsible for termination of elongation and the release of the final NRP product [41]. To do so, the TE domain performs one of the following reactions: the hydrolysis of the thioester bond between the 4' ppant arm and the assembled peptide chain, the macro-cyclization of the assembled peptide chain, or the oligomerization of several assembled peptide chains [42]. The TE domain presents as a Rossmann-like α/β hydrolase fold core domain, with an additional helix bundle called the "lid"; the lid has been observed in the "open" conformation, allowing access to the active site, and in the "closed" conformation, folding over the active site and shielding its content from the cellular environment [43, 44].

Following elongation by the NRPS chain, the NRP is presented to the active site of the TE domain by the PCP domain; the active site serine of the TE domain attacks the carbonyl carbon of the first residue of the NRP, releasing the 4' ppant arm as a leaving group in a nucleophilic substitution reaction [42]. As a result, the NRP product is tethered to the TE domain via the catalytic active site serine. The TE domain will then catalyze a second reaction, which may be oligomerization, cyclization or hydrolysis, following the same nucleophilic substitution mechanism [42].

1.2 The Tailoring Domains

A tailoring domain is an additional domain to the canonical C-A-PCP module that performs a specific reaction such as formylation, reduction, oxidation, methylation, halogenation or glycosylation, adding another layer of diversity to the NRP chemical space. Many NRPS include one or more tailoring domain, acting in *cis* (embedded in the assembly line) or in *trans* (as an independent soluble protein, most likely expressed as part of the NRPS gene cluster) [1]. True tailoring domain impart their modification co-synthetically, while domains that modify the substrate before its selection by the A

domain, or modify the final product of the assembly line after its release are sometimes also called tailoring domains/proteins [45]. The domains related to the C domain discussed in the C domain section are considered to be tailoring domains. The arrangement of the domains and the modular nature of NRPSs favor the horizontal addition of tailoring enzymes without disturbing existing domains, which are conserved through evolution if they enhance the properties of the resulting product. As such, the possibilities are limitless and the chemical diversity of NRPs is considerably broadened.

1.2.1 Ketoreductase domain

The ketoreductase (Kr) domain is responsible for the reduction of an α-keto-acid into an α-hydroxy-acid using NADPH as a co-factor. The α-hydroxy-acid may act as a nucleophile in the same way an α-amino-acid would, and be incorporated into the NRP chain through the formation of an ester bond in lieu of an amide bond [46]. We can identify ~30 distinct Kr domain-containing NRPSs through sequence-based prediction software, with about a dozen of them being the subject of biochemical studies. Prior to this thesis, the NRPS Kr domain still remained to be structurally characterized; however, it is highly similar in primary sequence to the Kr domains found in polyketide synthase (PKS). There are about 11 structures of PKS Kr domains currently deposited in the PDB. The structure of actinorhodin Kr domain in complex with NADPH and the inhibitor emodin was solved by Korman *et al*. to a resolution of 2.3Å (PDB:2RH4), revealing the specificities of substrate binding [47]. According to sequence-based prediction and given their analogy with PKS, the NRPS Kr domain is predicted to be composed of two domains, the active Kr and the pseudo-Kr (the later rendered inactive through evolution due to redundancy) [48]. Both sub-domains are composed of a Rossmann-like fold that allows for binding of NADPH, with an additional helix bundle involved in substrate binding. Through the ketoreduction reaction, a planar ketone is reduced to a tetrahedral alcohol, introducing a new chiral center. There are then two types of Kr domains following the nomenclature of PKSs, the “A type” produces and L-hydroxy substituent and the “B type” produces a D-hydroxy substituent [49]. The structure of the type A PKS Kr domain from amphotericin B synthase in complex with substrate and NADPH was solved by Zheng *et al*. to a resolution of 1.5Å (PDB:4DIF), and the structure of the type B PKS Kr domain from tylosin PKS was solved to a resolution of 1.95Å by Keatinge-Clay *et al*. (PDB:2Z5L).

1.2.2 Methyltransferase Domain

The methyltransferase (MT) domain is a ~25 kDa domain with a Rossmann-like fold which catalyzes the N-, C-, S- or O-methylation of the substrate using S-adenylmethinione (SAM) as a methyl-donor, as is the case for cyclosporin A shown in Figure 1 [50]. In addition to being essential to the function of some NRPs, methylation stabilizes the natural product and protects it from degradation. MT domains are usually inserted within the module (C-A-MT-PCP) and act in *cis*. Together with most Kr, oxidase and mono-oxygenase domains, the MT domain is often predicted to be inserted within the A domain, in between motif A8 and A9 [51]. The specificity of the MT domain is significantly affected by their binding partners and neighboring domains, and may act as a second substrate gatekeeper in addition to the A domain [52]. The structure of an MT domain, as part of the TioS_AMt didomain NRPS from the thiocoraline gene cluster, was solved by Mori *et al*. to a resolution of 2.9Å, showing that the interrupted A_{sub} domain folds into a bona fide domain, and that the activity of the A domain is not disturbed [53].

1.2.3 Formyltransferase Domain

In contrast to ribosomal synthesis, formylation of the first amino-acid is not required in all NRP synthesis. However, some NRPSs, such as linear gramicidin synthetase, present a formyltranferase (F) domain at the N-terminal of their initiation module (F-A-PCP). This domain formylates the substrate, using N^{10}-formyltetrahydrofolate as a co-factor, after it is loaded onto the PCP domain's 4' ppant arm by A domain. The PCP linker on its own is not long enough to allow the PCP domain to reach the A and the F domain active site [10, 54]. The A_{sub} domain has to rotate 180° and translocate 21Å, allowing the PCP domain to rotate 75° and translocate an impressive 61Å to reach the active site [10]. Moreover, the A_{sub} domain seems to be essential for efficient docking of the PCP domain, since the PCP and the F domain have a very limited binding surface (219$Å^2$) [10]. Contrary to what was observed for the MT domain, the F domain is inserted upstream to the A domain, and makes a fairly extensive interface with it, burying 831$Å^2$ of surface area [10].

1.2.4 Glycosyltransferase Domain

Glycosyltransferase (GT) domains are responsible for the addition of an activated sugar group (NDP-sugar) to a backbone or side chain, drastically altering the biophysical and chemical properties of the NRP; this reaction is regiospecific, stereospecific and stereoselective [55]. Bleomycin and vancomycin shown in Figure 1 are two examples of glycosylated NRPs.

1.2.5 Halogenase Domains

Halogenase domains are responsible for the addition of a halogen group (fluorine, chlorine, bromine or iodine) to an aliphatic carbon [56]. Halogenases often act in *trans* and are identified as part of an NRPS gene cluster, or, more commonly, an NRPS/PKS hybrid gene cluster. Halogenated NRPs are often toxic, and can have a very different function compared to their non-halogenated counterparts. Halogenases are particularly interesting for synthetic chemistry and the production of green chemicals for agriculture. For example, they can be used to introduce a halogen leaving group to a chemical intermediate in a regio- and stereospecific manner, which can then be used for nucleophilic substitution or elimination reactions. This process is also used *in vivo*: the CrpH domain, from cryptophycins cluster, is a soluble non-heme FAD-dependent halogenase responsible for the addition of one or two chlorines to a tyrosine residue, which can then be replaced by a methyl-oxide group by substitution [57]. The synthesis of vancomycin, shown in Figure 1, also involve a chlorination step [58].

1.2.6 Oxidase Domain

The oxidation (Ox) domain is a flavin mononucleotide dependent enzyme that catalyzes the oxidative dehydration reaction responsible for the formation of a double bond in the NRP. The Ox domain can be soluble and acting in *trans*, such as the one found in the bacillamide gene cluster [7]. They may also be inserted at various places within the modules and act in *cis*: the Ox domain from bleomycin synthetase BlmIII is inserted in the A domain [59]; the Ox domain from mtaC, part of the myzothiazole gene cluster, is inserted after the PCP domain as the final domain of the chain [60]. The myzothiazole synthetase includes another Ox domain on chain mtaD, disrupting the A domain as seen in the bleomycin synthetase [59]. Most Ox domains perform their reaction following the cyclo-dehydration of the substrate performed by the C_y domain [7]. As described earlier, the Cy domain is

able to perform a cyclodehydration reaction on a cycteine, a serine or a threonine residue and thusly produce a thiazoline, oxazoline or methyloxazoline five membered ring respectively [7]. The Ox domain may reduce the aliphatic carbon-carbon bond in between the hetero atoms to produce the aromatic compounds thiazole, oxazol and methyloxazole respectively, as seen in the anti-tumor compound bleomycin and the algicide compound bacillamide shown in Figure 1 [7, 59]. No structure of a *cis*-acting Ox domain disrupting the A domain or inserted after the PCP domain is currently available.

1.3 Macro Organization and Synthetic Logic of Nonribosomal Peptide Synthetases

1.3.1 The Canonical NRPS Module

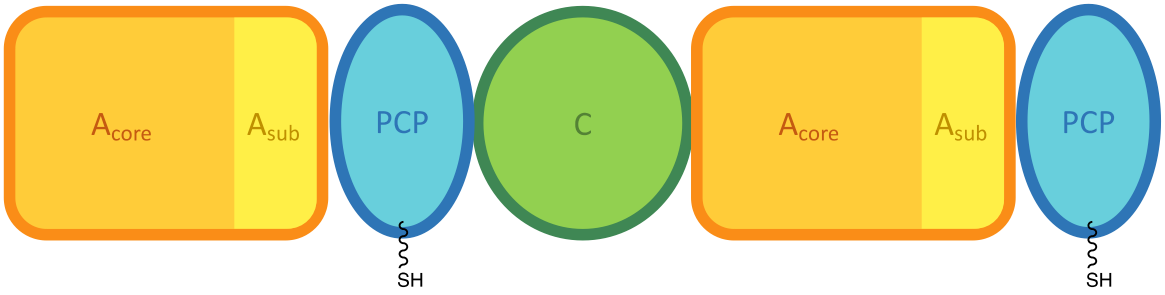


Figure 7: Domain architecture of a dimodular canonical NRPS. A domain composed of the A_{core} and A_{sub} subdomains shown in orange and yellow, PCP domain shown in blue and C domain shown in green.

The first module of an NRPS assembly line is the initiation module, which is responsible for selecting, adenylating and loading the first donor substrate onto the PCP domain's 4'ppant arm. A typical initiation module is made of an A domain and a PCP domain (A-PCP didomain). All subsequent modules downstream of the initiation module are elongation modules, responsible for selecting, adenylating and loading onto the 4' ppant arm the next residues for incorporation into the growing peptide chain. A typical elongation module is made of a C, A and a PCP domain (C-A-PCP tridomain) as shown in Figure 7.

1.3.2 The Strategy of Synthesis Employed by NRPS

During NRP synthesis, the growing chain is passed on from module to module by the PCP domains, each module being responsible for the incorporation of one residue, the specificity of which

is conferred but the A domain of the module. As such, a di-modular NRPS is expected to yield a dipeptide, a tri-modular NRPS is expected to yield a tripeptide, and an NRPS containing n modules is expected to yield a peptide which contains n amino acids. Moreover, since amino acid specificity is determined by the A domain, the templates that dictates the primary sequence of the NRP reside within the NRPS itself: the primary sequence of the NRP can be extrapolated from the modular sequence of the NRPS.

However, this is not always the case, as some NRPS may operate iteratively, engaging some or all of the modules more than once, producing longer products with repetitive motifs [61]. An example is SidD from the fungus *Aspergillus fumigatus*, which consists of one module followed by a PCP and a terminal C domain [8, 62]. It produces the extracellular siderophore fusarinine C, by linking three anhydromevalonyl-hydroxy-ornithine with ester bonds to produce the final depsipeptide [63]. The ferrichrome family of siderophores are cyclic hexa-peptides, typically consisting of three glycines and three hydroxamate ornithine residues, all synthesized by a two or three modular NRPS containing an additional PCP-C didomain for every iteration it is required to make [64].

Furthermore, stand-alone tailoring proteins can make a number of post-synthesis modifications after the NRP is released from the assembly line, making it hard to accurately predict the precise structure of the final product. Penicillin shown in Figure 1 and beauvericin are two examples of such NRPs. Ultimately, one must study the NRPS system within the host in order to elucidate the exact nature of the compound it produces, which can be challenging if said compound is only produced under specific conditions.

1.3.3 Intermodule and Interdomain Interactions and Interface

The 765Å^2 interface between the C and the A domain was initially proposed to be fixed, providing rigidity to form the overall structure of the NRPS [23]. It was also proposed that NRPS could adopt a higher order structure. Among the more popular theories, a model was proposed in which the C-A blocks would coil in an elliptical manner, with the smaller and more dynamic domains branching on either sides [65]. However, the elucidation of several structures showed the C-A block to be more malleable than previously thought: the C domain can adopt slightly different conformations and rotate

up to 20° with respect to the A domain, resulting in a shift of 30Å for some residues located in the N-terminus of the C domain [23]. The structure of DhbF (C_1-A_1-PCP_1-C_2-A_2-PCP_2-TE_2) dimodular NRPS was solved to a resolution of ~29Å electron microscopy (EM), the central A_1-PCP_1-C_2 segment to 2.1Å by X-ray crystallography, allowing for the visualization of inter-modular interactions [66]. In this structure, there is no significant interaction between an A_n domain and the $C_{(n+1)}$ domain of the next downstream module. Instead, the PCP_n is placed in between A_n and $C_{(n+1)}$, with the "front" active face of the PCP presenting the 4' ppant arm towards the A domain in a thiolation conformation, and the back face of the PCP pressed against the C domain with a surface area of 644 $Å^2$ [66]. EM data and 2D classification shows an extremely high flexibility around the hinge region in between modules, which does not support the model for higher order structures in multimodular NRPS, including the elliptical model [66]. Rather, it seems that NRPS follows the "beads on a necklace" model, were the modules (and the domains within the modules) adopt any and all conformation allowed without clashing with adjacent structures, with no higher order identifiable in the system studied [66].

Although no stable interactions able to restrict their motion were identified between NRPS modules, the transient interactions between domains and linker regions required for NRPS function have been observed at various point during the catalytic cycle [23]. During thiolation, a hydrophobic interaction takes place between the second helix of the PCP, and Ala259-Leu268 (in PA2212) of the A_{core} domain. Interestingly, the PCP domain interacts in a similar fashion with the C domain, where the same helix aligns parallel with another helix from the C domain. The docking of the PCP domain at the various active sites occurs largely through hydrophobic interactions. In contrast, dynamic interaction between the PCP and the A_{sub} domain mainly involved hydrogen bonding: positively charged residues from the A_{sub} domain (Arg, Gln and Asn), located in the loop region in between β_3 and β_4, engages in hydrogen bonding with backbone carbonyl group located in the first loop of the PCP domain (that is to say, the loop which contains the conserved serine that serves as the point of attachment of the 4' ppant arm) [14].

As discussed earlier, some domains such as the MT, Kr or Ox domains are lodged within the modules, predicted to be inserted between motifs A8 and A9 of the A domain [23]. In the case of the

TioS_Mt di-domain NRPS from the thiocoraline gene cluster, the MT domain is inserted in between β_{24} and α_{24} of the A domain [53]. The A domain conserves its overall structure and function in spite of the insertion of the MT domain. That said, the MT domain and the A_{sub} domain are pressed against each other, making extensive contact, with the slightly elongated α_{24} of the A domain acting as an anchor, not clearly belonging to one domain or the other [53]. From this structure, it can be inferred that the PCP domain must travel a distance of 60Å to reach the MT domain's active site after thiolation. More structures would be required to understand the dynamic cycle and rearrangement of the mobile parts of the module to accommodate the presence of the MT domain.

1.4 Polyketide Synthases and their Ketoreductase Domains

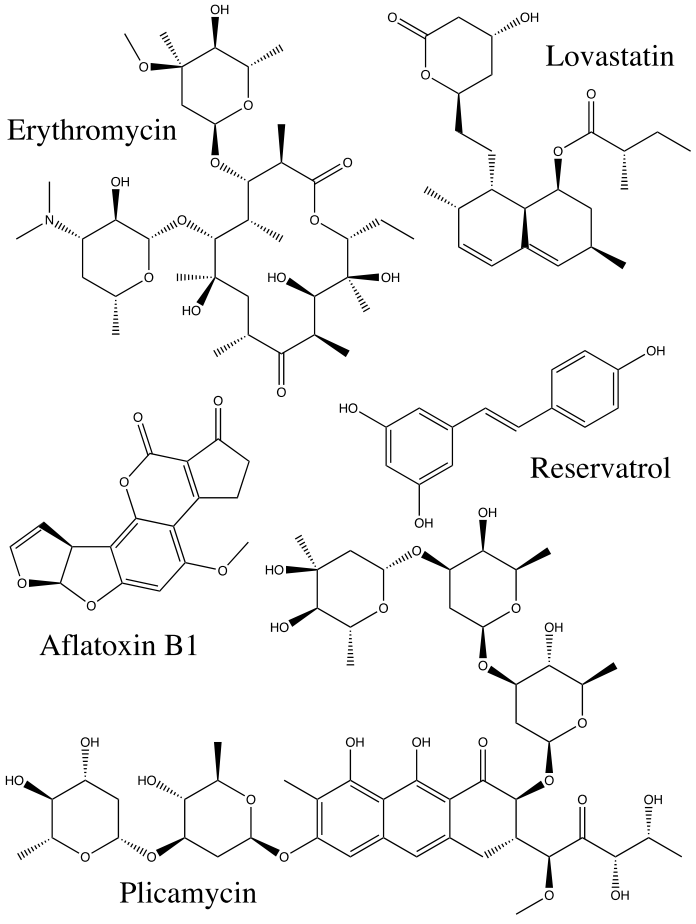


Figure 8: Examples of polyketides.

Polyketide synthases (PKS) are a group of large enzymatic complexes that synthetizes polyketide bioactive molecules. Polyketides (PK) are immensely structurally and chemically diverse, and they present a vast array of functions within the cell, as well as interesting pharmaceutical properties. To name a few examples of such compounds: the antibiotic erythromycin, the cholesterol-lowering drug lovastatin, the chemotherapeutic drug mithramycin, the antioxidant reservatrol and the carcinogenic toxin aflatoxin B1 shown in Figure 8 are part of the PK superfamily. The general organization of PKSs is akin to that of NRPSs: it is an assembly line, arranged in modules, each module being responsible for the

incorporation of one building block. The modules are composed of domains, each domain catalyzing a specific reaction, and the nascent PK chain is passed on along the assembly line for elongation by a carrier protein module with a prosthetic arm. Some PKSs, especially fungal ones, contain only one module and function in an iterative manner by adding several of the same building block to assemble the PK chain [67].

The principal domains commonly found in PKSs are: the acyltransferase domain (AT domain), the ketosynthase domain (KS domain), the acyl-carrier protein domain (ACP domain), the termination TE domain, and the modifying domains including the Kr domain, the dehydrogenase domain (DH domain), the MT domain and the enoylreductase domain (ER domain). The AT domain, analogous to the A domain in NRPS, selects the substrate (acetyl-coA or analog) and loads it onto the ACP domain. There is no need for adenylation since PKSs use acetyl-coAs as substrate (the acetyl moiety is already linked to a thiol group through a high-energy bond). The ACP domain is analogous to the PCP domain in NRPS, and is charged with channeling the substrate and presenting it to the active sites of the various domains. The KS domain, analogous to the C domain in NRPS, catalyzes the decarboxylative condensation reaction between the substrate and an acetyl group to form a β-keto acid. The Kr domain may reduce the β-keto-acid to a β-hydroxy acid in a stereoselective manner. The MT domain may add a methyl group to the α-carbon. The DH domain, following the Kr reaction, may catalyze the formation of a trans-double bond between the α and the β carbon, with elimination of the β-hydroxyl group, resulting in the formation of an enoyl group. The ER domain may reduce the α-β double bond created by the DH domain in a stereoselective manner, dictating the stereochemistry of the α-carbon. The sequential reactions of the Kr domain, the DH domain and the ER domain result in the formation of aliphatic α and β carbons. The TE domain terminates the reaction by hydrolysis or cyclization.

In the simplest terms, the synthetic logic of PKS is to elongate the chain by adding acetyl moieties to the growing PK, or a substrate analog containing an acetyl moiety, which are then modified in a way that is specific to the domain composition of the module charged with its incorporation. The PKS does not have access to a vast number of substrates like the NRPS does. The diversity of PKs comes from the fact that they are built virtually from scratch, two carbons at a time. Both carbons may be aliphatic with

both their stereochemistry controlled, the β carbon may hold a carbonyl group or a hydroxyl group of controlled stereochemistry, the α carbon could be aliphatic with control over the nature of the side chain and its stereochemistry, or it could be an α-β alkene. On top of that, other modifying "accessory enzymes" such as oxidases or halogenases could also add to the chemical diversity. As such, the PKS has very fine control over its product and may built an immense variety of natural compound practically de novo.

The PKS Kr domain and the NRPS Kr domain are very similar (25-35% identity). The PKS Kr domain is composed of a Rossmann-like fold made of seven β-sheet flanked by three α-helices on either side, with an additional lid over the active site formed by two helices. In PKSs, most Kr domains dimerize through helical dimerization elements and surface residues. Some PKS Kr domains are composed of two Kr subdomains: the active Kr domain and the pseudo-Kr domain, both with the same Rossmann-like fold with a lid structure. Only the active Kr domain has a catalytic activity [46]; the pseudo-Kr is a remnant domain with no catalytic activity, but conserves the same general structure as the active Kr domain [49]. The active Kr and pseudo-Kr are arranged relative to each other in a conformation identical to that of a dimer of active Kr domain observed in PKS, usually with a two-stand β sheet acting as a clasp maintaining the structure together [68]. The residues required for surface contact mimicked by the dual active/pseudo Kr domain and the helical dimerization element are identifiable from the sequence through prediction software [69]. The β-ketoacid building blocks of PKS show two distinct stereocenters: the α-carbon (with the orientation of the side chain) and the β-carbon (with the orientation of the hydroxyl group), both under the control of the Kr domain. PKS Kr domains are classified according to their stereospecificity and stereoselectivity: We call "type 1" the Kr domains that produce D-α-substituted substrates, and we call "type 2" the Kr domains that produce L-α-substituted substrates [70]. "A-type" is the Kr domains that are stereoselective for L-β-hydroxyl substituted substrates (S conformation), and we call "B-type" the Kr domains that are stereoselective for D-β-hydroxyl substituted substrates (R conformation) [68]. A-type PKS Kr domains are characterized by the presence of a tryptophan residue (W motif), located eight residues upstream of the catalytic tyrosine, and B-type PKS Kr domains are characterized by the presence of the Lys-Asp-Asp motif (LDD motif), located roughly 60 residues upstream of the catalytic tyrosine [70].

1.5 Bioengineering of Nonribosomal Peptide Synthetases

1.5.1 Antibiotic Resistance Crisis

The need for new tools to continue the research and development of new and more potent antibiotics is a growing concern among scientific communities and healthcare institutions alike. The ongoing antimicrobial resistance (AMR) crisis is described by the World Health Organization (WHO) as one of the biggest threats to our global health.

The U.S. Center for Disease Control and Prevention (CDC) reported 2 millions multidrug resistant (MDR) bacterial infections, resulting in 23,000 death yearly in the US, with a direct cost of $20 billion and an additional $35 billion in productivity losses [71]. The Canadian Institute of Health and Research (CIHR) reported that 2500,000 Canadians, or one in nine patients admitted to a hospital, contract an infection each year, 8,000 of which die of it [72]. In Europe, according to the Center for Infectious Disease Research and Policy, 670,000 suffer from AMR infections, resulting in 33,000 deaths yearly and a staggering 875,000 estimated years lost due to poor health [73]. On a global scale, over 700,000 people worldwide die of MDR bacterial infection every year [74]. If we do not find a solution to this crisis, statisticians estimate the death count to reach 10 million people worldwide by 2050 [75], which is more than the current deaths attributed to cancer (8.2M) and car accidents (1.25M) combined [74]. The associated cost of such a pandemic is estimated at 100 trillion dollars in health care cost, which higher than the GDP of the world [75]. Furthermore, the logistics of dealing with and containing the infected patients may not be feasible. In other words, AMR could be the end of our civilization as we know it, if we do not combine our effort and develop new tools to circumvent the crisis.

Between the lack of research and discovery of new antibiotics, and the difficulty of passing later stages of clinical trials, the number of novel antimicrobials being approved is in a dramatic decline [76]. No more than a couple of products are expected to reach the market in the coming five years [77]. It has become clear that we need a new tool for antibiotic discovery and development if we are to overcome the AMR crisis. We need antibiotics that are vastly different in shape or mechanism, to avoid the current resistance mechanisms, and we need more of them to choose from. Should the existing antibiotics remain unused for long enough, the bacterial population would lose its resistance due to

lack of pressure, making old antibiotics viable again. However, eliminating resistant populations requires a library of potent antibiotics to cycle through, in addition to global level of coordination and cooperation. Since its discovery in 1928, penicillin is estimated to have saved 200 million lives, and is considered to be one of the most prominent success of modern medicine, until resistance was first identified in 1965 [78]. The life span of an antibiotic before resistance arises is getting shorter and shorter: in contrast with penicillin, the antibiotic ceftaroline was first introduced in 2010, and resistance was first reported in 2011 [79].

The wider chemical space available to NRPS systems makes them particularly interesting as tools for drug development, especially when it comes to antibiotic resistance, as the assembly line can accommodate up to 500 different substrates. When taken together with the modular build of the NRPSs, the combinatorial possibilities are almost endless. If these vast possibilities are properly utilized, it could ensure that no single organism can evolve a mechanism for resistance towards all of the resulting possible antibiotics.

1.5.2 Challenges in Bioengineering of Recombinant NRPS

The modular nature of NRPS could be exploited for bioengineering of recombinant NRPS: a new assembly line, and therefore a new product, can be generated by selectively cutting and pasting together specific modules from different NRPS. Recombination by swapping, deleting or inserting one or several modules (C-A-PCP) has been successfully carried out, albeit at the cost of reducing the activity to 0.5-45% of the wild-type enzyme [80]. Generally, we observe that the loss of activity tends to be proportional to the amount of modification brought to the NRPS assembly line, and inversely proportional to the percentage of identity between the inserted module and the deleted module and/or the rest of the NRPS [80]. Moreover, experiment consisting in swapping the mobile PCP domain had very low success rate, even with high homology, suggesting that the channeling of the substrate along the assembly line relies on specific interdomain interactions that have not yet been sufficiently characterized and understood [80]. Special attention has been given to the linker regions, which typically have a significant impact in maintaining the fragile integrity of the assembly line. In some cases, deleting or inserting up to four amino acids could have a positive impact on the activity. In other cases,

the mutation of specific residues, or swapping of an entire linkers or hinge regions, could rescue the loss of activity after recombination [80]. Bringing altercations to only parts of a domain seem to be more successful: for example, the specificity of the A domain may be altered by using core motifs borrowed from another A domain, without disrupting the interdomain interactions or the mobile parts of the NRPS [80, 81]. Additional factors, such as the presence of tailoring domain embedded in the A domain, can also alter the specify and yield of the final product [52]. These findings suggest that we need more information about the specific interactions between NRPS domains, modules and linker regions to elucidate a more universal code for NRPS bioengineering [82].

1.6 Thesis Overview and Objectives

This thesis consists of two parts: the first part presents the structure of the initiation module (A-Kr) of the stratospherulide (Sts) NRPS from *Bacillus stratosphericus*; the second part consists of the structural and biochemical study of the individual domains that compose the module, as well as the investigation of some interesting features brought to light by the structure of the A-Kr module. We present the first structure of a Kr-containing NRPS module, the first structure of an NRPS Kr domain, the first structure of a keto-acid-selecting A domain, as well as the first report of a new NRPS domain, the pseudo-A_{sub} domain. Using this structural data, we will attempt to understand the organization of a Kr-containing NRPS and gain some insight as to how the enzyme is adapting to the use of a keto-acid.

How is the Kr domain incorporated within the NRPS assembly line? What structural changes, static and dynamic, have to occur to accommodate this new domain? How is the A domain able to select and adenylate a keto-acid instead of an amino acid? What is the role, structural or functional, of this new pseudo-A_{sub} domain? How does the NRPS Kr domain differ from a PKS Kr domain? How does a keto acid-selecting A domain differ from the canonical amino acid-selecting A domain typically found in NRPS?

2 Chapter II – Structural Insight into the Initiation Module A-KR-PCP of Stratospherulide Synthetase from *Bacillus stratosphericus*

2.1 Introduction

2.1.1 Stratospherulide Synthetase: a Depsipeptide Synthetase from *Bacillus stratosphericus*

Among the *Bacillus* family of bacteria is a microbe found high in the sky: the *Bacillus stratosphericus*. This organism was first identified in 2006; and is normally found in the stratosphere but brought down to Earth through atmospheric cycling processes [83]. It is mainly known for its surfactin synthetase and its ability to form biofilms useful for many purposes, including generating electricity [84]. It contains an NRPS we called stratospherulide synthetase (Sts), which bears very high similarity with the cereulide synthetase (Ces) from *Bacillus cereus* (37% identity), and the valinomycin synthetase (Vlm) from *Streptomyces* (33% identity).

2.1.2 Domain and Module Architecture of Stratospherulide Synthetase

As shown in Figure 9.a, stratospherulide synthetase (Sts) is a tetra-modular NRPS containing 15 domains separated into two protein chains, StsA and StsB. The initiation module of StsA contains an α-keto-acid-selecting A domain, a Kr domain and a PCP domain, together responsible for the addition of D-α-hydroxyl-isocaproic acid (αHIC) as first residue of the peptide chain. The second module of StsA contains a C domain, an α-amino-acid-selecting A domain, a PCP domain and an E domain, responsible for the addition of D-Valine as second residue of the peptide chain. Since we consider an initiation module to involve an A-PCP didomain, and an elongation module to involve a C-A-PCP tri-domain, the third module of Sts is then split between StsA which contains the C domain of the third module and StsB which contains an α-keto-acid-selecting A domain, a Kr domain and a PCP domain, together responsible for the addition of an unknown α-hydroxy-acid as the third residue of the peptide chain. The second module of StsB, the fourth and the final module of Sts, contains a C domain, an α-amino-acid-selecting A domain, a PCP domain and an TE domain, responsible for the addition of L-valine as the fourth residue of the peptide chain.

2.1.3 Depsipeptide Synthetases and their Products

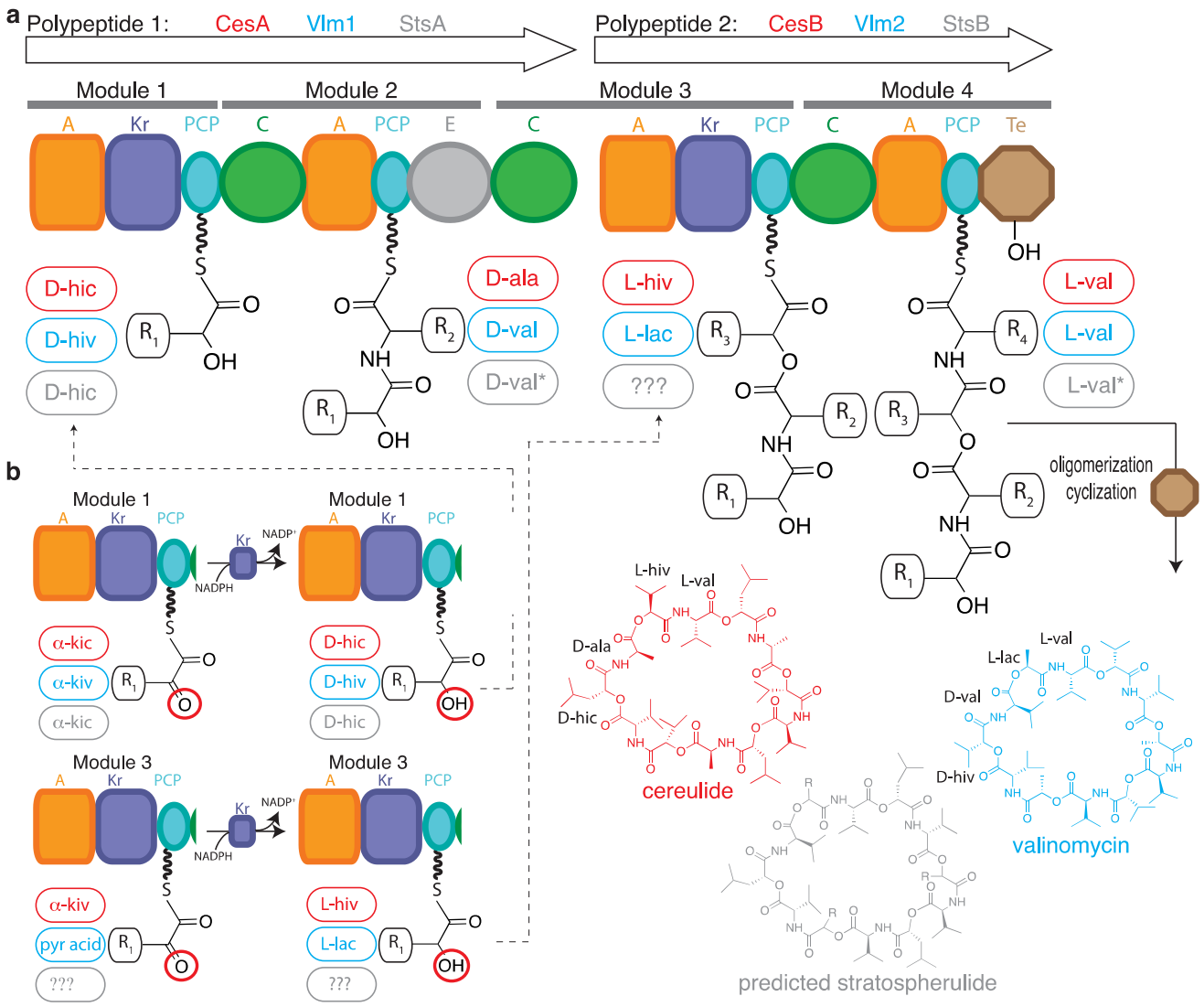


Figure 9: Domain architecture and synthetic logic of depsipeptide synthetases. (a) Module and domain architecture of depsipeptide synthetases. (b) Catalytic cycle of the Kr domain-containing modules 1 and 3 of depsipeptide synthetases.

A depsipeptide is a peptide chain in which one or more peptide bond is replaced by an ester bond, as is the case for cereulide, valinomycin and stratospherulide shown in Figure 9. This is made

possible in NRP synthesis by the incorporation of a Kr domain. The common branched amino acids leucine, valine, alanine and isoleucine are synthesized by transamination of α-ketoisocaproic acid (αKIC), α-ketoisovaleric acid (αKIV), pyruvate and α-ketoisoleucin (αKIL) respectively. Other α-ketoacids, such as α-ketoglutarate, are involved in amino acid synthesis as well as general metabolism [85]. The common α-ketoacids are therefore readily available in the cell [85]. In order to incorporate an α-hydroxyacid into the depsipeptide chain, an α-ketoacid is selected by the A domain and reduced to an α-hydroxyacid by the Kr domain. The α-hydroxyl group acts as a nucleophile in a manner identical to the α-amino group, resulting in the formation of an ester bond instead of an amide bond. This creates a product with very different properties, as the ester bonds confer different geometry and freedom of rotation.

As shown in Figure 9, depsipeptide synthetases characterized to date bears the same architecture: a D-hydroxy-acid specific initiation module containing a D-specific Kr domain followed by a D-amino-acid specific module containing an L-specific A domain and an E domain, the third module is an L-hydroxy-acid specific module containing a L-specific Kr domain (confirmed for Vlm and Ces), and finally a L-amino-acid specific termination module with a TE domain [86]. The TE domain is able to oligomerize three tetradepsipeptidyl moieties produced by the assembly line in a head-to-tail fashion to form a dodecadepsipeptide, before catalyzing macrocyclization to form the final cyclic-dodecadepsipeptide product [87].

2.1.4 Use of Chemical Probes and Inhibitor in Crystallography and Biochemistry

Chemical probes and inhibitors are often used in crystallography to trap an enzymatic complex in a specific conformation and obtain a snapshot of a particular step of the reaction cycle, or simply to reduce the flexibility of the complex to facilitate crystallization. Enzymes are expected to have a high affinity for their substrate and in some systems, substrate binding generally triggers an induced fit that slightly modifies the structure of the enzyme, potentially ordering some residues and/or relaxed loops. It is then common to co-crystallize an enzyme with its substrate in hope that it will induce crystallization or increase resolution, on top of enriching the structure by bringing some insight in terms of substrate binding and mechanism of reaction.

2.1.4.1 *In vitro modification with coenzyme A*

In *E. coli*, the 4' ppant arm is attached to the active serine of the PCP domain by the endogenous phosphopantetheinetransferase (PPTase), expressed by the *entD* gene. When growing a PCP-containing NRPS in the common expression strain *E. coli* BL21 (DE3), we obtain a mixed population of apo-NRPS (with no 4' ppant arm tethered to the PCP domain), holo-NRPS (with a 4' ppant arm tethered to the PCP domain) and holo-NRPS loaded with the cognate substrate or acetate. In order to obtain a homogenous population of holo-protein for crystallography or biochemical assays, the NRPS can be expressed in *E. coli* BL21 (DE3) *entD*$^-$ strain, where the *entD* gene has been knocked out. This results in a homogenous apo-NRPS population which can then be modified *in vitro* using coenzyme A (CoA) and the Sfp protein from *Bacillus subtilis*, a promiscuous PPTase [88], that tethers the phosphopantetheine moiety of CoA to the PCP domain, resulting in a homogenous population of holo-NRPS. Alternatively, the NRPS can be expressed in *E. coli* BL21 (DE3) BAP1, a strain in which the Sfp gene is integrated into the prp operon, inducible with Isopropyl-β-D-1-thiogalactopyranoside (IPTG) [89]: the NRPS and Sfp are co-expressed, resulting in a homogenous population of holo-NRPS. See Figure 10 for a schematic of the reaction.

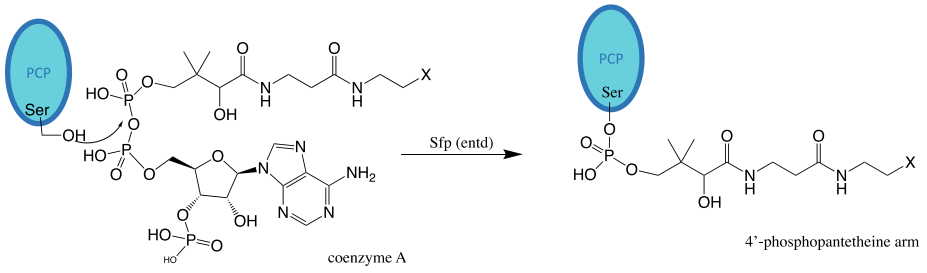


Figure 10: Enzymatic modification of the PCP domain with the 4' ppant arm using the Sfp protein. X=SH for coenzyme A or X=NH_2 for aminoacyl-coenzyme A.

2.1.4.2 *In vitro modification with aminoacyl-coA and substrate*

Most acyl-enzymes intermediate, formed through a thioester or a ester bonds, have a half-life in solution ranging from a few minutes to a few hours [90]. Such a time frame is not compatible with crystallography. In contrast, amide bonds (also known as peptide bonds) have a half-life of over 500

years, which provides a time frame much more suited to crystallography [91]. For this reason, it is useful to produce an NRPS with a terminal amino group at the end of the 4' ppant arm of the PCP domain instead of a thiol group, in order to visualize a substrate or analog loaded onto arm. Several analogs of coenzyme A have been developed, but most interesting to NRPS crystallography is the aminoacyl-coA [92]. The NRPS is produced in an *entD*$^-$ strain to obtain a pure solution of apo-NRPS. The holo-NRPS is obtained by incubating the enzyme with sfp and CoA, or a CoA analog such as aminoacyl-CoA. The strategy for synthesis of aminoacyl-CoA is to use the natural system for synthesis of CoA by purifying the *E. coli* enzymes pantothenate kinase (PanK), phosphopantetheine adenylyl transferase (PPAT) and dephospho-CoA kinase (DPCK) to recreate the synthesis *in vitro*. In *E. coli*, PanK phosphorylates the alcohol end of pantothenic acid (Vitamine B_5), at which point a cysteine is added to produce the thiol terminal end and decarboxylated by the phosphopantothenoylcysteine decarboxylase (PPC-DC) to produce 4' phosphopantetheine, which is then adenylated by PPAT to produce dephospho-CoA, when finally DPCK phosphorylates the 2'OH group of the adenyl moiety to produce the CoA. Fortunately, we can skip the addition of an amino acid and subsequent decarboxylation step by using an amino- or hydroxyl-pantetheine as starting material, which PanK is able to accept as a substrate and phosphorylate. PPAT may then add the adenyl moiety, followed by the phosphorylation the 2'OH group of the adenyl moiety by DPCK to produce hydroxyl- or aminoacyl-CoA, as shown in Figure 11.

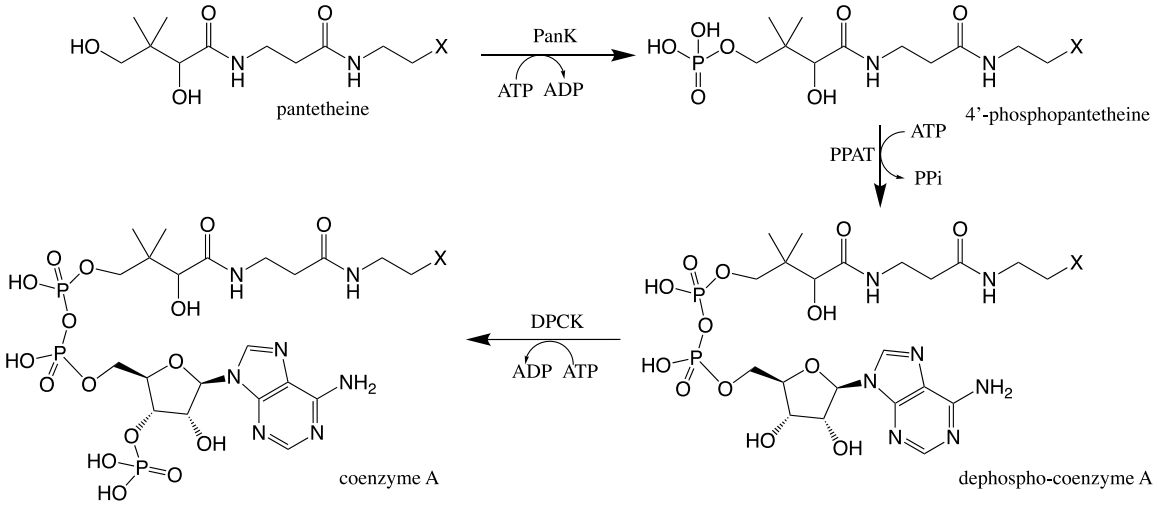


Figure 11: Coenzyme A analogs. X=SH for CoA or X=NH_2 for aminoacyl-CoA.

2.1.4.3 *Adenosine Vinylsulfonamide Inhibitors*

Enzymes are predicted to have a high affinity for the transition state of the reaction they catalyze, as stabilizing the transition state lowers the energy of activation of the reaction. For this reason, transition state analogs are often used as inhibitors due to their high affinity for the active site, and can even be used to "freeze" a transient interaction and trap an enzymatic complex in a specific conformation. As outlined in chapter 1, NRPSs are quite mobile, particularly the A_{sub} and PCP domains. For this reason, the adenosine vinylsufonamide inhibitor (AVSI) was designed to trap the complex in the thiolation conformation, where the substrate is loaded onto the PCP domain. The AVSI, shown in Figure 12, mimics the adenylated substrate with the sulfonyl group mimicking the phosphate group, while introducing a Michael acceptor (i.e. a double bond between the α- and the β-carbon bond to a carbonyl group, or carbonyl analog, the which β-carbon may readily accept a lone pair of electron). The thiol group of the PCP arm attacks the double bond and the sulfonyl group will accommodate the added electron pair through resonance. As a result, the 4' ppant arm is covalently linked to the AVSI, and the affinity of the A domain for the inhibitor is sufficient to trap the complex in the thiolation conformation. The AVSI is a potent inhibitor capable of completely abolishing the activity of the A domain and has been successfully used by many groups for the purpose of crystallography [14, 17, 66, 93-95].

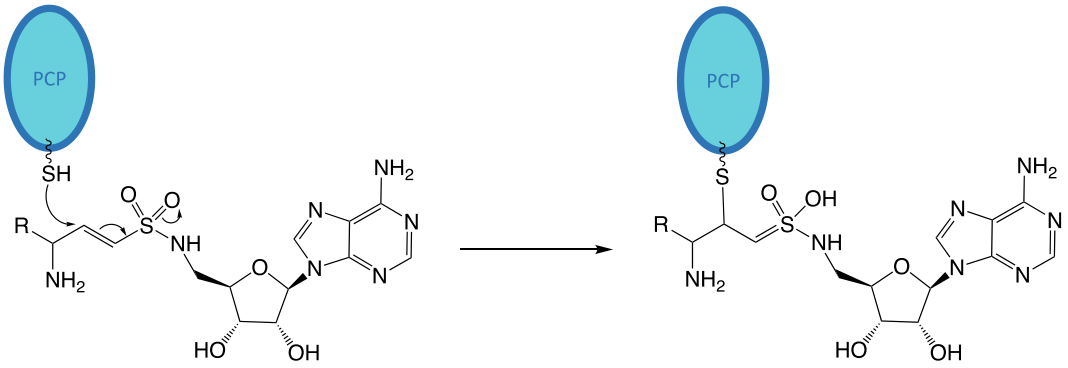


Figure 12: Adenosine vinylsulfonamide inhibitor reaction.

2.1.4.4 Small Molecules

Small molecules that are not meant to be covalently bound to the active site; they are often used in crystallography either to promote the formation of a certain conformation, or to gain insight on the reaction mechanism and substrate binding. They can either be co-crystallized with the protein or soaked into the crystal. Enzymes are expected to have a higher affinity for their substrate than their product, since substrates must be bound, and products must be released. For this reason, many inhibitors are designed to mimic the substrate, without being able to undergo the chemical reaction catalyzed by the target enzyme. For example, any enzymes that uses ATP can be inhibited by using a non-hydrolyzable ATP analogs. The NRPS A domain uses ATP to adenylate the substrate, resulting in the hydrolysis of the phosphoester bond that links the AMP moiety to the pyrophosphate moiety. If the oxygen of the phosphoester bond is replaced by a carbon, the bond is no longer hydrolysable without forming a highly instable carbanion, which is extremely unfavored and unlikely to happen. AMPcPP, shown in Figure 13, is often used for the crystallography of A domain-containing NRPS structures in order to visualize the substrate-binding phase of the catalytic cycle [22].

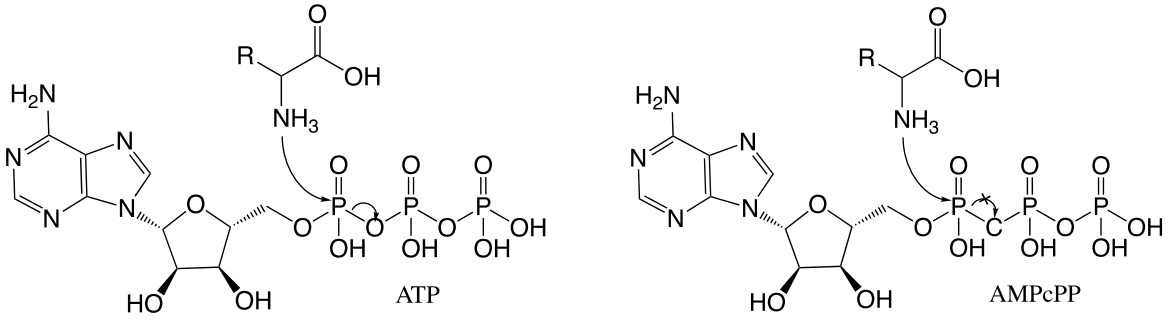


Figure 13: ATP and AMPcPP inhibitor [77]

2.2 Results

The goal of this chapter is to structurally characterize various combinations of the first module of StsA in order to visualize the specifics of the Kr domain incorporation within the NRPS module, and investigate the structural changes, static or dynamic, that are required for the accommodation of the tailoring domain.

To do this, I first set out to crystallize the A-Kr didomain. This structure would provide insight regarding the interdomain interactions between the Kr domain and the A domain and reveal whether or not the Kr domain is embedded within the A_{sub} domain as some predictions suggest. Moreover, excluding the flexible and mobile PCP domain of the initiation module may increase chances crystallization.

In parallel was carried out the crystallization of the A-Kr-PCP (or A-Kr-T) full initiation module, allowing for the visualization of the interactions not only with the A domain, but with the PCP domain (or T domain) as well, to gain insight regarding the incorporation of the Kr domain and the process of operation of a Kr-containing NRPS module. In an attempt to reduce the motion of the PCP domain and facilitate crystallization, the A-Kr-T protein was modified with aminoacyl-CoA and loaded with the αKIC substrate.

2.2.1 Crystallography of A-Kr Didomain

In order to create an A-Kr didomain construct, the boundaries of the Kr domain must be estimated: to facilitate crystallography, it is useful to cut out disordered loops, but any portion of the sequence that is essential for folding must be included. Moreover, the Kr domain was predicted to be inserted within the A domain, with A domain-specific sequence identified downstream of the Kr domain; in the absence of further knowledge on the matter, it is useful to create constructs with and without the estimated "linker region" between the last motif of the Kr domain and the first motif of the PCP domain. According to the prediction software and manual sequence analysis, the linker region between the Kr domain and the PCP domain was estimated to span between residue 1158 (with motif DAEM) and residue 1222 (motif FHEL). Two construct of the A-Kr didomain were successfully cloned in a dual affinity tag system (using the His-tag and CBP-tag, see methods section 2.4.1.): plasmid

pBac_Tandem_StsA_AKr_DAEM which includes the linker region (cut at the DAEM motif), and plasmid pBac_Tandem_StsA_AKr_FHEL which excludes the linker region (cut at the FHEL motif). The cloning of two other constructs, with slightly different choices of domain barriers, StsA_AKr_LIEE (cut at residue 1225) and StsA_AKr_TKAV (cut at residue 1176), was attempted as well without success.

Expression of StsA_AKr_DAEM was attempted twice without success. StsA_AKr_FHEL expressed well at 16°C overnight when induced with 100uM IPTG at OD 0.40-0.60, with a yield of 3.5-6mg of protein per gram of cell pellet after the affinity column chromatography purification steps. The MonoQ ion exchange column separated the protein sample into three oligomeric states: monomer, dimer and tetramer (as estimated by native-PAGE). The monomer is divided into either two or three peaks that appear identical in SDS-PAGE and native-PAGE. The yield of protein after the MonoQ ion exchange column was approximately 1mg of monomer per gram of cell pellet (30% recovery), and 0.13mg of dimer per gram of cell pellet (4% recovery), which bring the monomer:dimer ratio to 7.7:1 (or the dimer represents roughly 11% of the oligomeric population). For the purpose of crystallization, the dimer is further purified by size exclusion chromatography (SEC) to recover a solution of StsA_AKr_FHEL dimer in a single oligomeric state, with an estimated purity superior to 99% according to SDS-PAGE, and a final yield of 0.1mg of protein per gram of cell pellet. Further purification of the monomer was attempted as well, but the crystallization was unsuccessful using the monomer. The structural work described in this chapter was carried out using the dimer only.

Initial crystal hits were identified in four different conditions by sparse matrix screening: needle-shaped crystals in condition 1 (0.9M succinic acid, 1.5% PEG 2000 MME and 0.1M HEPES pH 7.0), needle-shaped crystals in condition 2 (20% PEG 3350 and 0.2M disodium phosphate), amorphous crystals in condition 3 (1.0M sodium malonate, 1.0% jeffamine and 0.1M HEPES pH 7.0) and quasi-crystals in condition 4 (17% PEG 3350 and 0.22M sodium citrate). After optimization of the precipitant concentration, conditions 1 and 2 gave rice-shaped crystals up to 250µm in length, condition 3 gave 200-300µm crystals of irregular shape, and condition 4 was discarded when no crystals formed after

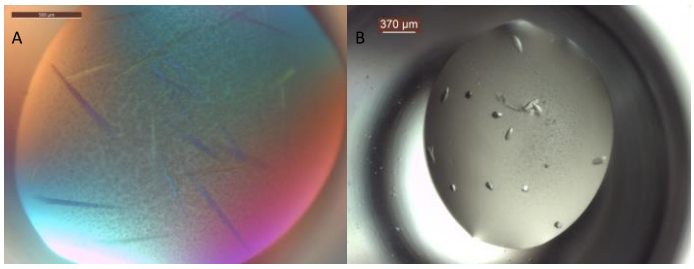


Figure 14: StsA_AKr_FHEL crystals (a) 700-900um needle-shaped crystals after third round of seeding grown in 25%PEG 3350 and 0.185M sodium phosphate (b) 50-100um rice-shaped crystals grown without seeding in 1.156M sodium malonate, 1.333% jeffamine, 0.1M HEPES pH 7.0 and 3%w/v trimethyl N-oxide dehydrate.

several rounds of optimization. Co-crystallization with 1mM NADPH was attempted without success. After pH-optimization, counter-ion optimization, detergent screen, additive screen and several rounds of seeding, the malonate/jeffamine condition (condition 3) yielded rice-shaped crystal with sharp or irregular edges of approximately 150µm in length, which diffracted to 19Å, shown in Figure 14.b.

To increase the yield of preparation of StsA_AKr_FHEL, the pBac_StsA_AKr_FHEL_(TEV)_His plasmid, using a single C-terminal His-tag, was successfully cloned and expressed in the same conditions as pBac_Tandem_StsA_AKr_FHEL. Removing the calmodulin column step shortened the preparation by one day and increase the yield after affinity column chromatography more than doubled to 12mg of protein per gram of cell pellet, with similar purity assessed by SDS-PAGE. After the MonoQ anion exchange purification step, 0.4mg of dimer per gram of cell pellet were recovered (4% recovery). The monomer to dimer ratio is identical to the one measured for the tandem construct. After SEC purification a solution of StsA_AKr_FHEL dimer in a single oligomeric state, with an estimated purity superior to 99% and a final yield of 0.2mg of protein per gram of cell pellet (or twice the yield obtained with the tandem construct) was obtained.

2.2.2 Crystallography of A-Kr-PCP Initiation Module

2.2.2.1 Crystallography of "StsA_AKrT"

The pBac_tandem_StsA_AKrT plasmid was successfully cloned. Crystallography of the A-Kr-PCP initiation module was first attempted using the StsA_AKrT_tandem construct, with both histidine and calmodulin binding protein (CBP) tags. The protein was grown in BL21 to yield a mix of apo, holo, and

holo states loaded with various substrate. The pBAC_tandem_StsA_AKrT expressed and purified similarly to pBAC_tandem_StsA_AKr described in section 2.2.1., yielding roughly 0.1mg per gram of cell pellet of StsA_AKrT dimer in a single oligomeric state, with an estimated purity superior to 99% (albeit as a mix of apo- and holo-protein, loaded or not with the substrate or acetate). Rice shaped crystals of size ranging from 50µm to 200µm were obtained and optimized in various conditions. The most promising ones were seen in condition 2 and condition 4. After exhausting all optimization possibilities in terms of crystallography, none of the crystals diffracted to a resolution higher than 30Å.

2.2.2.2 Crystallography of StsA_AKrT_CoA-NH-αKIC

The poor quality of diffraction of apo-AKrT protein crystals seemed to indicate that a homogenous population of protein was required to increase the resolution, which meant controlling the presence of the 4' ppant arm and loaded substrate. For this reason, the protein was grown in entd- in view of enzymatically modifying the protein with the 4' ppant arm *in vitro* with sfp and coA (or coA analogs). In order to increase the yield of preparation of StsA_AKrT, the construct was successfully cloned into a single C-terminal His tag construct, the pBac_StsA_AKrT_(TEV)_His. Given the high expression of StsA_AKrT, it seemed reasonable to attempt purification using only one affinity tag, placed at the C-terminus exclude truncated products. StsA_AKrT_(TEV)_His expressed well at 16°C overnight when induced with 10oµM IPTG at OD 0.40-0.60, with a yield of 15-20mg of protein per gram of cell pellet after the nickel column affinity chromatography purification steps. The yield of protein after the MonoQ ion exchange column was approximately 3mg of monomer per gram of cell pellet (25% recovery) and 1mg of dimer per gram of cell pellet (6% recovery). The dimer represented roughly 15% of the oligomeric population. For the purpose of crystallization, the dimer is modified with the aminoacyl-4' ppant arm using Spf and aminoacyl-coA, followed by substrate loading using αKIC and ATP. The modified StsA_AKrT_CoA-NH-αKIC dimer is further purified by two consecutive SEC steps, to recover a solution of StsA_AKrT_CoA-NH-αKIC dimer in a single oligomeric state, with an estimated purity of ~99% and a final yield of 0.3mg of protein per gram of cell pellet (which represents three times the yield obtained with the pBac_tandem_StsA_AKrT construct).

Sparse matrix with the StsA_AKrT_CoA-NH-αKIC protein yielded 32 hits, most of which grew at room temperature, including seven fully formed crystals, most of which were rice-shaped (similarly to the apo-protein crystals), but two of them were needles. The other crystals hits were either quasi-crystals, clusters or globular amorphous crystals. Out of the seven fully-formed crystal hits, three were non-redundant with the apo-protein hits described earlier and as such were more thoroughly optimized: condition 5 (0.1M Tris pH 8.5, 12% (v/v) glycerol and 1.5M ammonium sulfate), condition 6 (20mM magnesium chloride, 0.1M HEPES pH 7.5 and 22% (w/v) sodium polyacrylate 5100) and condition 7 (0.2M di-sodium tartrate and 20% (w/v) PEG 3550). The initial hits were rice crystals with heavy nucleation for condition 5, a single rice-shaped crystal with sharp edges for condition 6, and needle-shaped crystals for condition 7. After the first round of optimization in 24-well plates, 200µm box-like shaped crystals that diffracted to 9.7Å were obtained in condition 5. Single crystals with two distinct morphology, shown in Figure 15, that grew to 100-300µm on average but up to 700µm and diffracted to 12Å were obtained in condition 6. The crystals were very sensitive to cryoprotection and melted with glycerol and ethylene glycol. Using serial addition of sodium polyacrylate to a final concentration of 40% allowed for the crystals to be cryoprotected without melting. Finally, in condition 7, 50-100µm rice-shaped with edges or box-like crystals that diffracted to 8.9Å were obtained in 96-well plate, as they didn't grow well in 24-well plate. After several rounds of optimization of precipitant concentration and pH, additive and detergent screen, as well as optimization of the cryoprotection and dehydration protocols, the resolution did not improve. Co-crystallization and soaking with NADPH, αKIC, ATP and $MgCl_2$ was also attempted without success.

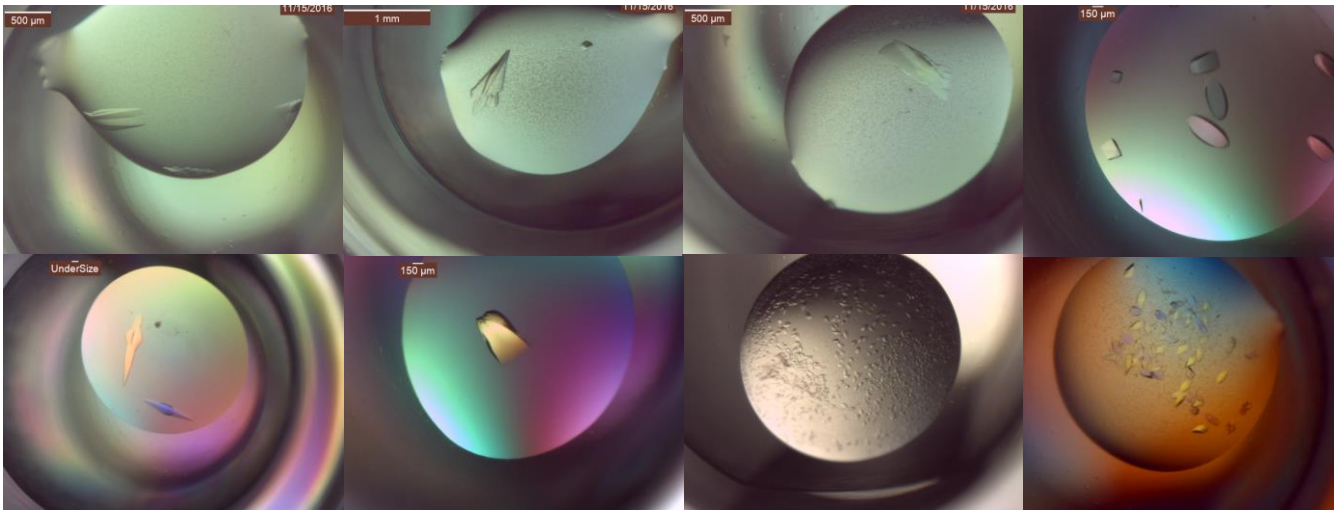


Figure 15: Morphologies of the StsA_AKrT_CoA-NH-αKIC crystals.

Since the resolution of AKrT-CoA-NH-αKIC crystals had plateaued, sparse matrix was attempted with seeding, using two different seed stocks made of the apo-crystals in condition 2 and 4. The sparse matrix with condition 4 seed stock yielded a few hits, with some rice-shaped crystals, but mostly clusters and amorphous crystals. The sparse matrix using a seed stock from condition 2 however yielded 46 hits; 38 of them were fully formed crystals, most of which were rice-shaped or rice-like, but some of them were more facetted and edgier, and 6 of them had a new morphology. Priority was given to the hits that were not redundant with the apo-crystals or the non-seeded ones. After the first round of optimization, we focused on the two most promising conditions: condition 8 (0.2M calcium acetate and 20% (w/v) PEG 3350) and condition 9 (0.49M sodium phosphate and 0.91M potassium phosphate). For condition 8, the initial hit was a 50µm birefringent facetted rice-like crystal, with sharp edges. After the first round of optimization were obtained 50-100µm crystals of the same morphology that diffracted to 10Å. For condition 9, the initial hit was a cluster of crystal, possibly with a new morphology. After the first round of optimization were obtained 100-200µm crystals, shaped like a football, that diffracted to 10Å. The resolution increased to 7Å for both conditions after optimization of the cryoprotection protocol. Through further optimization of the precipitant concentration and pH, a higher resolution of 6.7Å was obtained. The additive screen of condition 8 gave many exciting hits (10mM

phenol, 3% (w/v) galactose, 3% (v/v) ethylene glycol, 3% (w/v) D-glucose monohydrate, 5% (w/v) polyvinylpyrrolidone K15, 0.1M guanidine hydrochloride, 0.1M glycine, 10mM barium chloride dihydrate, 10mM calcium chloride dihydrate, 3% (w/v) sucrose, 3% (w/v) D-sorbitol, 3% (w/v) D-(+)-trehalose dihydrate, 1.2% (w/v) myo-inositol and 0.015mM CYMAL®-7), shown in Figure 16. All were reproduced in 24 well plates, with a similar shape of a facetted rice crystals with sharper edges. The additive screen for condition 9 also gave many interesting hits (0.1M sodium citrate tribasic dihydrate, 10mM strontium chloride hexahydrate, 10mM barium chloride dihydrate, 0.1M glycine and 10mM betaine hydrochloride). Co-crystallization with NADPH, ATP, αKIC and $MgCl_2$ was attempted in both conditions. Out of all these new crystals, many diffracted to a higher resolution: condition 8 crystals diffracted to 4.5Å with phenol additive, to 6.4Å with CYMAL®-7 and to 7.6Å with NADPH. The rest of the crystals in condition 8 diffracted to 8-12Å. Condition 9 crystals diffracted to 3.9Å with barium chloride and betaine additive, to 4.0Å with glycine additive and to 4.7Å with ATP, αKIC and $MgCl_2$ cocktail. The rest of the crystals in condition 9 diffracted to 6-7Å approximately. However, the crystals grown in condition 9 were twinned. The crystals grown in condition 10 were not twinned, but the processing of the 4.5Å data set obtained with phenol additive was attempted without success.

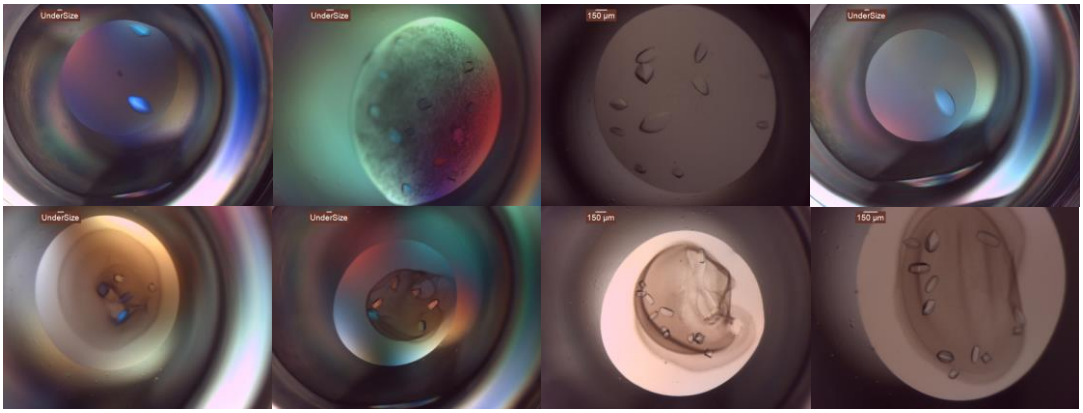


Figure 16: Crystals from additive screen of condition 10.

In an attempt to grow more non-twinned crystals, optimization of other seeded hits were attempted, including condition 10 (1.1M sodium chloride 0.1M HEPES pH 7.0, 0.5% (v/v) jeffamine ED-

2001, final pH 7.0), condition 11 (20mM magnesium chloride, 0.1M HEPES pH 7.5 and 22% (w/v) sodium polyacrylate 5100), condition 12 (0.2M potassium-sodium tartrate and 20% (w/v) PEG 3500), condition 13 (0.2M tri-lithium tartrate and 20% (w/v) PEG 3350) and condition 14 (0.2M di-sodium tartrate and 20% (w/v) PEG 3350), but none of them diffracted as well as crystals grown in condition 9 and 10.

Our focus shifted to condition 8 again, with the objective of increasing the resolution and obtaining a processable data set. However, the crystals were not reproducible, and the resolution dropped to 7Å. The resolution stayed stuck at 7Å after many rounds of optimization of crystallization condition and cryoprotection protocol. In an effort to optimize substrate loading and obtain a homogenous population of holo-protein modified with the substrate αKIC, the reaction mix was supplemented with buffer at pH 9.0 which improved the efficiency of the reaction, and a near-homogenous population of holo-protein modified with the substrate αKIC was obtained (monitored by LC-ESI-MS on intact protein, described in the methods section 2.4.3.). Using this new protein sample, the resulting crystals diffracted to 6-8Å consistently, or occasionally to 5Å for the best of them. After optimization of the cryoprotection and dehydration protocol (described in the methods section 2.4.7.), the crystals in condition 8 diffracted to 3.5Å and a non-twinned complete data set was collected. Even after optimizing the crystallization protocol, crystals were hard to reproduce. Growing the protein with seleno-methionine was attempted in order to calculate phases, and crystals that diffracted up to 6Å were obtained but the project was aborted when the structure was solved using omit-map sharpening and 4-fold non-crystallographic symmetry (NCS) domain averaging.

2.2.2.3 Crystallography of StsA_AKrT with αKIC-Adenosine Vinysulfonamide Inhibitor

In view of attempting to trap the A-Kr-PCP initiation module in the thiolation transition state, the αKIC-adenosine vinylsulfonamide (αKIC-AVSI) was synthetized by our collaborator John Colucci from Zamboni Chem Solution Inc. According to the LC-ESI-MS assay performed on intact protein, a 1000X excess of VSI over AKrT was needed for full modification, which is not amenable to a large-scale preparation, and such a high concentration of αKIC-AVSI resulted in the formation of multiple adducts. Sparse matrices with and without seeding, at room temperature and 4°C, were attempted three times following three sample preparations, but did not yield any fully formed crystals or any new morphology

worth optimizing, and were not reproducible. ICA-VSI was successfully synthetized by our collaborator John Colucci from Zamboni Chem Solution Inc. However, this inhibitor had no effect on the activity of the A domains of StsA_AKrT, as shown in by the adenylation assay.

2.2.3 Structure of A-Kr-PCP Initiation Module

A complete data set was obtained from a crystal of the A-Kr-PCP initiation module of StsA modified with aminoacyl-CoA and loaded with substrate αKIC, grown in condition 8 after optimization of the crystallization, cryoprotection and dehydration protocols. The crystal diffracted weakly with an $\langle I/\sigma_I \rangle$ of 2.0 at 3.9Å, 1.0 at 3.7Å and 0.33 at 3.4Å. Conventional techniques of molecular replacement using a homology model allowed us to place 4 Kr domains and 2 A_{core} domain, but the reamining 2 A_{core} domains as well as the A_{sub} domains which could not be placed, and the R_{free} was stalled at 44%. High-resolution data, albeit weak, were included and sharpening omit-map with 4-fold NCS domain averaging techniques were used to solve the structure to a resolution of 3.4Å to 3.9Å with a final R_{free} of 28%. The asymmetric unit of the crystal contains 4 copies of the StsA_AKrT construct, for a total of 4844 amino acids.

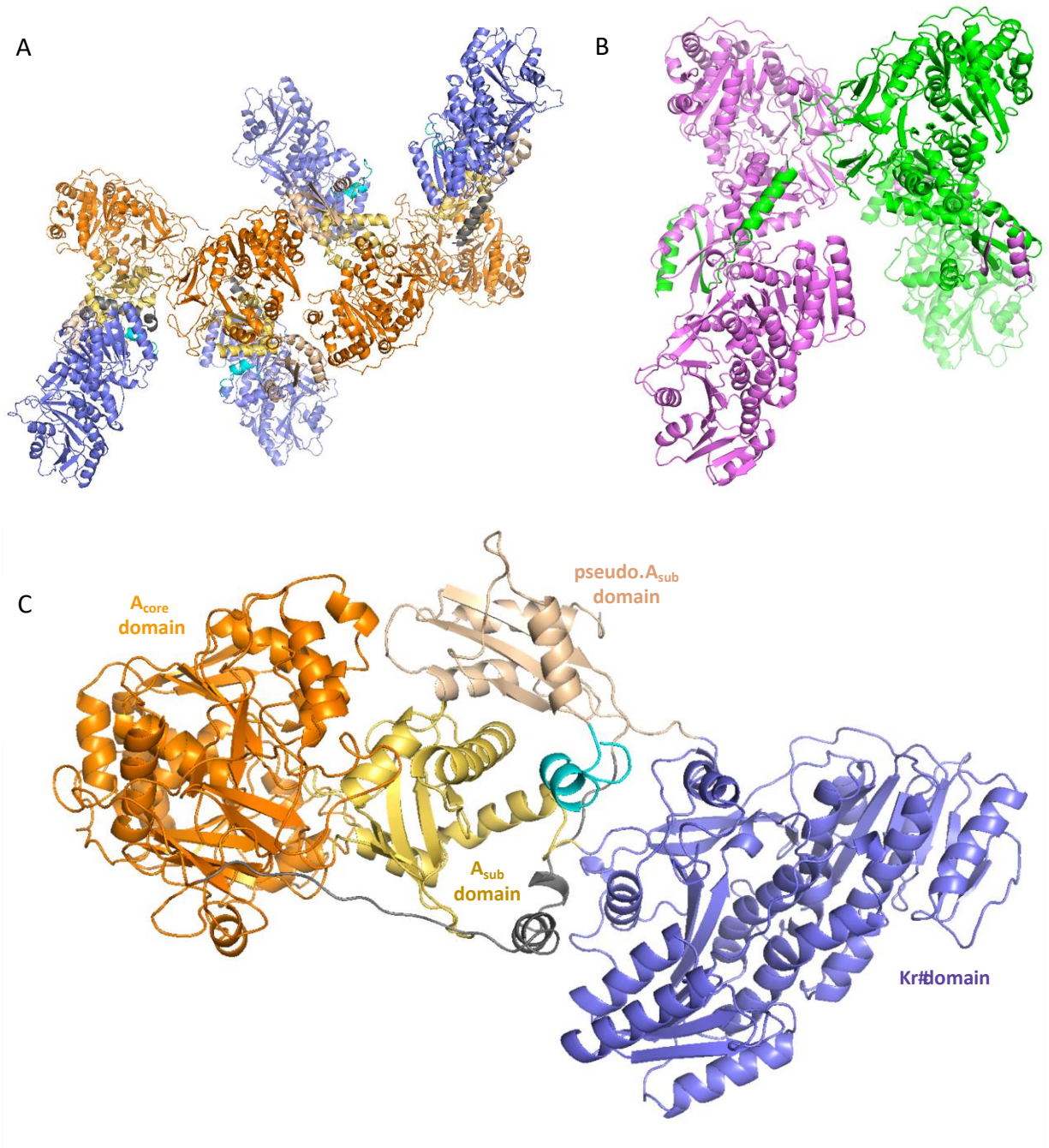


Figure 17: Structure of A-Kr partial initiation module (a) A-Kr dyad of dimers, or full content of the asymmetric unit (b) A-Kr Dimer; one molecule is shown in green, the other in purple to visualize the 3D domain swap (c) Model for the monomer. A_{core} domain shown in orange; A_{sub} domain shown in yellow; pA_{sub} domain shown in light pink; Kr domain shown in slate blue; first helix of the PCP domain shown in light blue.

The final model contains 1211 residues out of the 1318 residues of the crystallized StsA_AKrT protein. PCP domain is not ordered, as well some loop regions. The asymmetric unit contains four molecules, organized as a dyad of dimers (Figure 17.a.). The monomers composing the dimers are related by pseudo-C2 symmetry, and the dimers composing the dyad are related by pseudo-C2 symmetry as well (Figure 17.a and Figure 17.b). Excluding the disordered PCP domain, the module shows 4 separate domains: The A_{core} domain, the A_{sub} domain, the Kr domain and a novel domain which we called the pseudo-A_{sub} (pA_{sub}) domain. The A domain is in the closed conformation, and both A_{sub} and A_{core} subdomains show the same overall structure as observed for a typical α-amino acid selecting A domain. The structure shows that the A domain is uninterrupted by the Kr domain. The overall structure of the Kr domain is similar to the one observed for PKS Kr; it is composed an active Kr subdomain facing towards the A domain, and a pseudo-Kr subdomain facing the C domain downstream. The Kr domain is inserted between the A_{sub} domain and the pseudo-Asub domain (pA_{sub} domain). The pA_{sub} is interrupted at the linker region between the first and second β-strand, dividing it in two parts located 1112 residues apart in the 2D sequence: the N-terminal portion of the pA_{sub} contains the first α-helix and the first β-strand, and is located upstream of the A domain in the primary sequence; the C-terminal portion of the pA_{sub} contains the last two β-strands and α-helices, and is located downstream of the Kr domain in the primary sequence. Moreover, in the StsA_AKrT structure, the pA_{sub} of a given molecule consists of its own C-terminal portion of the pA_{sub} complemented with the N-terminal portion of the pA_{sub} of its dimeric partner, through a 39 residues-long strand-swap illustrated in Figure 17.b. In Figure 17 is shown a model for the StsA_AKrT monomer, in which the pA_{sub} is complemented by the N-terminal portion from the same molecule. The pA_{sub} domain has the same overall structure as the A_{sub} domain, with the exception that the conserved β-hairpin loop in the hinge region between the A_{core} and the A_{sub} is missing in the pA_{sub} (it is not needed since the sequence of StsA starts with the N-terminal portion of the pA_{sub}, there is then nothing to link the pA_{sub} to at the N-terminal region), and helices α_1 and α_2 are slightly shifted. The Kr domain is continuous with the A_{sub} at the N-terminus through a seven residue-long linker, and with the pA_{sub} at the C-terminus through a six residue-long linker. The Kr and the A domains are both intact and stand as independent structures, undisturbed by the presence of the other. The domains are present in an elongated conformation relative to one another, with no

significant surface contact or residue interaction identifiable. Similarly, the Kr is making no surface contact with the pA_{sub}. However, the Kr domain is making significant contacts with the linker region between the N-terminal portion of the pA_{sub} and the A_{core} of its dimeric partner, involved in the strand swap previously described. The active sites of the A domain and the Kr domain are located 70Å apart in the conformation observed in the StsA_AKrT structure. The pAsub domain is located adjacent to the A_{sub} domain, in between the A and Kr domain as shown in Figure 17. It does not disturb the architecture of the other domains, as it is loosely tethered to the Kr domain in a way that allows for significant mobility and flexibility.

2.3 Discussion

2.3.1 On the crystallography of StsA_A-Kr didomain

The StsA_AKrT structure shows that the linker region between the Kr domain and the PCP domain to span between residues 1156 and 1227. Our predictions for the location of the linker region were then fairly accurate, with an error margin of 5 amino acids, but did not account for the presence of the pA_{sub}. The StsA_AKr_FHEL construct included the C-terminal portion of the pA_{sub}, giving it the potential to form a fully folded pA_{sub}. This may have hindered crystallization as the pA_{sub} likely has an extensive freedom of motion. The crystallography of the A-Kr didomain may have been more successful with a complete deletion of the pA_{sub}, reducing flexibility and allowing for the formation of a more ordered crystal lattice. However, in the StsA_AKrT structure, the pA_{sub} makes significant contact with the A_{core} of the neighboring molecule within the asymmetric unit and may be required for crystallization of the A-Kr didomain as well.

2.3.2 On the crystallography of StsA_AKrT initiation module

The number of crystal hits obtained, morphology of crystals, reproducibility and optimization potential of the StsA_AKrT crystal drastically increased once the presence of the 4’ ppant arm and substrate loaded was controlled for, and a homogenous population of pure protein in a single oligomeric state modified with αKIC substrate stably loaded onto the aminoacyl-4’ ppant arm was obtained. Switching from a tandem tag construct to a single C-terminal His-tag construct was

consequential to the elucidation of the structure, since it greatly improved the yield of preparation and simplified the procedure without compromising the final purity. The purity was maintained with a single tag system because StsA_AKrT expresses very well in *E. coli* with no major contamination from cleavage products. Seeding improved the quality of the crystals and was one of the critical steps towards improving the resolution of the crystals such that the diffraction data could be adequately processed. Additionally, optimization of the cryoprotection and simultaneous dehydration protocol was paramount to the elucidation of the structure. The resolution improved from 8Å to 3.8Å after dehydration, and thusly accomplished what numerous rounds of optimization of the crystallization conditions could not. Finally, the iterative omit-map sharpening process combined with 4-fold NCS averaging was crucial to the elucidation of the structure of A-Kr(-PCP) initiation module of Sts.

2.3.3 On the use of α-ketoacid AVSI

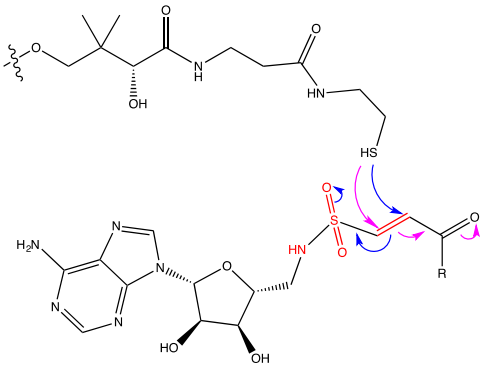


Figure 18: Dual Michael acceptor group in α-ketoacids adenosine vinylsulfonamide inhibitor

The use of αKIC-AVSI for crystallography was not practical: many types of side-products and multiple adducts were visible on the LC-ESI-MS trace after modification. This is probably due to the introduction of another Michael acceptor as shown in Figure 18, leading to many possible outcomes for this reaction. Thus, it is probably not advisable to use α-ketoacid-AVSI. Consequently, the ICA-AVSI was synthetized, but the compound was not able to inhibit the activity of StsA_AKrT during adenylation assays, suggesting that the α-substituent of the substrate may be important for proper binding to the active site.

2.3.4 On the structural model for StsA_AKrT initiation module

The structure A-Kr partial initiation module of StsA was elucidated to a resolution of 3.8Å. Through this structure can be visualized for the first time an α-ketoacid-selecting A domain and an NRPS Kr domain. Interestingly, the structure also revealed a surprising feature: a 39 amino-acids-long strand

swap between dimeric partners, which enables structures located over 1000 amino-acids apart in the sequence to come together through a domain swap to form the pseudo-A_{sub} domain, as shown in Figure 19. The pA_{sub} a novel domain uncharacterized and untheorized prior to this study.

The A domain shows the same overall structure and morphology as α-amino acid-selecting A domains [10, 13, 14, 26]. However, we can identify some varying features: The A_1 domain or StsA contains an additional loop spanning from residue 426 to 465, located 8 residues downstream of the Stachelhaus residue Gly324, which is 4 amino acids longer than in amino acid-selecting A domains. This loop seems to induce a slight shift in helix Pro389 to Pro401, which in turn induces a significant shift of helix Ala358 to Gly370 to protrude away from the A_{core} domain. Additionally, the second turn of the helix spanning from residue Ile305 (equivalent to the Stachelhauss residue Asp235) to residue Thr319 contains an additional amino acid with respect to the typical α-amino acid-selecting A domains. The resulting wider helix turn seemingly pushes the adjacent helix (residues 275 to 289), placing it in a slightly shifted position. In the StsA_AKr structure, we can also see that the loop spanning from residue Phe426 to Gly436 is 4 residues longer than the typical length of such a loop in amino acid-selecting A domains. Based on primary sequence alignment, these varying features are expected to be conserved among α-ketoacid-selecting A domains. At the current resolution, we can only comment on the overall structure and large-scale conformation of the backbone. A higher resolution model in complex with substrate is needed in order to investigate the mechanism of α-substituent specificity.

The Kr domain shows the same overall structure as a PKS Kr domain, with an active Kr subdomain and a pseudo-Kr subdomain [52, 68, 96, 97]. In order to make any conclusions about the mechanism of ketoreduction, a higher resolution structure in complex with substrate is needed. A specific inhibitor that would allow us to visualize the arm presenting the substrate to the active site, following a mechanism of inhibition similar to the AVSI, would be needed to confirm the mechanism of stereospecificity proposed for PKS Kr domains [97, 98]. Additionally, various substrates and inhibitors could be used to obtain snap shots of the protein in different conformations (i.e. adenylation, thiolation, ketoreduction and condensation) and visualize the complete catalytic cycle of the Kr domain-containing initiation module of Sts.

2.3.5 On accommodation of the Kr domain within the module

The Kr domain is not pressed against the A domain as was reported for the TioS A-MT didomain NRPS from the thiocoraline gene cluster [53]. Rather, the Kr and the A domains are undisturbed by one another and the Kr domain makes no direct contact with the A_{core}, the A_{sub} or the pA_{sub}. The linker region presenting the N-terminal portion of the pA_{sub}, which is placed in the cleft between the Kr and the other domains, mediates the interface by making contact with both the Kr and the A domain. The A_{core}, A_{sub} and Kr domain aligns on the same axis in a linear conformation; in other words, the Kr domain is "branching" from the C-A scaffold. In the conformation observed, the Kr domain does not significantly clash with other domains when superimposed with di-modular NRPS structures in the adenylation state, the thiolation state and the condensation state determined by X-ray crystallography and electron microscopy [66, 99]. In the conformation observed, the active site of the Kr domain is located 70Å way from the active site of the A domain, which represents a greater distance than the length of the 4' ppant arm (~17Å). Structural rearrangements are then expected to occur, possibly involving the pA_{sub} and/or the Kr domain, in order to achieve the ketoreduction conformation.

2.3.6 On the module architecture

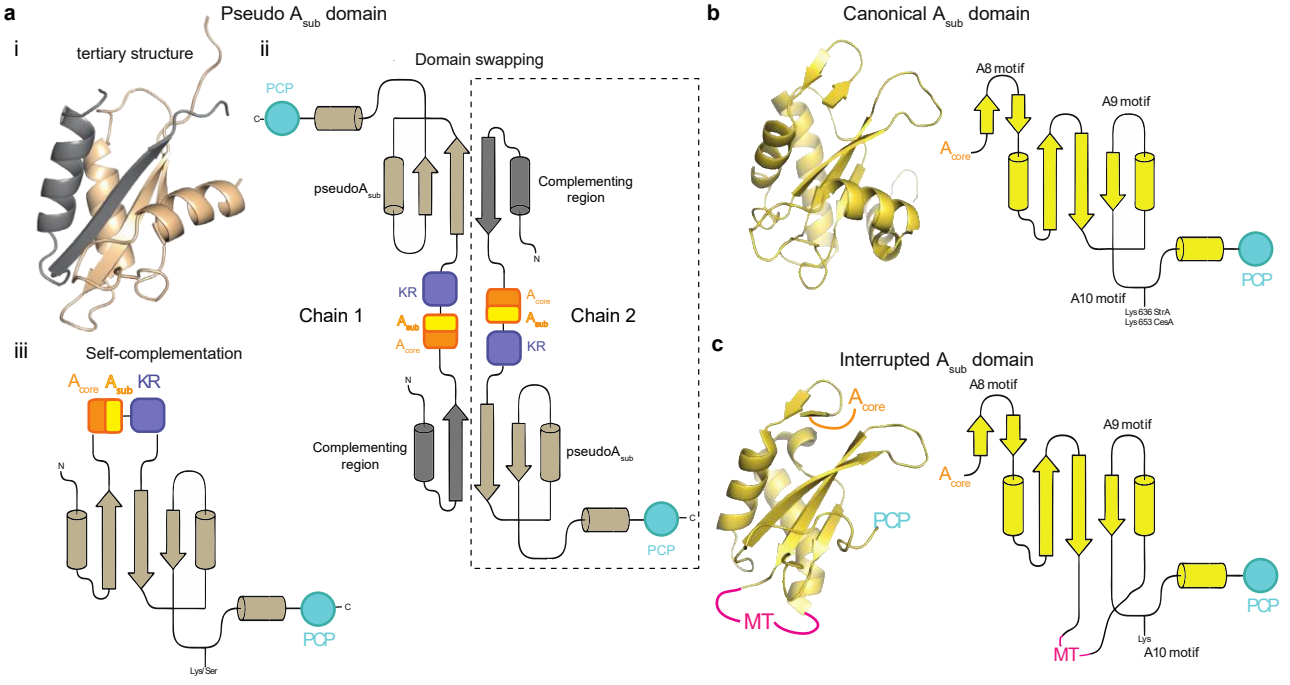


Figure 19: pA_{sub} 3D domain swap and dimerization of Sts. (a) Diagram for the formation of the pA_{sub} by 3D domain swap and model for self-complementation (b) Structure and diagram of the A_{sub} domain (c) Structure and diagram of the interrupted A_{sub} domain from the TioS_AMt didomain [53]

Sequence-based analysis predicted the presence of A_{sub} domain motifs downstream of the Kr domain, so the Kr was expected to the inserted within the A_{sub} between motifs A8 and A9 as was also predicted for other tailoring domain [59], and confirmed for the MT domain [53]. The extra sequence, due to the presence of two halves of the A_{sub} domain on either side of the Kr domain, was thought to have evolved to accommodate the Kr domain within the module. Instead, the A_{sub} and the rest of the A domain is not disrupted, the Kr domain is inserted immediately downstream of the A domain and the C-terminal portion of the pA_{sub} is located within the 64 amino acid-long region between the Kr domains and the PCP domain as shown in Figure 19. Such partial domains or pseudo-domains, believed to be evolution remnants, have been reported before in NRPS, such as the A domain-like remnant found in SidD from *Aspergillus fumigatus* [100]. It is not surprising that evolutionary insertion and deletions should be common in NRPSs given their modular and domain-based nature, and some of those remnant features might give us some clues about the evolutionary process that brought us the contemporary NRPS. The presence of a pA_{sub} indicates that the common ancestor of Sts and the other depsipeptide synthetase probably included a whole A domain at the N-terminal position, which was deleted through evolution. Our hypothesis is that a pre-existing A_1-Kr_1 initiation module, from another NRPS, inserted in the middle of the A_{sub} of the A_2-PCP_2 initiation module of the ancestor of Sts, resulting in the following architecture: A_{core1}-Nt_A_{sub1}-A_{core2}-A_{sub2}-Kr_2-CtA_{sub1}-PCP_1. Following the insertion, the A_{core1} subdomain became vestigial or redundant, and was deleted through evolution; the 39 amino acid-long linker region between Nt_A_{sub1} and A_{core2} evolved to allow the N-terminal and C-terminal portion of A_{sub1} to complemente each other and form the pA_{sub} domain. The result would then be the Nt_A_{sub1}-A_{core2}-A_{sub2}-Kr_2-CtA_{sub1}-PCP_1 module architecture observed in StsA.

2.3.7 On the pAsub

In the conformation visualized with StsA_AKr structure, the pAsub is located in between the Kr and the A domains, above the A_{sub} with respect to the A-Kr extended axis and does not disturb the other domains. By superimposing the StsA_AKr structure with the structure of the initiation module F-A-PCP of linear gramicidin in the open and closed conformation [23], we can see that the pA_{sub} does not clash with the A_{sub} and its trajectory from the open to the closed conformation. The pA_{sub} as observed in the current structure does however occupy the same space as the C domain of di-modular NRPSs observed in the thiolation conformation [66, 99]. We would then expect to see a rearrangement of the pA_{sub} domain, along with the PCP and A_{sub} domains, to achieve the thiolation conformation in Sts, although we were not able to obtain this structure by using AVSI. From this structure alone, is it difficult to make conclusions about the role of the pA_{sub} domain. It may have a catalytic function or fulfil a more structural role; it may be required for the accommodaton the Kr domain within the NRPS module, through dynamic or static interactions, or it could also simply be a remnant of an ancestral domain that has yet to be deleted by evolution. The structure of StsA_AKrT in complex with substrate, as well as the structure of StsA_A-Kr-PCP-C would shed some light onto the specifics of the interactions of the pA_{sub} with the other domains, as well as its possible involvement in the catalytic cycle.

2.3.8 On 3D domain swapping

To the best of our knowledge, Sts is the only NRPS reported to dimerize in solution. Since Kr domains often dimerize in PKS, the dimerization of StsA_AKrT was expected to occur through the dimerization of the Kr domains, in a manner analogous to PKSs, but a dimerization element could not be identified in the primary sequence of Sts. The structure of StsA_AKrT revealed that dimerization does not occur through the Kr domains, but through the formation of the pA_{sub} by domain swap, as shown in Figure 19. As a result, the dimeric partners are linked via the pA_{sub} domain at the center of the structure, with the both A domains and Kr domains extending on either side in a manner resembling the wings of a butterfly, as shown in Figure 17.a. A similar 3D domain swap was first reported in 1962 by Crestfiled *et al.*, with their work on the RNase A [101]. Hundreds of proteins characterized today contain a domain swap that commonly induces some degree oligomerization, including the Cro repressor from bacteriophage λ, the chicken citrate synthase, the human interleukin-5 or the human

cyclin dependent kinase CksHs1 involved in cell cycle regulation [102]. Assays performed on native samples using size exclusion column chromatography and native-PAGE show that the majority of StsA_AKrT is present as a monomer in solution. Both halves of the pA_{sub} may or may not be stable on their own, but one might propose that the module may be more stable with a fully formed pA_{sub}. The linker between the A domain and the N-terminal portion of the pA_{sub} is roughly 137Å long, and the distance between the N-terminus of the A domain and the N-terminal portional of the pA_{sub} is roughly 84Å without clashes. In other words, the linker is long enough to allow for the N-terminal portion of the pA_{sub} to complement the C-terminal portion of the pA_{sub} of the same molecule, thus forming a monomer A-Kr-PCP module as shown in Figure 17.c. This system must be studied further in order to understand the specifics of the mechanism of dimerization. We propose that if the linker were shorter, the intra-molecular complementation of the pA_{sub} could become less favored, and so could be controlled the equilibrium of oligomeric populations in solution.

2.4 Material and Methods

2.4.1 Cloning

The StsA_AKrT gene was synthesized by GenScript and cloned into pBAC_Tandem_forward vector (pET21-derived vector containing a C-terminal CBP tag and an N-terminal octa-histidine tag, respectively preceded and followed by a tobacco etch virus (TEV) protease cleavage site), to yield the pBac_Tandem_StsA_AKrT construct. StsA_AKrT gene was amplified from the pBac_Tandem_StsA_AKrT construct. The PCR amplicon was digested and cloned into pBac_Ct_His (pET21-derived vector containing a C-terminal octa-histidine tag preceded by a TEV protease cleavage site) using T7-ligase to produce the pBac_StsA_AKrT_(TEV)_His construct. In order to produce the StsA_AKr construct, the domain boundaries between the Kr domain and the PCP domain was estimated using a pool of information gathered from pfam analysis [103], secondary structure prediction on CLC Main Workbench 7 [104], secondary structure prediction using JPred [105], and manual alignment with a list of A-PCP with characterized domain boundaries. StsA_AKr gene was amplified from the pBac_Tandem_StsA_AKrT construct. The PCR amplicon was digested and ligated into cloned into

pBac_Tandem_forward and pBac_Tandem_Ct_His using T7 ligase to produce the pBac_Tandem_StsA_AKr construct and the pBac_StsA_AKr_(TEV)_His construct respectively.

2.4.2 Expression and purification

2.4.2.1 Expression and Purification of StsA_AKr_(Tandem) and StsA_AKrT_(Tandem)

StsA_AKrT_(Tandem) and StsA_Akr_(Tandem) were expressed and purified by nickel column affinity chromatography according to the same protocol as for StsA_AKrT_(TEV)_His (detailed in section 2.4.2.2.). The protein was then loaded onto the calmodulin resin packed chromatography column (G Biosciences) pre-equilibrated with nickel column binding buffer. The column was washed with five column volume of nickel column binding buffer and eluted with 2.5 column volumes of calmodulin column elution buffer (50mM Tris-HCl pH 8.0, 250mM NaCl, 2mM EGTA, 10% Glycerol and 2mM βME). The purity of the protein was assessed by SDS-PAGE. The affinity tags were cleaved, and the protein was dialysed according to the same protocol as for StsA_AKrT_(TEV)_His, except that 1mg of TEV protease was used per 10mg of protein. The protein was collected and run through the reverse-nickel column followed by a reverse calmodulin column, both columns pre-equilibrated with nickel column binding buffer, collecting the protein in the flow through. The protein was further purified by anion exchange chromatography and SEC according to the same protocol as for StsA_AKrT_(TEV)_His. The purity of the protein was assessed by SDS-PAGE and the oligomeric state of the protein was assessed by native-PAGE. The protein was either used immediately for crystallization, or flash frozen in liquid nitrogen and conserved at -80°C at 15mg/ml in 20% glycerol.

2.4.2.2 Expression and Purification of StsA_AKrT_(TEV)_His for crystallography

The StsA_AKrT_(TEV)_His construct was heterologously over-expressed in *Escherichia coli* BL21 (DE3) endD$^-$ cells (lacking the gene coding for enterobactin synthetase component D, responsible for the addition *in vivo* of the 4'-phosphopantetheine arm (4' ppant arm) to the PCP domain of apo-NRPS required for their function, using coenzyme A) [88]. 1L of Luria broth (LB) media supplemented with 35μg/ml kanamycin was inoculated with 10ml overnight culture and grown at 37°C with 220rpm shaking until the culture reached an optical density at 600nm (OD_{600}) of 0.5. Protein expression was induced with 0.1mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) and incubated at 16°C with 220rpm

shaking for 16-18 hours. Cell pellets were recovered by centrifugation at 5000rpm for 25 minutes at 4°C, and resuspended in nickel column binding buffer (50mM Tris-HCl pH 8.0, 250mM NaCl, 2mM $CaCl_2$, 10% Glycerol and 2mM β-mercaptoethanol (βME)) supplemented with 1mM phenylmethanesulfonyl fluoride (PMSF), 1 cOmplete™ EDTA-free protease inhibitor cocktail (Roche) and deoxyribonuclease I (BioShop). The resuspended cells were lysed by sonication (8 minutes 45 seconds; 20 seconds on/59 seconds off; 50% amplitude) and clarified by centrifugation at 18500rpm for 30 minutes at 4°C. The lysate supernatant was loaded onto a series of two 5ml HiTrap IMAC FF column (GE Healthcare) charged with Ni^{2+} pre-equilibrated with nickel column binding buffer. The column was washed with 5 column volumes of nickel column binding buffer and eluted with 5 to 10 column volumes of nickel column elution buffer (50mM Tris-HCl pH 8.0, 500mM NaCl, 4mM $CaCl_2$, 150mM imidazole, 10% glycerol and 2mM βME). The protein was quantified by NanoDrop 2000 (Fisher Scientific) and analysed by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE). The protein was incubated with 1mg of TEV protease per 50mg of protein, packed into 12K MWCO dialysis tubing (FisherBrand) and dialyzed at 4°C for 16h against 2L of dialysis buffer (50mM Tris-HCl pH 8.0, 2mM CaCl, 10mM NaCl, 10% glycerol and 2mM βME). The protein was collected and run through the nickel column pre-equilibrated with nickel column binding buffer, collecting the protein in the flow through. The protein was packed into the 150ml superloop (GE Healthcare) and loaded onto the MonoQ HR 16/10 anion exchange column (GE Healthcare) pre-equilibrated with 2% MonoQ elution buffer (25mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 8.0, 500mM NaCl, 10% Glycerol and 2mM βME) in MonoQ binding buffer (25mM HEPES pH 8.0, 10% Glycerol and 2mM βME). The column was equilibrated with two column volume of 2% MonoQ elution buffer in MonoQ binding buffer and washed with 42% MonoQ elution buffer in MonoQ binding buffer. The protein was eluted with a gradient from 42% to 75% MonoQ elution buffer in MonoQ binding buffer over 24 column volumes. The purity of the protein was assessed by SDS-PAGE and the oligomeric state of the protein was assessed by native-PAGE. The fractions containing the dimer protein were pooled together and quantified by NanoDrop 2000 (Fisher Scientific). The dimeric protein was concentrated to 15-30mg/ml using 30K MWCO Amicon Ultra-15 filtration and incubated for 4-8 hours, or overnight, at room temperature with 60mM 4'-phosphopantetheinyl transferase Sfp (plasmid provided by Dr Christian Chalut, CNRS), 10mM

magnesium chloride and 2.4mM aminoacyl coenzyme A [106]. The presence of the aminoacyl-4'ppant arm was assessed by LC-ESI-MS. The dimeric holo-protein was loaded onto the HiLoad 16/60 Superdex 200 size exclusion chromatography (SEC) column (GE Healthcare) pre-equilibrated with SEC buffer (25mM HEPES pH 8.0, 100mM NaCl and 0.2mM tris-(2-carboxyethyl)-phosphine (TCEP)) and eluted with one column volume of SEC buffer. The purity of the protein was assessed by SDS-PAGE and the oligomeric state of the protein was assessed by native-PAGE. The fractions containing the dimeric holo-protein were pooled together and concentrated by filtration to 15-30mg/ml. The protein was incubated at room temperature for 16 hours with 200mM HEPES pH 9.2, 13mM adenosine triphosphate (ATP) and 20mM α-keto-isocaproic acid (α-KIC). The modification of the 4' ppant arm with α-KIC was assessed by LC-ESI-MS. The sample was run through the SEC column as previously performed. The purity of the modified dimeric protein was assessed by SDS-PAGE. If necessary, another SEC was run according to the same protocol. The pure modified protein was used immediately for crystallization. If needed, excess protein was flash frozen in liquid nitrogen and conserved at -80°C in 20% glycerol.

2.4.2.3 Expression and Purification of StsA_AKr and StsA_AKrT_Bmdb_C3

StsA_AKr_(TEV)_His and StsA_AKrT_Bmdb_C3 were expressed and purified according to the same protocol as for StsA_AKrT_(TEV)_His, except that they may or may not be modified with the 4' ppant arm, they are not modified with α-KIC and they only require one run of SEC column to achieve and acceptable level of purity. The protein was either used immediately for crystallization or adenylation assay, or flash frozen in liquid nitrogen and conserved at -80°C at 15mg/ml in 20% glycerol.

2.4.2.4 Expression and Purification of TEV Protease

TEV protease was heterologously over-expressed in *Escherichia coli* Rosetta 2 (DE3) cells (EMD Millipore). 1L of Terrific Broth (TB) media supplemented with 42mM magnesium chloride, 34μg/ml chloroamphenicol and 100μg/ml ampicillin was inoculated with 10ml overnight culture and grown at 37°C until the culture reached an OD_{600} of 1.0. Protein expression was induced with 1mM IPTG and incubated at 16°C for 16-18 hours. The protein was purified by nickel column chromatography according to the same protocol as for LgrA_FA, except for the composition of TEV nickel column binding buffer (50mM Tris-HCl pH 8.0, 200mM NaCl, 25mM Imidazole, 10% glycerol and 2mM βME) and TEV

nickel column elution buffer (50mM Tris-HCl pH 8.0, 200mM NaCl, 250mM Imidazole, 10% glycerol and 2mM βME). The protein was concentrated if needed using 30K MWCO Amicon Ultra-15 filtration and further purified by SEC according to the same protocol as for StsA_AKrT_(TEV)_His, except for the composition of the SEC buffer (25mM Tris-HCl pH 7.5, 200mM NaCl, 2mM EDTA, 10% glycerol and 0.25mM TECEP). The protein was flash frozen in liquid nitrogen and stored at -80°C in 1mg aliquot in 10% glycerol.

2.4.2.5 Expression and Purification of Sfp Protein

Sfp plasmid was provided by Dr Christian Chalut (CNRS) and expressed according to the same protocol as for StsA_AKrT_(TEV)_His. SFP was purified by nickel column affinity chromatography followed by size exclusion chromatography according to the same protocol as for StsA_AKrT_(TEV)_His. The protein was flash frozen in liquid nitrogen and stored at -80°C at 4.5mg/ml in 10% glycerol.

2.4.2.6 Expression and Purification of Aminoacyl-coenzyme A synthesis enzymes

Plasmids pBac_PANK, pBAc_PPAT and pBac_DPCK were obtained from Dr Janice Reimer (Schmeing lab) and expressed according to the same protocol as for StsA_AKrT_(TEV)_His. PANK and DPCK were purified by nickel column chromatography according to the same protocol as for StsA_AKrT_(TEV)_His, except for the composition of the nickel column binding buffer (50mM Tris-HCl pH 7.5, 250mM NaCl, 2mM $CaCl_2$ and 2mM βME) and the nickel column elution buffer (50mM Tris-HCl pH 7.5, 500mM NaCl, 2mM $CaCl_2$, 500mM Imidazole and 2mM βME), and the protein was eluted with a gradient from 0% nickel column elution buffer to 100% nickel column elution buffer over 10 column volumes. The purity of the proteins was assessed by SDS-PAGE. The protein solutions were dialysed overnight according to the same protocol as for StsA_AKrT_(TEV)_His, except for the composition of the dialysis buffer (50mM Tris-HCl pH 7.5, 500mM NaCl, 2mM $CaCl_2$ and 2mM βME). The protein solutions were flash frozen in liquid nitrogen and stored at -80°C at 10mg/ml in 10% glycerol. PPAT was purified according to the same protocol as for PANK and DPCK, except that 10% glycerol was added to all buffers.

2.4.3 LC-ESI-MS on Intact Protein

LC-ESI-MS on intact protein was developed by Dr. Diego Alonzo, and was performed in this thesis and as described by Huguenin-Dezot *et al.* [11] using a deconvolution range of ±30kDa around the molecular weight of the protein under investigation.

2.4.4 Enzymatic Synthesis of Aminoacyl-Coenzyme A

The aminoacyl-coenzyme A (aminoacyl-CoA) used in this thesis was chemically synthetized and purified by Dr. Janice Reimer, using the synthesis enzymes PanK, PPAT and DPCK expressed and purified by Claudia Alonso. The strategy for synthesis of aminoacyl-CoA is to highjack the natural system for synthesis of CoA by cloning, expressing and purifying the three synthesis enzymes PanK, PPAT and DPCK, and allowing them to perform their sequential reaction using aminoacyl-pantothenic acid as starting material instead of the natural substrate pantothenic acid. The phosphoenolpyruvate/pyruvate kinase (PEP/PK) regeneration system was used in complement with pyrophosphatase. The "all-in-one-pot" 30ml reaction mix containing 20 mM KCl, 10 mM $MgCl_2$, 50 mM Tris pH 9, 3 mM aminopant, 30 mM ATP, 9.6mg PanK, 9.6mg PPAT, 9.6mg DPCK, 30mM PEP, 8 Units/ml PK enzyme (Sigma) and 6 Units/ml protein pyrophosphatase (Roche), adjusted to pH 9.0 was incubated in a 37°C water bath for 2 hours. An additional 9.6mg aliquot of DPCK was added to the reaction mix along with 2ml 100mM ATP, the pH is adjusted to 9.0 and the reaction was placed in a 37°C water bath to incubate overnight. The reaction was spun down using a 10kDA spin cut-off filter, collecting the flow through, in order to remove the enzymes from the reaction mix.

2.4.5 Vinyl-Sulfonamide Inhibitor

The αKIC and ICA vinyl-sulfonamide inhibitor described in this thesis were kindly provided by John Colucci and Zamboni Chem Solution INC.

2.4.6 Crystallography and Diffraction Data

The crystallization of A-KR-PCP proved challenging and required several iterations to obtain crystals of sufficiently high diffraction quality. The initial crystallization conditions were acquired by sparse matrix screens (Qiagen) in 96-well sitting drop format with apo or amino-ppant modified A-KR-PCP pure dimer sample. Despite thorough optimization of these initial conditions no improvement in

morphology or diffraction quality was obtained. A method for sparse-matrix microseeding resulted in a number of new conditions and following iterative rounds of optimization, including additive and detergent screens, streak microseeding, and switching to a vapor diffusion hanging-drop setup, a condition with single crystals diffraction to approximately 7 Å was obtained [107]. When advanced cryoprotection methods were employed, the resolution was improved to 3.4Å in diffraction experiments.

The structure of StsA A-KR-PCP was obtained from crystals of purified A-KR-PCP dimer protein incubated for 16 h at room temperature with Sfp (1:1.22 molar ratio), 10 mM $MgCl_2$, and 3 mM amino-CoA (kindly provided by Janice Reimer) [92]. Phosphopantetheinylation was verified by LC-MS and the sample was incubated for 16 hours in the presence of 20 mM α-KIC and 13 mM ATP in 200mM HEPES pH 9.2. Following LC-MS analysis the sample was purified by SEC (Superdex-200 16/60). The fractions containing the pure protein for crystallization were recovered and concentrated to 15.5 mg/mL. Hanging drop vapor diffusion crystallization with three 4 μL drops per coverslip and a 500 μL reservoir solution of 0.167M sodium succinate, 0.25% v/v PEG 2000 MME, 17mM HEPES pH 7.0, 17mM HEPES pH 8.0, 67mM sodium chloride, 6% v/v PEG-3350, and 0.105 M calcium acetate was achieved. In order to make the drop, 4 solutions were added in a specific order to a master mix. 4 μL of solution A (1M sodium succinate, 1.5% v/v PEG 2000 MME, 0.1M HEPES pH 7.0, 0.025M HEPES pH 8.0 and 0.1M NaCl), 8 μL solution B (18% v/v PEG-3350 and 0.314 M calcium acetate), 2.4μl of 0.1M phenol solution (Hampton Additive Screen solution n#27), 2.4 μL of solution C (50mM αKIC, 50mM ATP and 100mM MgCl) and 9.6 μL of protein solution at 15mg/ml was added in this order. The master mix was thoroughly mixed and the added to three 4 μL drops on two separate coverslips. These were sealed on top of two separate wells and crystals appeared within minutes and grew to their final size in 24 hours.

The cryoprotection was performed by sequential addition of increasing volumes and percentages of ethylene glycol of three times 7 μL of 7%, three times 7 μL of 14%, three times 7 μL of 21%, three times 7 μL of 28% v/v. This happened in solutions with identical contents to the equilibrated crystallization drop with a five-minute wait time for equilibration after each addition. After the stepwise addition of cryoprotectant, 86 μL of solution was removed from the drop, and 86 μL of 28% v/v ethylene

glycol was added to. To allow for crystal dehydration, the cryo-protected drop was exposed for 30 to 45 minutes to the environment at room temperature before flash freezing under a liquid nitrogen stream. Data was collected at 100 K at the APS-NECAT 24-ID-C beamline with a Dectris PILATUS 6MF pixel array detector.

2.4.7 Data Processing and Structure Elucidation

Data were indexed and integrated in the space group $P2_12_12_1$ with the program DIALS [108] or iMOSFLM [109] and scaled with the program AIMLESS [110]. The molecular replacement ensemble search model was generated using SWISSMODEL [111] from structures of domains showing high levels of homology with the A and KR domains identified with HHsearch [112]. An initial MR solution with 4 KR domains was obtained. Subsequently, 2 A_{core} domains were placed through iterative runs of AUTOBUILD [113]. The model refined to a R_{free} of 42% and stalled there despite numerous rounds of modelling and refinement. A combination of iterative steps of manual model building in Coot [114] on B-factor sharpened- and NCS-averaged maps improved the model. The iterative omit-map sharpening and non-crystallographic symmetry averaging processes were conceptualized and performed by Jimin Wang, as described the manuscript in preparation *On sharpening omit-maps for extraction of weak structural information* by Wang *et al*. [115]. 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. B-factor sharpening required manual iterative optimization. Refinements were performed using Refmac and phenix.refine [110, 113]. The final model refined to a R_{free} of 28% using data to 3.4 Å resolution. Once the higher-resolution A domain structure (presented in chapter III) was determined, it was used to guide refinement of the A-KR-PCP structure (presented in this chapter).

3 Chapter III – Characterization of Separate Domains from the Initiation Module A-KR-PCP of *Bacillus stratosphericus*

3.1 The Ketoacid-Selecting Adenylation Domain of the Initiation Module of Stratospherilude Synthetase

3.1.1 Introduction – Linear gramicidin and its biosynthesis

Linear gramicidin and its biosynthesis has been well studied since its discovery in the 1940s [116]. It is a linear pentadecapeptide capable of forming membrane channels in Gram-positive bacteria. Gramicidin A presents an alternating sequence of L- and D-amino acid, shown in Figure 20, granting the compound its unique ability to form a β-helix-like structure [117].

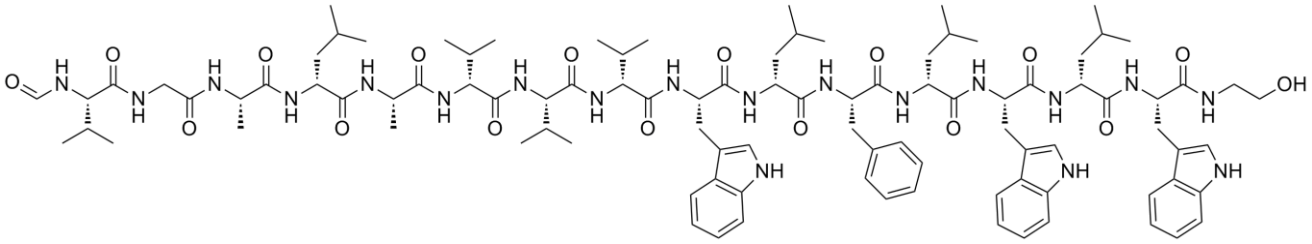


Figure 20: The structure of linear gramicidin A. The primary sequence of linear gramicidin A is Formyl-L-Val-Gly-L-Ala-D-Leu-L-Ala-D-Val-L-Val-D-Val-L-Trp-D-Leu-L-Phe-D-Leu-L-Trp-D-Leu-L-Trp-ethanolamine. The N-terminal formyl group and C-terminal ethanolamine group protect linear gramicidin A from degradation by proteases.

The initiation module of linear gramicidin synthetase A (LgrA) was structurally and functionally characterized by Reimer *et al.* [10]. The initiation module contains an F domain, an A domain and a PCP domain arranged in a F-A-PCP architecture. The A domain selects and adenylates an L-Valine before loading it onto the PCP domain. The substrate is then presented to the F domain and formylated using 10-formyl-tetrahydrofolate (N^{10}-formyl-THF) to form L-formyl-valine. The formylated substrate is then brought to the donor site of the C domain for condensation. The first adenylation domain of LgrA

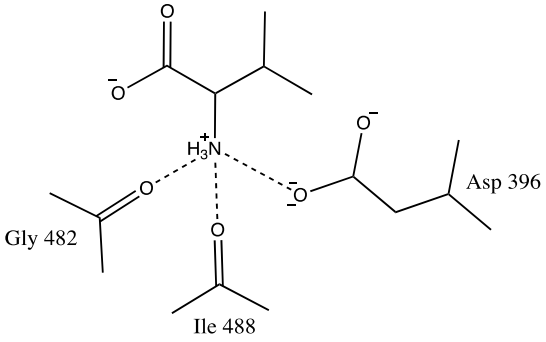


Figure 21: α-substituent specific active site residues of the first A domain of LgrA.

follows the Stachelhaus rule for coordination of the substrate through the α-substituent: the α-amino group of the substrate is coordinated by the side chain of Asp395 (1AMU Asp235), as well as the backbone carbonyl of Gly482 (1AMU Gly324) and Ile488 (1AMU Ile330), as shown in Figure 21.

The mechanism of specificity of α-amino acid-selecting A domains for the α-substituent (i.e. the α-amino group in the case of α-amino acids) of their substrate was characterized by Stachelhaus *et al.* and confirmed by many structures of NRPS A domains in complex with their substrate [10, 13, 80, 118]. However, the mechanism of specificity of α-ketoacid-selecting A domains for the α-substituent (i.e. the α-ketone group in the case of α-ketoacids) of their substrate remains unknown. The resolution of the structure of the A-Kr-PCP initiation module of Sts allows for the visualization of the overall domain arrangement and large-scale conformation of the backbone, but a higher resolution model of the A domain in complex with its substrate is needed in order to unravel the mechanism of α-substituent specificity, and potentially use this knowledge to develop new drugs including ester bonds as well as peptide bonds.

3.1.2 Results

The objective for this section is to obtain a high-resolution structure of the first A domain of StsA in complex with substrate, and to investigate the mechanism for α-substituent specificity in NRPS A domain through biochemical assays, using the first A domain of LgrA and StsA as models for an α-amino acid-selecting A domain and an α-ketoacid-selecting A domain respectively.

3.1.2.1 Crystallography of StsA_A

The goal for this section is to obtain a high-resolution structure of the α-ketoacids-selecting A domain of StsA in complex with its cognate substrate α-KIC in order to investigate the mechanism of α-substituent specificity in α-ketoacids-selecting A domains. To do so, the pBac_StsA_A_(TEV)_His

plasmid was successfully sub-cloned. StsA_A expresses well with a yield of 1.5mg of protein per gram of cell pellet in the nickel column eluate, with an estimated purity of 95% based on SDS-PAGE. Approximately 1mg of protein per gram of cell pellet is recovered after running the protein solution through the MonoQ and the SEC, in a single oligomeric state with an estimated purity of 99%.

Figure 22: StsA_A crystals.

After sparse matrix, a 100µm square-based-bipyramidal-shaped crystals (see morphology in Figure 22) formed after 6 months in the condition 0.1M NaCl, 0.1M HEPES pH7.5 and 1.6M ammonium sulfate, and diffracted to 1.89Å. After several rounds of optimization, crystals were grown in 2 days in the precipitation solution 1.65M $AmSO_4$ and 22.5% glycerol, with 6-AHA or TMAO additive. The protein was co-crystallized with αKIC, but also αKIV, αKIL and ICA. Several complete data sets were obtained, and the StsA_A structure in complex with adenylated-αKIC was solved with a resolution of 2.5Å.

The overall structure of the A domain is the same as the one observed in the 3.5Å resolution StsA_AKr structure. The structure StsA_A was superimposed with the structure of the phenylalanine-activating A domain from gramicidin S (PDB:1AMU) in order to identify the highly conserved A domain motifs in StsA, the sequences of which are shown in the second column of Table 2. The sequences of these motifs in StsA were used to identify the highly conserved motifs in other α-ketoacids-selecting A domains based on primary sequence alignment; the consensus sequences thusly identified are shown in the third column of Table 2.

Table 2: Highly conserved A domain motifs and consensus in α-ketoacid selecting A domains. Sequence in StsA corresponding the A domain motifs A1 to A9 were identified based on 3D structure alignment with the phenylalanine-activating A domain from gramicidin S (PDB:1AMU) [13]. Consensus was determined based on the primary sequence alignment of StsA A domain with characterized α-ketoacid-selecting A domains, including Ces, Vlm and CrpD [6, 11, 57].

Highly conserved A domain motifs in α-amino acid-selecting A domains	Sequence in StsA A domain	Consensus in α-ketoacid-selecting A domain
(A1) LTYxEL	QSYAQL	LSYAxL
(A2) LKAGxAYVPID	CVLGGFVPTLVS	CVLGGFIPVPIx
(A3) LAYxxYTSGTTGxPKG	SVFFILTSGSTGMPKC	xAKIIxTSGSTGxPKG
(A4) FDxS	HIGG	HVGG
(A5) NxYGPTE	PAFGMSE	PxWGMSE
(A6) GELxIxGxGLARGYW	RVQIKGPTTMKGY	xLQ(I/V)KGxTVxxGY
(A7) YKTGDQ	FHTGDLG	FDTGDL(G/A)
(A8) GRxDxQVKIRGxRVELEEVE	GREKDMIIINGKNYHNYEIE	GRxKDIIIINGVNYxNxEIE
(A9) LPxYMIP	KMGLSAS	KxGISxS
(A10) NGKIDR	SGKIER	IGKIQR

As seen in Table 2, some motifs (such as A10 and A3) are very well conserved between α-amino acids- and α-ketoacids-selecting A domains; others, such as motifs A1, A2, A6, A7 and A8, are fairly similar, with some differences conserved among α-ketoacids-selecting A domain. In contrast, the motifs A4 and A5, located near the substrate and interacting with it, are highly conserved in terms of sequence among α-ketoacid-selecting A domains, and are vastly different from the consensus established in α-amino acid-selecting A domains, making them regions of interest regarding the mechanism of α-substituent specificity.

3.1.2.2 Stachelhaus Residues and Side Chain Specificity of StsA_A

In order to investigate the substrate specificity of the α-ketoacid-selecting A domains in more details, we set out to determine the side-chain specificity of StsA_A by performing adenylation assays

using cognate substrate αKIC (α-ketoisocaproic acid or α-keto-leucine) as well as αKIV (α-ketovaline), αKIL (α-keto-isoleucine) and pyruvate (α-ketoalanine), the results of which are shown in Figure 23.a.

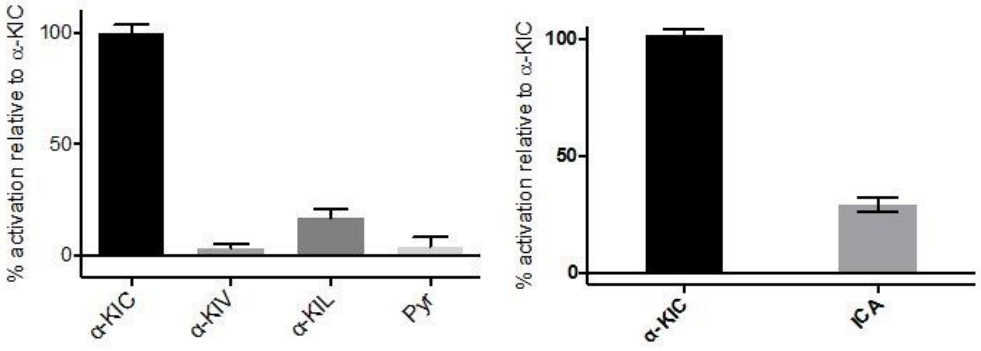


Figure 23: Side-chain specificity of StsA_A (a) Initial rate of adenylation of StsA_A with αKIC, αKIV, αKIL and pyruvate (b) Initial rate of adenylation of StsA_A αKIC and ICA.

As seen in Figure 23.a., StsA_A is able to adenylate αKIL, albeit at a lower rate than cognate substrate αKIC. StsA_A does not show a significant level of adenylation activity using αKIV and pyruvate as substrate.

In order to investigate the affinity of StsA-A for the α-substituent of its substrate αKIC, adenylation assays were performed using ICA as a substrate (i.e. αKIC with no α-keto substituent), as shown in Figure 23.b. StsA_A is able to adenylate ICA, albeit at a lower rate than αKIC, suggesting that the removal of the α-keto substituent does not completely abolish binding but drastically reduces it.

In order to study the binding residues of the active site of α-ketoacid-selecting A domains in more details, the Stachelhaus residues (described in more detail in Chapter I) in StsA_A were identify by structural superposition with the phenylalanine-activating A domain from gramicidin S (PDB:1AMU); the Stachelhaus residues found in other α-ketoacid-selecting A domains were identified through primary sequence alignment with StsA, the results of which are shown in Table 3.

Table 3: Extension to the Stachelhaus code for aliphatic α-ketoacid-selecting A domains. The position of the Stachelhauss residue was determined based on 3D alignment of the StsA A domain structure with the phenylalanine-activating A domain from gramicidin S (PDB:1AMU) [13]. The nature of the corresponding Stachelhauss residues in characterized α-ketoacid-selecting A domains from Ces, Vlm and CrpD [6, 11, 57] was determined based on primary sequence alignment with StsA A domain.

	Position in 1AMU	235	236	239	278	299	301	322	330	331	517
α-ketoacids	Residue in 1AMU	D	A	W	T	I	A	A	I	C	K
	Stra_A1 (αKIC)	I	G	M	W	I	G	A	S	A	K
	CesA_A1 (αKIC)	V	G	M	W	V	G	T	S	G	K
	CrpD_A2 (αKIC)	V	A	Y	F	L	G	S	S	G	K
	CesB_A3 (αKIV)	V	G	M	W	V	A	V	S	G	K
	Vlm_A1 (αKIV)	A	A	M	W	I	A	V	S	G	K
	Vlm_A3 (Pyr)	V	V	H	W	I	A	E	N	M	K
α-amino acids	Position in 1AMU	235	236	239	278	299	301	322	330	331	517
	Residue in 1AMU	D	A	W	T	I	A	A	I	C	K
	Leu (1)	D	A	W	F	L	G	N	V	V	K
	Leu (2)	D	A	W	L	Y	G	A	V	M	K
	Leu (3)	D	G	A	Y	T	G	E	V	V	K
	Leu (4)	D	A	F	M	L	G	M	V	F	K
	Val (1)	D	A	F	W	I	G	G	T	F	K
	Val (2)	D	F	E	S	T	A	A	V	Y	K
	Val (3)	D	A	W	M	F	A	A	V	L	K
	Ala	D	L	L	F	G	I	A	V	L	K

As shown in Table 3, the Stachelhaus residues differ between α-ketoacids and α-amino acids in the following way: at position 235, the conserved Asp residue (involved in direct binding with the α-substituent) is replaced by a branched amino acid for α-ketoacids, most likely Val. At position 239, the bottom of the side chain cleft is capped by a Met residue for α-ketoacids, instead of a Trp residue for

α-amino acids. At position 330 and 331, located in close proximity to the α-substituent, can be found a Ser and a Gly respectively, conserved among α-ketoacid-selecting A domains.

3.1.2.3 The α-substituent-specific Stachelhaus residues and supporting residues

Through primary sequence alignment and structural superposition were identified three key regions of the A domain's active site interacting with the α-substituent of the substrate: Asp235, Gly324 and Ile330, along with their supporting residues. In this section, each region will be investigated in more details, using the first A domain of LgrA and StsA as models for an α-amino acid-selecting A domain and an α-ketoacid-selecting A domain respectively.

The first region, surrounding residue Asp235 in 1AMU, was studied in LgrA and StsA, as shown in Figure 24.

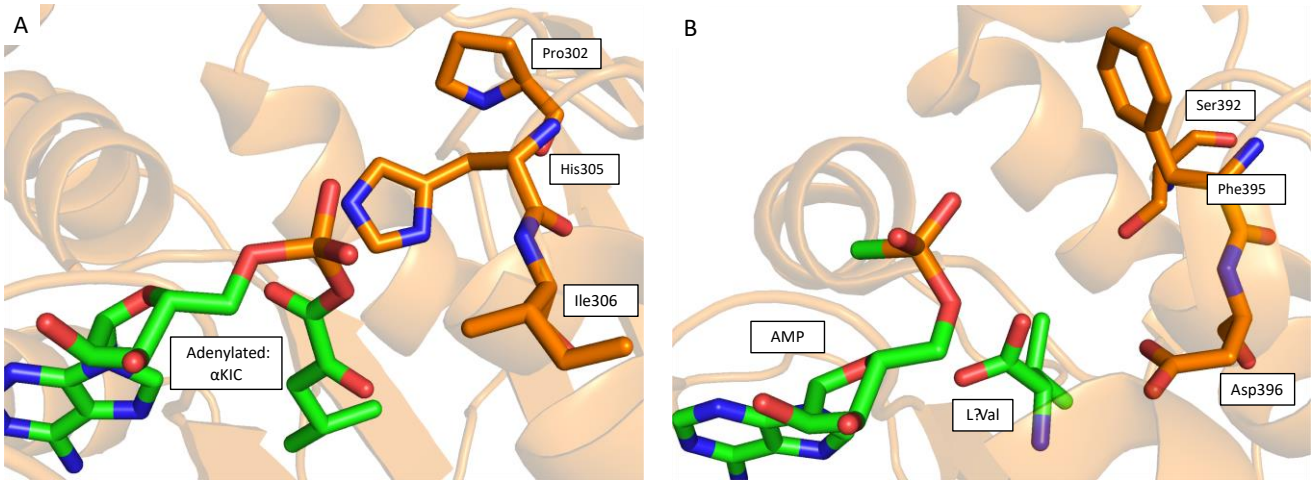
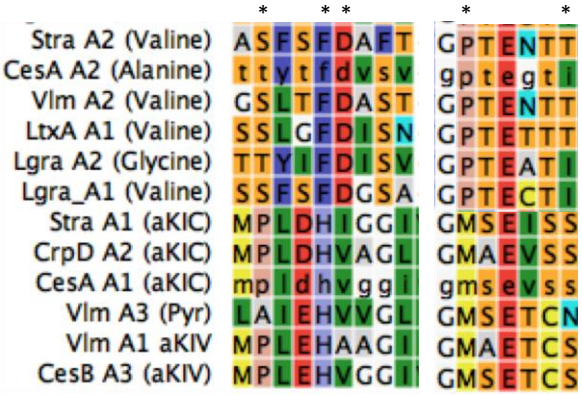


Figure 24: Interaction of Stachelhaus residue Asp235 and supporting residues with α-ketoacid αKIC in StsA and with α-amino acid L-Val in LgrA. (a) StsA residue Ile306 with supporting residues His305 and Pro302 (b) LgrA residue Asp396 with supporting residues Phe395 and Ser392.

Residue Asp396 in LgrA_FA (1AMU Asp325) aligns with residue Ile306 in StsA (Figure 24). The neighboring residue in StsA His305 aligns with residue Phe395 in LgrA_FA, and is rotated towards the substrate. Additionally, the conserved Ser392 in LgrA_FA is mutated to Pro302 in StsA, which seems to be bending the loop region and possibly influences the position of residues His305 and Ile306 in StsA. to His mutation, the Asp to Ile -selecting A domains.



e-based alignment of regions of the α-substituent of the acid and α-ketoacid selecting A ith * from left to right are the rA) / Pro302 (StsA); residues s305 (StsA); residues Asp396 sA); residues Pro483 (LgrA) / dues Ile448 (LgrA) / Ser 420 (StsA).

The second region, surrounding residue Gly324 in 1AMU, was studied in LgrA and StsA, as shown in Figure 26.

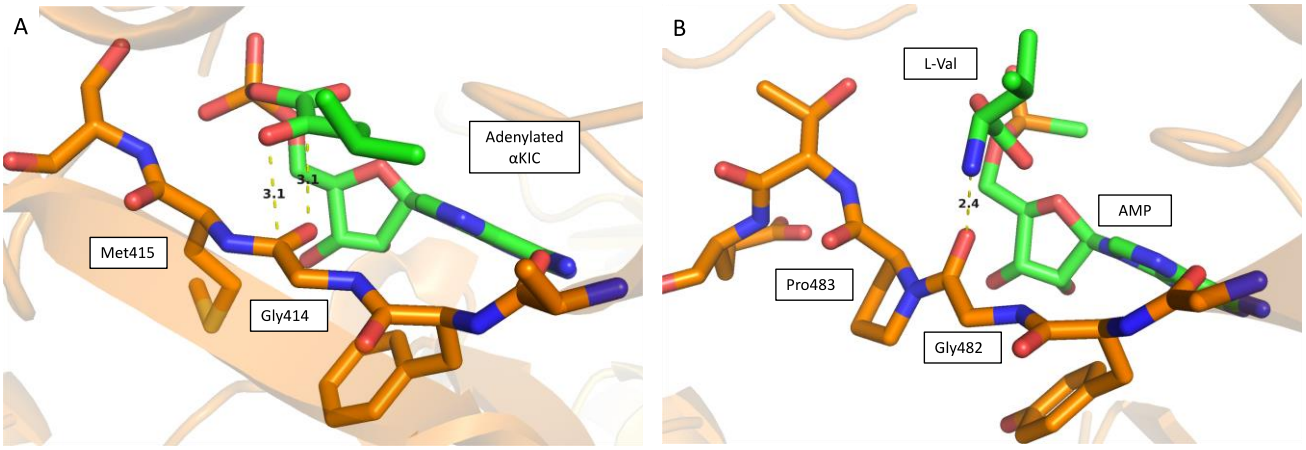


Figure 26: Interaction of Stachelhaus residue Gly 324 with α-ketoacid αKIC in StsA and with α-amino acid L-Val in LgrA. (a) StsA residue Gly414 with supporting residue Met415 (b) LgrA residue Gly482 with supporting residue Pro483.

The Stachelhaus residue Gly324 is conserved in StsA with Gly414, and LgrA with Gly482. However, the backbone carbonyl of Gly414 in StsA is rotated about ~45° with respect to the backbone carbonyl of Gly482 in LgrA and Gly324 in 1AMU, as shown in Figure 26. In this conformation, the backbone carbonyl of Gly414 in StsA interacts with the α-carbonyl group of αKIC through a reciprocal carbonyl-carbonyl interaction. The neighboring residue in StsA Met415 aligns with residue Pro483 in LgrA_FA, and seems to be bending the backbone in this region to allow for Gly414 to coordinate the substrate. Pro483 and Met415 are conserved among α-amino acid-selecting and α-ketoacid-selecting A domains respectively (Figure 25).

The third region, surrounding residue Ile330 in 1AMU, was studied in LgrA and StsA, as shown in Figure 27.

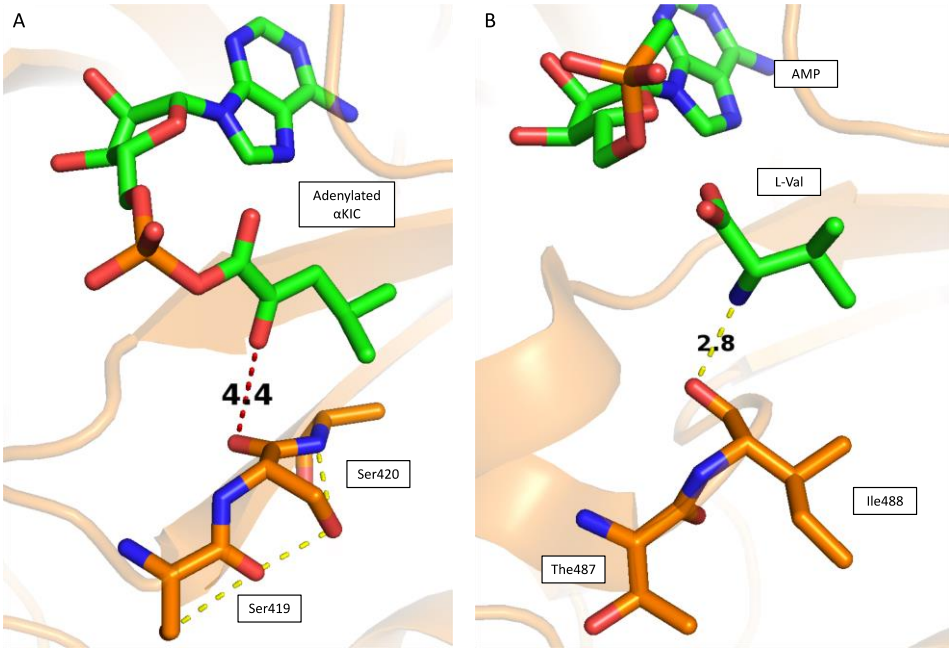


Figure 27: Interaction of Stachelhaus residue Ile330 with α-ketoacid αKIC in StsA and with α-amino acid L-Val in LgrA. (a) StsA residue Ser420 with supporting residue Ser419 (b) LgrA residue Ile488 with supporting residue Thr487.

Residue Ile488 in LgrA_FA (1AMU Ile330) aligns with residue Ser420 in StsA (Figure 27). The hydroxyl group of Ser420 engages in hydrogen-bonding with the backbone carbonyl of the residue directly upstream, Ser419. As a result, the backbone carbonyl of Ser420 is rotated ~80° away from the substrate with respect to residue Ile488 in LgrA_FA structure (PDB:5ES6). The Ser variation is conserved among α-ketoacid-selecting A domains (Figure 25).

3.1.2.4 α-Substituent Specificity of the A domain

The goal of this section is to investigate the mechanism of α-substituent specificity in NRPS A domains through biochemical assay, by focusing on the three regions of interest described above (i.e. Asp235, Gly324 and Ile330, along with their supporting residues).

In order to characterize α-substituent specificity of the wild-type CesA, StsA and LgrA A domains, adenylation assays were performed on StsA_AKrT and CesA_AKrT using L-leucine and αKIC as substrate, and on LgrA_FA using L-valine and αKIV, as shown in Figure 28.

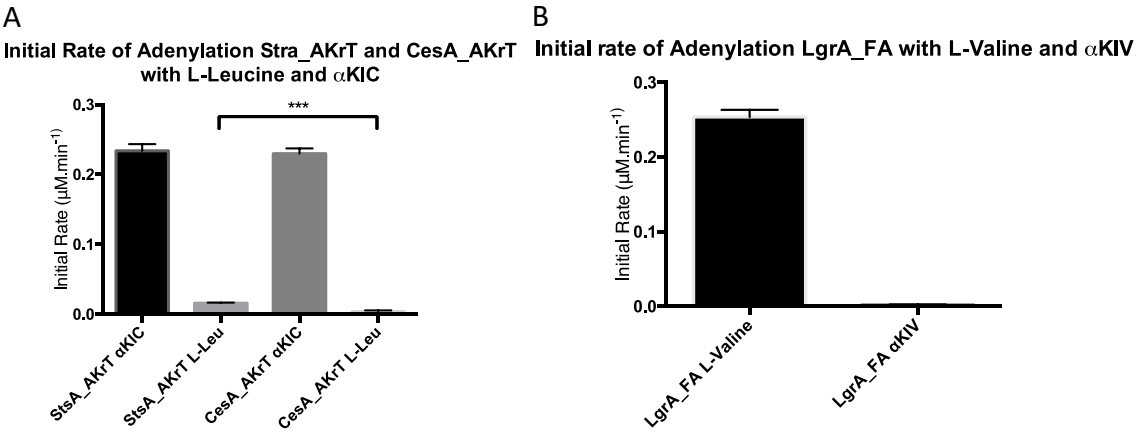


Figure 28: Wild type StsA, LgrA and CesA A domain specificity for the α-substituent of their substrate (a) Adenylation assay on StsA_AKrT and CesA_AKrT using cognate substrate αKIC and α-amino equivalent L-leucine. (b) Adenylation assay on LgrA_FA using cognate substrate L-valine and α-keto equivalent αKIV

Figure 28 shows CesA_AKrT and LgrA_FA to have a very high specificity for their cognate substrate (i.e. αKIC and L-valine, respectively) and do not present a significant level of adenylation activity for the non-cognate α-substituent version of their substrate (i.e. L-leucine and αKIV respectively). StsA_AKrT adenylates cognate substrate αKIC with a similar rate from CesA; however, StsA_AKrT is also able to adenylate the α-amino version of its substrate (i.e. L-leucine) with a rate corresponding to 6.8% of the adenylation rate measured using cognate substrate αKIC.

In order to investigate the role of Asp235 and putative supporting residues regarding the specificity of the A domain for the α-substituent of its substrate, constructs StsA_AKrT_H305F, StsA_AKrT_I306D, StsA_AKrT_H305F/I306D, LgrA_FA_F395H/D396V and LgrA_FA_F395H/D396V/S392P were successfully sub-cloned and purified with the same purity and approximate yield as their wild-type counterparts. Constructs LgrA_FA_F395H, LgrA_FA_D396V were successfully sub-cloned, but expression trials were not attempted for these constructs. This series of constructs are called the Asp mutants. The affinity of the Asp mutants for the cognate and non-cognate α-substituent version of their substrate was investigated by adenylation assay shown in Figure 29.

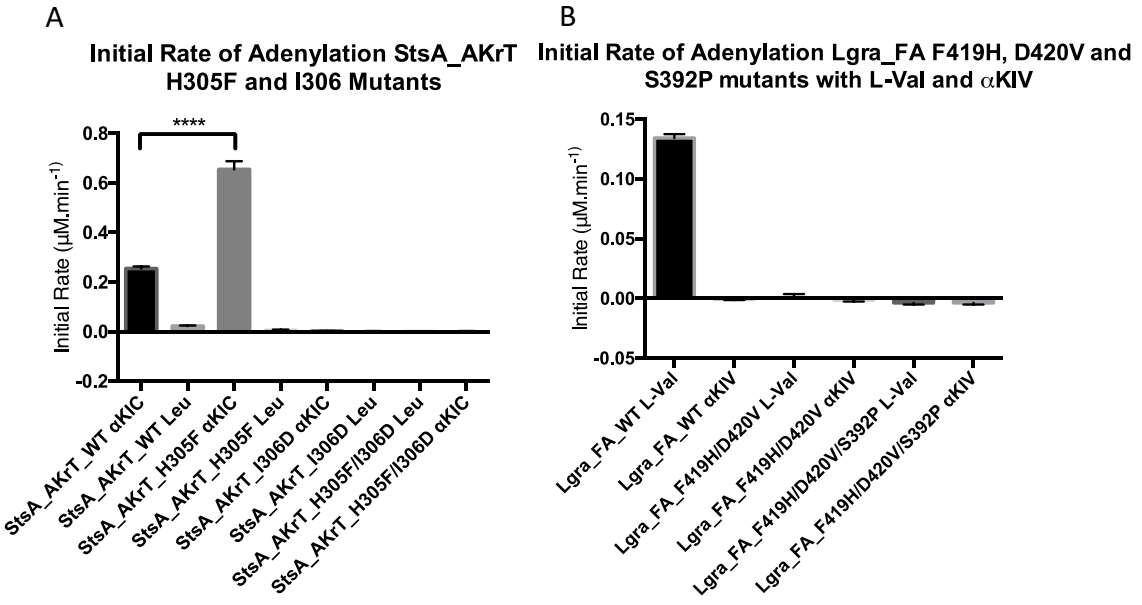


Figure 29: Relative adenylation activity of LgrA and StsA A domain Asp mutants with cognate substrate and α-keto or α-amino counterpart. (a) Adenylation assay on StsA_AKrT_WT, StsA_AKrT_H305F, StsA_AKrT_I306D and StsA_AKrT_H305F/I306D with cognate substrate αKIC and α-amino counterpart L-leucine. (b) Adenylation assay on LgrA_FA_WT, LgrA_FA_F395H/D396V and LgrA_FA_F395H/D396V/S392P with cognate substrate L-valine and α-keto counterpart αKIV.

None of the StsA_AKrT Asp mutants have a significant level of adenylation activity for the non-cognate L-leucine (Figure 29). However, StsA_AKrT_H305F mutants presents a rate of adenylation significantly superior to that of the wild type using cognate substrate αKIC. Furthermore, none of the LgrA_FA Asp mutants have a significant level of adenylation activity for the non-cognate αKIV substrate.

In order to investigate the role of Gly414 and putative supporting residue in specificity of the A domain for the α-constituent of its substrate, constructs StsA_AKrT_M415P, StsA_AKrT_I306D/M415P, StsA_AKrT_H305F/I306D/M415P, LgrA_FA_P483A, LgrA_FA_D396V/P483A, LgrA_FA_F395H/D396V/P483A, LgrA_FA_P483M, LgrA_FA_D396V/P483M and LgrA_FA_F395H/D396V/P483M were successfully sub-cloned and purified with purity and approximate yield as their wild-type counterparts. This series of constructs are called the Gly mutants. The affinity

of the Gly mutants for the cognate and non-cognate α-substituent version of their substrate was investigated by adenylation assay shown in Figure 30.

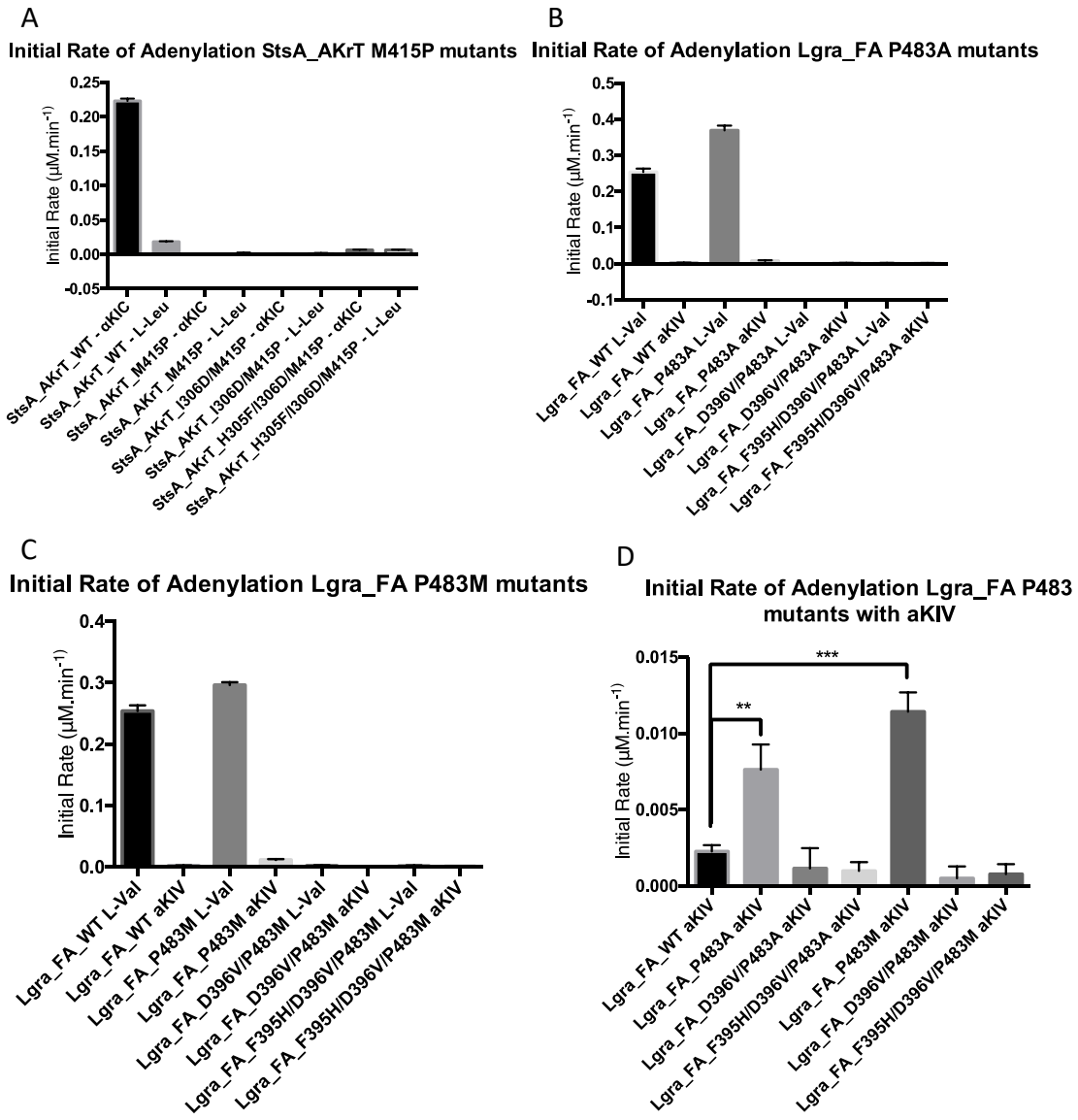


Figure 30: Relative adenylation activity of LgrA and StsA A domain Gly mutants with cognate substrate and α-keto or α-amino counterpart. (a) Adenylation assay on StsA_AKrT_WT, StsA_AKrT_M415P, StsA_AKrT_I306D/M415P and StsA_AKrT_H305F/I306D/M415P with cognate substrate αKIC and α-amino counterpart L-leucine. (b) Adenylation assay on LgrA_FA_WT, LgrA_FA_P483A, LgrA_FA_D396V/P483A and LgrA_FA_F395H/D396V/P483A with cognate substrate L-valine and α-keto

counterpart αKIV. (c) Adenylation assay on LgrA_FA_WT, LgrA_FA_P483M, LgrA_FA_D396V/P483M and LgrA_FA_F395H/D396V/P483M with cognate substrate L-valine and α-keto counterpart αKIV. (d) Adenylation assay on LgrA_FA_WT and full LgrA_FA Gly mutant series with non-cognate αKIV.

None of the StsA_AKrT Gly mutants have a significant level of adenylation activity for the non-cognate L-leucine (Figure 30). However, the LgrA_FA_P483A and LgrA_FA_P483M mutants have a non-zero initial rate of adenylation for non-cognate substrate αKIV, and have a significantly higher level of activity for non-cognate substrate αKIV than the wild type (3.4% and 5% of the activity of StsA_AKrT with αKIC for LgrA_P483A and LgrA_P483M respectively). It is worth noting that the LgrA_FA_P483A and LgrA_FA_P483M mutants also have a significantly higher level of activity for cognate substrate L-Valine than the wild type, with 45.5% and 16.7% increase respectively. The other LgrA_FA Gly mutants do not present a significant level of activity for cognate substrate L-Valine.

3.1.2.5 Crystallography of LgrA_FA_P483M

LgrA_FA_P483M was crystallized and soaked with αKIV substrate according to the protocol described by Reimer et al. [10]. After rigid body refinement using LgrA_FA_WT (PDB 5ES6) as a model, unbiased density for AMPcPP and αKIV was revealed. Although the structure is still under refinement, the backbone carbonyl of Gly482 is visibly rotated ~45° as seen in StsA.

3.1.3 Discussion

3.1.3.1 Side Chain Specificity of the Adenylation Domain of Stratospherulide Synthetase

The A domain of StsA has a fairly high side chain specificity: it may adenylate αKIL, with ~20% of the rate of adenylation of the cognate substrate αKIC, and cannot adenylate pyruvate or αKIV.

The Stachelhaus code is not entirely applicable to α-ketoacids selection. The A_{sub} domain catalytic Lys517 however is conserved in α-ketoacid-selecting A domains, and position 236 is in accordance with the Stachelhaus code (for example, the pyruvate-selecting A domain presents a bulkier group at this position since pyruvate doesn't have a gamma-carbon); positions 278, 299, 301 and 322

follow the Stachelhaus code as well. However, the conserved Asp235, which interacts directly with the α-substituent of the substrate, is mutated to a branched aliphatic residue in α-ketoacid-selecting A domains. This was surprising, since we expected a hydrogen-bond donor to occupy the position and interact with the α-keto group. Positions 330 and 331 are occupied by a serine and a glycine, respectively, in most of the α-ketoacid-selecting A domains aligned. The bottom of the side chain-binding cleft is capped by a Met residue for keto-selecting A domains at position 239 instead of a Trp residue. It is difficult to determine a side-chain specific code for α-ketoacid-selecting A domains at this time, due to the limited number of sequence available for alignment.

3.1.3.2 The ATP-Binding Residues of the Adenylation Domain of Stratospherulide Synthetase

The substrate-binding residues of the active site of the α-amino acid-selecting A domain from gramicidin S (PDB:1AMU) and the α-ketoacid-selecting A domain from StsA are shown in Figure 31.

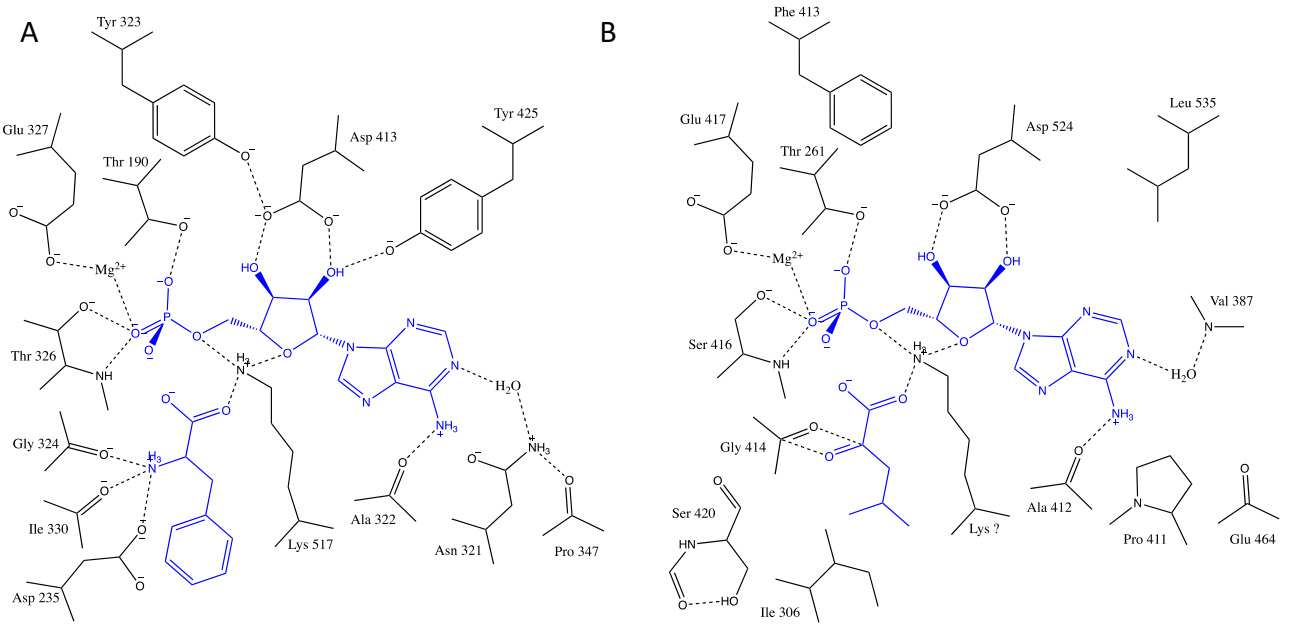


Figure 31: Catalytic and substrate-binding residues of 1AMU α-amino acid-selecting A domain and StsA α-ketoacid-selecting A domain. (a) 1AMU (b) StsA

As shown in Figure 31, the coordination of the ribose moiety of ATP is conserved in StsA, but the added stability provided by the supporting Tyr residues is lost. Asp413 in 1AMU and Asp524 in StsA

both coordinate the 2' and 3' hydroxyl group of the ribose, but Tyr 425 and Tyr 323 in 1AMU are mutated to Leu535 and Phe413, respectively, in StsA.

As shown in Figure 31, the strategy for the coordination of the adenine moiety of ATP is conserved in StsA, but carried out by different active site residues. Ala332 in 1AMU is conserved with Ala412 in StsA, and coordinates the 6-amino group in the same way. However, Asn321 in 1AMU is mutated to Pro411 in StsA, losing the ability to coordinate the 1-N of adenine through a water molecule as seen in 1AMU. Instead, the water molecule is coordinated by the backbone amide of Val387 in StsA.

The coordination of the α-phosphate of ATP is conserved in StsA (Figure 31): interaction with Thr326, Thr190 and Glu327 in 1AMU is conserved with in Ser416, Thr261 and Glu417 respectively in StsA. Unfortunately, the catalytic Lys residue is not visible in the StsA structure.

3.1.3.3 Specificity of the Adenylation Domain of Stratospherulide Synthetase for the α-Substituent of its Substrate

StsA_AKrT can adenylate ICA and leucine with 40% and 6.8% of the rate of adenylation measured for αKIC, respectively. This indicates that StsA_A presents a somewhat loose specific for the α-substituent of its substrate. The fact that it can so easily accommodate ICA shows that the interaction with the α-carbonyl group of αKIC is not crucial for substrate binding. We can then expect most of evolution of the active site to have occurred in order avoid a negative interaction due to the introduction of a partially negatively charged carbonyl group, instead of a positively charged amino group, and the number of positive interactions between the α-carbonyl group of αKIC and the active site residues to be limited.

Through structure alignment and study of the structure of the A domain of StsA were identified three regions of interest involved in the binding of the α-substituent of the substrate, as shown in Figure 32: in the α-amino acid-selecting A domain of LgrA, the α-amino group is coordinated by the side chain of Asp396 backbone carbonyl groups of Gly482 and Ile448; in contrast, in the α-ketoacid-selecting A domain of StsA, the α-ketone is coordinated by the backbone carbonyl of Gly414 through a reciprocal carbonyl-carbonyl interaction, while the interaction with the other two residues is lost. The rest of this section discusses the three regions of interest in more details.

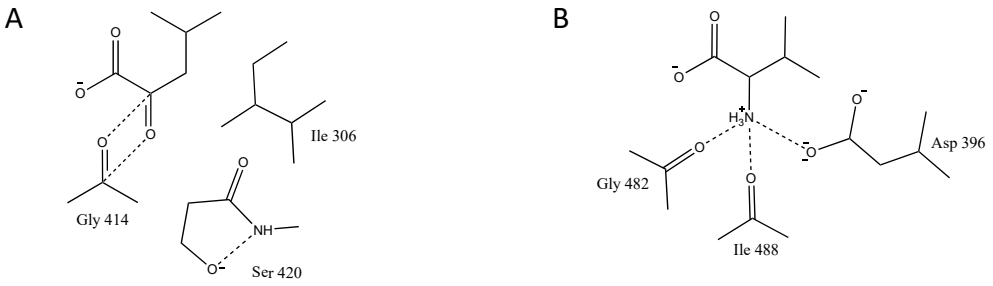


Figure 32: Schematics of the mechanism of α-substituent specificity in StsA and LgrA. (a) StsA (b) LgrA

3.1.3.3.1 The Asp/Ile and Supporting Residue

Asp235 is replaced by Ile306 conservatively among α-ketoacid-selecting A domains. This was surprising since Ile is hydrophobic, it is then curious to see it so close to the α-ketone. Our hypothesis was that this simply removes a negative interaction between the negatively charged Asp and the partially negatively charged carbonyl oxygen. However, the Ile to Asp mutation in StsA did not provide the A domain with the ability to adenylate α-amino acids, and the reverse Asp to Val mutation in LgrA did not allow the LgrA_A domain to adenylate α-ketoacids. Therefore, the Asp to Ile mutation in α-amino acid-selecting A domains is not the driving force for selectivity of the α-substituent. The Ile to Asp mutation did however abolish activity of StsA_A for its cognate substrate αKIC, so the absence of Asp is one of the conditions for adenylation of α-ketoacids. As a future experiment, the Asp could be mutated to Asn: the amide group may act as a hydrogen donor and coordinate the α-ketone; it may increase the rate of adenylation of α-ketoacid selecting A domains (which is significantly lower than that of most α-amino acid selecting A domains).

The putative supporting residue Phe234 is conservatively replaced by His305 in α-ketoacid-selecting A domains. In the absence of a high-resolution structure of the A domain of StsA in complex with substrate at the time of the experiment, we wondered if the nitrogen atoms of the His305 side chain could potentially interact with the α-ketone. But the His to Phe mutation, combined with the Ile to Asp mutation or not, did not allow Sts_AKrT to adenylate leucine, nor did the reverse mutation allow LgrA_FA to adenylate αKIV. Surprisingly, the adenylation assay showed that the His to Phe mutation on

its own increased the adenylation activity of StsA for αKIC. We can conclude from this that the Phe to His mutation is likely to be a deleterious by-product of the rapid evolution of the A domain, pressured to accommodate an α-ketoacid as a substrate. The Ile305 and its putative partner His306 are not responsible for the α-ketoacid specificity to the A domain.

3.1.3.3.2 The Gly and Supporting Residues

In the structure of the A domain from gramicidin S (PDB:1AMU), the backbone carbonyl of Gly324 interacts with the α-amino group of the substrate through hydrogen-bonding. Gly324 is conserved among α-amino acid-selecting A domains including LgrA_FA, but it is also conserved among α-ketoacid selecting A domain, including StsA, with Gly414. However, Gly324 is immediately followed by Pro325, which is conserved among in α-amino acid-selecting A domains, whereas Gly414 in StsA is immediately followed by Met415, which is conserved among in α-ketoacid-selecting A domains. This Pro to Met mutation conveys some flexibility to the backbone: in the StsA_A structure, the backbone carbonyl of Gly424 is rotated ~45° compared to LgrA, allowing it to engage in a reciprocal carbonyl-carbonyl interaction with the α-ketone, placing the partially positive carbons exactly 3.1Å away from the partially negative oxygens in a perfectly rectangular conformation. Such carbonyl-carbonyl interactions are comparable in strength and stability to hydrogen bonding [119]. In contrast, the backbone carbonyl of Gly324 is pointing up in 1AMU, allowing to engage in hydrogen-bonding with the α-amino group of the substrate.

The Met to Pro mutation, combined or not the mutation of Ile306 and His305 to Asp and Phe respectively, did not provide StsA_AKrT with the ability to adenylate α-amino acids; in the contrary, it abolished the limited ability of the wild type StsA_AKrT to adenylate L-leucine. However, the single P483A mutation allowed LgrA_FA_P483A to adenylate αKIV to 3.4% of the initial rate of StsA_AKrT with cognate substrate αKIC. The single mutation P483M was even more successful, and allowed LgrA_FA_P483M to adenylate αKIV to 5% of the initial rate of StsA_AKrT with cognate substrate αKIC. It is worth noting that the P483A and P483M point mutation also increased the adenylation activity of LgrA_FA for its cognate substrate L-valine. In order to understand if this mutation specifically allows LgrA_FA to adenylate α-ketoacids, or if it simply reduces the overall specificity of substrate of LgrA_FA

(allowing partial adenylation activity for non-cognate substrate regardless of their nature), it would be useful to investigate the specificity of LgrA_FA_P483M/A for other non-cognate substrates.

In the LgrA_FA_P483M structure in complex with αKIV and AMPcPP, we can see that the P483M mutation allows the backbone carbonyl of Gly506 to rotates ~45°, as observed in StsA_A, allowing for the formation of an imperfect carbonyl-carbonyl interaction with the α-ketone group of αKIV. Therefore, the absence of a proline at this position is essential for the A domain to select and α-ketoacid. However, the reverse mutation in StsA did not provide it with the ability to select an α-amino acid.

3.1.3.3.3 The Ile/Ser Residue

The conserved Stachelhaus residue Ile330 in 1AMU is mutated to Ser420 in StsA, which is conserved among α-ketoacid selecting A domains. In 1AMU, the backbone carbonyl of Ile330 is pointing towards the α-amino group of the substrate (in the "up" position), engaging in hydrogen-bonding. In StsA, the backbone carbonyl of Ser420 is rotated ~80° away from the substrate (in the "down" position), avoiding a repulsive interaction with the α-carbonyl group of αKIC. In the LgrA_FA structure elucidated by Reimer *et al.*, the backbone chain is somewhat flexible in that region, and the backbone carbonyl of Ile488 (1AMU Ile330) is observed the "up" position (as seen in 1AMU) in the closed conformation, and in the "down" position (as seen in StsA) in the open conformation [10]. However, the Ser mutation in StsA may lock the carbonyl group in the "down" position, reducing the entropic cost of docking of the substrate.

For future directions, it would be interesting to see if mutating Ile488 to a serine in the LgrA_FA_P483M construct could increase its affinity for αKIV. Additionally, the Asp396 in the LgrA_FA_P483M construct could be mutated to an Asn: the hydrogen-bond donor may interact positively with the α-carbonyl group of the substrate and increase the affinity of LgrA_FA for αKIV.

3.1.4 Material and Methods

3.1.4.1 Cloning

3.1.4.1.1 Cloning of pBac_StsA_A

The StsA_A gene was amplified from pBac_StsA_AkrT_(TEV)_His. PCR product was digested with NotI (NEB) and NdeI (NEB) restriction enzymes ligated into pBac_(TEV)_His using (pre-digested with the same restriction enzymes) using T7 ligase (NEB) to yield the plasmid pBac_StsA_A.

3.1.4.1.2 Cloning of StsA_AkrT A-Domain Mutants

Plasmids pBac_StsA_AkrT_H305F_(TEV)_His, pBac_StsA_AkrT_I306D_(TEV)_His and pBac_StsA_AkrT_H305F/I306D_(TEV)_His mutants were generated by site-directed mutagenesis from pBac_StsA_AkrT_(TEV)_His using primers StsA_AkrT_H305F_1 and StsA_AkrT_H305F_2, primers StsA_AkrT_I306D_1 and StsA_AkrT_I306D_2, and primers StsA_AkrT_H305F/I306D_1 and StsA_AkrT_H305F/I306D_2 respectively (Table __).

Plasmids pBac_StsA_AkrT_M415P_(TEV)_His, pBac_StsA_AkrT_I306D/M415P_(TEV)_His and pBac_StsA_AkrT_H305F/I306D/M415P_(TEV)_His were generated by site-directed mutagenesis from pBac_StsA_AkrT_(TEV)_His, pBac_StsA_AkrT_I306D_(TEV)_His and pBac_StsA_AkrT_H305F/I306D_(TEV)_His respectively, using primers StsA_AkrT_M415P_1 and StsA_AkrT_M415P_2 (see Table of Primers).

3.1.4.1.3 Cloning of LgrA_FA A-Domain Mutants

Plasmids pBac_LgrA_FA_F395H, pBac_LgrA_FA_D396V pBac_LgrA_FA_F395H/ D396V and LgrA_FA_F395H/D396V/S392P mutants were generated by site-directed mutagenesis from pBac_LgrA_FA using primers LgrA_FA_F395H_1 and LgrA_FA_F395H_2, primers LgrA_FA_D396V_1 and LgrA_FA_D396V_2, primers LgrA_FA_F395H/D396V_1 and LgrA_FA_F395H/ D396V_2, and primers LgrA_FA_F395H/D396V/S392P_1 and LgrA_FA_F395H/D396V/S392P_2 respectively (see Table of Primers). Plasmid pBac_LgrA_FA_F395H/D396V/S392P was generated by site-directed mutagenesis from pBac_LgrA_FA_F395H/D396V using primers LgrA_FA_F395H/ D396V/S392P_1 and LgrA_FA_F395H/ D396V/S392P_2 (see Table of Primers).

Plasmids pBac_LgrA_FA_P483A, pBac_LgrA_FA_D396V/P483A and pBac_LgrA_FA_F395H/D396V/P483A mutants were generated by site-directed mutagenesis from pBac_LgrA_FA, pBac_LgrA_FA_D396V and pBac_LgrA_FA_F395H/D396V respectively, using primers LgrA_FA_P483A_1 and LgrA_FA_P483A_2 (see Table of Primers).

Plasmids pBac_LgrA_FA_P483M, pBac_LgrA_FA_D396V/P483M and pBac_LgrA_FA_F395H/D396V/P483M mutants were generated by site-directed mutagenesis from pBac_LgrA_FA, pBac_LgrA_FA_D396V and pBac_LgrA_FA_F395H/D396V respectively, using primers LgrA_FA_P483M_1 and LgrA_FA_P483M_2 (see Table of Primers).

3.1.4.2 Expression and Purification

3.1.4.2.1 Expression and Purification of StsA_A.

StsA_A was expressed and purified by nickel column chromatography according to the same protocol as for StsA_AkrT_(TEV)_His. The composition of the nickel column binding buffer was changed however (50mM Tris-HCl pH 8.0, 250mM NaCl, 10% glycerol and 2mM βME) and the nickel column elution buffer (50mM Tris-HCl pH 8.0, 100mM NaCl, 150mM Imidazole, 10% glycerol and 2mM βME), and the column was washed with 17% elution buffer before elution. After cleaving the His-tag with the TEV protease, the protein was run through the nickel column, collecting the flow-through, before being further purified by anion exchange chromatography and SEC according to the same protocol as for StsA_AkrT_(TEV)_His. The purity of the protein was assessed by SDS-PAGE and the oligomeric state of the protein was assessed by native-PAGE. The protein was either used immediately for crystallization, or flash frozen in liquid nitrogen and stored at -80°C at 15mg/ml in 20% glycerol.

3.1.4.2.2 Expression and Purification StsA_AkrT_H305F, StsA_AkrT_I306D, StsA_AkrT_H305F/I306D StsA_AkrT_M415P, StsA_AkrT_I306D/M415P and StsA_AkrT_H305F/I306D/M415P.

StsA_Akr_(TEV)_His, StsA_AkrT_H305F, StsA_AkrT_I306D, StsA_AkrT_H305F/I306D and StsA_AkrT_Bmdb_C3 are expressed and purified according to the same protocol as for StsA_AkrT_(TEV)_His, except that they may or may not be modified with the 4' ppant arm, they are not modified with α-KIC and they only require one run of SEC column to achieve and acceptable level of

purity. The protein was either used immediately for crystallization or adenylation assay, or flash frozen in liquid nitrogen and stored at -80°C at 15mg/ml in 20% glycerol.

3.1.4.2.3 Expression and Purification of LgrA_FA_F395H, LgrA_FA_D396V, LgrA_FA_F395H/D396V, LgrA_FA_P483A, LgrA_FA_D396V/P483A, and LgrA_FA_F395H/D396V/S392P LgrA_FA_F395H/D396V/P483A, LgrA_FA_P483M, LgrA_FA_D396V/P483M and LgrA_FA_F395H/D396V/P483M

The LgrA_FA A domain mutants were expressed according to the same protocol as for StsA_AkrT_(TEV)_His, except that expression was induced using 0.5mM IPTG. The protein was purified by nickel column chromatography according to the same protocol as for StsA_AkrT_(TEV)_His, except for the composition of LgrA nickel column binding buffer (50mM Tris-HCl pH 7.0, 150mM NaCl, 2mM Imidazole and 0.25mM TECEP) and LgrA nickel column elution buffer (50mM Tris-HCl pH 7.0, 150mM NaCl, 250mM Imidazole and 0.25mM TECEP), and the protein was eluted with a gradient from 0% to 100% elution buffer over five column volumes. The affinity tags were cleaved with TEV protease and the protein was dialyzed according to the same protocol as for StsA_AkrT_(TEV)_His, except for the composition of the dialysis buffer (20mM Tris-HCl pH 8.0, 50mM NaCl and 0.25mM TECEP). After running the reverse nickel column according to the same protocol as for StsA_AkrT_(TEV)_His, the protein was loaded onto the HiTrap Q HP anion exchange chromatography column pre-equilibrated with anion exchange binding buffer (20mM Tris-HCl pH 8.0 and 0.25mM TECEP) and eluted with a gradient from 0% to 100% anion exchange elution buffer (20mM Tris-HCl pH 8.0, 1M NaCl and 0.25mM TECEP) in anion exchange binding buffer over 10 column volumes. The purity of the protein was assessed by SDS-PAGE. The protein was concentrated by filtration and further purified by SEC according to the same protocol as for StsA_AkrT_(TEV)_His, except for the composition of the SEC buffer (20mM Tris-HCl pH 7.0, 150mM NaCl and 0.25mM TECEP). The protein was concentrated by filtration to 12mg/ml, flash frozen in liquid nitrogen and stored at -80°C in 25% glycerol.

3.1.4.3 LC-ESI-MS

The LC-ESI-MS assay was performed as described in Chapter II section 2.4.3

3.1.4.4 Adenylation Assay

The adenylation activity of the various enzymes described in this thesis were carried out by measuring the pyrophosphate production of said enzymes using the EnzCheck® assay (Thermo Fisher Scientific). 100 µL reactions mix were prepared, containing 2 mM substrate (αKIC, αKIV, αKIL, ICA, L-Leucine or L-Valine), 1 mM ATP, 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, 10% glycerol and 10µl of 10X reaction buffer (1.5M NaCl, 200mM HEPES 7.0, 100mM MgCl for LgrA enzymes and 1M NaCl, 250mM HEPES pH 8.0 and 100mM MgCl for StsA enzymes and Br-NADPH synthesis enzymes). The reactions were added to a 96-well clear-bottom Costar® plates (Sigma), and the reactions were initiated by adding 15uM of the enzyme under investigation, and the absorbance of the reaction mix at 360 nm was immediately recorded over time. Reactions were performed in triplicate. Initial rates were calculated by linear regression and plotted using the program GraphPad Prism 6.

3.1.4.5 Crystallography of StsA_A and Diffraction Data Collection

A solution of pure StsA_A in SEC buffer (20mM Tris-HCl pH 7.0, 150mM NaCl and 0.25mM TECEP) supplemented with 2mM AMPcPP, 2mM αKIC and 10mM $MgCl_2$ was used as protein solution for crystallization. Sparce matrix screening using The Classics Suite, The Classics II Suite, and The PEGS Suite (Qiagen) were automatically set using the NT8® drop setter for protein crystallization (Formulatrix®) upon 96 well low volume 3 sitting drops INTELLI-PLATES® (MiTeGen). 40µl of custom precipitant solution was added to each well, and the NT8® drop setter was programmed to add 0.2µl of protein solution followed by 0.2µl of precipitant solution to each drop. The 96-well plate is then sealed and placed in a controlled environment at either 4°C or room temperature in the ROCK IMAGER® incubator (Formulatrix®) with visible light and UV-images programed to be taken daily. In this manner, the initial hit condition was identified: 0.1M NaCl, 0.1M HEPES pH7.5 and 1.6M ammonium sulfate, with a single crystal that grew over 7 months into a 100um square-based bipiramidal shape. The crystal diffracted to 1.89Å and showed no signs of twinning, although some anomalies in the data made it difficult to fully process the data set and elucidate the structure. The crystallization condition was then optimized to obtain fully grown crystals within a more feasible time frame. Some of the optimization process

included: precipitant concentration and pH, protein concentration, co-crystallization substrate concentration, additive and detergent screen (Hampton), sparse matrix screening using the Ammonium Sulfate Suite (Hampton), seeding and drop volume. The final crystals were obtained with 4µl drops made of 2µl precipitant solution (1.65M $AmSO_4$ and 22.5% glycerol), 1.6µl protein at 12.5mg/ml supplemented with 2mM AMPcPP, 2mM αKIC and 10mM $MgCl_2$, and 0.4µl of additive solution (either 30%w/v 6-amino-hexanoic acid or 30%w/v trimethylamine N-oxide dihydrate), sealed to allow vapor diffusion in a 24-well sitting drop crystallization plate (MiTeGen) with 500µl of precipitant solution in the reservoir. Crystals grew to 200µm square-based bipiramidal shape in two days. Trays were also set according to the same protocol using αKIL, αKIV and ICA at 6mM as substrate for co-crystallization. Crystals were cryoprotected using a stabilization solution supplemented with increasing concentration of ethylene glycol: two times 2µl of 7%w/v ethylene glycol, two times 2µl of 14%w/v ethylene glycol followed by two times 2µl of 28%w/v ethylene glycol were added to the drop, allowing for 5min of equilibration time between additions. After looping and cryoprotection, crystals were flash frozen and stored in liquid nitrogen. A complete data set at 2.5Å resolution was collected.

3.1.4.6 Data Processing and Structure Elucidation

Data sets were indexed and integrated using iMosflm [109], scaled using CCP4 Aimless program suite [110], and refined with Phenix [113] using the excised StsA_A domain from the 3.5 Å resolution StsA_Akr model described in Chapter II as initial refinement model.

3.1.5 Supplementary Tables

3.1.5.1 List of Constructs

StsA_A
StsA_AkrT_H305F
StsA_AkrT_I306D
StsA_AkrT_H305F/I306D
StsA_AkrT_M415P

StsA_AkrT_I306D/M415P
StsA_AkrT_H305F/I306D/M415P
LgrA_FA_F395H
LgrA_FA_D396V
LgrA_FA_F395H/D396V
LgrA_FA_F395H/D396V/S392P
LgrA_FA_P483A
LgrA_FA_D396V/P483A
LgrA_FA_F395H/D396V/P483A
LgrA_FA_P483M
LgrA_FA_D396V/P483M
LgrA_FA_F395H/D396V/P483M

3.1.5.2 Table of Primers

StrA_AKRT_H305F_1	GCCGCTGGATTTTATTGGCGGTATC
StrA_AKRT_H305F_2	GATACCGCCAATAAAATCCAGCGGC
StrA_AKRT_I306D_1	CCGCTGGATCATGATGGCGGTAT
StrA_AKRT_I306D_1	ATACCGCCATCATGATCCAGCGG
StrA_AKRT_H305F_I306D_1	GCCGCTGGATTTTGATGGCGGTATC
StrA_AKRT_H305F_I306D_2	GATACCGCCATCAAAATCCAGCGGC
Stra_AKrT_M415P_1	CCGGCATTCGGCCCGTCAGAAATCTCAT
Stra_AKrT_M415P_2	ATGAGATTTCTGACGGGCCGAATGCCGG
LgrA_FA_F395H_1	GTTTATTCCAGTTTTTCCCATGACGGGTCTGCACTGGATATA
LgrA_FA_F395H_2	TATATCCAGTGCAGACCCGTCATGGGAAAAACTGGAATAAAC
LgrA_FA_D396V_1	GTTTATTCCAGTTTTTCCTTTGTCGGGTCTGCACTGGATATA
LgrA_FA_D396V_2	TATATCCAGTGCAGACCCGACAAAGGAAAAACTGGAATAAAC

LgrA_FA_F395H/D396V_1	GTTTATTCCAGTTTTTCCCATGTCGGGTCTGCACTGGATATA
LgrA_FA_F395H/D396V_2	TATATCCAGTGCAGACCCGACATGGGAAAAACTGGAATAAAC
LgrA_FA_F395H/D396V/S392P_1	GTTGGTTTATTCCCCTTTTTCCCATGTCG
LgrA_FA_F395H/D396V/S392P_2	CGACATGGGAAAAAGGGGAATAAACCAAC
LgrA_FA_P483A_1	GGGTACGGGGCGACAGAGTGC
LgrA_FA_P483A_2	GCACTCTGTCGCCCCGTACCC
LgrA_FA_P483M_1	CGGGTACGGGATGACAGAGTGCA
LgrA_FA_P483M_2	TGCACTCTGTCATCCCGTACCCG

3.2 The Pseudo-Adenylation Sub-Domain of the Initiation Module of Stratospherulide Synthetase

3.2.1 Introduction

3.2.1.1 The Adenylation Sub-Domain of Amino-Acid-Selecting A domains

As detailed earlier in Chapter I, the adenylation domain is a roughly 60kDa globular domain, consisting of a 45kDa A_{core} and a 12kDa A_{sub} subdomains. The A_{sub} domain, together with the PCP domain, are the most dynamic and mobile parts of the NRPS assembly line. The cycle starts with the A_{sub} domain in the open conformation, allowing the substrates to bind the active site residues from the A_{core} domain. The A_{sub} domain then rotates 30° towards the active site to achieve the closed conformation, allowing the catalytic Lys517 to complete the active site of the A domain. After adenylation, the A_{sub} domain performs a 140° rotation to achieve the thiolation state and allow for the PCP domain to access the active site, essentially facing the other way allowing Lys434, located on the opposite face of the A_{sub} domain with respect to Lys517, to complete the active site of the A domain for thiolation [120]. Following thiolation, the PCP domain translocate 45Å and rotates 75° to reach the active site of the C domain and achieve the condensation conformation [23]. If the module contains a tailoring domain, the A_{sub} domain and the PCP domain must travel extraordinary distances to reach the additional active site before condensation. For example, to allow for the formylation of the substrate after thiolation, the A_{sub} domain migrates 21Å and rotates a full 180°, while the PCP domain migrates 61Å and rotates 75° in order to present its substrate to the active site of the tailoring domain [10, 120].

3.2.1.2 Pseudo Adenylation Sub-Domain of Stratospherulide synthetase

The structure of the A-Kr-PCP initiation module of Sts presented in Chapter II revealed the presence of a novel domain, structurally similar to the A_{sub} domain, which was called the pseudo-A_{sub} (pA_{sub}) domain. The pA_{sub} domain of StsA is separated in two portions, located over 1000 amino acids apart in the primary sequence, brought together through a domain swap between dimeric partners. This study is the first to report this existence of such a domain, which raises the question of how common the pA_{sub} domain may be in NRPSs, and if it may be found in other Kr domain-containing NRPS

modules. The presence of the pA_{sub} domain, disrupted by the A-Kr didomain, suggests that an evolutionary event involving a pre-existing A-Kr didomain inserting within the A_{sub} domain of the initiation module of a second NRPS may have occurred, followed by the deletion of the initial A_{core} of the disrupted module and the modification of the initial A_{sub} domain into a pA_{sub} domain due to redundancy. It is still unclear whether the pA_{sub} domain may hold a structural or catalytic function, or if it is simply an evolutionary remnant, the likes of which has been reported before in NRPSs [100].

If evolutionary remnants are not uncommon in NRPSs, Sts is the only NRPS known to dimerize in solution, to the best of my knowledge. Sts was initially thought to dimerize through the Kr domains, by comparison with PKSs; instead, the formation of the pA_{sub} domain through a domain swap between dimeric partners seemingly induces dimerization. However, Sts is mostly present as a monomer in solution. The linker region between the A domain and the N-terminal portion of the pA_{sub} domain, involved in the domain swap, is long enough to allow for the N-terminal portion of the pA_{sub} domain to complement the C-terminal portion of the pA_{sub} domain of the same module, allowing for the formation of a monomer, which is favored in solution since the intra-molecular strand-swap is entropically preferred. If such is the mechanism of oligomerization in StsA, the relative prevalence of the oligomeric population may be manipulated by altering the length of the swapped strand (i.e. the inter-molecular domain swap, and the formation of a dimer, may be favored if the linker region were substantially shortened). Sequence-based analysis revealed that proline residues are often found in the hinge region on either side of the linker region involved in domain swapping. The implication of the hinge region prolines and the specifics of the mechanism of oligomerization remains unknown.

In the conformation visualized in the A-Kr-PCP structure described in Chapter II, the pA_{sub} domain is predicted to clash with the C domain in the thiolation conformation [66, 99]. We can then expect extensive conformational rearrangements to occur, involving the pA_{sub} domain along with the PCP and A_{sub} domains, in order to achieve the thiolation conformation, and possibly the ketoreduction and condensation conformations as well.

3.2.2 Results

The goal for this section is to investigate the pA_{sub} domain of StsA in further details in order to answer a few outstanding questions: is this novel domain unique to Sts, or can it be found in related NRPSs as well? Does the pA_{sub} hold a structural or catalytic function, if it holds any function at all? Finally, we aim to shed some light on the mechanism of domain swapping and dimerization of the NRPS through the formation of the pA_{sub} domain.

3.2.2.1 Prevalence of pA_{sub} Domains in Kr-Containing NRPSs

The sequences of known Kr domain-containing NRPS modules were aligned with the goal of investigating the presence of pA_{sub} domain-like sequence motifs in the N-terminal region of the A domain of the Kr domain-containing module, and in the linker region between the Kr domain and the PCP domain, where the pA_{sub} domain is found in StsA, the results of which are shown in Figure 33.a.

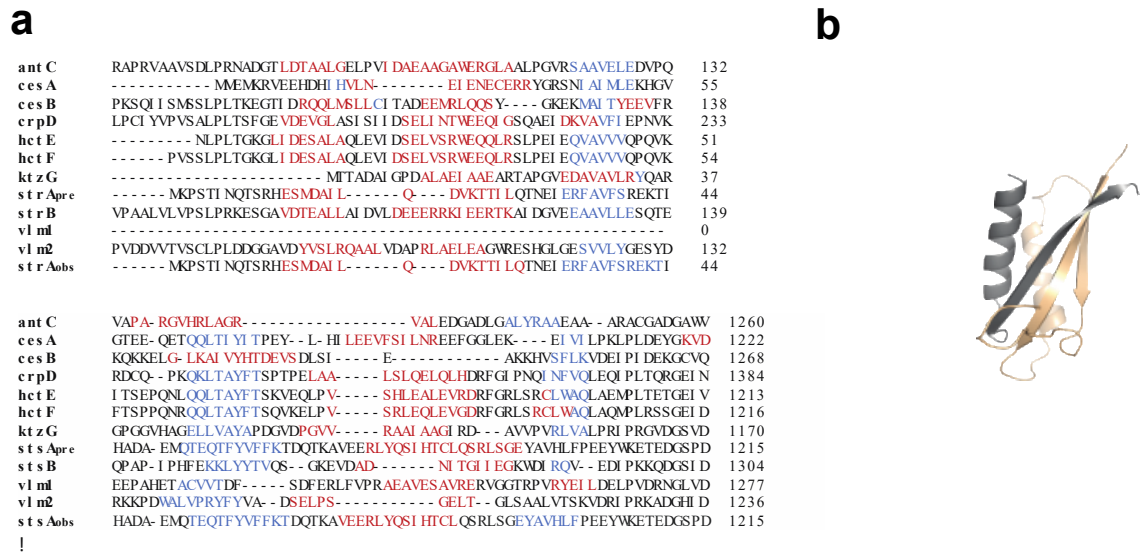


Figure 33: Alignment of putative pA_{sub} regions in Kr domain-containing NRPSs (a) Alignment of putative the N-terminal (upper alignment) and C-terminal (lower alignment) portions of the pA_{sub} in Kr domain-containing NRPS modules (i.e. CesA and CesB from the cereulide synthetase [6], Vlm1 and Vlm2 from the valinomycin synthetase [11], CrpD from the cryptophycins gene cluster [57], HtcE and HctF from the hectochlorin gene cluster [121], AntC from the antimycin gene cluster, as well as the analogs SmlB from the JBIR-06 and NatB from the neo-antimycin synthetase [122], and KtzG from the kutznerides cluster

[123]). Predicted α-helices are shown in red and predicted β-stands are shown in blue (b) Structure of the pA_{sub} domain of StsA; the N-terminal portion is shown in dark grey and the C-terminal portion is shown in light pink.

The alignment of known Kr domain-containing NRPS modules (Figure 33.a) revealed the presence of conserved motifs expected to fold into an α-helix and a β-strand located at the N-terminal region of the A domain of the Kr domain-containing module in most of the sequences aligned, which corresponds to the secondary structures expected for the N-terminal portion of the pA_{sub} domain. The alignment of known Kr domain-containing NRPS modules revealed the presence of conserved motifs expected to fold into a β-strand followed by an α-helix and another β-strand in most of the sequences aligned, located in the linker region between the Kr domain and the PCP domain, which corresponds to the secondary structures expected for the C-terminal portion of the pA_{sub} domain.

3.2.2.2 The Oligomerization of StsA Through the Formation of the pA_{sub} Domain

In order to investigate the mechanism of domain swapping and dimerization of StsA through the formation of the pA_{sub} domain, we set out to produce a series of construct of StsA_AKrT modified with a deletion of an increasing number of residues in the center of the linker region between the A domain and the N-terminal portion of the pA_{sub} domain, which is involved in the strand swap forming the full pA_{sub} domain.

Plasmids pBac_StsA_AKrT_Δ63-67 (with a deletion of 5 amino acids in the center swapped strand), pBac_StsA_AKrT_Δ57-71 (15 amino acids deletion in the center of the swapped strand), and pBac_StsA_AKrT_Δ45-82 (full deletion of the swapped strand), as well as plasmid pBac_StsA_AKrT_P46A_P48A_P82A_P83A (Pro to Ala mutations in the hinge regions of the swapped strand) were successfully cloned and purified similarly to the wild-type. The percentage of dimer as part of the complete oligomeric population of protein isolated after the MonoQ anion exchange column was measured for each construct and compared to the wild-type, as shown in Table 4.

Table 4: **Percentage of dimer among the oligomeric population measured for the StsA_AKrT wild-type, as well as the swapped strand deletion constructs and the Pro to Ala mutation constructs in the hinge regions of the swapped strand.** The dimeric and monomeric population, as well as higher oligomers, were quantified after the MonoQ anion exchange purification. Each construct was purified once, the results below are then single points of measurement.

Construct	% Dimer
StsA_AKrT_WT	15
StsA_AKrT_Δ63-67	25
StsA_AKrT_Δ57-71	35
StsA_AKrT_Δ45-82	60
StsA_AKrT_P46A_P48A_P82A_P83A	25

We can see in Table 4 that the percentage of dimer present in solution gradually increases as the length of the swapped strand decreases. The oligomerization profile is significantly altered for the full linker deletion construct (StsA_AKrT_Δ45-82), in which case the percentage of dimer exceeds that of monomer, making the dimer the predominant population in solution. The mutation of the proline residues found in the hinge regions of the swapped strand to alanine residues had the effect of increasing the percentage of dimer from 15% for the wild-type to 25% for the proline mutation construct (StsA_AKrT_P46A_P48A_P82A_P83A).

As had been observed in the wild-type protein, the monomer is divided into two peaks on the MonoQ ion exchange column. The stability over time of these two monomeric populations and the dimeric population was investigated using StsA_AKrT_WT, StsA_AKrT_Δ63-67 and pBac_StsA_AKrT_P46A_P48A_P82A_P83A. The monomer peak 1 and 2 were pooled separately, incubated at 4°C for a period of 24h before applying the samples on the MonoQ column again. Each peak eluted at the same salt concentration as they did before the 24h incubation period, indicating that the two monomeric populations and the dimeric population are stable overtime.

3.2.2.3 *Catalytic Lysine Mutants of the pA_{sub} Domain*

In order to investigate the catalytic function of the pA_{sub} domain, the sequence of the pA_{sub} domain from StsA in the region corresponding to the catalytic lysine (Lys517 in 1AMU) was aligned with the sequence of the putative pA_{sub} domains from other Kr domain-containing NRPSs identified in section 3.2.2.1., the results of which is shown in Figure 34.a.

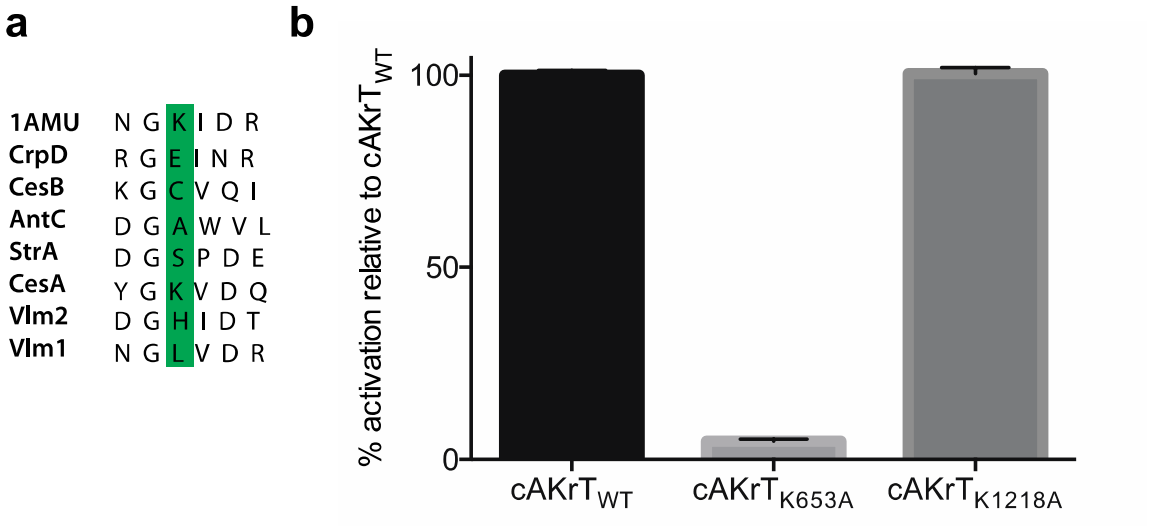


Figure 34: Catalytic role of Lys1218 of the putative pA_{sub} domain of CesA (a) Alignment of the catalytic lysine region of the pA_{sub} domain from StsA with the putative-pA_{sub} domains from related Kr domain-containing NRPSs (b) Adenylation assay of CesA_AKrT_WT, CesA_AKrT_K653A (mutation of the A_{sub} domain catalytic Lys to Ala) and CesA_AKrT_K1215A (mutation of the pA_{sub} domain catalytic Lys to Ala) with cognate substrate αKIC.

The local sequence alignment (Figure 34.a) and the structural alignment of StsA_AKrT structure with 1AMU, shows that the catalytic lysine (Lys517 in 1AMU) is changed to a serine in StsA. Moreover, the catalytic lysine is changed to another residue in most of the Kr domain-containing NRPSs used for this alignment, with the exception of CesA. For this reason, we decided to study the catalytic role of the putative pA_{sub} domain of CesA by mutating the catalytic Lys of the A_{sub} and pA_{sub} domains to alanine residues, before performing adenylation assays to investigate the effect of these mutations on the

catalytic activity of the A domain. Plasmids pBac_CesA_AKrT_K653A (mutation of the catalytic Lys of the A_{sub} domain) and pBac_CesA_AKrT_K653A (mutation of the catalytic Lys of the pA_{sub} domain) were successfully sub-cloned. The constructs CesA_AKrT_WT (wild-type CesA), CesA_AKrT_K653A and CesA_AKrT_K653A were successfully expressed and purified into stable and homogenous solution of pure monomeric protein. Adenylation assay using cognate substrate αKIC was performed on all three constructs, the results of which are shown in Figure 34.b.

The mutation of the catalytic Lys to an Ala in the A_{sub} domain of CesA_AKrT abolishes the adenylation activity of the A domain, whereas mutation of the catalytic Lys to an Ala in the pA_{sub} domain of CesA_AKrT has no effect on the adenylating activity of the A domain (Figure 34.b), suggesting that the catalytic Lys of the A_{sub} domain of the initiation module of CesA is required for the catalytic activity of the A domain but the same Lys found in the putative pA_{sub} domain of CesA is not.

3.2.2.4 Partial and Full pAsub Domain Deletion Constructs.

In order to investigate the structural role of the pA_{sub} domain of StsA, we decided to produce a series of constructs, including partial and full deletion of the pA_{sub} domain, in order to investigate the effect of such deletions on the structural integrity of the module, as well as the effect it may have on the catalytic function of the domains which composes the pA_{sub} domain-containing module.

Plasmids pBac_StsA_AKrT_ Δ1-83_(TEV)_His (deletion of the N-terminal portion of the pA_{sub} domain) and pBac_StsA_AKrT_Δ1-83_ Δ1162-1225_(TEV)_His (full deletion of the pA_{sub} domain) were successfully sub-cloned. The StsA_AKrT_Δ1-83_(TEV)_His construct did not express as a soluble protein after three attempts, suggesting that the N-terminal portion of the pA_{sub} may be required for folding. The pBac_StsA_AKrT_Δ1-83_ Δ1162-1225_(TEV)_His expresses well in *E. coli* with a yield of 2mg of protein per gram of cell pellet after the nickel column, suggesting that the presence of the pA_{sub} domain is not required for proper folding of the pA_{sub} domain-containing initiation module of StsA. Following the nickel column purification step, SDS-PAGE performed on the protein eluate showed the presence of a significant amount of a 67kDa protein contaminant, presumed to be a Kr-PCP didomain cleavage product according to LC-ESI-MS on intact protein (not shown), although this was not confirmed by tryptic digest. The full pA_{sub} domain deletion construct was purified to an estimated purity of 99%

according to SDS-PAGE with some difficulty (see methods section 3.2.4.2.3. for more details). Adenylation assay using cognate substrate αKIC was performed on the purified StsA_AKrT_Δ1-83_ Δ1162-1225 construct, the results of which are shown in Figure 35.

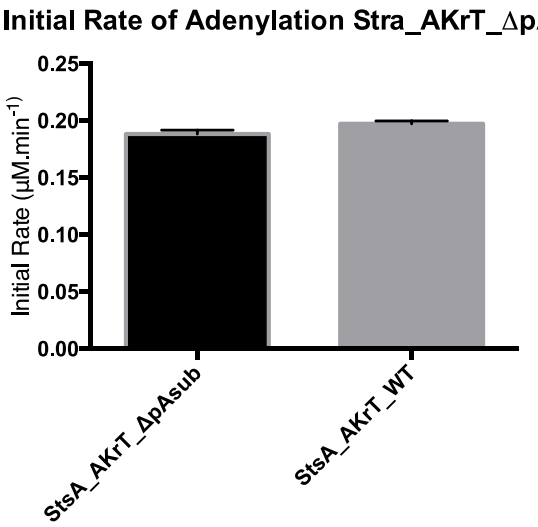


Figure 35: Adenylation assay for the wild-type and full pA_{sub} domain deletion construcuts in StsA_AKrT performed with cognate substrate αKIC.

The full pA_{sub} domain deletion construct of StsA_AKrT initiation module adenylates αKIC with a similar initial rate as the wild type (Figure 35), suggesting that the presence of the pA_{sub} domain is not required for the adenylation activity of the A domain of the module.

3.2.3 Discussion

3.2.3.1 Prevalence of pA_{sub} domains in NRPSs

Most of the Kr domain-containing NRPS modules investigated in this study through primary sequence alignment showed the presence of motifs corresponding to the N-terminal and C-terminal portions of the pA_{sub} domain, located in the N-terminal region of the A domain and the linker region between the Kr domain and the PCP domain respectively, suggesting that most Kr-containing NRPS modules are expected to include a fully folded pA_{sub} domain as observed in StsA. We propose that the common ancestor to most contemporary Kr domain-containing NRPS modules may have been formed thought the addition of a pre-existing A-Kr didomain into a canonical NRPS module, disrupting the A

domain between β_2 and β_3 of the A_{sub}, which evolved into the pA_{sub} due to redundancy after elimination of the A_{core} domain of the disrupted module, as proposed in section 2.3.6.

3.2.3.2 Oligomerization of the initiation module of StsA

The oligomeric populations are stable over time, suggesting that the formation of the pA_{sub}, though a domain swap or not, is fixed and does not allow for dynamic exchange between the dimeric and monomeric populations. The oligomeric profile of the protein in solution can be manipulated to some extent by altering the length of the linker region in between the N-terminal portion of the pA_{sub} and the A domain, since the percentage of dimer increases as the length of the swapped strand decreases. This suggests that the formation of a monomer is favored with the wild-type because the intra-molecular domain swap is entropically preferred, but the formation of the dimer may become more favored as the swapped strand reduces in length and it becomes more difficult for the N-terminal portion of the pA_{sub} domain to reach the C-terminal portion and complete it. When linker region involved in the domain swap is fully deleted, we still observe 40% of monomer in solution; in this case, parts of the N-terminal region of the A domain may unfold to allows for the formation of a pA_{sub} domain in the monomeric state, or the protein may not be fully folded. The mutation of the proline residues found in the hinge region of the swapped strand to alanine residues slightly increase the percentage of dimer, although it is difficult to make any definitive conclusion since the construct was only purified once; this suggest that the prolines, often found in the hinge region of the linker region involved in a domain swap, may favor the inter-molecular strand swap and promote the oligomerization.

3.2.3.3 Catalytic role of the pA_{sub} domain

For most putative pA_{sub}-containing NRPS modules investigated in this study through primary sequence alignment, the catalytic lysine (corresponding to the Lys517 of the A_{sub} domain in 1AMU) of the putative pA_{sub} domain is changed to another residue, the nature of which is not conserved. This suggest that the putative pA_{sub} domain identified in most Kr domain-containing NRPS modules may not be catalytically active. The catalytic lysine is however present in the putative pA_{sub} of CesA. The mutation of that lysine to an alanine does not abolish or reduce the adenylation activity of CesA_AKrT,

suggesting that it is not required for adenylation and that the pA_{sub}, unlike the canonical A_{sub} domain, does not have a catalytic role in adenylation.

3.2.3.4 Structural role of the pA_{sub} domain

The A-Kr-PCP initiation module of StsA was not able to fold properly upon deletion of the N-terminal portion of the pA_{sub} domain, but it was however able to fold into a soluble state upon deletion of the full pA_{sub} domain, suggesting that the initiation module may be structurally stable in the absence of the pA_{sub} domain, but destabilized into insoluble aggregates should the pA_{sub} domain not be fully folded. Moreover, the initiation module is able to adenylate its cognate substrate αKIC in the absence of the pA_{sub} domain. This information, combined with our conclusions from section 3.2.3.3. above, suggests that the pA_{sub} domain is not involved in the adenylation reaction, and does not hold a catalytic or structural role essential to the selection and activation of the substrate.

Several questions regarding the role of the pA_{sub} domain remain unanswered: could it be engaging in any form of static or dynamic inter-domain interactions in order to facilitate the thiolation or ketoreduction reactions, or is it simply and evolution remnant devoid of any function due to redundancy? Answers to these questions may be provided thought the study of the A-Kr-PCP module of StsA with deletion of the full pA_{sub} domain, following thiolation and ketoreduction, by LC-ESI-MS on intact protein as described in this thesis.

3.2.4 Material and Methods

3.2.4.1 Cloning

3.2.4.1.1 Cloning of the catalytic lysine mutants, the linker deletion constructs and the hinge region proline mutants.

Plasmids pBac_CesA_AKrT_K653A and pBac_CesA_AKrT_K1218A were generated by site-directed mutagenesis from pBac_CesA_AKrT_WT using primers CesA_AKrT_K653A_1 and CesA_AKrT_K653A_2, and primers pBac_CesA_AKrT_K1218A_1 and pBac_CesA_AKrT_K1218A_2 respectively (see Table of Primers). Plasmid pBac_StsA_AKrT_K636A was generated by site-directed

mutagenesis from pBac_StsA_AKrT_WT using primers pBac_StsA_AKrT_K636A_1 and pBac_StsA_AKrT_K636A_2 (see Table of Primers).

Plasmid pBac_StsA_AKrT_Δ63-67 was generated by site-directed mutagenesis from pBac_StsA_AKrT_WT using primers pBac_StsA_AKrT_Δ63-67_1 and pBac_StsA_AKrT_Δ63-67_2 (see Table of Primers).

Plasmid pBac_StsA_AKrT_Δ57-71 was generated by site-directed mutagenesis from pBac_StsA_AKrT_Δ63-67 using primers pBac_StsA_AKrT_Δ57-71_1 and pBac_StsA_AKrT_Δ57-71_2 (see Table of Primers).

pBac_StsA_AKrT_Δ45-82 plasmid was generated by site-directed mutagenesis from pBac_StsA_AKrT_Δ57-71 using primers pBac_StsA_AKrT_Δ45-82_1 and pBac_StsA_AKrT_Δ45-82_2 (see Table of Primers).

pBac_StsA_AKrT_P46A_P48A and pBac_StsA_AKrT_P82A_P83A mutants were generated by site-directed mutagenesis from pBac_StsA_AKrT_WT using primers pBac_StsA_AKrT_P46A_P48A_1 and pBac_StsA_AKrT_P46A_P48A_2, and primers pBac_StsA_AKrT_P82A_P83A_1 and pBac_StsA_AKrT_P82A_P83A_2 respectively (see Table of Primers).

3.2.4.1.2 Cloning of the pAsub half domain and full domain deletion constructs.

The StsA_AKrT_ Δ1-83 gene was amplified from pBac_StsA_AKrT_(TEV)_His. PCR product was digested and ligated into pBac_(TEV)_His using (pre-digested with the same restriction enzymes) using T7 ligase (NEB) to yield the plasmid pBac_StsA_A.

pBac_StsA_AKrT_Δ1-83_ Δ1162-1225_(TEV)_His was generated by *in vivo* homologous recombination cloning. Fragment 1 was amplified from pBac_StsA_AKrT_Δ1-83_(TEV)_His using primers StsA_AKrT_Δ1162-1225_1_1 and StsA_AKrT_Δ1162-1225_1_2, and fragment 2 was amplified from pBac_StsA_AKrT_(TEV)_His using primers StsA_AKrT_Δ1162-1225_2_1 and StsA_AKrT_Δ1162-1225_2_2 (see Table of Primers). The PCR amplicons were treated with DpnI restriction enzyme (NEB) and transformed into 10β competent *Escherichia coli* cells (NEB) where the fragments were condensed

by endogenous homologous recombination to yield the pBac_StsA_AKrT_Δ1-83_ Δ1162-1225_(TEV)_His plasmid.

3.2.4.2 Expression and Purification

3.2.4.2.1 Expression and purification of the catalytic lysine mutants and the hinge region proline mutants.

The catalytic lysine of the pseudo-A_{sub} mutants StsA_AKrT_K636A, CesA_AKrT_K653A and CesA_AKrT_K1218A and StsA_AKrT_K636A, and the proline to alanine mutants in the N-terminal and C-terminal region of the linker between the Kr domain and the N-terminal portion of the pseudo-A_{sub} construct StsA_AKrT_P46A/P48A/P82A/P83A, were expressed and purified according to the same protocol as for StsAKr described in 2.3.2.3.

3.2.4.2.2 Expression and purification of the pAsub linker deletion constructs

The linker between the N-terminal portion of the pseudo-A_{sub} deletion constructs, StsA_AKrT_Δ63-67, StsA_AKrT_Δ57-71 and StsA_AKrT_Δ45-82 were expressed and purified according to the same protocol as for StsA_AKrT_(TEV)_His described in 2.3.2.2.

3.2.4.2.3 Expression and purification of the pAsub half domain and full domain deletion constructs.

Expression and purification of the StsA_AKrT_ Δ1-83 construct was attempted following the same protocol as for StsA_AKrT_WT without success.

StsA_AKrT_pseudoA_{sub}deletion was expressed according to the same protocol as for StsA_AKrT_(TEV)_His. The protein was purified by nickel column chromatography according to the same protocol as for StsA_AKrT_(TEV)_His, except that the protein was eluted with a gradient from 0 to 100% elution buffer over 10 column volumes. The purity of the protein was assessed by SDS-PAGE, which revealed that approximately 60% of the eluate is contaminated by a 67kDa protein that co-elutes with StsA_AKrT_Δ1-83_ Δ1162-1225_(TEV)_His. After the reverse nickel column, the protein was further purified by anion exchange column chromatography according to the same protocol as for StsA_AKrT_(TEV)_His, except for the composition of the MonoQ binding buffer (25mM HEPES pH8.0, 100mM NaCl, 10% glycerol and 2mM BME) and MonoQ elution buffer (25mM HEPES pH8.0, 500mM NaCl, 10% glycerol and 2mM BME), and that the column was washed with 30% MonoQ elution buffer

to eliminate the contaminants, before being re-equilibrated in MonoQ binding buffer and eluting the column with a gradient from 0% to 75% MonoQ elution buffer. The protein was collected with ~30% recovery and an estimated purity of ~60-70% based on SDS-PAGE. The protein was further purified by SEC according to the same protocol as for StsA_AKrT_(TEV)_His, except that the protein was concentrated using a HiTrap column due to the fact that the protein seemed to get trapped in spin filters (with ~30-40% recovery after the concentration step when performed using spin filters). The purity of the protein was assessed by SDS-PAGE and the oligomeric state of the protein was assessed by native-PAGE. The protein was either used immediately or flash frozen in liquid nitrogen and stored at -80°C at 15mg/ml in 20% glycerol.

3.2.4.3 LC-MS

The LC-MS assay was performed as described in Chapter II section 2.4.3.

3.2.4.4 Adenylation Assay

The adenylation assay was performed as described in Chapter III section 3.1.4.4.

3.2.5 Supplementary Tables

3.2.5.1 List of Constructs

CesA_AKrT_K653A
CesA_AKrT_K1218A
StsA_AKrT_K636A
StsA_AKrT_Δ63-67
StsA_AKrT_Δ57-71
StsA_AKrT_Δ45-82
StsA_AKrT_P46A/P48A/P82A/P83A
StsA_AKrT_ Δ1-83
StsA_AKrT_Δ1-83_ Δ1162-1225

3.2.5.2 List of Primers

CesA_AKrT_K653A_1	5' ctttaaaacggatagtggggcaataacgagaaatgcatttc 3'
CesA_AKrT_K653A_2	5' gaaatgcatttctcgttattgccccactatccgttttaaag 3'
CesA_AKrT_K1218A_1	5' ccgcttgatgaatatggtgcagtagatcaaactcgtttgg 3'
CesA_AKrT_K1218A_2	5' ccaaacgagtttgatctactgcaccatattcatcaagcgg 3'
StrA_AKrT_K636A_1	5' gaaaaccagctctggcgcaattgaacgtgccca 3'
StrA_AKrT_K636A_2	5' tgggcacgttcaattgcgccagagctggttttc 3'
StsA_AKrT_Δ63-67_1	5' CTGGATAAACACATTGAACAGGCCATCCAGACCCA 3'
StsA_AKrT_Δ63-67_2	5' TGGGTCTGGATGGCCTGTTCAATGTGTTTATCCAG 3'
StsA_AKrT_Δ57-71_1	5' TGTCGCACCTGTTTCTGCAAAAAGCAGTTCCGG 3'
StsA_AKrT_Δ57-71_2	5' CCGGAACTGCTTTTTGCAGAAACAGGTGCGACA 3'
StsA_AKrT_Δ45-82_1	5' CGCGAAAAAACGATCCCAGCCCTGAGCTAC 3'
StsA_AKrT_Δ45-82_2	5' GTAGCTCAGGGCTGGGATCGTTTTTTCGCG 3'
StsA_AKrT_P46A_P48A_1	5' GAAAAAACGATCCGTGCGCGCGCGTATCATCTGTCGC 3'
StsA_AKrT_P46A_P48A_2	5' GCGACAGATGATACGCGCGCGCACGGATCGTTTTTTC 3'
StsA_AKrT_P82A_P83A_1	5' CTCTGGCCGACATGGCAGCAGCCCTGAGCTACG 3'
StsA_AKrT_P82A_P83A_2	5' CGTAGCTCAGGGCTGCTGCCATGTCGGCCAGAG 3'
Stra_AKrT_Δ1162-1225_1_1	TTGTTAACAGCAAAAATGAA
Stra_AKrT_Δ1162-1225_1_2	CAGATCCGCTACCTGATTCGGTTTGCATTTC
Stra_AKrT_Δ1162-1225_2_1	CGAATCAGGTAGCGGATCTGGTATGGATGC
Stra_AKrT_Δ1162-1225_2_2	TTCATTTTTGCTGTTAACAA

3.3 The Ketoreductase Domain of the Initiation Module of Stratospherulide Synthetase

3.3.1 Introduction

3.3.1.1 Prevalence of Ketoreductase Enzymes

Short-chain dehydrogenases/reductases and related enzymes (SDR) is a large super-family of proteins with a wide variety of functions, including over 46,000 identified proteins across species, more than 70 genes in human alone, and over 300 crystal structures deposited in the PDB [124]. In spite of the relatively low level of sequence identity between members of the SDR family (10-30%), their 3D structures are well conserved [125]. The core component of SDR is composed of a Rossmann-like fold comprised of a seven-stranded β-sheet flanked by three α-helices on either side, with an two helices over the active site [125].

3.3.1.2 Significance of Ketoreductase Enzymes

Enzymes are often preferred over chemical catalysts for biotechnological applications for various reasons: they are natural biocatalysts which can easily be extracted from natural sources, they present a regio-, enantio- and chemo-selectivity which can hardly be matched by chemical catalysts, they offer a high fidelity of product, and they produces quick reactions with high yield in large scale synthesis, using mild conditions and generally safe natural substrates, without producing harmful by-products [126]. Ketoreductases are the most commonly used class of enzymes in the pharmaceutical industry [127]. They are prized for their ability to perform a stereoselective reaction on a stereospecific substrate, resulting in the formation of an alcohol with a novel chiral center, using NADPH (a safe biological compound which can be extracted from natural sources) as a co-factor [128]. The oxidized NADP+ produced by the ketoreduction reaction can be safely recycled back into NADPH, most commonly using the glucose dehydrogenase (GDH) as a biocatalyst and glucose as substrate, producing solely the unharmful green by-product gluconic acid in the process [128]. As such, ketoreducatases are unparalleled for large scale synthesis, and can produce key intermediates for the synthesis of number crucial pharmaceutical compounds [129]: such as the antibiotic sulopenem (Iterum Therapeutics plc) [130], the widely used asthma medication montelukast (Singulair, Merck & Co., Inc.) [131], the anti-

retrovirals atazanavir (Teyetaz) and fosamprevanir, used in prevention and treatment of HIV infections [132], and the statin medication atorvastatin (Lipitor, Pfizer Inc.) [133], ranked highest-selling drug for seven consecutive years and considerably reducing the burden on our healthcare system for its use in cardiovascular diseases prevention and treatment [134]. The advantage of using ketoreductases for the pharmaceutical industry is well established, and novel methods, such as immobilization, are being developed to facilitate their use [135].

3.3.1.3 Mechanism for Enzymatic Catalysis of the Ketoreduction Reaction

The proposed mechanism for enzymatic reduction of a ketone using NADPH is depicted in Figure 36.A: The active site is composed of the catalytic triad Ser-Tyr-Lys [136], which interact with the substrate directly, and extended by a conserved Asn to form the active site tetrad [137]. The Tyr is activated into a catalytic base by a hydrogen bonding network involving the NADPH ribose hydroxyl group, the Lys and the Asn, which pulls the electron density away from the Tyr active proton and lowers its pKa [137]. The Tyr activates the ketone into an alkoxide carbocation intermediate, which is stabilized by the Ser, lowering the energy of activation for the hydride shift from position 4 of the nicotinamide ring of the NADPH cofactor to the carbonyl carbon [136].

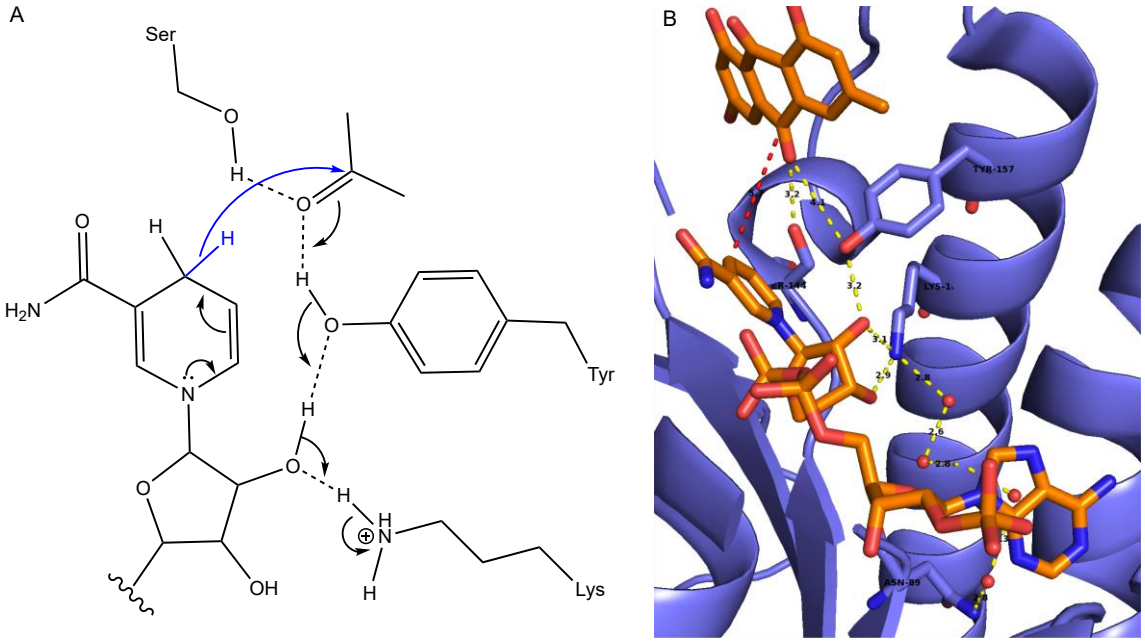


Figure 36: Mechanism of enzymatic ketoreduction using NADPH (A) Schematics of the mechanism of enzymatic ketoreduction in short-chain dehydrogenase/reductase (B) Active site of the Kr domain of StsA with NADPH and Emodin inhibitor modeled in by superposition with actinorhodin ketoreductase in complex with substrate and inhibitor (PDB:2RH4) [138].

3.3.1.4 Mechanism of Stereoselectivity of Enzymatic Ketoreduction in Polyketide Synthetase Ketoreductase Domain

Excised PKS Kr domains conserve their stereoselectivity, even after being introduced to a new PKS system, suggesting that Kr domains present intrinsic stereoselectivity for the reduced hydroxyl group [139]. "A-type" Kr domains are stereoselective for L-hydroxyl substituted substrates (S conformation), and "B-type" Kr domains are stereoselective for D-hydroxyl substituted substrates (R conformation) [140]. A model for the mechanisms of stereoselectivity was proposed for PKS Kr domains [97]. The cofactor NADPH is imbedded in its binding pocket in such a way as to expose the pro-S face hydride (labeled H_s in Figure 36) for transfer with the substrate [97]. This is conserved in A-type and B-type ketoreductases, suggesting that the stereoselectivity of the hydroxyl product is determined by the orientation in which the substrate bound to the 4' ppant arm is presented to the pro-S hydride by the

ACP domain (ACP domains in PKSs are equivalent to PCP domains in NRPSs) [97]. Alignments using over 30 different type A and type B PKS Kr domains demonstrated the presence of a conserved tryptophan (W motif) located 8 residues upstream from the catalytic tyrosine for A-type PKS Kr domains, and a conserved leucine-aspartate-aspartate (LDD) motif located about 55 residues upstream of the catalytic tyrosine for B-type PKS Kr domains [98].

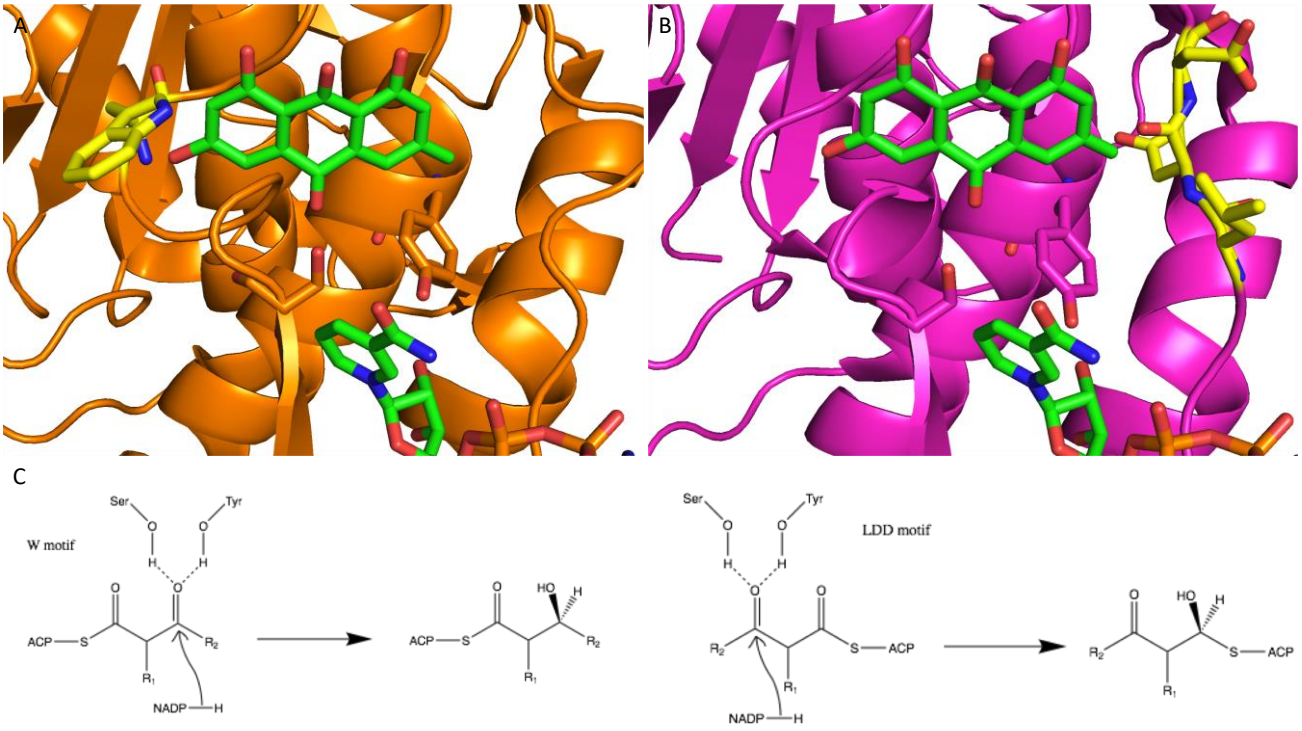


Figure 37: Model for mechanism of stereoselectivity in PKS Kr domains. (A) A type PKS Kr domain from hedamycin PKS (PDB:3SJU) [141] (B) B-type PKS Kr domain from tylosin PKS (PDB:2Z5L) [140] (C) Model for the mechanism of stereoselectivity in PKS Kr (A type W motif results in S-conformation on the left, B type LDD motif results in R-conformation on the right). Emodin inhibitor is modeled in by superposition with Actinorhodin ketoreductase in complex with substrate and inhibitor (PDB 2RH4) [138]. Emodin inhibitor and NADP+ co-factor shown in green. W motif and LDD motif residues shown in yellow.

As shown in Figure 37, the W motif (A-type) and the LDD (B-type) motif are located on either side of the Tyr/Ser duo (which interacts with the substrate directly), and are believed to guide

the 4’ ppant arm to present the substrate: if the W motif (A-type) is present adjacent to the serine, the 4’ ppant arm will enter the active site from the side of the W motif, resulting in an L-hydroxyl (S conformation). If the LDD (B-type) motif is present adjacent to the tyrosine residue, the 4’ ppant arm will enter the active site from the side of the LDD, resulting in a D-hydroxyl (R conformation) [97, 98]. The 4’ ppant arm bearing the substrate can enter the active site via two separate cavities shown in Figure 38: one on the W-motif (A-type) side of the active site, and another one on the LDD-motif (B-type) side of the active site.

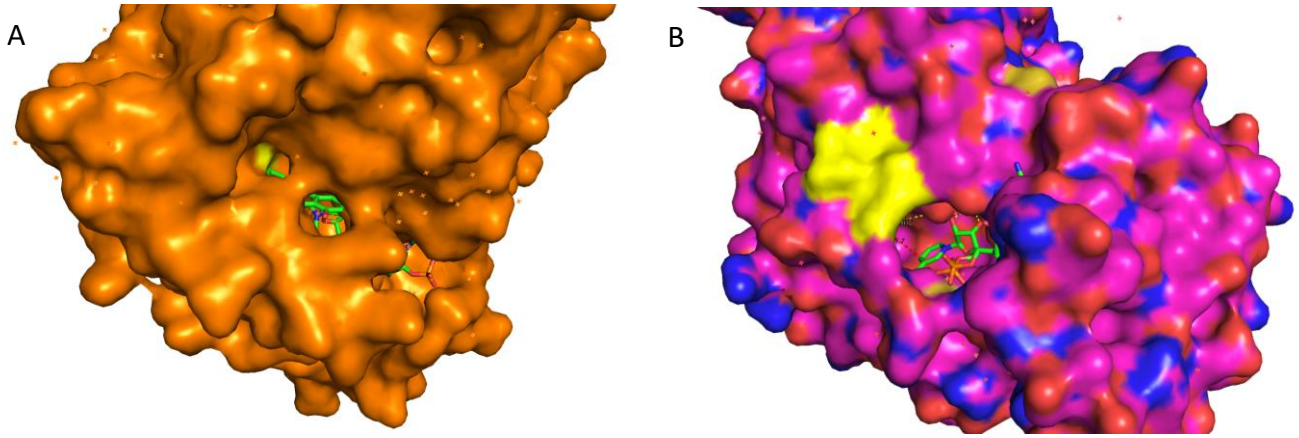


Figure 38: S/R-specific cavities in A-type vs B-type PKS K domains. (A) A type PKS Kr domain from hedamycin PKS (PDB:3SJU) [141] showing the S-conformer specific cavity (B) B-type PKS Kr domain from tylosin PKS (PDB:2Z5L) [140] showing the R-conformer specific cavity shown on the right. NADP+ co-factor shown in green.

3.3.1.5 Bromo-NADP⁺ Inhibitor Design and Proposed Synthesis

The most common cause for lack of diffraction is disorder within the crystal. For this reason, crystallographers often use chemical probes to trap the enzymatic complex in a certain conformation, reducing flexibility and disorder within the crystal. Most active sites have highest affinity for the transition state of the reaction they are catalyzing, since the transition state must be stabilized to reduce the energy of activation and facilitate the reaction. For this reason, it is reasonable to assume that a transition state analog may be a good inhibitor for an enzyme, trapping it in the conformation

required for the reaction to occur. We followed this theory to design an inhibitor of Kr domains, which could help us obtain a better resolution and/or add the PCP domain to our current model of the Kr domain-containing initiation module of StsA. Our model for the transition state of the ketoreduction reaction, based the model for the mechanism of enzymatic ketoreduction using NADPH (shown in Figure 36.A), is depicted in Figure 39.C: The electron density from the hydride is distributed between the position 4 of the nicotinamide ring and the carbonyl carbon to be reduced, which are expected to be located about 5Å apart (Figure 36.B). We designed the inhibitor bromo-$NADP^+$ (Figure 39.E) to mimic this transition state (Figure 39.C), in order to trap the complex in the conformation analogous to the transition state of the ketoreduction reaction (Figure 39.D).

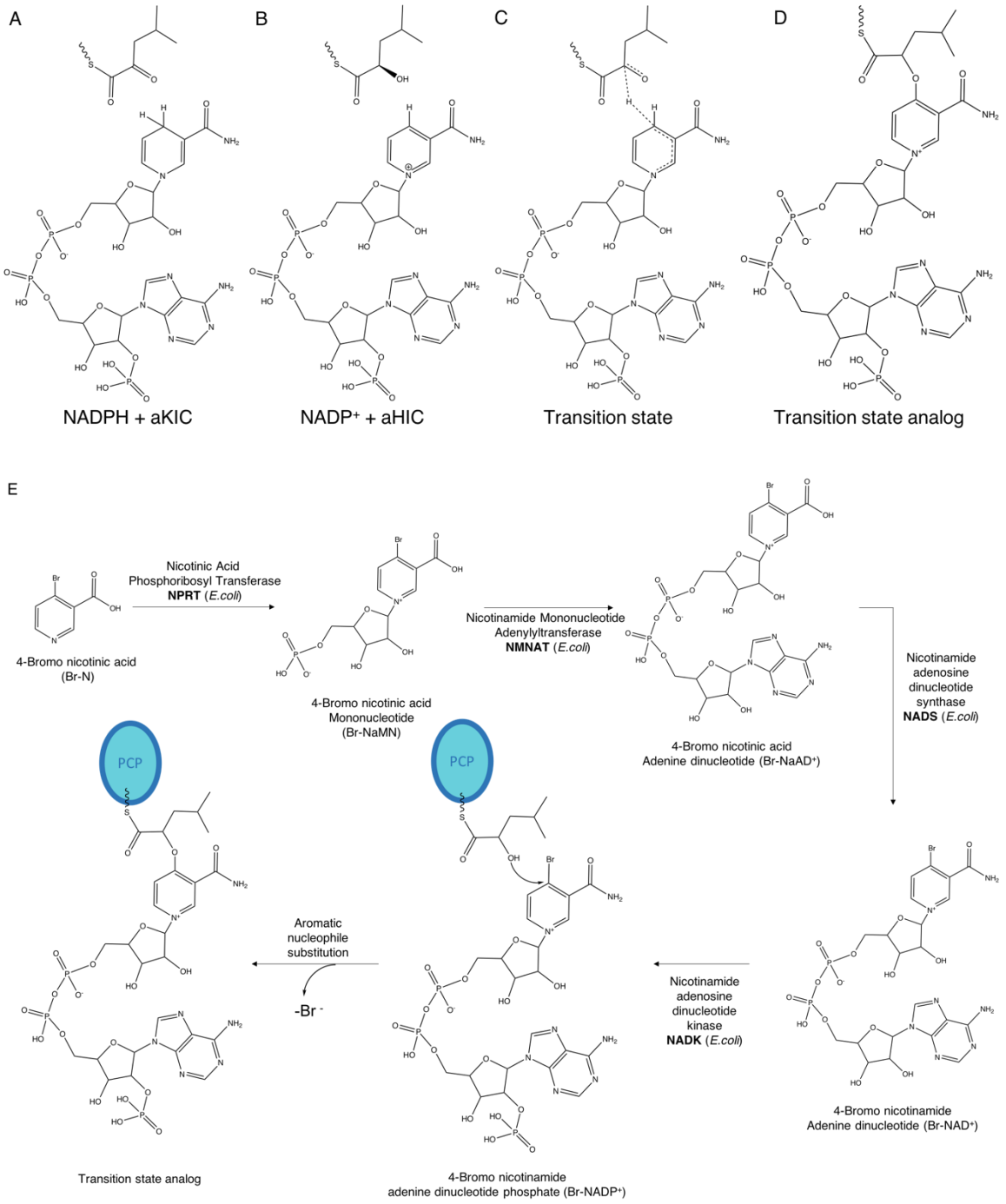


Figure 39: Design and synthesis of Br-NADP+ inhibitor to trap the Strs_AKrT complex into a conformation analogous to the ketoreduction transition state (A) Kr domain substrates αKIC bound to the 4' ppant arm and NADPH (B) Kr domain products αHIC bound to the 4' ppant arm and NADP+ (C) Model for the

transition state of the ketoreduction reaction (D) Proposed chemical analog for the transition state of the ketoreduction reaction. (E) Proposed strategy for the enzymatic synthesis of the Br-NADP+ inhibitor.

We propose a strategy for enzymatic synthesis of the Br-$NADP^+$ inhibitor, depicted in Figure 39.E. The nicotinamide phosphoribosyl transferase (NPRT) from *E. coli* [142] may be used to condense the original compound 4-bromo-nicotinic acid (Sigma) with phosphoribosylpyrophosphate (PRPP) (Sigma), to yield the product 4-bromo-nicotinic acid mononucleotide (Br-NaMN). The *E. coli* NPRT has an optimum pH of 7.5, and surprisingly requires ATP for its function, unlike other phosphoribosyltransferases [142]. In the structure of NPRT from *Thermoplasma acidiphilum* in complex with NaMN (PDB:1YTK), the catalytic cleft is relatively shallow about the nicotinic ring, and there is no significance steric hindrance around the 4' position, suggesting that the NPRT may be able to accommodate a bromine at the 4' position of the nicotinic ring [143].

Br-NaMN may be adenylated using ATP into 4-bromo-nicotinic acid adenine dinucleotide (Br-$NaAD^+$) by the *E. coli* enzyme nicotinamide mononucleotide adenyltransferase (NMNAT), which presents higher activity for NaMN than it does for nicotinamide mononucleotide (NMN), but can accommodate both substrates [144]. The reaction is typically carried out at pH 8.0, with 2mM MgCl and 2mM ATP, at 37°C [145]. In the structure of NMNAT from *E. coli* in complex with $NaAD^+$ (PDB:1K4M), the product is imbedded vertically in a relatively planar catalytic pocket, with the 4' position of the nicotinic ring pointing outwards, suggesting that the nicotinamide monoclucleotide adenylyltransferase (NMAT) may be able to accommodate a bromine at the 4' position of the nicotinic ring [146]. The resulting Br-$NaAD^+$ may be converted to Br-NAD^+ using the *E. coli* enzyme NAD^+ synthetase, using NH_4^+ at pH 7.5, for 30 minutes at 30°C [147]. As seen in the structure of the NAD synthase from *E. coli* in complex with $NaAD^+$ (PDB:1WXG), there is a tryptophan at position 254 that might cause some steric hindrance with a bromine at position 4' of the nicotinic [148].

Alternatively, Br-NMN can be adenylated using ATP into 4-bromo-nicotinamide adenine dinucleotide (Br-NAD) by the *E. coli* enzyme NAD-repressor (NadR). The NADR enzyme is a transcriptional repressor of the genes involved in the *de novo* synthesis of NAD and its salvage route,

as well as a repressor of the transport of NMN; interestingly, NadR has been shown to present NMAT activity as well [149]. NadR has an optimum pH of 8.6 and requires divalent cation to operate such as Mg^{2+}, but an 4-fold increase in activity was noted when using Ni^{2+} at 1mM [149]. NadR is the only bacterial NMNAT which is highly specific for the amidated form of the nicotine ring, and adenylates NMN 170 times faster than is does NaMN [144, 149]. In the structure of NadR from Haemophilus influenzae in complex with NAD (PDB:1LW7), the product is imbedded into a deep catalytic pocket, where is a tryptophan at position 149 that might cause some steric hindrance with a bromine at position 4' of the nicotinic ring [150].

Br-NAD^+ (or Br-$NaAD^+$) may be phosphorylated into 4-bromo-nicotinic acid adenine dinucleotide phosphate Br-$NADP^+$ (or Br-$NaADP^+$, respectively,) using the NAD kinase from *E. coli* [151]. The NADK was shown to be most efficient at pH 7.5 at 60°C, and present a very high specificity for NAD over NADH (furthermore, the presence of NADH or NADPH strongly inhibits the activity of the NADK) [151]. The activity of *E. coli* NADK for $NaAD^+$ has not been investigated to my knowledge. NaADP is present *in vivo* and acts as a Mg^{2+}-mobilizing messenger , but it is presumed to be made by exchanging the nicotinamide ring from NADP with nicotinic acid by the ADP-ribosyl cyclase [152] The structure of NADK has been solved in complex with substrate and/or product from *Archaeoglobus fulgidus* (PDB:1Z0U), from *Listeria monocytogenes* (PDB:5EJI), from *Mycobacterium tuberculosis* (PDB:1Y3I) and in human (PDB:3PFN), among other organisms.

The Br-$NADP^+$ inhibitor thusly synthetized may either be coupled *in vitro* with the reduced α-hydroxyacid attached to coenzyme A by aromatic nucleophilic substitution [153], and subsequently attached to the PCP domain of the NRPS using SFP, which has been shown to accommodate various compounds attached to the 4' ppant arm [154]. Alternatively, the Br-$NADP^+$ inhibitor can be incubated with the NRPS already modified with the 4' ppant arm bound to the reduced α-hydroxyacid to allow the aromatic nucleophilic substitution to happen *in situ*. The later method might be preferable, since we expect the active site to bring the inhibitor and the reduced α-hydroxyacid in close proximity, making the reaction more favorable.

3.3.2 Results

3.3.2.1 Stereoselectivity of StsA Kr domain

My goal for this section was to investigate the mechanism of stereoselectivity of NRPS Kr domains. To do so, I decided to compare the Kr domain of StsA, along with other NRPS Kr domains, with characterized PKS Kr domains in order to determine whether or not the model for stereoselectivity in PKSs could be extended to NRPS Kr domains as well.

The structure of StsA A-Kr-PCP initiation module was superimposed with the B-type PKS Kr domain from the tylosin PKS (PDB:2Z5L) [140], in order to compare the area of the LDD motif associated with stereoselectivity in B-type PKS Kr domains, as shown in Figure 40.

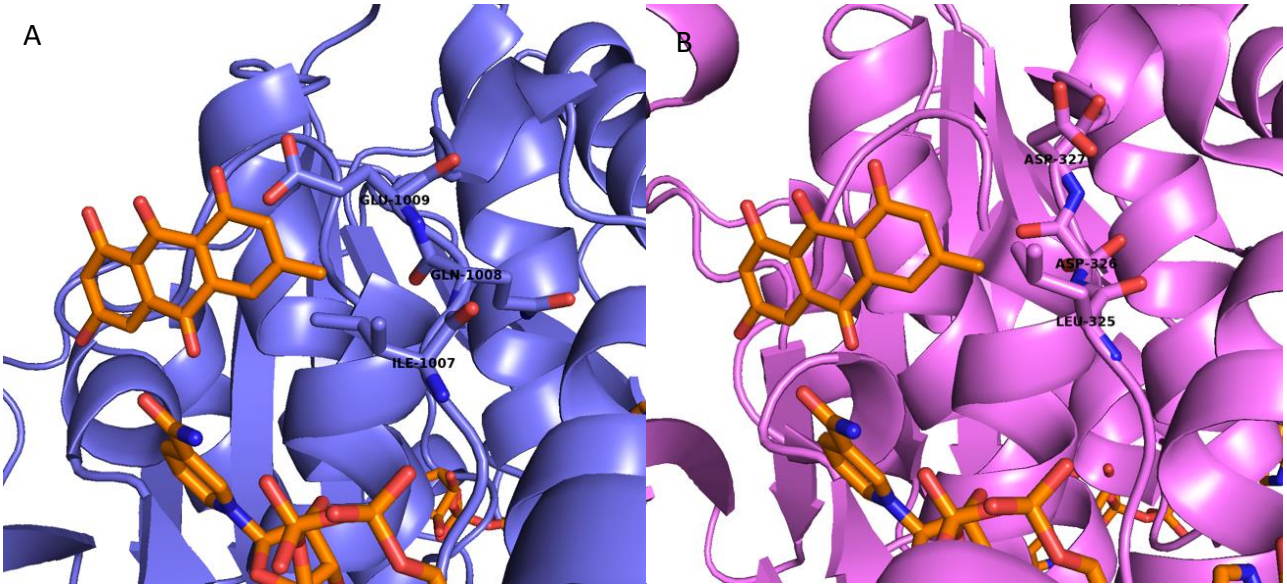


Figure 40: StsA Kr domain 3D domain alignment B-type PKS Kr. (A) IQE motif in StsA Kr domain (B) LDD motif in B-type PKS Kr domain from tylosin PKS (PDB:2Z5L) [140]. Emodin inhibitor modeled in by superposition with actinorhodin ketoreductase in complex with substrate and inhibitor (PDB 2RH4) [138].

The LDD motif in B-type PKS Kr aligns in 3D with the motif Ile-Gln-Glu (IQE) in the Kr domain of StsA (Figure 40). The sequence of the StsA Kr domain was aligned with the sequences of characterized NRPS Kr domains, shown in Figure 41.

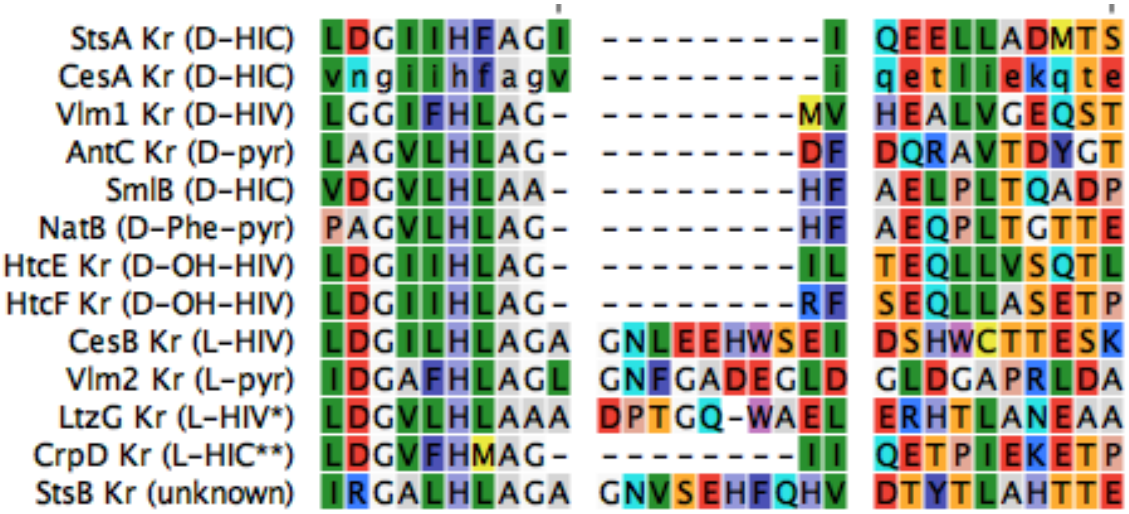


Figure 41: Sequence-based alignment of A-type and B-type characterized NRPS Kr domain. CesA Kr domain produces D-HIC and CesB Kr domain produces L-αHIV [6]. Vlm1 Kr domain produces D-αHIV and Vlm2 Kr domain produces L-pyruvate [11]. CrpD Kr domain is predicted to produce L-αHIC [57]. HtcE and HctF part of the hectochlorin biosynthetic gene cluster both contain a Kr domain predicted to produce L-dihydroxy-αHIV [121]. AntC from antimycin gene cluster contains a Kr domain predicted to produce D-pyruvate; analogs SmlB Kr domain from JBIR-06 synthase produces D-αHIC, and NatB Kr domain from neo-antimycin synthetase produces D-phenyl-pyruvate [122]. KtzG Kr domain from kutznerides cluster produces L-dihydroxy-αHIV [123].

The IQE putative motif, which aligns with the LDD motif in B-type PKS Kr domains (Figure 41), in not conserved in other NRPS Kr domains. However, the third residue of the motif, the Glu (E) residue located 3 to 4 residues downstream of the conserved LDGILHLAG motif, is conserved among B-type NRPS and absent in A-type NRPS (with the exception of AntC which instead contains an Asp residue immediately downstream). In the structure of StsA_AKrT, this Glu residue aligns with the third Asp residue of the LDD motif found in B-type PKS Kr domains.

The A-type NRPS Kr domains contain an additional 9-amino acid long sequence immediately downstream of the conserved LDGILHLAG motif, which is absent in B-type NRPS (Figure 41). The Kr domain of CrpD is predicted to product an L-α-hydroxyacid (L-αHIC), but seems to follow the sequence pattern of B-type NRPS Kr domains (Figure 41). The stereoselectivity of the Kr domain of StsB is unknown, but seems to follow the sequence pattern of A-type NRPS Kr domains.

3.3.2.2 Affinity of StsA Kr domain for NADPH cofactor.

My goal for this section was to investigate the relative affinity of the Kr domain of StsA for NADPH and related molecules, in order to gather fundamental knowledge on the binding affinity of NRPS Kr domains and potentiate this knowledge to the development of specific ligands and inhibitors.

The intensity of fluorescence emission of NADPH and NADH (λ_{exc}=340 nm, λ_{emm}=445 nm) is enhanced when bound to the active site of KR domains. It is then possible to study the binding of co-factors to the active site of the Kr domain of StsA by fluorescence. The binding of $NADP^+$ may be investigated by competition with NADPH. Following this logic, the affinity of the Kr domain of StsA for NADH and $NADP^+$ was investigated by a fluorescence assay, the results of which are shown in Figure 42.

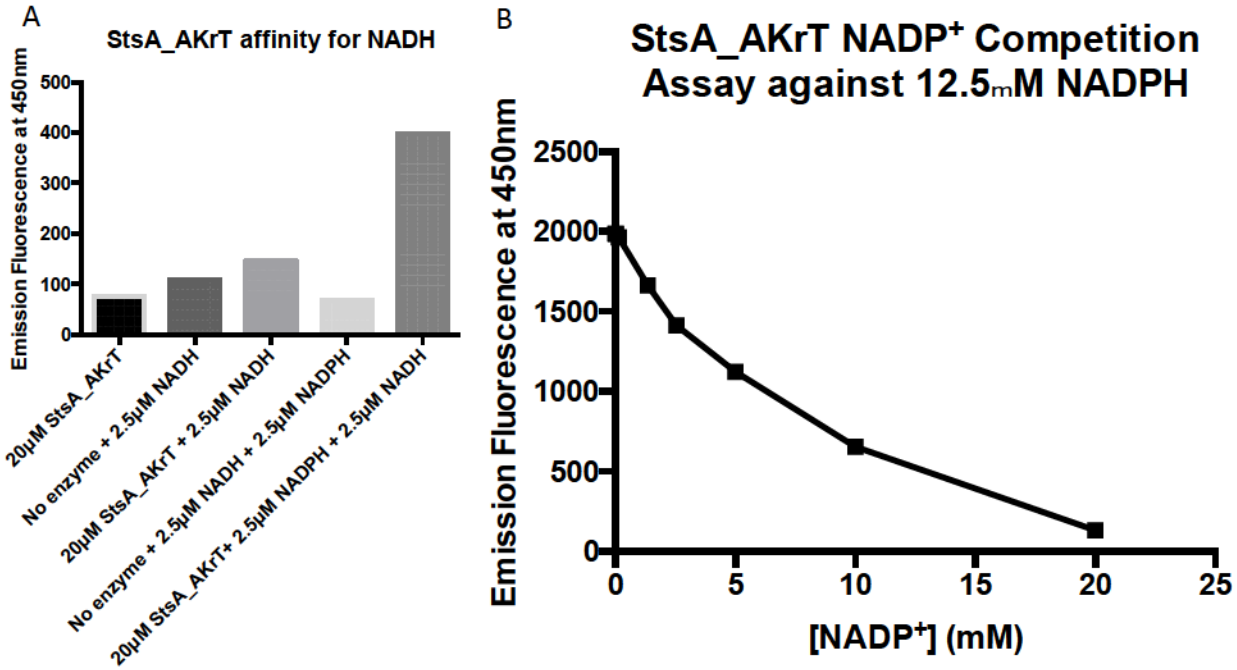


Figure 42: Affinity of the Kr domain of StsA for NADH and NADP+. (A) Fluorescence binding assay of NADH to the active site of the Kr domain of StsA. (B) Fluorescence competition assay of NADP+ against NADPH for binding of the active site of the Kr domain of StsA.

There is no significant increase in fluorescence when 10 μM of StsA_AKrT is added in excess (8-fold) to 2.5μM of NADH compared to the controls of enzyme with no NADH and 2.5μM NADH with no enzyme, (Figure 42). In contrast, there is a significant increase in fluorescence when 2.5μM NADPH is added to a mixture of 20μM StsA_AKrT with 2.5μM NADH. There is then no significant binding of NADH to the active site of the Kr domain of StsA at the concentrations investigated.

The affinity of the Kr domain of StsA for $NADP^+$ was investigated by competition assay against 12.5μM of NADPH mixed with 12.5μM of StsA_AKrT enzyme. There is no significant drop in fluorescence when adding $NADP^+$ in concentration ranging from 10μM to 145μM (Figure 42). A 100-fold excess of $NADP^+$ over NADPH is required for the fluorescence to start decreasing (7.8% decrease), and a 1600-fold excess of $NADP^+$ is required to completely abolish the fluorescence. The Kr domain of StsA has a low affinity for $NADP^+$ which can only compete with NADPH for the active site when added in extreme excess.

3.3.2.3 Enzymatic synthesis of Br-NADPH inhibitor

My goal for this section was to obtain a proof-of-concept for the mechanism of inhibition of ketoreductases by the Br-NADPH inhibitor and to initiate its synthesis.

In order to estimate whether or not the α-hydroxyl group of αHIC is able to attack the Br-NADPH and substitute the Br, the nucleophilic substitution reaction was attempted using bromo-nicotinic acid (Br-NA), and glycine as a nucleophile. To this end, 15mM of 4-Br-NA and 30mM glycine were stirred in aqueous solution with 50mM ammonium bicarbonate at pH 10.6. After 1h incubation at room

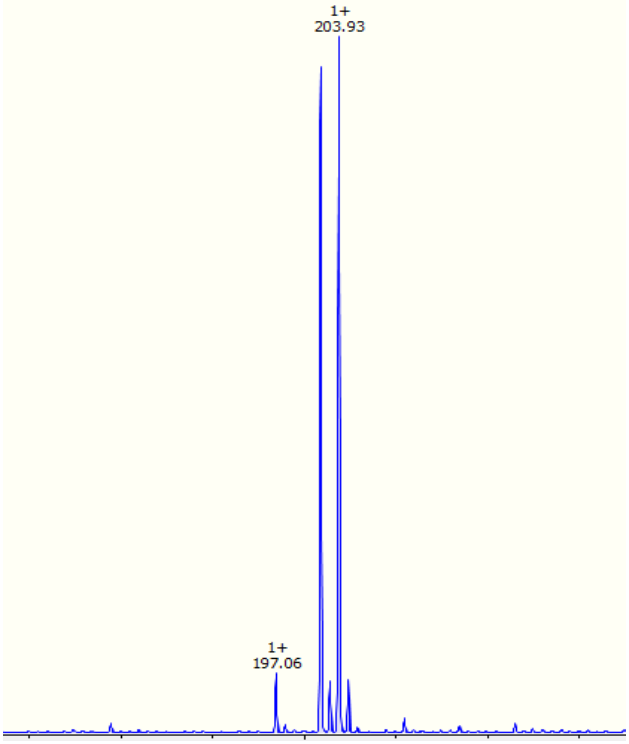


Figure 43: Mass spectrometry trace showing the product of aromatic nucleophilic substitution of glycine onto Br-nicotinic acid. Br-Nicotinic acid weighs 202g/mol and glycine-nicotinic acid weighs 197g/mol.

temperature, the sample was analyzed by direct injection mass spectrometry. A small amount of product with the expected mass was formed (Figure 43).

The synthesis enzymes NPRT, NADR and NADK were successfully cloned, over-expressed and purified. The identities of all three enzymes were confirmed by LC-ESI-MS. The first reaction with NRPT was attempted using nicotinic acid (NA) in order to optimize the synthesis and purification protocol. An adenylation assay was performed using NA and ATP with NPRT, suggesting that the reaction goes to 64% completion with NA after an overnight incubation at room temperature. However, the product was not easily purifiable by LC using the Agilent® ZORBAX Extend-C18 column.

3.3.3 Discussion

3.3.3.1 On the affinity of StsA Kr domain for NADPH cofactor.

The Kr domain of StsA has a very high affinity for its co-factor NADPH. NADH did not bind the active site of the Kr domain, even when the latter was present in large excess. It is worth noting that NADH and NADPH are found in similar concentrations in most cellular tissue according to a recent study using a novel extraction method to measure nicotinamide nucleotides [155], it is then not surprising that the NRPS Kr domain would have evolved a preference for the anabolic co-factor NADPH over the catabolic NADH. Moreover, the Kr domain shows show a high preference for NADPH over $NADP^+$ in the competition assay: this is not surprising since most enzymes have higher affinity for their substrate than

for their product. NADPH is present in 2 to 10 fold excess over $NADP^+$ in most tissues [155], which further drives the release of $NADP^+$ once the reaction is complete and allows for a new NADPH molecule to bind the active site.

3.3.3.2 On the mechanism for stereoselectivity

When superimposing the structure of the StsA Kr domain with B-type PKS Kr domains, the LDD motif (marker for B-type specificity in PKS Kr domains) aligns with the IQE motif in StsA. According to sequence-based alignment with NRPS Kr domains, the Glu residue of the IQE putative motif is conserved among B-type NRPS Kr domains. This suggests that the LDD motif in PKS Kr domains might be replaced by the Glu residue (with diversified supporting residue) in NRPS Kr domains, but the limited number of sequences of characterized NRPS Kr domains available for alignment prevents us from making any definitive conclusions at this time. The alignment of A-type NRPS Kr domains revealed the presence of an additional 9-amino acid-long sequence immediately upstream of the LDD sequence; due to its location, this added sequence may be blocking the access to the active site via the D-conformer-specific cavity.

The structure of the A-type Kr domain from amphotericin PKS in complex with a 4' ppant arm analog loaded with substrate was elucidated by Liu *et al.* to a resolution of 1.8Å [52]. In this structure, the arm indeed presents the substrate through the S-specific (or L-specific) cavity, on the side of the W motif. To the best of my knowledge, this structure is the only example of a Kr domain in complex with the arm presenting its substrate to the active site. In order to confirm the mechanism for stereoselectivity of Kr domains, the structure of a B-type Kr domain in complex with the 4' ppant arm presenting the substrate would be useful. To aid this, the proposed Br-NADPH inhibitor may be used.

3.3.3.3 On the Br-NADPH Inhibitor

The Br-NADPH inhibitor is designed to trap the NRPS or PKS complexes in the ketoreduction conformation, by making a covalent bond between the 4' position of the nicotinamide ring of NADPH and the α-hydroxyl group of the substrate loaded onto the 4'ppant arm. The nucleophilic substitution reaction on Br-NA is possible using glycine as a nucleophile, suggesting that the Kr domain may indeed be inhibited following the proposed mechanism. Moreover, the product of the NPRT reaction, Br-

NaMN, and the final product Br-NADPH both contain a pyridinium ion due to the addition of the phosphoribose moiety. The pyridinium ion is expected to be 1000-fold more reactive than the pyridine equivalent found in Br-NA. The pyridinium ion is able to accommodate the additional lone pair of electrons during to the nucleophilic attack and stabilize the transition state, making the reaction much more favorable. As final product of the synthesis, the Br-NADPH inhibitor is then expected to be sufficiently reactive to act as a suicide inhibitor and trap the substrate loaded onto the arm when presented to the active site of the Kr domain by the PCP domain. Purification of the synthesis intermediate is expected to be more amenable to optimization using a polar LC column.

3.3.3.4 On the need for non-oxidizable NADPH analog and the design of a 1-C-nicotinamide adenosine dinucleotide phosphate inhibitor

In order to visualize the specific interactions in the Kr domain active site with the substrate, a higher resolution structure in complex with NADPH and αKIC is needed. A construct for the Kr domain of StsA was successfully cloned, over-expressed in *E. coli* and purified for crystallography (results not shown). Crystals were obtained by sparse matrix, but they were not easily optimizable to diffract to a sufficient resolution for structure elucidation. Co-crystallization with substrate analog may help to improve the quality of the crystals. Since there is no non-oxidizable version of NADPH available to the best of my knowledge, many PKS Kr domain structures were obtained by co-crystallizing the enzyme with the substrate to be reduced and $NADP^+$. The enzyme is stalled in the substrate binding state since $NADP^+$ is the product of the reduction reaction, not the substrate.

Using an inhibition by product mechanism is not ideal in this case for several reasons. First, the cofactor NADPH has been shown to bind the active site first, inducing a slight conformational change in the active site of the Kr domain of FabG [156]. Once the active site is ordered by NADPH binding, the substrate binds the catalytic pocket directly on top of the cofactor [96]. Therefore, the structure of a Kr in complex with $NADP^+$ and substrate does not capture the native substrate binding state. Second, the nicotine ring from $NADP^+$ and NADPH are substantially different in terms of structure (the nicotinamide ring is flat in $NADP^+$, and adopts a tensed boat conformation in NADPH), electron density (the nitrogen of the nicotinamide ring is positively charged in $NADP^+$, when it bears a lone pair of electron in NADPH)

and in terms of biophysical and chemical properties (resonance, pi-stacking interaction, hydrogen bonding, etc.). $NADP^+$ is aromatic and more hydrophilic than NADPH. As a result, $NADP^+$ and NADPH are expected to interact with the active site residues and the substrate in very different ways, making $NADP^+$, in my opinion, a poor structural analog for NADPH in the context of co-crystallization. Finally, we have shown that the Kr domain of StsA has a much higher affinity for NADPH than is does for $NADP^+$. Most enzymes have higher affinity for their substrate than their product, but this property may be even more important in NRPSs, since the various conformational changes required for synthesis are not triggered by ATP (or GTP) hydrolysis as is the case for ribosomal synthesis. Rather, it seems that the cycle is driven solely by tethered Brownian motion and the affinity of the various active sites binding pockets for the cognate residue they are specific for. These findings prove that inhibition by product using $NADP^+$ is not ideal in order to obtain a structure of the substrate bindings state and visualize the specific interaction between active site the residues and the substrates.

For this reason, we propose the design of a 1-C-nicotinamide adenine dinucleotide phosphate (C^1-NADPH) inhibitor as non-oxidizable mimic of NADPH, shown in Figure 44, to be used for crystallization.

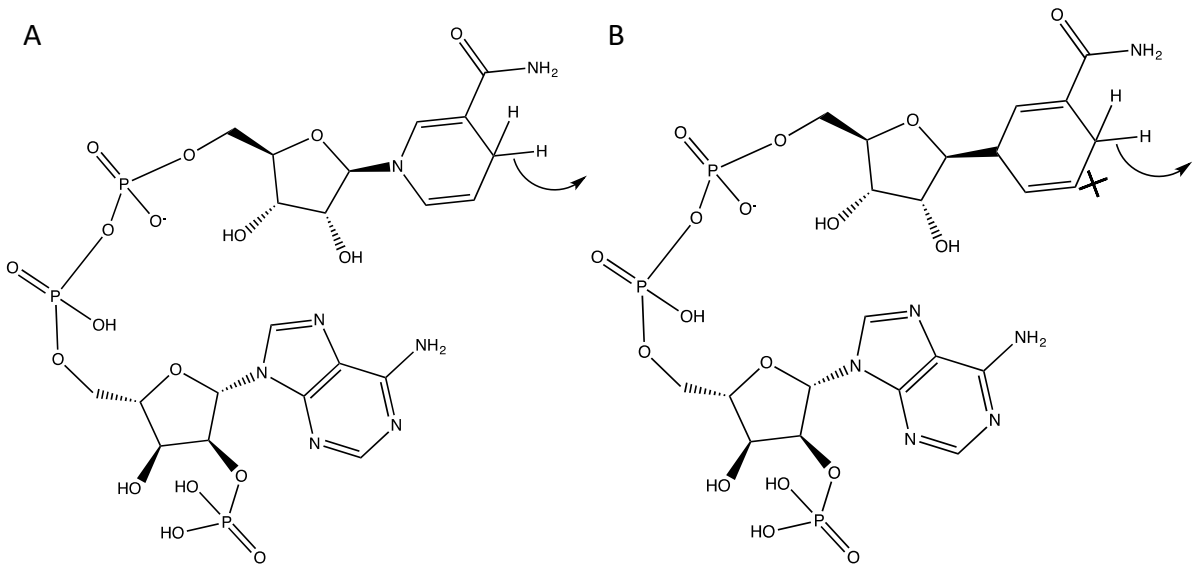


Figure 44: 1-C-nicotinamide adenine dinucleotide phosphate (C^1-NADPH) inhibitor (A) Schematics of NADPH chemical structure (B) Schematics of C^1-NADPH chemical structure.

NADPH is able to donate a hydride by elimination: the nitrogen of the nicotinamide ring bears a lone pair of electrons, which can be shared with the adjacent carbon to form a double bond between N^1 and C^2; as a result, the double bond between C^2 and C^3 is relayed to C^3 and C^4, chasing out the hydride group which is eliminated as a leaving group. If the N^1 nitrogen is replaced by a carbon, the geometry of the molecule is roughly conserved but its chemical properties are dramatically altered: the carbon bears no lone pair of electrons, the hydride cannot be eliminated and donated to the carbonyl carbon of the substrate for the ketoreduction reaction to occur. The C^1-NADPH would then mimic the co-factor structurally without having the ability to be oxidized, and act as a substrate analog inhibitor which could be used for crystallography or biochemical assays, in a manner analogous to AMPcPP.

3.3.4 Material and Methods

3.3.4.1 Cloning of the Br-NADPH Synthesis Enzymes

Nicotinic acid phosphoribosyl transferase (NPRT), NAD repressor (NADR) and NAD kinase (NADK) were amplified from *E. coli* genomic DNA and cloned into pBAC_Tandem_forward by ligation-independent cloning (LIC). *E. coli* genomic DNA was purified using MasterPure™ DNA purification kit (Epicentre®) according to the manufacturer's recommendations. For ligation-independent cloning, the LIC_ pBAC_Tandem_forward vector (pBAC_Tandem_forward vector containing the LIC insertion sequence TCGACGACGACAAGGCCTTGGCTCTCCTCA instead of the multiple cloning site, including the restriction site for StuI AGGCTC) is digested with StuI restriction enzyme (NEB) and treated with the T4 polymerase (NEB) (the T4 DNA polymerase acts as a nuclease in the condition where the complementary nucleotide to be added is not present in the media) with 25mM deoxythymidine tri-phosphate (dTTP) for 30 minutes at room temperature to expose the vector LIC overhangs. The pncB gene (coding for NPRT) was amplified from *E. coli* genomic DNA. The amplified insert was treated with the T4 polymerase in the presence of 25mM deoxyadenosine tri-phosphate (dATP) to expose the insert LIC overhangs complementary to the vector LIC overhangs. 1μl of vector is incubated with 2μl of insert for 5 minutes at room temperature before adding 1μl of 25mM Ethylenediaminetetraacetic acid (EDTA)

and transforming into10β competent *E. coli* cells (NEB), to yield the pBac_NPRT plasmid. pBac_NADR and pBac_NADK were generated by LIC as described for pBAC_NPRT.

3.3.4.2 Expression and Purification of Br-NADPH Synthesis Enzymes

NPRT, NADR and NADK were purified by nickel column chromatography according to the same protocol as for StsA_AKrT_(TEV)_His, except for the composition of the nickel column elution buffer (50mM Tris-HCl pH 8.0, 200mM NaCl, 500mM Imidazole and 2mM βME), and the protein was eluted with a gradient from 0% nickel column elution buffer to 100% nickel column elution buffer over 10 column volumes. The purity of the proteins was assessed by SDS-PAGE. The protein solutions were further purified by SEC using the HiLoad 16/60 Superdex 75 column (GE Healthcare) according to the same protocol as for StsA_AKrT_(TEV)_His. The purity of the proteins was assessed by SDS-PAGE. The protein was flash frozen in liquid nitrogen and stored at -80°C in 1mg aliquot in SEC buffer.

3.3.4.3 LC-MS

The LC-MS assay was performed as described in Chapter II section 2.4.3.

3.3.4.4 Adenylation Assay

The adenylation assay was performed as described in Chapter III section 3.1.4.4.

3.3.4.5 NADPH Consumption Assay

To investigate the ketoreduction activity of the Kr domain, we measured the NADPH consumption by single turn-over of NADPH using a fluorescence assay. The intensity fluorescence emission of NADPH (λ_{exc}=340 nm, λ_{emm}=445 nm) is enhanced when bound to the active site of KR domains. By adding the enzyme in excess with respect to NADPH, we ensure that all the NADPH present in the reaction mix is bound to the active site of the Kr domain (thus enhancing its fluorescence activity). We can then monitor NADPH consumption by measuring the loss of fluorescence. 100µL reactions were prepared containing 25mM HEPES pH 8.0, 100mM NaCl, 2mM $MgCl_2$, 2.5µM NADPH and 9µM enzyme, and were placed in a 96-well black-bottom plates (Corning). The reactions were started by adding 20 µL of a 5X cocktail containing 10mM α-KIC and 10mM ATP. The fluorescence intensity (λ_{exc}=340 nm, λ_{emm}=445 nm) was recorded using Molecular Devices plate reader. Data was analyzed and plotted using the program GraphPad Prism 6.

3.3.4.6 NADPH Competition Assay

The NADPH competition assay was performed following the same protocol as for the NADPH consumption assay, except that NADPH was added first, followed by various concentration of $NADP^+$. $NADP^+$ des not emit fluorescence at this wavelength upon binding to the active site of the Kr domain. Binding of the competitor is measured by loss of fluorescence.

4 Chapter IV – Conclusion

4.1 On crystallography

The crystallography of the initiation module of Sts was extremely challenging and required a combined nine years of work and invaluable collaboration to be seen to completion. Regarding the crystallography of NRPSs in general, it is essential to finely control the presence of the 4' ppant arm and the substrate loaded onto it. The development of an LC-ESI-MS protocol powerful enough to monitor these modifications, developed by Dr. Diego Alonzo, was crucial to obtain a quality of sample sufficient for crystallization. Moreover, the importance of post-crystallization treatment should not be underestimated: the optimization of the protocol for cryopreservation and dehydration drastically improved the resolution of the crystals from 7.0-8.0Å to 3.8Å, when all options for optimization of the crystallization protocol were exhausted. Finally, the iterative omit-map sharpening process combined with 4-fold NCS averaging optimized and developed by our collaborator Dr. Jimin Wang considerably improved the level of details displayed on the StsA A-Kr partial initiation module structure.

4.2 On the general architecture of the Kr containing initiation module of Sts NRPS

The StsA_AKrT structure presented here contains novel features never seen before: the first structure of a Kr-containing NRPS module, which shows how the Kr domain interacts with the other domains and is incorporated within the module, the first structure of an NRPS Kr domain and the structure of the first ketoacid-selecting A domain. Contrary to what was predicted, the A domain in uninterrupted by the Kr domain and the two domains make virtually no surface contact. The structure also revealed a surprising feature: the pA_{sub}, a novel domain interrupted by the A-Kr didomain, dividing it in two parts located 1112 residues apart in the primary sequence, brought together through a 39 amino-acids-long strand swap between dimeric partners, which induces dimerization. The Kr domain is positioned on top of the "C-A scaffold" and does not clash with the other domains, but the pA_{sub} is located at the same position as the C domain during thiolation. As such, we can expect significant conformational changes to occur to achieve all the subsequent steps of the synthesis of stratospherulide.

4.3 On the A domain

The α-ketoacid-selecting A domain characterized in this study displays the same overall structure and general characteristics as a α-amino acid-selecting A domain. However, the higher resolution structure of the α-ketoacid-selecting A domain of Sts presented in this study, combined with sequence-based alignment, highlighted several differences in highly conserved motifs and substrate binding residues. Interestingly, a proline mutation was identified in the StsA A domain, which accommodates the α-ketoacid substrate by distorting the neighboring residues into a conformation that allows for positive interactions with the substrate. Remarkably, the reverse mutation conferred onto the α-amino acid-selecting A domain of LgrA the ability to select an α-ketoacid as a substrate. More research is needed to fully understand the specifics of the mechanism of selection of α-ketoacids by NRPS A domains to potentiate their use for the development of novel natural products. To this end, special attention should be given to a conserved serine residue which appears to be involved in accommodating the α-ketoacid by avoiding a repulsive interaction.

4.4 On the pAsub and oligomerization

The pA_{sub} domain of StsA displays the same overall structure as the active A_{sub} of StsA and other NRPS modules. However, the pA_{sub} holds no catalytic function and is not required for the proper folding of the initiation module, nor is it required for its adenylation activity. More research is required to fully understand the implications, if any, of the pA_{sub} in later stages of synthesis, such as thiolation, ketoreduction and condensation. Dimerization of Sts is governed by a domain swap of the N-terminal portion of the pA_{sub} between dimeric partners, mediated by a 39-amino-acid long linker region. The composition and relative weight of the oligomeric populations can be controlled to some extent by controlling the length of the strand swapped between dimeric partners. It is still not clear whether the pA_{sub} is involved in the complex dynamic interactions observed during NRP synthesis, or if it is simply a remnant of an evolutionary deletion.

4.5 On the Kr domain

The NRPS Kr domain characterized in this study displays the same overall structure and general characteristics as its PKS cousins. Preliminary results seem to indicate that the mechanism for

stereoselectivity of the product is similar in PKSs and NRPSs, but more research is required to fully understand the specifics of this process. In particular, a high-resolution structure of both B-type and A-type NRPS Kr domains in complex with the 4' ppant arm of the PCP domain presenting the α-ketoacid substrate to the active site is needed in order to validate the model for stereo selectivity in PKS and NRPS Kr domains. To this end, we propose the design of two inhibitors designed to trap the Kr domain substrate in its binding phase, and to trap the NRPS Kr-containing module in the ketoreduction state.

4.6 Final statement

NRPSs are truly elegant and remarkable systems able to produce an immense variety of products. This feature makes them invaluable as a tool for research and development in our fight against global threats to our way of life, such as the antibiotic resistant crisis. In this study, we began to characterize an NRPS module able to incorporate an α-ketoacid into the peptide chain, resulting in the formation of an ester bond instead of the typical amide bond. The ester bond has a different energy, geometry, length and rotational freedom than the amide bond. The ability to incorporate α-ketoacids instead of α-amino acids provides another layer of diversity of substrate accessible to the NRPS. The stereoselectivity of product have yet to be fully characterized; this work hopefully sets the first stone towards a greater understanding of Kr domain-containing NRPSs and how to harness their potential towards the development of new drugs and green chemicals so desperately needed.

References

1. Finking, R. and M.A. Marahiel, *Biosynthesis of Nonribosomal Peptides.* Annual Review of Microbiology, 2004. **58**(1): p. 453-488.
2. Strieker, M., A. Tanovic, and M.A. Marahiel, *Nonribosomal peptide synthetases: structures and dynamics.* Curr Opin Struct Biol, 2010. **20**(2): p. 234-240.
3. Wang, H., et al., *Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(25): p. 9259-9264.
4. McErlean, M., J. Overbay, and S. Van Lanen, *Refining and expanding nonribosomal peptide synthetase function and mechanism.* J Ind Microbiol Biotechnol, 2019. **46**(3-4): p. 493-513.
5. Strieker, M., A. Tanovic, and M.A. Marahiel, *Nonribosomal peptide synthetases: structures and dynamics.* Curr Opin Struct Biol, 2010. **20**(2): p. 234-40.
6. Alonzo, D.A., N.A. Magarvey, and T.M. Schmeing, *Characterization of cereulide synthetase, a toxin-producing macromolecular machine.* PloS one, 2015. **10**(6): p. e0128569-e0128569.
7. Bloudoff, K., et al., *Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis.* Proceedings of the National Academy of Sciences of the United States of America, 2017. **114**(1): p. 95-100.
8. Oide, S., et al., *NPS6, encoding a nonribosomal peptide synthetase involved in siderophore-mediated iron metabolism, is a conserved virulence determinant of plant pathogenic ascomycetes.* The Plant cell, 2006. **18**(10): p. 2836-2853.
9. Ahmed, E. and S.J. Holmstrom, *Siderophores in environmental research: roles and applications.* Microb Biotechnol, 2014. **7**(3): p. 196-208.
10. Reimer, J.M., et al., *Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase.* Nature, 2016. **529**(7585): p. 239-42.
11. Huguenin-Dezot, N., et al., *Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid.* Nature, 2019. **565**(7737): p. 112-117.
12. Keating, T.A., et al., *The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains.* Nat Struct Biol, 2002. **9**(7): p. 522-6.
13. Conti, E., et al., *Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S.* Embo j, 1997. **16**(14): p. 4174-83.
14. Mitchell, C.A., et al., *Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains.* Biochemistry, 2012. **51**(15): p. 3252-63.
15. Ehmann, D.E., et al., *The EntF and EntE adenylation domains of Escherichia coli enterobactin synthetase: Sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates.* Proceedings of the National Academy of Sciences, 2000. **97**(6): p. 2509-2514.
16. Stachelhaus, T., et al., *Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain.* J Biol Chem, 1998. **273**(35): p. 22773-81.

17. Drake, E.J., et al., *Structures of two distinct conformations of holo-non-ribosomal peptide synthetases.* Nature, 2016. **529**(7585): p. 235-8.
18. Walsh, C.T., *Insights into the chemical logic and enzymatic machinery of NRPS assembly lines.* Nat Prod Rep, 2016. **33**(2): p. 127-35.
19. Weber, T. and M.A. Marahiel, *Exploring the domain structure of modular nonribosomal peptide synthetases.* Structure, 2001. **9**(1): p. R3-9.
20. Schmelz, S. and J.H. Naismith, *Adenylate-forming enzymes.* Curr Opin Struct Biol, 2009. **19**(6): p. 666-71.
21. Stachelhaus, T., H.D. Mootz, and M.A. Marahiel, *The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases.* Chem Biol, 1999. **6**(8): p. 493-505.
22. Challis, G.L., J. Ravel, and C.A. Townsend, *Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains.* Chem Biol, 2000. **7**(3): p. 211-24.
23. Reimer, J.M., et al., *Piecing together nonribosomal peptide synthesis.* Current Opinion in Structural Biology, 2018. **49**: p. 104-113.
24. Gulick, A.M. and C.C. Aldrich, *Trapping interactions between catalytic domains and carrier proteins of modular biosynthetic enzymes with chemical probes.* Nat Prod Rep, 2018. **35**(11): p. 1156-1184.
25. Schmitt, E., et al., *Crystal structure of aspartyl-tRNA synthetase from Pyrococcus kodakaraensis KOD: archaeon specificity and catalytic mechanism of adenylate formation.* Embo j, 1998. **17**(17): p. 5227-37.
26. Osman, K.T., et al., *Crystal structure of Bacillus cereus D-alanyl carrier protein ligase (DltA) in complex with ATP.* J Mol Biol, 2009. **388**(2): p. 345-55.
27. Du, L., Y. He, and Y. Luo, *Crystal structure and enantiomer selection by D-alanyl carrier protein ligase DltA from Bacillus cereus.* Biochemistry, 2008. **47**(44): p. 11473-80.
28. Hur, G.H., C.R. Vickery, and M.D. Burkart, *Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology.* Nat Prod Rep, 2012. **29**(10): p. 1074-98.
29. Scaglione, A., et al., *Structure of the adenylation domain Thr1 involved in the biosynthesis of 4-chlorothreonine in Streptomyces sp. OH-5093-protein flexibility and molecular bases of substrate specificity.* Febs j, 2017. **284**(18): p. 2981-2999.
30. Du, L., Y. He, and Y. Luo, *Crystal structure and enantiomer selection by D-alanyl carrier protein ligase DltA from Bacillus cereus.* Biochemistry, 2008. **47**(44): p. 11473-11480.
31. Osman, K.T., et al., *Crystal structure of Bacillus cereus D-alanyl carrier protein ligase (DltA) in complex with ATP.* J Mol Biol, 2009. **388**(2): p. 345-355.
32. Yonus, H., et al., *Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains.* J Biol Chem, 2008. **283**(47): p. 32484-32491.
33. Challis, G.L. and J.H. Naismith, *Structural aspects of non-ribosomal peptide biosynthesis.* Curr Opin Struct Biol, 2004. **14**(6): p. 748-56.
34. Bloudoff, K. and T.M. Schmeing, *Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity.* Biochim Biophys Acta Proteins Proteom, 2017. **1865**(11 Pt B): p. 1587-1604.

35. Di Lorenzo, M., et al., *Tandem heterocyclization domains in a nonribosomal peptide synthetase essential for siderophore biosynthesis in Vibrio anguillarum.* (1572-8773 (Electronic)).
36. Gao, X., et al., *Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain.* Nat Chem Biol, 2012. **8**(10): p. 823-30.
37. Gaudelli, N.M., D.H. Long, and C.A. Townsend, *beta-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis.* Nature, 2015. **520**(7547): p. 383-7.
38. Haslinger, K., et al., *X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis.* Nature, 2015. **521**(7550): p. 105-9.
39. Stachelhaus, T., A. Huser, and M.A. Marahiel, *Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases.* Chem Biol, 1996. **3**(11): p. 913-21.
40. Beld, J., et al., *The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life.* Nat Prod Rep, 2014. **31**(1): p. 61-108.
41. Miller, B.R. and A.M. Gulick, *Structural Biology of Nonribosomal Peptide Synthetases.* Methods in molecular biology (Clifton, N.J.), 2016. **1401**: p. 3-29.
42. Akey, D.L., et al., *Structural basis for macrolactonization by the pikromycin thioesterase.* Nature chemical biology, 2006. **2**(10): p. 537-542.
43. Bruner, S.D., et al., *Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE.* Structure, 2002. **10**(3): p. 301-10.
44. Samel, S.A., et al., *The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide.* J Mol Biol, 2006. **359**(4): p. 876-89.
45. Sundaram, S. and C. Hertweck, *On-line enzymatic tailoring of polyketides and peptides in thiotemplate systems.* Curr Opin Chem Biol, 2016. **31**: p. 82-94.
46. Calderone, C.T., et al., *A ketoreductase domain in the PksJ protein of the bacillaene assembly line carries out both alpha- and beta-ketone reduction during chain growth.* Proc Natl Acad Sci U S A, 2008. **105**(35): p. 12809-14.
47. Zwahlen, R.D., et al., *Bacterial MbtH-like Proteins Stimulate Nonribosomal Peptide Synthetase-Derived Secondary Metabolism in Filamentous Fungi.* ACS Synth Biol, 2019.
48. Labby, K.J., S.G. Watsula, and S. Garneau-Tsodikova, *Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis.* Nat Prod Rep, 2015. **32**(5): p. 641-53.
49. Keatinge-Clay, A.T., *The structures of type I polyketide synthases.* Natural Product Reports, 2012. **29**(10): p. 1050-1073.
50. Dittmann, J., et al., *Mechanism of cyclosporin A biosynthesis. Evidence for synthesis via a single linear undecapeptide precursor.* J Biol Chem, 1994. **269**(4): p. 2841-6.
51. Xu, F., et al., *Modified substrate specificity of a methyltransferase domain by protein insertion into an adenylation domain of the bassianolide synthetase.* J Biol Eng, 2019. **13**: p. 65.
52. Liu, C., et al., *Substrate-bound structures of a ketoreductase from amphotericin modular polyketide synthase.* J Struct Biol, 2018. **203**(2): p. 135-141.
53. Mori, S., et al., *Structural basis for backbone N-methylation by an interrupted adenylation domain.* Nature Chemical Biology, 2018. **14**(5): p. 428-430.

54. Miller, B.R., et al., *Analysis of the linker region joining the adenylation and carrier protein domains of the modular nonribosomal peptide synthetases.* Proteins, 2014. **82**(10): p. 2691-702.
55. Kamra, P., R.S. Gokhale, and D. Mohanty, *SEARCHGTr: a program for analysis of glycosyltransferases involved in glycosylation of secondary metabolites.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W220-5.
56. Cadel-Six, S., et al., *Halogenase genes in nonribosomal peptide synthetase gene clusters of Microcystis (cyanobacteria): sporadic distribution and evolution.* Molecular biology and evolution, 2008. **25**(9): p. 2031-2041.
57. Magarvey, N.A., et al., *Biosynthetic characterization and chemoenzymatic assembly of the cryptophycins. Potent anticancer agents from cyanobionts.* ACS Chem Biol, 2006. **1**(12): p. 766-79.
58. Li, T.L., et al., *Biosynthesis of the vancomycin group of antibiotics: characterisation of a type III polyketide synthase in the pathway to (S)-3,5-dihydroxyphenylglycine.* Chem Commun (Camb), 2001(20): p. 2156-7.
59. Du, L., et al., *An oxidation domain in the BlmIII non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in Streptomyces verticillus ATCC15003.* FEMS Microbiology Letters, 2000. **189**(2): p. 171-175.
60. Silakowski, B., et al., *New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of Stigmatella aurantiaca DW4/3-1.* J Biol Chem, 1999. **274**(52): p. 37391-9.
61. Westphal, K.R., et al., *Fusaoctaxin A, an Example of a Two-Step Mechanism for Non-Ribosomal Peptide Assembly and Maturation in Fungi.* Toxins (Basel), 2019. **11**(5).
62. Stack, D., C. Neville, and S. Doyle, *Nonribosomal peptide synthesis in Aspergillus fumigatus and other fungi.* Microbiology, 2007. **153**(Pt 5): p. 1297-306.
63. Schrettl, M., et al., *Distinct roles for intra- and extracellular siderophores during Aspergillus fumigatus infection.* PLoS pathogens, 2007. **3**(9): p. 1195-1207.
64. Oide, S., et al., *Individual and combined roles of malonichrome, ferricrocin, and TAFC siderophores in Fusarium graminearum pathogenic and sexual development.* Front Microbiol, 2014. **5**: p. 759.
65. Marahiel, M.A., *A structural model for multimodular NRPS assembly lines.* Natural Product Reports, 2016. **33**(2): p. 136-140.
66. Tarry, M.J., et al., *X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture.* Structure, 2017. **25**(5): p. 783-793.e4.
67. Chooi, Y.-H. and Y. Tang, *Navigating the fungal polyketide chemical space: from genes to molecules.* The Journal of organic chemistry, 2012. **77**(22): p. 9933-9953.
68. Zheng, J., et al., *Structural and functional analysis of A-type ketoreductases from the amphotericin modular polyketide synthase.* Structure, 2010. **18**(8): p. 913-22.
69. Zheng, J., et al., *The missing linker: a dimerization motif located within polyketide synthase modules.* ACS chemical biology, 2013. **8**(6): p. 1263-1270.

70. Bonnett, S.A., et al., *Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a cis-alkene.* Chemistry & biology, 2013. **20**(6): p. 772-783.
71. Gelband, H., et al., *The state of the world's antibiotics 2015.* Wound Healing Southern Africa, 2015. **8**(2): p. 30-34.
72. *http://healthycanadians.gc.ca/publications/drugs-products-medicaments-produits/antibiotic-resistance-antibiotique/alt/pub-eng.pdf.*
73. *http:///news-perspective/2018/11/european-study-33000-deaths-year-resistant-infections*.
74. *WHO Report 2015: Data tables (official report)*. Retrieved 2016-01-26].
75. *amr-review.org*.
76. Boucher, H.W., et al., *Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America.* Clinical Infectious Diseases, 2009. **48**(1): p. 1-12.
77. *Antibacterial Agents in Clinical Development*. Available from: https://apps.who.int/iris/bitstream/handle/10665/258965/WHO-EMP-IAU-2017.11-eng.pdf.
78. Alharbi, S.A., et al., *What if Fleming had not discovered penicillin?* Saudi journal of biological sciences, 2014. **21**(4): p. 289-293.
79. Ventola, C.L., *The antibiotic resistance crisis: part 1: causes and threats.* P & T : a peer-reviewed journal for formulary management, 2015. **40**(4): p. 277-283.
80. Winn, M., et al., *Recent advances in engineering nonribosomal peptide assembly lines.* Nat Prod Rep, 2016. **33**(2): p. 317-47.
81. Wang, M. and H. Zhao, *Characterization and Engineering of the Adenylation Domain of a NRPS-Like Protein: A Potential Biocatalyst for Aldehyde Generation.* ACS Catalysis, 2014. **4**(4): p. 1219-1225.
82. Brown, A.S., et al., *Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines.* Nat Prod Rep, 2018. **35**(11): p. 1210-1228.
83. Shivaji, S., et al., *Bacillus aerius sp. nov., Bacillus aerophilus sp. nov., Bacillus stratosphericus sp. nov. and Bacillus altitudinis sp. nov., isolated from cryogenic tubes used for collecting air samples from high altitudes.* Int J Syst Evol Microbiol, 2006. **56**(Pt 7): p. 1465-73.
84. Zhang, J., et al., *Enhanced electricity production by use of reconstituted artificial consortia of estuarine bacteria grown as biofilms.* Environ Sci Technol, 2012. **46**(5): p. 2984-92.
85. Walser, M., et al., *Synthesis of essential amino acids from their alpha-keto analogues by perfused rat liver and muscle.* The Journal of clinical investigation, 1973. **52**(11): p. 2865-2877.
86. Makarasen, A., K. Yoza, and M. Isobe, *Higher structure of cereulide, an emetic toxin from Bacillus cereus, and special comparison with valinomycin, an antibiotic from Streptomyces fulvissimus.* Chemistry–An Asian Journal, 2009. **4**(5): p. 688-698.
87. Suwan, S., et al., *Structure of cereulide, a cyclic dodecadepsipeptide toxin from Bacillus cereus and studies on NMR characteristics of its alkali metal complexes including a conformational structure of the K+ complex.* Journal of the Chemical Society, Perkin Transactions 1, 1995(7): p. 765-775.

88. Chalut, C., et al., *The nonredundant roles of two 4'-phosphopantetheinyl transferases in vital processes of Mycobacteria.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(22): p. 8511-8516.
89. Pfeifer, B.A., et al., *Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli.* Science, 2001. **291**(5509): p. 1790-2.
90. Yang, W. and D.G. Drueckhammer, *Understanding the relative acyl-transfer reactivity of oxoesters and thioesters: computational analysis of transition state delocalization effects.* J Am Chem Soc, 2001. **123**(44): p. 11004-9.
91. Radzicka, A. and R. Wolfenden, *Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases.* Journal of the American Chemical Society, 1996. **118**(26): p. 6105-6109.
92. Mishra, P.K. and D.G. Drueckhammer, *Coenzyme A Analogues and Derivatives: Synthesis and Applications as Mechanistic Probes of Coenzyme A Ester-Utilizing Enzymes.* Chem Rev, 2000. **100**(9): p. 3283-3310.
93. Sundlov, J.A., et al., *Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains.* Chem Biol, 2012. **19**(2): p. 188-98.
94. Sundlov, J.A. and A.M. Gulick, *Structure determination of the functional domain interaction of a chimeric nonribosomal peptide synthetase from a challenging crystal with noncrystallographic translational symmetry.* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 8): p. 1482-92.
95. Miller, B.R., et al., *Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture.* J Biol Chem, 2016. **291**(43): p. 22559-22571.
96. Keatinge-Clay, A.T. and R.M. Stroud, *The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases.* Structure, 2006. **14**(4): p. 737-748.
97. Reid, R., et al., *A model of structure and catalysis for ketoreductase domains in modular polyketide synthases.* Biochemistry, 2003. **42**(1): p. 72-9.
98. Caffrey, P., *Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases.* Chembiochem, 2003. **4**(7): p. 654-7.
99. Reimer, J.M., et al., *Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility.* Science, 2019. **366**(6466).
100. Reiber, K., et al., *The expression of selected non-ribosomal peptide synthetases in Aspergillus fumigatus is controlled by the availability of free iron.* FEMS Microbiol Lett, 2005. **248**(1): p. 83-91.
101. Crestfield, A.M., W.H. Stein, and S. Moore, *On the aggregation of bovine pancreatic ribonuclease.* Arch Biochem Biophys, 1962. **Suppl 1**: p. 217-22.
102. Liu, H., et al., *[Comparison of sequences of the hemagglutinin gene and phylogenetical analysis of H9 subtype avian influenza viruses isolated from some regions in China].* Wei Sheng Wu Xue Bao, 2002. **42**(3): p. 288-97.
103. Finn, R.D., et al., *Pfam: the protein families database.* Nucleic acids research, 2014. **42**(Database issue): p. D222-D230.
104. www.qiagenbioinformatics.com.

105. Cuff, J.A., et al., *JPred: a consensus secondary structure prediction server.* Bioinformatics, 1998. **14**(10): p. 892-3.
106. Sanchez, C., et al., *Cloning and characterization of a phosphopantetheinyl transferase from Streptomyces verticillus ATCC15003, the producer of the hybrid peptide-polyketide antitumor drug bleomycin.* Chem Biol, 2001. **8**(7): p. 725-38.
107. D'Arcy, A., et al., *Microseed matrix screening for optimization in protein crystallization: what have we learned?* Acta Crystallogr F Struct Biol Commun, 2014. **70**(Pt 9): p. 1117-26.
108. Winter, G., et al., *DIALS: implementation and evaluation of a new integration package.* Acta crystallographica. Section D, Structural biology, 2018. **74**(Pt 2): p. 85-97.
109. Battye, T.G., et al., *iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 271-81.
110. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta crystallographica. Section D, Biological crystallography, 2011. **67**(Pt 4): p. 235-242.
111. Schwede, T., et al., *SWISS-MODEL: An automated protein homology-modeling server.* Nucleic acids research, 2003. **31**(13): p. 3381-3385.
112. Fidler, D.R., et al., *Using HHsearch to tackle proteins of unknown function: A pilot study with PH domains.* Traffic (Copenhagen, Denmark), 2016. **17**(11): p. 1214-1226.
113. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta crystallographica. Section D, Biological crystallography, 2010. **66**(Pt 2): p. 213-221.
114. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
115. Jimin Wang, D.A.A., Clarisse Chiche-Lapierre, T. Martin Schmeing, *On sharpening omit-maps for extraction of weak structural information.* 2019.
116. J., H.R.D.D.R., *Fractionation of the bactericisal agent from cultures of a soil bacillus. .* Journal of Biological Chemistry, 1940(132(2)): p. 791-792.
117. Sarges, R. and B. Witkop, *GRAMICIDIN A. V. THE STRUCTURE OF VALINE- AND ISOLEUCINE-GRAMICIDIN A.* J Am Chem Soc, 1965. **87**: p. 2011-20.
118. Yonus, H., et al., *Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains.* J Biol Chem, 2008. **283**(47): p. 32484-91.
119. Rahim, A., et al., *Reciprocal carbonyl-carbonyl interactions in small molecules and proteins.* Nat Commun, 2017. **8**(1): p. 78.
120. Reimer, J.M., et al., *Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase.* Nature, 2016. **529**: p. 239-239.
121. Ramaswamy, A.V., C.M. Sorrels, and W.H. Gerwick, *Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium Lyngbya majuscula.* J Nat Prod, 2007. **70**(12): p. 1977-86.
122. Awakawa, T., et al., *Reprogramming of the antimycin NRPS-PKS assembly lines inspired by gene evolution.* Nature communications, 2018. **9**(1): p. 3534-3534.
123. Fujimori, D.G., et al., *Cloning and characterization of the biosynthetic gene cluster for kutznerides.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(42): p. 16498-16503.

124. Persson, B., et al., *The SDR (short-chain dehydrogenase/reductase and related enzymes) nomenclature initiative.* Chemico-biological interactions, 2009. **178**(1-3): p. 94-98.
125. Oppermann, U., et al., *Short-chain dehydrogenases/reductases (SDR): the 2002 update.* Chemico-Biological Interactions, 2003. **143-144**: p. 247-253.
126. Nakamura, K., et al., *Recent Developments in Asymmetric Reduction of Ketones with Biocatalysts*. Vol. 14. 2003. 2659-2681.
127. Noey, E.L., et al., *Origins of stereoselectivity in evolved ketoreductases.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(51): p. E7065-E7072.
128. Moore, J.C., et al., *Advances in the Enzymatic Reduction of Ketones.* Accounts of Chemical Research, 2007. **40**(12): p. 1412-1419.
129. Bornscheuer, U.T., et al., *Engineering the third wave of biocatalysis.* Nature, 2012. **485**: p. 185.
130. Brenek, S.J., et al., *Development of a Practical and Convergent Process for the Preparation of Sulopenem.* Organic Process Research & Development, 2012. **16**(8): p. 1348-1359.
131. Bollikonda, S., et al., *An Enantioselective Formal Synthesis of Montelukast Sodium.* The Journal of Organic Chemistry, 2015. **80**(8): p. 3891-3901.
132. Honda, Y., et al., *New approaches to the industrial synthesis of HIV protease inhibitors.* Org Biomol Chem, 2004. **2**(14): p. 2061-70.
133. Zhu, D., C. Mukherjee, and L. Hua, *'Green' synthesis of important pharmaceutical building blocks: Enzymatic access to enantiomerically pure α-chloroalcohols*. Vol. 16. 2005. 3275-3278.
134. Zarganes-Tzitzikas, T., C.G. Neochoritis, and A. Dömling, *Atorvastatin (Lipitor) by MCR.* ACS medicinal chemistry letters, 2019. **10**(3): p. 389-392.
135. Devine, P.N., et al., *Extending the application of biocatalysis to meet the challenges of drug development.* Nature Reviews Chemistry, 2018. **2**(12): p. 409-421.
136. Jörnvall, H., et al., *Short-chain dehydrogenases/reductases (SDR).* Biochemistry, 1995. **34**(18): p. 6003-6013.
137. Filling, C., et al., *Critical residues for structure and catalysis in short-chain dehydrogenases/reductases.* J Biol Chem, 2002. **277**(28): p. 25677-84.
138. Korman, T.P., et al., *Inhibition kinetics and emodin cocrystal structure of a type II polyketide ketoreductase.* Biochemistry, 2008. **47**(7): p. 1837-47.
139. Kao, C.M., et al., *Alcohol Stereochemistry in Polyketide Backbones Is Controlled by the β-Ketoreductase Domains of Modular Polyketide Synthases.* Journal of the American Chemical Society, 1998. **120**(10): p. 2478-2479.
140. Keatinge-Clay, A.T., *A tylosin ketoreductase reveals how chirality is determined in polyketides.* Chem Biol, 2007. **14**(8): p. 898-908.
141. Javidpour, P., et al., *Structural and biochemical studies of the hedamycin type II polyketide ketoreductase (HedKR): molecular basis of stereo- and regiospecificities.* Biochemistry, 2011. **50**(34): p. 7426-39.
142. Wubbolts, M.G., et al., *Variation of cofactor levels in Escherichia coli. Sequence analysis and expression of the pncB gene encoding nicotinic acid phosphoribosyltransferase.* J Biol Chem, 1990. **265**(29): p. 17665-72.
143. Shin, D.H., et al., *Crystal structure of a nicotinate phosphoribosyltransferase from Thermoplasma acidophilum.* J Biol Chem, 2005. **280**(18): p. 18326-35.

144. Magni, G., et al., *Enzymology of NAD+ synthesis.* Adv Enzymol Relat Areas Mol Biol, 1999. **73**: p. 135-82, xi.
145. Mehl, R.A., C. Kinsland, and T.P. Begley, *Identification of the Escherichia coli nicotinic acid mononucleotide adenylyltransferase gene.* Journal of bacteriology, 2000. **182**(15): p. 4372-4374.
146. Zhang, H., et al., *Crystal structures of E. coli nicotinate mononucleotide adenylyltransferase and its complex with deamido-NAD.* Structure, 2002. **10**(1): p. 69-79.
147. Willison, J.C. and G. Tissot, *The Escherichia coli efg gene and the Rhodobacter capsulatus adgA gene code for NH3-dependent NAD synthetase.* Journal of bacteriology, 1994. **176**(11): p. 3400-3402.
148. Jauch, R., et al., *Structures of Escherichia coli NAD synthetase with substrates and products reveal mechanistic rearrangements.* J Biol Chem, 2005. **280**(15): p. 15131-40.
149. Raffaelli, N., et al., *The Escherichia coli NadR regulator is endowed with nicotinamide mononucleotide adenylyltransferase activity.* Journal of bacteriology, 1999. **181**(17): p. 5509-5511.
150. Singh, S.K., et al., *Crystal structure of Haemophilus influenzae NadR protein. A bifunctional enzyme endowed with NMN adenyltransferase and ribosylnicotinimide kinase activities.* J Biol Chem, 2002. **277**(36): p. 33291-9.
151. Kawai, S., et al., *Molecular characterization of Escherichia coli NAD kinase.* Eur J Biochem, 2001. **268**(15): p. 4359-65.
152. Lin, W.K., et al., *Synthesis of the Ca(2+)-mobilizing messengers NAADP and cADPR by intracellular CD38 enzyme in the mouse heart: Role in beta-adrenoceptor signaling.* J Biol Chem, 2017. **292**(32): p. 13243-13257.
153. Stenlid, J.H. and T. Brinck, *Nucleophilic Aromatic Substitution Reactions Described by the Local Electron Attachment Energy.* The Journal of Organic Chemistry, 2017. **82**(6): p. 3072-3083.
154. Yin, J., et al., *Site-specific protein labeling by Sfp phosphopantetheinyl transferase.* Nat Protoc, 2006. **1**(1): p. 280-5.
155. Lu, W., et al., *Extraction and Quantitation of Nicotinamide Adenine Dinucleotide Redox Cofactors.* Antioxid Redox Signal, 2018. **28**(3): p. 167-179.
156. Price, A.C., et al., *Cofactor-induced conformational rearrangements establish a catalytically competent active site and a proton relay conduit in FabG.* Structure, 2004. **12**(3): p. 417-28.