Cyclodipeptide synthases : towards understanding their catalytic mechanism and the molecular bases of their specificity

Yan Li

UNIVERSITÉ PARIS-SUD 11

ECOLE DOCTORALE :

INNOVATION THÉRAPEUTIQUE : DU FONDAMENTAL A L'APPLIQUÉ

PÔLE : INGENIERIE DES PROTEINES ET CIBLES THERAPEUTIQUES

DISCIPLINE :

STRUCTURE, FONCTION ET INGENIERIE DES PROTEINES

ANNÉE 2009 – 2012 SÉRIE DOCTORAT N° 1185

THÈSE DE DOCTORAT

soutenue le 26/09/2012

par

Yan LI

Titre:

Les cyclodipeptide synthases : vers la compréhension de leur mécanisme catalytique et des bases moléculaires de leur spécificité

Directeur de thèse : Dr Muriel GONDRY Chercheur CEA (CEA, Saclay)

Composition du jury :

Président du jury :	Pr Philippe MINARD	
Rapporteurs :	Pr Yves MECHULAM	Directeur de recherche CNRS (Ecole Polytechnique, Palaiseau)
	Pr Olivier PLOUX	Professeur (Ecole nationale supérieure de chimie de Paris)
Examinateurs :	Pr Sylvie REBUFFAT	Professeur (Muséum national d'Histoire naturelle, Paris)
	Pr Philippe MINARD	Professeur d'Université (Université Paris-Sud 11, Orsay)

Résumé

Les cyclodipeptides et leurs dérivés, les dicétopipérazines (DKP), constituent une large classe de métabolites secondaires aux activités biologiques remarquables qui sont essentiellement synthétisés par des microorganismes. Les voies de biosynthèse de certaines DKP contiennent des synthases de cyclodipeptides (CDPS), une famille d'enzymes récemment identifiée. Les CDPS ont la particularité de détourner les ARNt aminoacylés de leur rôle essentiel dans la synthèse protéique ribosomale pour les utiliser comme substrats et ainsi catalyser la formation des deux liaisons peptidiques de différents cyclodipeptides. Le travail de thèse présenté dans ce manuscrit a pour objectif de caractériser la nouvelle famille des CDPS. Dans un premier temps, la caractérisation tant structurale que mécanistique de la première CDPS identifiée, AlbC de *Streptomyces noursei*, est présentée. Puis, les résultats obtenus avec trois autres CDPS, chacune de ces enzymes ayant des caractéristiques adéquates pour approfondir l'étude de la famille des CDPS, sont décrits. Ainsi, la CDPS Ndas_1148 de *Nocardiopsis dassonvillei* a permis d'étendre nos connaissances sur les bases moléculaires de la spécificité des CDPS. La CDPS AlbC-IMI de *S.* sp. IMI 351155 est un bon modèle pour analyser l'interaction de chacun des deux substrats nécessaires à la formation d'un cyclodipeptide. Enfin, la caractérisation de la CDPS *Nvec*-CDPS2 chez l'animal *Nematostella vectensis* a permis de fournir le premier exemple d'enzyme d'origine animale impliquée dans la synthèse peptidique non ribosomale.

Abstract

Cyclodipeptides and their derivatives, the diketopiperazines (DKPs), constitute a large class of secondary metabolites with noteworthy biological activities that are mainly synthesized by microorganisms. The biosynthetic pathways of some DKPs contain cyclodipeptide synthases (CDPSs), a newly defined family of enzymes. CDPSs hijack aminoacyl-tRNAs from their essential role in ribosomal protein synthesis to catalyze the formation of the two peptide bonds of various cyclodipeptides. The aim of the work presented in this thesis manuscript is to characterize the CDPS family. At first, the structural and mechanistic characterization of the first identified CDPS, AlbC of *Streptomyces noursei*, is presented. Then, the results obtained with three other CDPSs, each of which having suitable properties to increase our understanding of the CDPS family, are described. The CDPS Ndas_1148 of *Nocardiopsis dassonvillei* extends our knowledge of the molecular bases of the CDPS specificity. The CDPS AlbC-IMI of *S.* sp. IMI 351155 is a good model to analyze the interaction of each of the two substrates required for the formation of a cyclodipeptide. Finally, the characterization of the CDPS *Nvec*-CDPS2 from *Nematostella vectensis* provides the first example of enzymes of animal origin involved in nonribosomal peptide synthesis.

Mots-clés :

cyclodipeptide synthase (CDPS), diketopiperazine (DKP), nonribosomal biosynthesis, peptide bond, aminoacyl-tRNA, secondary metabolite

Intitulé du laboratoire :

Equipe Enzymologie et Biosynthèse Peptidique Non Ribosomale
Laboratoire de Toxinologie Moléculaire et Biotechnologies (LTMB)
Service d'Ingénierie Moléculaire des Protéines (SIMOPRO)
CEA Saclay, 91191 Cedex, Gif-sur-Yvette

PÔLE : INGENIERIE DES PROTEINES ET CIBLES THERAPEUTIQUES
UNIVERSITÉ PARIS-SUD 11
UFR «FACULTÉ DE PHARMACIE DE CHÂTENAY-MALABRY »
5, rue Jean Baptiste Clément
92296 CHÂTENAY-MALABRY Cedex

To my dear family!

Acknowledgements

It would not have been possible to achieve this doctoral thesis without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

First of all, I thank my committee members for their valuable suggestions and feedback in evaluating my research.

I would like to thank the International PhD Program of the Life Sciences division of the CEA (Irtelis) for the doctoral fellowship. I thank the CEA sector "Molecular Engineering of Proteins (SIMOPRO)" which received me to accomplish my thesis. I also thank the sector for their powerful experimental apparatus without the access to which, I can not image how to accomplish this thesis work in three years.

I am heartily grateful to my PhD supervisor, Dr Muriel Gondry, for her encouragement, supervision and help in every aspect of the work during my thesis. She taught me how to be a scientific researcher. I deeply admire her for her rigorous scientific approach and rich knowledge in the domain.

I sincerely thank my colleagues of the laboratory. I will always be grateful for having worked with this excellent research team for three years. I deeply thank Mireille Moutiez, Pascal Belin and Jérôme Seguin for the kind transfer of their scientific knowledge to me. It is a good memory when we shared time together in the laboratory and in the coffee corner. I will particularly never forget the kind and timely help of Pascal Belin who always reached me out a helping hand when I felt helpless. I sincerely thank him for having helped me rebuild confidence. I would like to thank Guillaume Grach for his precious advice and friendship. I also thank our technician Cédric Masson for having helped me purify some proteins and Sandrine Braud for her Photoshop help. I thank Ludovic Sauguet, Jean-Baptiste Charbonnier and Marie-Hélène Le Du for the structure of AlbC which made a solid foundation for my thesis work. I also thank

Mathieu Fonvieille for his kind help, advice, encouragement and all products he synthesized for my work.

I am thankful to Robert Thai for his training and help in analysis with mass spectrometry. I thank the technician Steven Dubois for his kind help in mass spectrometry and other sample analyses. I also thank Alain Lecoq for the synthesis of some chemical molecules and his personal advice during my thesis. I am grateful for the help of Fabrice Beau and Fannely Berthon on analysis with Beta-Imager. Additionally, I feel lucky for having harvested so much friendship in SIMOPRO. I specially thank Bertrand Czarny and Cathrine Nury for their encouragement and help. Bertrand, as what you said, I will miss your office-scaring and all our funny discussions. Catherine, thank you for having given me a lift so often during my thesis. I will miss the days we spent together in SIMOPRO.

I would like to acknowledge our collaborators for their excellent experimental contribution and scientific interest, especially Jean-Luc Pernodet, Cécile Martel and Karine Tuphile.

I also thank my friends (Bo Lu, Jun Han, Lin Xia, Sandrine Ragu…too many to list her but you know who you are!) for providing support and friendship that I need. I particularly thank my dear friend Yue Jiao for her amazing friendship and everything she did for me. I consider her as my sister. I also thank my landlady Mireille Verneau whom I consider as my family in France. I thank her for her delicious desserts, her amazing Christmas dinner and all countless help.

Finally, I give my deep gratitude to my family especially my grandfather Shukun Li, my mother Ping Zhang, my father Guangjian Li and my dear fiancé Bo Gao. Without your support and encouragement, I would not stick to the present. I know you miss me so much just like I miss you. You have done so much for me and now, I think it is time to give my love back. I love you forever.

Table of contents

Abbreviations

aaRS	Aminoacyl-tRNA synthetase
aa-tRNA	Aminoacyl-tRNA
AHL	N-acylhomoserine lactone
ATP	Adenosine triphosphate
BBB	Blood–brain barrier
CAT	Chloramphenicol acetyltransferase
CDO	Cyclodipeptide oxydase
CDPS	Cyclodipeptide synthase
cFL	Cyclo(L-Phe-L-Leu)
DKP	Diketopiperazine
EDTA	Ethylenediaminetetraacetic acid
EF	Elongation factor
EIC	Extracted ionic current
GNAT	GCN5-related N-acetyltransferase
GSH	Glutathione
HAT	Histone acetyltransferase
HPLC	High-performance liquid chromatography
IF	Initiation factor
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
mRNA	Messenger RNA
NOS	Nitric oxide synthase
NRPS	Nonribosomal peptide synthetase
PCP	Peptidyl carrier protein
PMSF	Phenylmethylsulfonyl fluoride
PTC	Peptidyl transferase center
PVDF	Polyvinylidene fluoride
RF	Release factor
RNase P	Ribonuclease P
RRF	Ribosome recycling factor
SDS-PAGE	Sodium dodecyl sulfatepolyacrylamide gel electrophoresis
TCA	Trichloroacetic acid
TCEP	Tris(2-carboxyethyl)phosphine
TE	Thioesterase
TFA	Trifluoroacetic acid
TLC	Thin layer chromatography
TRH	Thyrotropin-releasing hormone
tRNA	Transfer RNA
UV	Ultraviolet
Δ	Double bond between carbons Cα and Cβ of an amino acid

Introduction

Natural products have been the source of most of the active ingredient of medicines. Analysis of the sources of new and approved drugs for the treatment of human diseases indicates that natural products always play a highly significant role in the drug discovery and development process. Almost half of the drugs approved since 1994 are based on natural products. In the past decade, with the development of the combinatorial chemistry, many pharmaceutical companies put an emphasis on high-throughput screening of synthetic libraries and decrease the research into natural products (Harvey 2008; Li and Vederas 2009). This has been because of the perceived disadvantages of natural products, such as difficulties in access and supply, and complexities of natural product chemistry since many effective natural compounds cannot be easily obtained by the chemical synthesis pathway. Nevertheless, the rapid development of biotechnologies in recent years renews the interest of natural products in the drug discovery. Untapped biological resources, biological screening methods, robotic separation with structural analysis, metabolic engineering, and synthetic biology offer exciting technologies for new natural product drug discovery (Li and Vederas 2009). A significant number of natural product drugs are actually produced by microbes. Numerous microbial metabolites have been found to have interesting pharmaceutical activities such as antimicrobial and antitumor activities.

In the bioactive natural compounds, there is an important class of molecules called "diketopiperazines (DKPs)" which consist of cyclodipeptides and their derivatives. They are commonly biosynthesized by a large variety of organisms, including mammals (De Carvalho and Abraham 2012). The ability of microorganisms to produce DKPs is widespread and published data have shown that about 90% of Gram-negative bacteria produce them (Fenical 1993). In view of the potential medical value of such molecules, research on DKPs is always an active domain. However, a lot of DKPs with complex modifications cannot be easily obtained through the

chemical synthesis approach. It is thus important to decipher their biosynthetic pathways. Some cyclodipeptides have been reported to be synthesized by dedicated nonribosomal peptide synthetases (NRPSs) which are large multimodular biocatalysts. The structure and the catalytic mechanism of NRPSs will specially be described in the section "bibliographic studies".

The research team named "Enzymology and non ribosomal peptide biosynthesis", directed by Dr Muriel Gondry (CEA/DSV/iBiTec-S/SIMOPRO), is devoted to studies on the biosynthetic pathways and mechanisms of DKPs. At the beginning, in collaboration with another team directed by Jean-Luc Pernodet (IGM, CNRS UMR 8621, Université Paris-Sud 11), they isolated the biosynthetic pathway of albonoursin (cyclo(ΔPhe-ΔLeu)) (**Figure 1A**) which is a DKP produced by *Streptomyces noursei* and *Streptomyces albulus* with remarkable bioactivities like antibacterial and antitumor activities (Gondry et al. 2001; Lautru et al. 2002). Like most secondary metabolite genes in *Streptomyces* species, the albonoursin biosynthetic genes are clustered. The DNA fragment has a length of 3.8 kb containing four genes named *albA*, *albB*, *albC* and *albD*. However, subsequent studies showed that only *albA*, *albB* and *albC* are necessary for the biosynthesis of albonoursin (**Figure 1B**). *albA* and *albB* code the cyclic dipeptide oxidase (CDO), while *albC* codes a small enzyme of 28 kDa, called AlbC. Firstly, the precursor of albonoursin, cyclo(Phe-Leu) or indicated as cFL, is synthesized by AlbC; then the CDO catalyzes the formation of α,β-unsaturated residues of cyclo(Phe-Leu) to generate albonoursin; the latter is released to the culture medium after synthesis (**Figure 1C**) (Lautru et al. 2002).

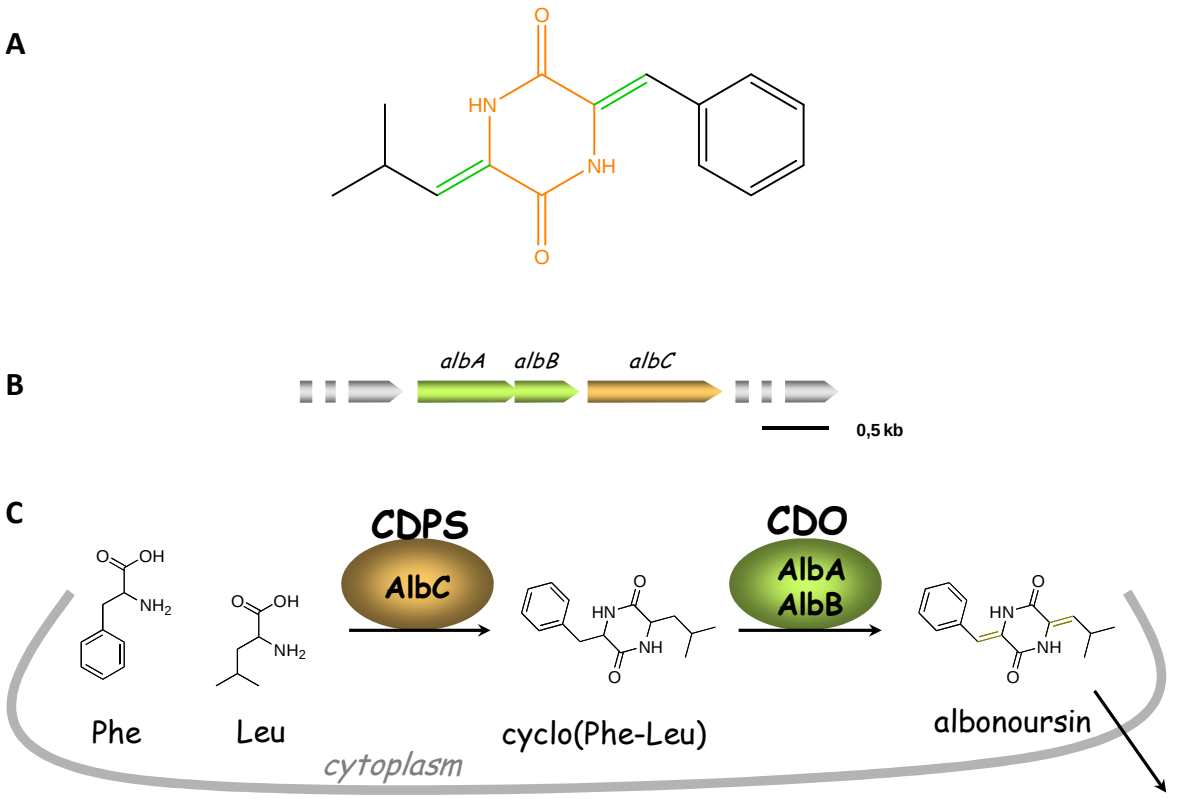


Figure 1: The albonoursin and its biosynthesis (Lautru et al. 2002). (A) Structure of albonoursin, with the DKP skeleton shown in orange and the two α,β-dehydrogenations in green. (B) The gene cluster composed of *albA*, *albB* and *albC*, and responsible for the biosynthesis of albonoursin. (C) Schema of the biosynthetic pathway of albonoursin.

AlbC catalyzes the formation of cyclodipeptides but it is unrelated to NRPSs or other known proteins. In the next few years, other similar proteins to AlbC were successively discovered in various bacterial phyla. Until 2009, eight related proteins from different bacterial organisms (**Figure 2**) had been identified and characterized (Gondry et al. 2009). They are all composed of 216-249 amino acid residues, 13 of which are conserved among them. In addition, further biological analysis demonstrated that they all use aminoacyl-tRNAs (aa-tRNAs) as substrates to catalyze the formation of cyclodipeptides. These proteins thus form a family of tRNA-dependent peptide bond-forming enzymes dedicated to the formation of cyclodipeptides. They are named "cyclodipeptide synthases (CDPSs)" (Gondry et al. 2009). However, the primary characterization showed that those CDPSs do not synthesize the same cyclodipeptides (**Figure 3**). All the eight CDPSs were expressed in *E. coli*. Their culture supernatants were analyzed by LC-MS/MS which is HPLC

coupled to mass spectrometry in order to characterize the cyclodipeptides synthesized. The results showed that AlbC produced twelve cyclodipeptides, including the principal products cFL and cFF. Thus, AlbC can incorporate into cyclodipeptides various nonpolar residues, such as phenylalanine, leucine, tyrosine and methionine, and to a much lesser extent alanine and valine. Indeed, the ten possible cyclodipeptides composed of phenylalanine, leucine, tyrosine and methionine are all synthesized in detectable amounts by AlbC. Almost all of the compounds produced by other CDPSs are combinations of the same four amino acids, with the restriction that cyclodipeptides synthesized by Rv2275 always contain tyrosine, and those synthesized by the other CDPSs almost always contain leucine (**Figure 3**).

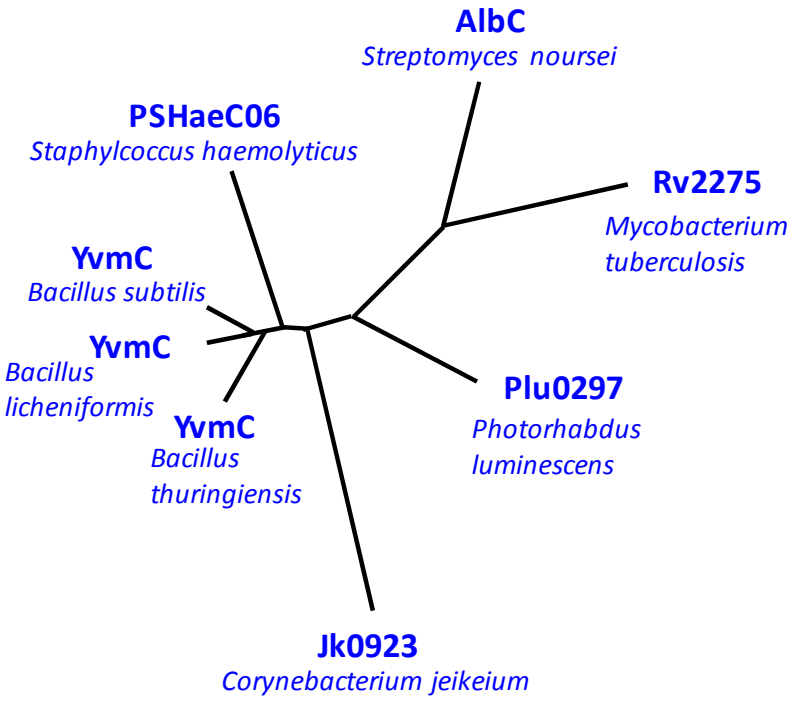


Figure 2: Eight characterized CDPSs to date (in bold) from different bacterial phyla shown in the form of phylogenetic tree.

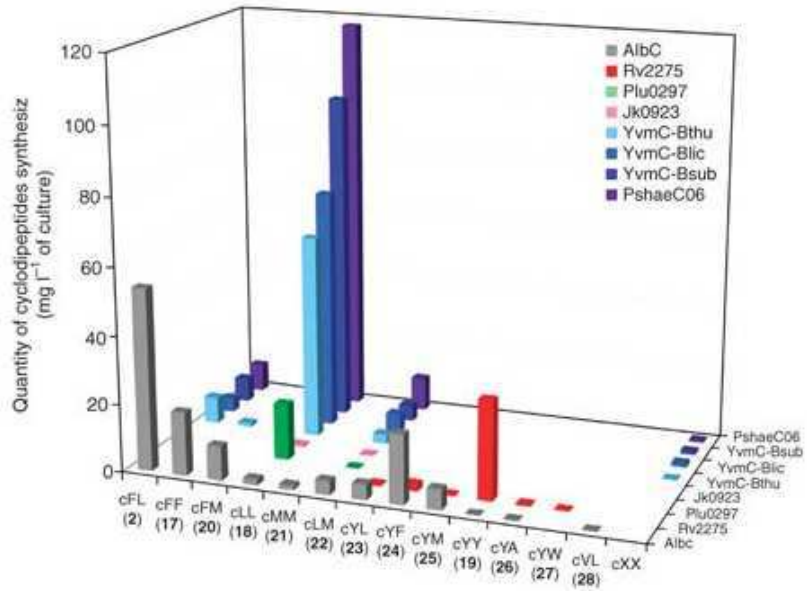


Figure 3: Histogram of the amounts of the various cyclodipeptides synthesized by eight recombinant CDPSs in *E. coli* (Gondry et al. 2009). cXX corresponds to an unidentified cyclodipeptide of the same molecular mass as cLL.

Cyclodipeptide formation is often the first step in the synthesis of more complex DKPs that are obtained after tailoring reactions (Gardiner et al. 2004; Loria et al. 2008). This appears also to be the case for the CDPS-synthesizing cyclodipeptides because the CDPS genes in prokaryotes are generally organized into operon-like structures and some of the proteins encoded by these operons could play a role in modifying the cyclodipeptide. So far, three proteins encoded by genes closely linked to CDPS genes have been experimentally characterized and shown to have three different cyclodipeptide-tailoring activities: α,β-dehydrogenation, DKP ring oxidation and C-C aryl coupling (**Figure 4**) (Belin et al. 2012). One of the three proteins is the CDO having the α,β-dehydrogenation activity. As previously described, it is involved in oxidizing the cFL in the biosynthetic pathway of albonoursin (**Figure 1** and **Figure 4A**). Many of the CDPS genes identified in databases are closely linked to a gene encoding a cytochrome P450 enzyme (P450). P450s constitute a superfamily of heme-containing monooxygenases that dissociate molecular oxygen to catalyze numerous reactions on a wide range of structurally diverse molecules (Guengerich 2001; Isin and Guengerich 2007). The first evidence of a role for a P450 in a CDPS-dependent pathway was provided by the identification of

the *yvmC-cypX* gene cluster in the pulcherrimin synthesis in *B. subtilis* (**Figure 4B**) (Tang et al. 2006). The CDPS YvmC catalyzes the formation of cyclo(Leu-Leu) (cLL) (Gondry et al. 2009), which is subsequently converted into pulcherriminic acid by CypX; the pulcherriminic acid associates with iron (III) to generate pulcherrimin. CypX having the cyclodipeptide-tailoring activity belongs to the P450 family. The recent biochemical and structural characterization of CYP134A1 (= CypX) has provided insight into the oxidation of the DKP ring of cyclodipeptides (Cryle et al. 2010). The third cyclodipeptide-tailoring identified is CYP121 involved in the biosynthetic pathway of mycocyclosin isolated from *Mycobacterium tuberculosis* (**Figure 4C**). Genetically, CYP121 is linked to the CDPS Rv2275 in an operon-like structure (Cole et al. 1998; Roback et al. 2007). The two proteins constitute the metabolic pathway of the DKP mycocyclosin. Rv2275 synthesize cyclo(Tyr-Tyr) (cYY) which is then modified by CYP121 having C-C aryl coupling activity.

Bioinformatic analyses of the genetic environment of putative CDPSs have enabled to identify new putative cyclodipeptide-tailoring enzymes such as methyl-transferases, oxidoreductases, 2-oxoglutatarate-dependent oxygenases, acyl-CoA or peptideligases, and hypothetical proteins not yet related to any known function. These proteins may actually catalyze alternative modifications of the side chains of the residues constituting the cyclodipeptide or the cyclodipeptide ring itself (Belin et al. 2012).

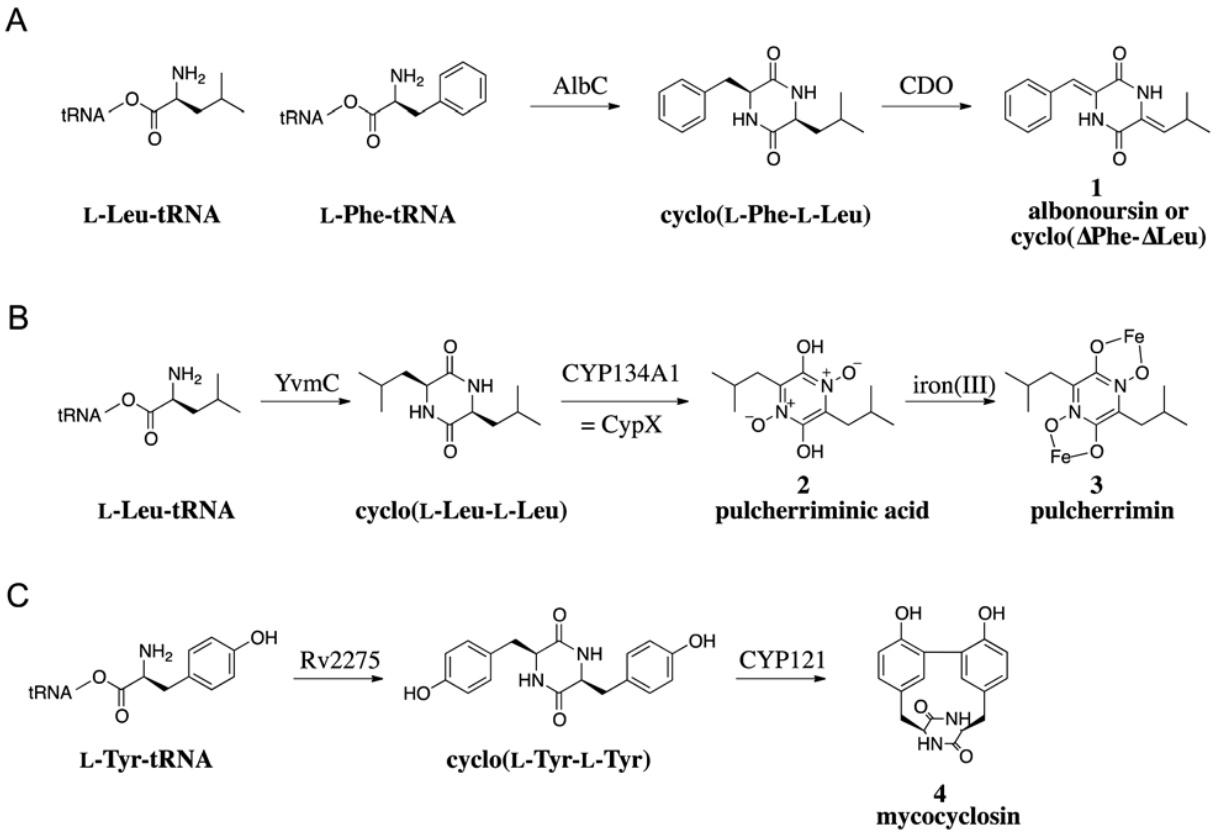


Figure 4: The CDPS-dependent biosynthetic pathways of albonoursin 1 (A), pulcherrimin 3 (B) and mycocyclosin 4 (C) (Belin et al. 2012).

Nevertheless, our knowledge on the CDPS family and the CDPS-dependent biosynthetic pathways of DKPs is still very poor. It is important to understand their molecular mechanisms in order to, in the long run, apply this knowledge to the production of molecules that have new or improved biological and pharmacological activities.

My Ph.D project mainly consists of characterization of some CDPSs of interest, elucidation of their molecular bases on catalysis and on substrate specificity, as well as preliminary work on the catalytic process of CDPSs. This work can be achievable thanks to the existence of new CDPSs in the nature, some of which are good study models.

This thesis manuscript will begin with bibliographic studies on DKPs and the known peptide bond-forming biocatalysts (**Chapter 1**). The result part will be

divided into four chapters (**Chapters 2-5**). The **chapter 2** will be dedicated to a published work on structural and mechanic characterization of AlbC, which provides insight into the interaction between the CDPS and its aa-tRNA substrates, as well as the catalytic mechanism. In the **chapter 3**, I will introduce a new CDPS Ndas_1148 from *Nocardiopsis dassonvillei* characterized in the laboratory. The characteristics of Ndas_1148 allow us to better understand the molecular bases of the substrate specificity. In the **chapter 4**, I will introduce another CDPS, named AlbC-IMI, freshly identified from *Streptomyces* sp. IMI 351155. AlbC-IMI is used in my work as a CDPS model to study the catalytic mechanism. The **chapter 5** concerns our published work on *Nvec*-CDPS2 which is the first eukaryotic CDPS identified in the sea anemone *Nematostella vectensis*. Finally, we will conclude the whole work and present the perspectives of studies on CDPSs.

1 BIBLIOGRAPHIC STUDIES

1.1 Natural Diketopiperazines (DKPs)

Cyclodipeptides and their derivatives DKPs have been detected in a variety of natural resources. They constitute a large class of secondary metabolites synthesized predominantly by microorganisms. Recently, the interest in these compounds has significantly increased because of their diverse and remarkable bioactivities (Prasad 1995; Martins and Carvalho 2007; Huang et al. 2010), such as antibacterial (Magyar et al. 1996; Cain et al. 2003; Kohn and Widger 2005), antifungal (Ström et al. 2002; Musetti et al. 2007), antiviral (Rodriguez and Carrasco 1992), antitumor (Kanoh et al. 1999; Williams et al. 1999; Kanzaki et al. 2004; Jia et al. 2005), immunosuppressive (Waring and Beaver 1996) and anti-inflammatory (Minelli et al. 2012) activities. Some DKPs are found to play physiological roles like in quorum-sensing (Holden et al. 1999; Degrassi et al. 2002; Park et al. 2006; Li et al. 2011; Ortiz-Castro et al. 2011), in plant-growth regulatory (Ortiz-Castro et al. 2011), and in central nervous system (Minelli et al. 2009). In this part, I will introduce the DKP family, their physiological roles, their biological and pharmacological activities, and some applications or potential applications of DKPs in medical and pharmaceutical domains.

1.1.1 DKP family

DKPs are a class of cyclic organic compounds and characterized by a common motif: the DKP nucleus, which is heterocycle piperazine-2.5-dione, also known as dioxopiperazine. This nucleus presents two cis amide bonds. The general structure of DKPs is shown in **Figure 5**.

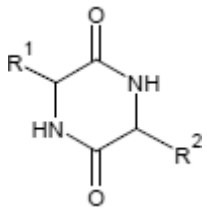


Figure 5: General structure of DKPs. R^1 and R^2 represent variable lateral chains.

When the groups R1 and R2 correspond to side chains of amino acids, these DKPs are named cyclodipeptides. They are from the cyclization of two amino acids resulting from the formation of two peptide bonds. The cyclodipeptides can be of cis or trans conformation, depending on whether the constituting amino acids are of identical or different configuration.

1.1.1.1 Natural abundance of DKPs

DKPs are ubiquitous in nature. They are produced in numerous prokaryotic and eukaryotic organisms forming a large family of natural products although these molecules are a relatively unexplored class of bioactive peptides.

A lot of DKPs have been isolated in prokaryotes of various species like *Streptomyces noursei* (Khokhlov and Lokshin 1963), *Pseudomonas aeruginosa* (Jayatilake et al. 1996), *Lactobacillus reuteri* (Li et al. 2011) and *Salinispora arenicola* (Schultz et al. 2008). Besides, DKPs are also produced by plants and by several eukaryotic microorganisms such as yeasts, lichens and fungus (Prasad 1995). Some examples are shown in **Table 1**. Numerous DKPs are isolated from marine microorganisms, sponges, see stars, tunicates (ascidians), and red algae (Huang et al. 2010). Finally, one DKP, cyclo(His-Pro), has been shown to be present in mammals. This DKP was demonstrated to be present in human (hypothalamus, stomach and esophagus), monkey (hypothalamus and spinal cord), rat and mouse (hypothalamus, cerebellum and cortex) (Prasad 1995).

Diketopiperazine	Species	Common Name	Reference
Cyclo(Pro-Leu)†	*Rosellinia necatrix*	Fungus	34
	Aspergillus fumigatus	Fungus	76
Cyclo(Pro-Val)	*Rosellinia necatrix*	Fungus	34
	Aspergillus ochraceus	Fungus	85
	Metarrhizum ansiopha	Fungus (mold)	85
Cyclo(Pro-Phe)	*Rosellinia necatrix*	Fungus	34
Cyclo(Phe-Phe)	*Penicillium nigricans*	Fungus (mold)	25
	Streptomyces noursei		30
Cyclo(Ala-Leu)	*Aspergillus niger*	Fungus (black mold)	33
Cyclo(Pro-Tyr)	*Alternaria alternata*	Fungus	132
Cycloserine dimer	*Streptomyces orchidaceus*	Actinomycetes	79
Cyclo(Pro-Trp)	*Penicillium brevicompactum*	Fungus (mold)	141
Cyclo(D-Ala-*N*-methyl-Leu)	*Beauveria nivea*		44
Picroroccellin	*Rocella fuciformis*	Lichen	52
Zizyphine and zizyphinine	*Zizyphus oenoplia*	Shrub	131, 208
Thodotorulic acid	*Rhodotorula pilimanae*	Fungus (yeast)	10
Dimerumic acid, Coprogen B, and coprogen	*Fusarium dimerum*	Fungus	43, 81
Albonoursin	*Streptomyces noursei*	Actinomycetes	83
Amphomycin	*Streptomyces canus*	Actinomycetes	26, 57

Table 1: Some simple cyclodipeptides that occur naturally in the protist and plant kingdoms (Prasad 1995).

A significant structural diversity of DKPs is associated with the variety of their natural sources, which I will detail in the following section.

1.1.1.2 Structural diversity of DKPs

Structures of DKPs vary a lot from simple cyclodipeptides to very complicated derivatives. Some examples are shown in **Figure 6**. In most cases, amino acids incorporated in DKPs are of L-configuration. In this thesis, the configuration of amino acids constituting the cyclodipeptides will not be specified except for those of D-configuration. All amino acids are not similarly incorporated to form cyclodipeptides. They are often composed of hydrophobic and aromatic amino acids especially the leucine (cyclo(Leu-Pro)), the valine (cyclo(ΔAla-Val) (**Figure 6A**) (Holden et al. 1999), the proline, the tyrosine (cyclo(Pro-Tyr)) (Holden et al. 1999), the phenylalanine (cyclo(Phe-Pro)) (Li et al. 2011), and the tryptophan (cyclo(Trp-Phe)). Several charged and polar amino acids are also present in DKPs, such as the serine (gliotoxine (**Figure 6B**)), and the histidine (cyclo(His-Pro)). Several unusual amino acids are also found

in some DKPs, such as the δ–hydroxyl-leucine in cyclomarazines A and B (**Figure 6C**), and the norvaline (Nva) in cyclo(4-methyl-D-Pro-Nva) (**Figure 6D**) (Adamczeski et al. 1995).

Besides the different types of amino acids, various chemical modifications to cyclodipeptides also enrich the diversity of DKPs. The modification can be introduced to DKP nuclei or to their side chains. The most common chemical modifications are the methylation like cyclo(4-methyl-D-Pro-Nva) (**Figure 6D**), the hydroxylation like bicyclomycin (**Figure 6E**), the nitration like the phytotoxin thaxtomin A (**Figure 6F**), and the presence of double bonds like albonoursin (**Figure 6G**).

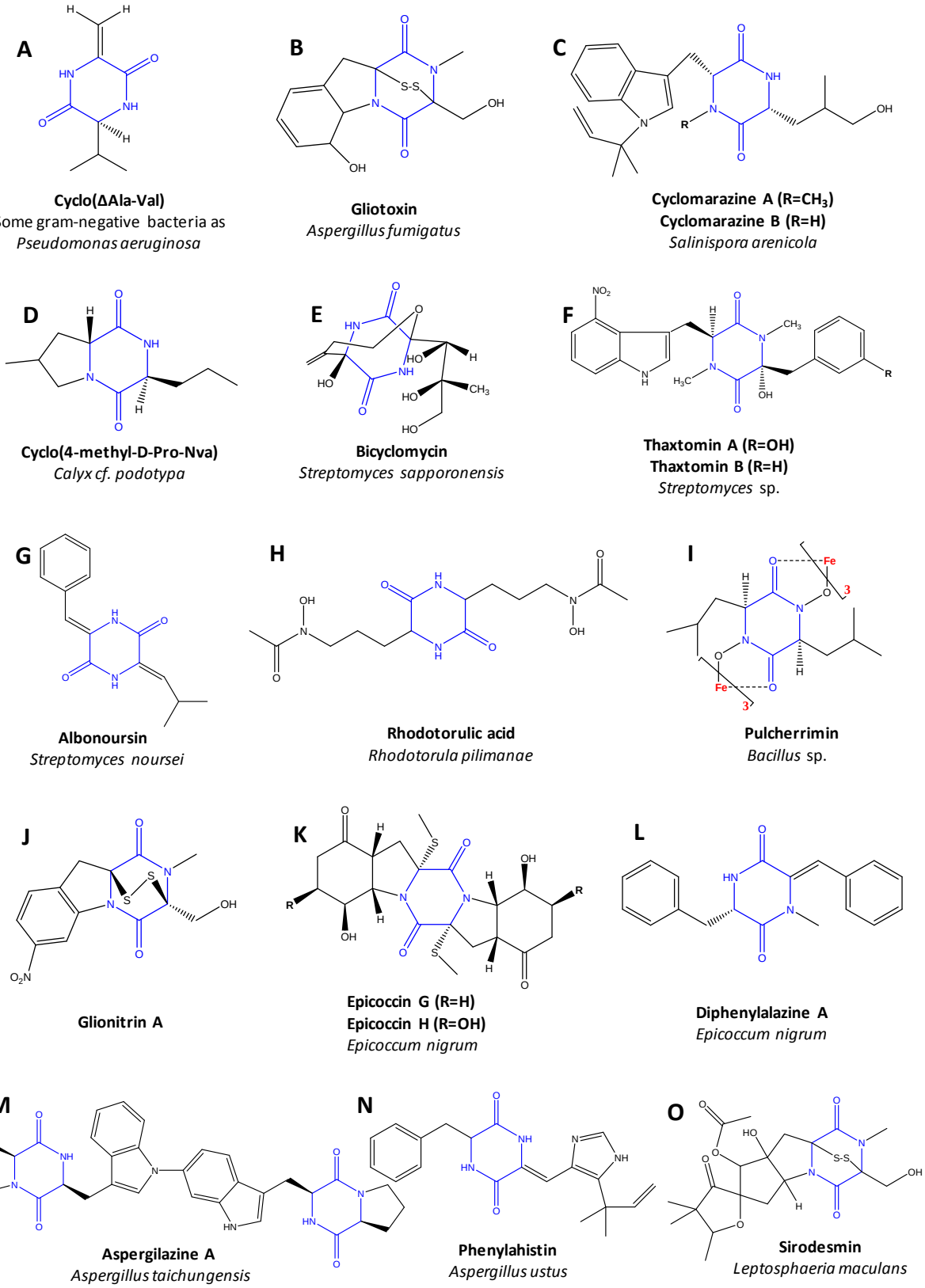


Figure 6: Examples of DKP structures. The DKP-producing species are indicated in italics.

1.1.1.3 Physiological roles of DKPs

Although the DKPs are ubiquitous in nature and more and more DKPs are being discovered, knowledge to their physiological roles is still limited. In this part, I will introduce several known physiological roles of some DKPs.

1.1.1.3.1 Cell-to-cell communication: Quorum sensing

Communication between cells *via* diffusible chemicals is a general phenomenon virtually found in all living organisms. It has been intensively studied in bacteria in the last two decades. One of the best-known examples is quorum sensing. A number of bacteria associated with eukaryotic hosts employ quorum sensing systems to sense their population density thereby modulating the expression of sets of genes involved in physiological responses associated with survival, propagation, and/or virulence (Salmond et al. 1995; de Kievit and Iglewski 2000; Henke and Bassler 2004).

In Gram-negative bacteria, the most universal cell-cell signaling mechanism occurs *via* the production and response to a class of small diffusible molecules called *N*-acylhomoserine lactones (AHLs). However, DKPs are identified in cell-free culture supernatants of some Gram-negative bacteria. Cyclo(ΔAla-Val) and cyclo(Pro-Tyr) were found in *Pseudomonas aeruginosa*, *Proteus mirabilis*, *Citrobacter freundii* and *Enterobacter agglomerans* (cyclo(ΔAla-Val) only) (Holden et al. 1999). Although these two DKPs were absent from *Pseudomonas fluorescens* and *Pseudomonas alcaligenes*, a third DKP, cyclo(Phe-Pro) was isolated from both pseudomonas (Holden et al. 1999). The three DKPs were revealed capable of activating or antagonizing a LuxR-based AHL biosensor or other LuxR-based quorum-sensing systems. Although the physiological role of these DKPs has yet to be established, their activity suggests the existence of cross talk among bacterial signalling systems (Holden et al. 1999). Furthermore, the study of G. Degrassi et al. demonstrated that plant growth-promoting *Pseudomonas putida* WCS358 produced at least four different cyclodipeptides (cyclo(Pro-Tyr), cyclo(Pro-Leu), cyclo(Phe-Pro), and cyclo(Val-Leu)), and some of them potentially cross-talked with the quorum sensing LuxI and LuxR homologs (Degrassi et al. 2002). It was found that three DKPs (cyclo(Pro-Val),

cyclo(Phe-Pro), and cyclo(Pro-Tyr)) from *Pseudomonas aeruginosa* were involved in plant growth promotion by this symbiotic bacterium (Ortiz-Castro et al. 2011). The observation that quorum-sensing-regulated bacterial production of DKPs modulates auxin signaling and plant growth promotion establishes an important function for DKPs mediating prokaryote/eukaryote transkingdom signaling (Ortiz-Castro et al. 2011). In addition, cyclo(Phe-Pro) was found produced by *Vibrio vulnificus* and related *Vibrio* spp. (*V. cholera, V. parahaemolyticus, and V. harveyi*) (Park et al. 2006). *Vibrio vulnificus* is an opportunistic human pathogen that causes severe wound infection and primary septicemia (Strom and Paranjpye 2000). The study suggests that cyclo(Phe-Pro) is a signal molecule controlling the expression of genes important for the pathogenicity of *Vibrio* spp. by activating the quorum sensing bioindicator (Park et al. 2006). If methods could be developed to interfere with quorum sensing systems of Gram-negative pathogens, a novel means of controlling their pathogenicity might be possible (Hartman and Wise 1998).

In Gram-positive bacteria are also isolated DKPs participating in interspecies cell-to-cell communication. In the work of Jingru Li et al., they showed that the human vaginal isolate *Lactobacillus reuteri* RC-14 produced cyclo(Pro-Pro) and cyclo(Tyr-Pro) as the signaling molecules that are able to interfere with the staphylococcal quorum sensing system *agr*, a key regulator of virulence genes (Li et al. 2011). Their work contributes to a better understanding of interspecies cell-to-cell communication between *Lactobacillus* and *staphylococcus*, and provides a unique mechanism by which endogenous or probiotic strains may attenuate virulence factor production by bacterial pathogens.

1.1.1.3.2 Virulence of pathogenic microorganisms

So far, certain DKPs are found to be involved in the virulence of pathogenic microorganisms such as thaxtomin phytotoxins and gliotoxin.

Thaxtomin phytotoxins, first reported in 1989, are cyclic dipeptides (2,5-diketopiperazines) formed from the condensation of 4-nitrotrytophan and phenylalanine groups. Individual thaxtomins differ only in the presence or absence

of *N*-methyl and hydroxyl groups and their respective substitution sites. The great interest in the thaxtomins derives mainly from their established roles as virulence factors in the common scab of potato disease and their apparent ability to inhibit cellulose synthesis in developing plant cells. Common scab is an economically important disease that is caused by *Streptomyces* species, which attack growing tubers through immature lenticels and wound sites. All these *streptomyces* species produce the phytotoxin thaxtomin which could be thaxtomin A (**Figure 6F**) or another member of the thaxtomin family (Loria et al. 2008). Numerous studies have demonstrated that the thaxtomin family is directly associated with the potato common scab disease. Firstly, thaxtomin is present in the tissues of infected plants and, once extracted and added in healthy tissues, induces symptoms of the common scab (King et al. 1992). Then, Healy et al. found that thaxtomin A production was abolished in biosynthesis pathway disruption mutants which were avirulent on potato tubers. Moreover, introduction of the thaxtomin synthetase cosmid into a mutant restored both pathogenicity and thaxtomin A production, demonstrating a critical role for thaxtomins in pathogenesis (Healy et al. 2000). In short, the generation of thaxtomins by common scab-causing species from diverse geographic areas of the world and the quantitative relationship established between phytotoxin production and virulence overwhelmingly supports the concept of these toxins as pathogenicity determinants (King and Calhoun 2009).

Gliotoxin (**Figure 6B**) is an epipolythiodioxopiperazine toxin that is made by the filamentous fungus *Aspergillus fumigatus.* This molecule has several remarkable biological activities. One of them under study is its contribution to the virulence of *A. fumigates* (Spikes et al. 2008). In the study of Spikes et al., they tested for gliotoxin production and virulence in different animal models. The results showed that gliotoxin production correlated positively with virulence in a nonneutropenic mouse model of invasive pulmonary aspergillosis and a *Drosophila melanogaster* model of aspergillosis (Spikes et al. 2008).

1.1.1.3.3 Iron-binding DKPs

It is known that all life forms have an absolute requirement for iron to maintain their metabolism. For aerobic organisms the concentration of free, aqueous ferric ion is limited to 10^{-18} M at neutral pH due to the insolubility of $Fe(OH)_3$. This is the driving force for the excretion by microbes of strong iron-chelating agents (siderophores) and the expression of high-affinity transport systems, which provide a reliable cellular iron supply. Studies showed under iron-deficient growth conditions the yeast *Rhodotorula pilimanae* excreted vast amounts of rhodotorulic acid, a cyclodipeptide of δ-*N*-acetyl-L-δ-*N*-hydroxyornithine (**Figure 6H**) which is able to tightly bind iron (Müller et al. 1985).

Another DKP capable of binding iron is pulcherriminic acid. Fruit-borne *Metschnikowia pulcherrima* is yeast, which can protect fruits against postharvest rot caused by some postharvest pathogens due to its antifungal effects. *M. pulcherrima* produces a red pigment called pulcherrimin (**Figure 6I**) which is a large complex formed nonenzymatically from pulcherriminc acid and ferric ions. The antimicrobial activity of *M. pulcherrima*may is attributed to the formation of pulcherrimin which depletes the iron in the medium and creates an environment unsuitable for growth of microbes that require iron for growth since iron is essential for the growth of many microorganisms (Sipiczki 2006).

1.1.1.3.4 Effect on the central nervous system

Cyclo(His-Pro) is the first cyclodipeptide shown to be endogenous to the mammalian brain. It is present in central nervous system, body fluids, anuran skins, and gastrointestinal system (Prasad 1995; Minelli et al. 2009). This DKP is widely studied for its effects on the central nervous system. It might produce analgesia (Prasad 1995). Modulation of prolactin secretion, thermoregulation, and stereotypical behavior are three biological activities of cyclo(His-Pro) that seem to share a common dopaminergic mechanism (Prasad 1995).

In conclusion, the physiological role of DKPs in organisms that produce these compounds remains poorly documented. However, in most cases, these molecules are principally studied for their biological and pharmacological activities.

1.1.1.4 Biological and pharmacological activities of DKPs

As freshly mentioned above, DKPs are widely studied because of their numerous biological and pharmacological activities such as antibacterial, antiviral, antifungal, antitumor, anti-inflammatory, and immunosuppressive activities. Now, we will present, through some examples, these remarkable activities of DKPs.

1.1.1.4.1 Antibiotic DKPs

Bicyclomycin (**Figure 6E**) is a DKP isolated from *streptomyces sapporonensis* in 1972 then from *streptomyces aizunensis* in 1973 (Bradley et al. 1996). This molecule is a clinically useful antibiotic exhibiting activity against a broad spectrum of Gram-negative bacteria such as *Escherichia coli*, *Klebsiella*, *Salmonella*, *Shigella* and *Citrobacter*, and against the Gram-positive bacterium, *Micrococcus luteus*. Bicyclomycin has been used to treat diarrhea in humans and bacterial diarrhea in calves and pigs and is marketed by Fujisawa (Osaka, Japan) under the trade name Bicozamycin® (Kohn and Widger 2005). The structure of bicyclomycin is unique among antibiotics and studies showed that it employed a novel mode of action by inhibition of the RNA transcription termination factor Rho in *Escherichia coli* (Zwiefka et al. 1993; Kohn and Widger 2005). Rho is a hexameric RNA/DNA helicase/translocase that terminates transcription of select genes in bacteria. Bicyclomycin can disrupt the Rhomolecular machinery thereby giving rise to a catastrophic effect caused by the untimely overproduction of proteins not normally expressed constitutively, thus leading to a toxic effect on the cell (Kohn and Widger 2005). A recent study demonstrated that the sensibility of *E. coli* to bicyclomycin could be altered by deletions of different types of genes (Tran et al. 2011). Up to now, bicyclomycin is the only known selective inhibitor of Rho.

Glionitrin A (**Figure 6J**) is a new DKP disulfide with antibiotic-antitumor activity isolated from coculture of a mine drainage-derived *Sphingomonas* bacterial strain and a mine drainage-derived *Aspergillus fumigatus* fungal strain. Glionitrin A displayed significant antibiotic activity against a series of microbes including methicillin-resistant *Staphylococcus aureus*. Besides, an *in vitro* cytotoxicity assay revealed that glionitrin A had potent submicromolar cytotoxic activity against four human cancer cell lines: HCT-116, A549, AGS, and DU145 (Park et al. 2009).

1.1.1.4.2 Antiviral DKPs

A well-known antiviral DKP is gliotoxin (**Figure 6B**). This molecule has been mentioned above about its contribution to the virulence of pathogenic microorganisms. Besides its physiological role, gliotoxin is found to be a potent inhibitor of poliovirus RNA synthesis *via* its inhibition to the activity of the poliovirus polymerase $3D^{pol}$ *in vitro* (Rodriguez and Carrasco 1992). Gliotoxin is the first inhibitor reported of this viral enzyme. The toxicity of gliotoxin is due to the presence of a disulphide bridge, which can inactivate proteins *via* reaction with thiol groups, and to the generation of reactive oxygen species by redox cycling (Gardiner et al. 2005).

Moreover, three DKPs (**Figure 6K-L**) extracted from the fungus *Epicoccum nigrum* showed inhibitory effects on HIV-1 replication in C8166 cells (Guo et al. 2009). Recently, a novel DKP dimer, aspergilazine A (**Figure 6M**), dimerized by two DKP units *via* a rare N-1 to C-6 linkage, was isolated from the marine-derived fungus *Aspergillus taichungensis*. This DKP dimer showed a weak activity against influenza A (H1N1) virus (Cai et al. 2012).

1.1.1.4.3 Antifungal DKPs

Ten years ago, Ström et al. firstly reported the production of the antifungal DKPs, cyclo(Phe-Pro) and cyclo(Phe-*trans*-4-OH-Pro), by lactic acid bacteria (Ström et al. 2002). Recently, a new antifungal compound cyclo(Leu-Leu) was identified from *Lactobacillus plantarum* AF1, which was isolated from kimchi, a traditional Korean

food which is a well-known lactic acid-fermented vegetable product (Yang and Chang 2010). In this study, soybeans treated with different concentration of culture supernatant of *Lb. plantarum* AF1 could partially even totally inhibit the growth of *Aspergillus flavus*, which often germinates in stored cereal grains. Up to now, the mechanism of inhibition is still not clear. The end products from kimchi lactic acid bacteria, like cyclo(Leu-Leu), may be a promising alternative to chemical preservatives as a potential biopreservative which prevent fungal spoilage and mycotoxin formation in food and feed (Yang and Chang 2010).

Moreover, three other antifungal DKPs, cyclo(Phe-4-hydroxy-Pro), cyclo(Leu-4-hydroxy-Pro) and cyclo(Ala-4-hydroxy-Pro), were extracted from broth culture of the grapevine endophyte *Alternaria alternata*. The three DKPs demonstrate real effectiveness in inhibiting the fungus *Plasmopara viticola* sporulation which causes the grapevine downy mildew, one of the most destructive diseases affecting this crop (Musetti et al. 2007).

1.1.1.4.4 Antitumor DKPs

More and more DKPs have been demonstrated to have antitumor activity. However, their modes of action are found very varied. Several examples will be introduced in this part to illustrate their antitumor activity.

As previously mentioned (section **1.1.1.4.1**), glionitrin A was also shown to have antitumor activity. An *in vitro* cytotoxicity assay revealed that glionitrin A had potent submicromolar cytotoxic activity against four human cancer cell lines: HCT-116, A549, AGS, and DU145 (Park et al. 2009).

The DKP phenylahistin (**Figure 6N**) is a cell cycle inhibitor produced by *Aspergillus ustus*. Phenylahistin exhibits antitumor activity against eight tumor cell lines *in vitro*, and against P388 leukemia and Lewis lung carcinoma cells *in vivo* (Kanoh et al. 1999). The mechanism of action of phenylahistin is not very clear but it was elucidated that phenylahistin arrested cells in mitosis by inhibiting tubulin polymerization (Kanoh et al. 1999).

Recently, a new DKP disulfide, deoxyapoaranotin, was separated from *Aspergillus* sp. KMD 901 and found to have direct cytotoxic and apoptosis-inducing effects towards HCT116 colon cancer cell lines (Choi et al. 2011).

1.1.1.4.5 Other biological activities and applications of DKPs

Besides the biological activities detailed above, DKPs exhibit some other activities. For example, cyclo(His-Pro) is an *in vivo* anti-inflammatory compound by modulating NF-κB and Nrf2 signalling (Minelli et al. 2012). Its cytoprotective/anti-inflammatory effects can be ascribed to the cross-talk between the suppression of NF-κB signalling and the activation of Nrf2-EpRE/ARE pathway, the former depressing the pro-inflammatory response, the latter enhancing the antioxidant defensive response (Minelli et al. 2012). Some DKPs exhibit immunosuppressive activity like gliotoxin (**Figure 6B**) (Grovel et al. 2006). Gliotoxin is an immunosuppressive cytotoxin produced by numerous environmental or pathogenic fungal species. It is thought to play a role in the *A. fumigatus* virulence by facilitating fungal growth and colonization of host tissue through induction of a local or generalized immunosuppression (Grovel et al. 2006). Cyclo(Trp-Trp) and cyclo(Trp-Pro) specifically block the calcium channels therefore show an interest in the treatment of cardiovascular disorders (Milne et al. 1998). Furthermore, several mono-*N*-methylated and di-*N*-methylated DKPs were demonstrated to be able to help the passage of baicalin and dopamine across the blood−brain barrier (BBB) by passive diffusion. Thereby, the DKPs or cyclodipeptide scaffolds can be considered a novel family of brain delivery systems (BBB-shuttles) to transport to the brain drugs and other cargos that cannot cross the BBB unaided (Teixidó et al. 2007).

DKPs display numerous biological and pharmacological activities. These properties make it possible to foresee their use in the treatment of pathologies.

In the first part of this chapter, I introduced the DKP family, their natural abundance and structural diversity, some physiological roles, and their biological activities which make them a family of molecules important to the research for their interest in various applications and treatment. In the study, it is important to find out

how these DKPs are produced by organisms. So, in the next part, we will talk about the biosynthesis of DKPs.

1.1.2 Biological mechanisms of DKP formation

DKPs are commonly biosynthesized from amino acids by different organisms, including mammals, and are considered to be secondary functional metabolites or side products of terminal peptide cleavage. Up to now, several biosynthetic pathways of DKPs have been deciphered. In general, these mechanisms can be classified as two types: nonenzymatic pathways and enzymatic pathways.

1.1.2.1 Nonenzymatic pathways of DKP formation

Cyclo(His-Pro) has important implications in neurophysiological functions and is an endogenous cyclic dipeptide that exists throughout the central nervous system, peripheral tissues, and body fluids. In mammals, the DKP cyclo(His-Pro) is derived from the nonenzymatic cyclization of the thyrotropin-releasing hormone (TRH, pGlu-His-Pro) after cleavage by pyroglutamate aminopeptidase. Miyashita et al. proved that cyclo(His-Pro) can emanate from a direct predecessor, that is TRH-Gly (pGlu-His-Pro-Gly), form of TRH by pyroglutamate aminopeptidase action, not through TRH formation. TRH-Gly is firstly transformed by pyroglutamate aminopeptidase to His-Pro-Gly. Nonenzymatic conversion of His-Pro-Gly to cyclo(His-Pro) then occurs (**Figure 7**) (Miyashita et al. 1993). The proline induces constraints which promote cis conformation of the peptide bond between the histidine and the proline, thereby faciliting the cyclization to generate the DKP nucleus.

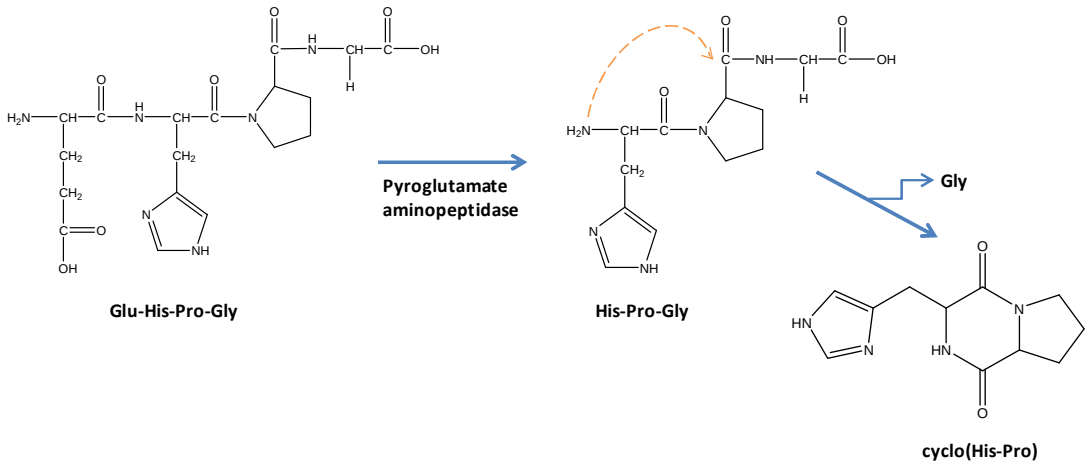


Figure 7: Mechanism of the formation of cyclo(His-Pro) in mammals (Miyashita et al. 1993).

However, in microorganisms, all the known mechanisms of DKP formation involve enzymes.

1.1.2.2 Enzymatic pathways of DKP formation

DKPs are commonly considered to be secondary metabolites in organisms. Some protease enzymes, such as dipeptidyl peptidases, cleave the terminal ends of proteins to generate dipeptides which naturally cyclize to form DKPs. There exist two known important enzyme families which can catalyze the formation of DKPs: nonribosomal peptide synthetases (NRPSs) and cyclodipeptide synthases (CDPSs). Here I will briefly illustrate with examples the roles of the two enzyme families in biosynthesis of DKPs. The structures and the catalytic machanisms of NRPSs are detailed in this chapter (section **1.2.2.1**) whereas the results of studies on CDPSs are presented in the introduction and the **chapters 2-5** because my thesis work is on CDPSs.

1.1.2.2.1 Formation of DKPs involving NRPSs

NRPSs are large multimodular biocatalysts that utilize complex regiospecific and stereospecific reactions to assemble structurally and functionally diverse peptides that have important medicinal applications. NRPSs are not enzymes dedicated only to the DKP formation. In contrast, they commonly catalyze the

formation of multiple peptide bonds. However, some dipeptides, usually containing proline, are sometimes prematurely released then cyclize to form cyclodipeptides as side products of the reaction catalyzed by NRPSs. One of examples is the formation of cyclo(D-Phe-Pro) as side product during the nonribosomal assembly of **tyrocidine A** by NRPSs (Schwarzer et al. 2001). The antibiotic tyrocidine A is a cyclic decapeptide synthesized nonribosomally by three NRPSs, TycA, TycB, and TycC, which consist of a total 10 modules, each being responsible for the incorporation of one amino acid into the final product (Schwarzer et al. 2001). Phenylalanine and proline are the first two amino acids incorporated by TycA. During the second reaction of condensation, as the dipeptide D-Phe-Pro is covalently fixed to enzyme, the side product cyclo(D-Phe-Pro) is formed due to the cis conformational constraint of proline which facilitates the nonenzymatic cyclization thereby the premature release of this DKP (**Figure 8**) (Schwarzer et al. 2001).

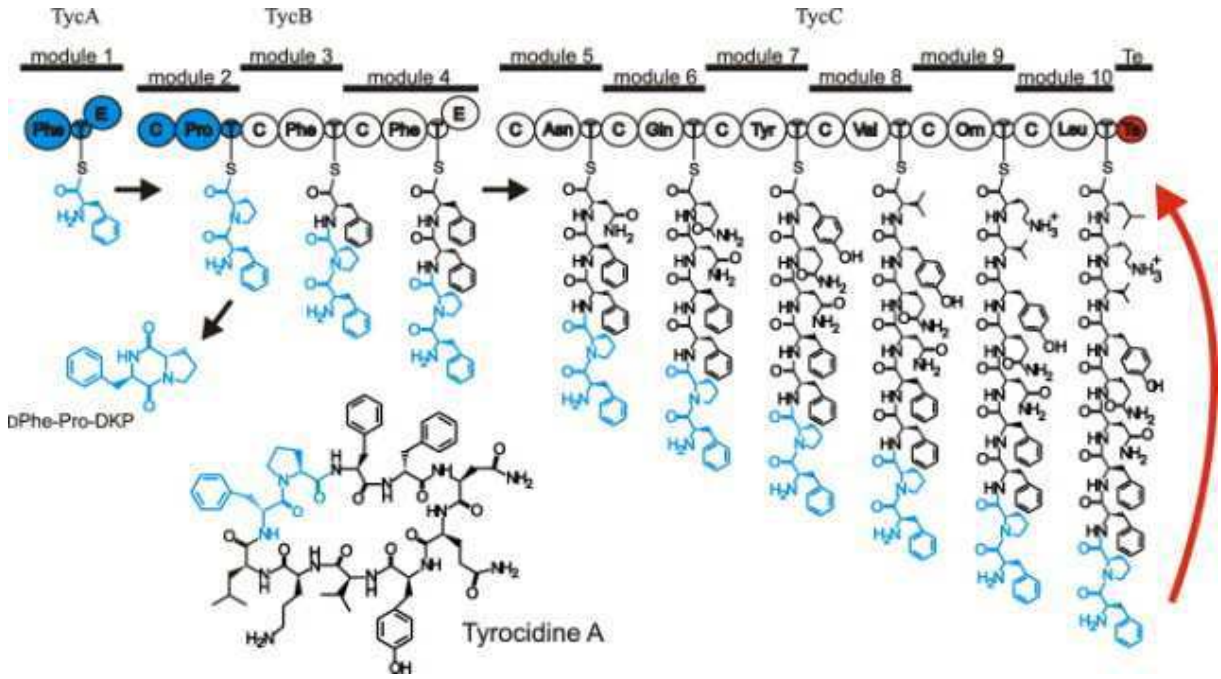


Figure 8: Formation of the side product cyclo(D-Phe-Pro) during the synthesis of tyrocidine A catalyzed by NRPSs (Schwarzer et al. 2001). Three NRPSs, TycA, TycB, and TycC, act in concert to synthesize the cyclic decapeptide from the amino acid precursors. TycA comprises one module, TycB three, and TycC six modules, each of which is responsible for the incorporation of a cognate amino acid into the growing chain. The Te domain (red) at the last module of TycC catalyzes peptide cyclization and thereby release of the final product. The D-Phe-Pro intermediate bound to the second module is chemically unstable and is released as a side product as the cyclo(D-Phe-Pro).

Although the majority of DKPs deriving from NRPSs harbor a proline residue at the second amino acid position, such as cyclo(D-Phe-Pro) mentioned above, and are proposed to form spontaneously resulting from conformational constraints induced by the proline residue (Stachelhaus et al. 1998; Stachelhaus and Walsh 2000; Schultz et al. 2008), there are some exceptions. This is the case of the production of **cyclomarazine A** (**Figure 6C**) during the biosynthesis of cyclic heptapeptides cyclomarin A by a 7-module NRPS from the marine bacterium *Salinispora arenicola* (**Figure 9**) (Schultz et al. 2008). On the subject of the formation of cyclomarazine A, N-methylation of the second residue incorporated induces constraints which promote cis conformation of the peptide bond between the two residues, thereby faciliting the formation of cyclomarazine A.

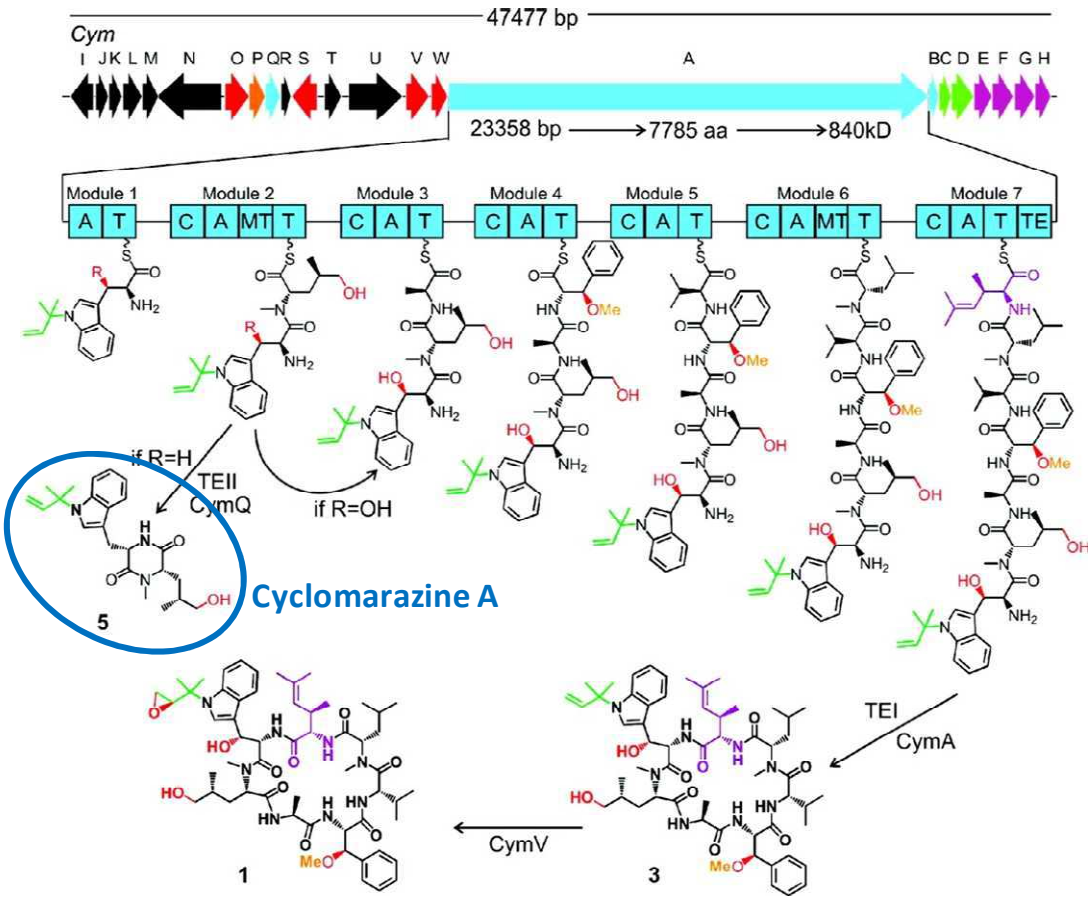


Figure 9: Biosynthetic gene cluster organization of *cym* and proposed biosynthesis of cyclomarin A (1) and cyclomarazine A (5) (Schultz et al. 2008). Each arrow represents the direction of transcription of an ORF and is color coded to signify enzyme function which is further reflected chemically. NRPS-related genes are colored blue with enzymatic domain abbreviations as follows: A, adenylation; T, thiolation (peptidyl carrier protein); C, condensation; MT, methyltransferase; and TE, thioesterase.

There exist some NRPSs dedicated to the biosynthesis of cyclodipeptides. Several biosynthetic pathways of DKPs involving NRPSs have been isolated in recent years. The biosynthetic pathway of the phytotoxin **thaxtomin A** (**Figure 6F**) is firstly described in 2000 (Healy et al. 2000). Biosynthesis of this compound involves conserved NRPSs, TxtA and TxtB, encoded by the *txtA* and *txtB* genes (Healy et al. 2000). TxtA and TxtB are responsible for production of the *N*-methylated cyclic dipeptide backbone of the toxin, and a P450 monooxygenase named TxtC is required for post-cyclization hydroxylation steps (Healy et al. 2002). A nitric oxide synthase (NOS) is suggested to function in the nitration of thaxtomin (Loria et al. 2008) (**Figure 10**).

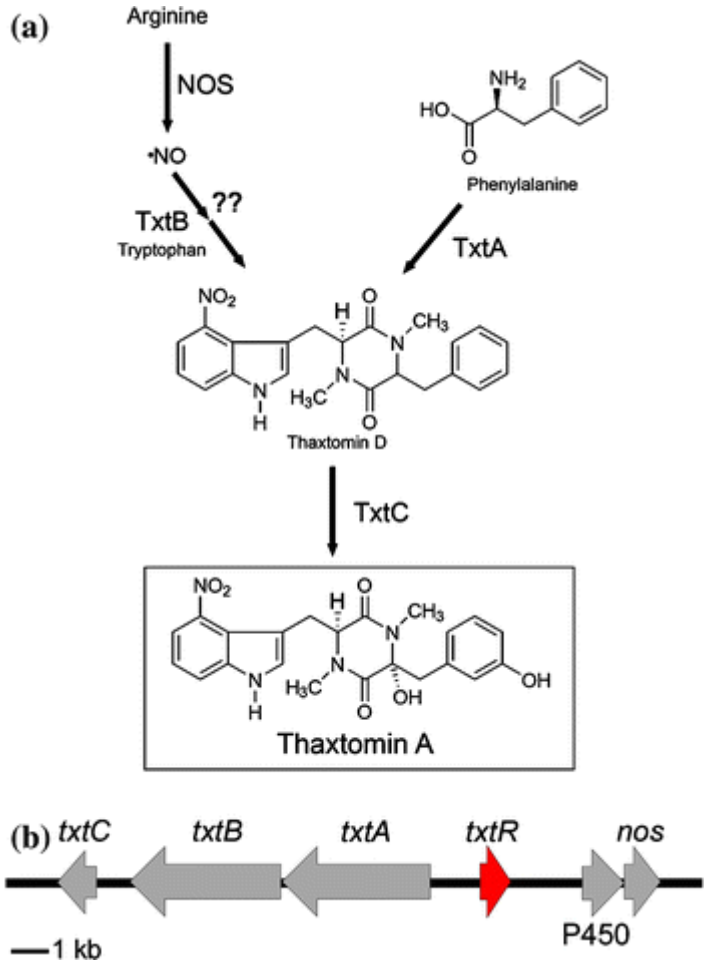


Figure 10: Biosynthesis pathway of thaxtomin A in *Streptomyces acidiscabies* (Loria et al. 2008). (a) The biosynthesis of thaxtomin A involves NRPSs (TxtA and B), a P450 monooxygenase (TxtC), and a nitric oxide synthase (NOS). (b) The genes known or predicted to be involved in thaxtomin A biosynthesis are clustered together on the chromosome. Biosynthetic genes are shown in gray, while regulatory genes are indicated in red.

Another NRPS dedicated to the DKP synthesis is a dimodular NRPS identified within a gene cluster of *Aspergillus fumigates*, aspecies reported to produce fumitremorgins and other prenylated alkaloids (Maiya et al. 2006). Fumitremorgins are tremorgenic mycotoxins, members of a group of prenylated indole alkaloids derived from tryptophan and proline. Some compounds of this group have interest in the development of anticancer drugs such as tryprostatin B which is a mammalian cell-cycle inhibitor (Zhao et al. 2002). It was found that in *A. fumigates*, the gene Afu8g00170 encoded the NRPS brevianamide synthetase. The DKP brevianamide F (cyclo(Trp-Pro)), synthesized by this NRPS, is the precursor of a variety of fungal prenylated alkaloids with biological activity, including fumigremorgins A, B and C and tryprostatin B (Maiya et al. 2006). Some modification enzyme genes are clustered with Afu8g00170 to modify brevianamide F thereby to produce varied prenylated alkaloids. For example, Afu8g00170 is clustered with the gene Afu8g00210 encoding a prenyltransferase, which is able to prenylate brevianamide F to form tryprostatin B (Grundmann and Li 2005; Maiya et al. 2006). There are other examples of biosynthesis of DKPs involving NRPSs such as those of sirodesmin (**Figure 6O**) and gliotoxin (**Figure 6B**), which are both secondary metabolites made by fungi. The biosynthetic gene cluster of sirodesmin is identified in *Leptosphaeria maculans* (Gardiner et al. 2004) (**Figure 11**). It is composed of 18 genes among which the gene *sirP* encodes a two-module NRPS involved in the biosynthesis of a precursor of sirodesmin, named phomamide (Gardiner et al. 2005) (**Figure 12**). Besides, a similar gene cluster is identified in the gliotoxin-producing fungus *Aspergillus fumigatus*by genome database searches (Gardiner et al. 2004) (**Figure 11**). It is composed of 12 genes, many of which are similar to those in the *L. maculans* sirodesmin biosynthetic gene cluster. The gene *gliP* is a homologue of *sirP,* thereby is predicted to encode a NRPS responsible for the synthesis of a precursor of gliotoxin, that is cyclo(Phe-Ser) (Gardiner et al. 2005) (**Figure 12**).

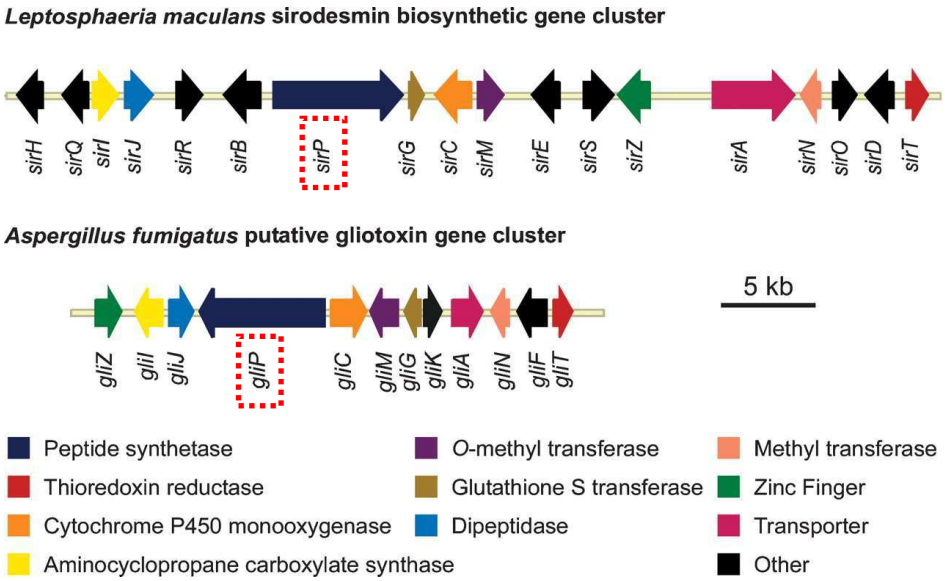


Figure 11: Comparison of the sirodesmin and gliotoxin biosynthetic gene clusters from *Leptosphaeria maculans* and *Aspergillus fumigatus*, respectively (Gardiner et al. 2005). *sirP* and *gliP* encoding NRPSs involved in biosynthesis of precursors of sirodesmin and gliotoxin respectively are encircled in red dotted lines. Genes with obvious homologues in the clusters are coloured. The 'other' category contains genes encoding cytochrome P450 monooxygenases (GliF, SirB, SirE), a prenyl transferase (SirD), an acetyl transferase (SirH), epimerases (SirQ, SirR, SirS), an oxidoreductase (SirO) and a hypothetical protein (GliK).

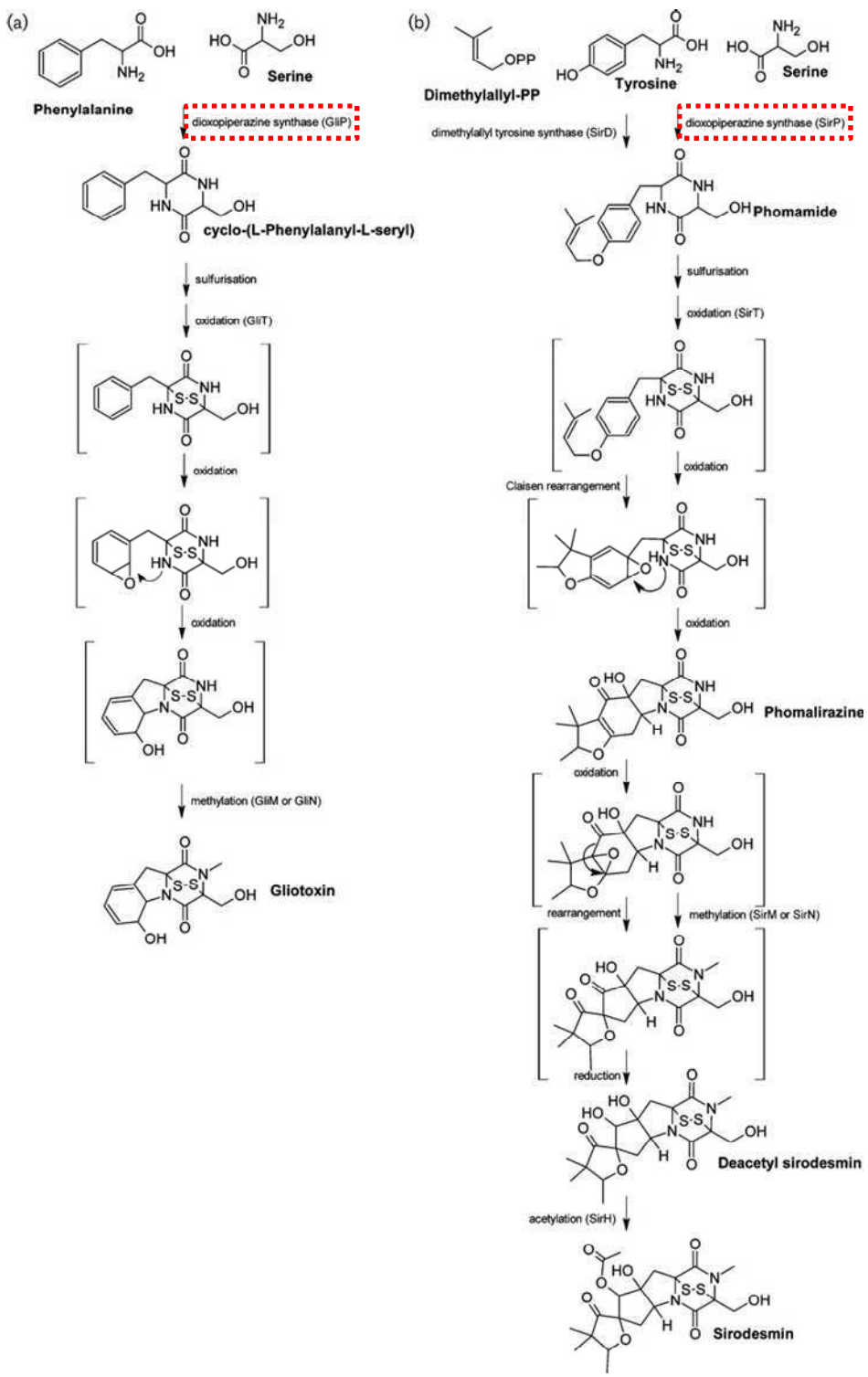


Figure 12: Predicted biosynthetic pathways for (a) gliotoxin and (b) sirodesmin (Gardiner et al. 2005). The two NRPSs GliP and SirP are encircled by dotted red lines.

Besides NRPSs, in the introduction part of this manuscript, we have presented another enzyme family called CDPSs dedicated to the biosynthesis of cyclodipeptides.

1.1.2.2.2 Formation of cyclodipeptides by CDPSs

As described in the introduction of the manuscript, the CDPS family is specially involved in the synthesis of cyclodipeptides. Unlike NRPS which is often a huge enzyme with multi-modules, the CDPS is always a small enzyme composed of only about 200 amino acid residues. They are very different from NRPSs or other known enzymes. This thesis is based on studies of the CDPS family. Before our dissection of CDPSs, it is necessary to describe all the known biological macromolecules that are able to catalyze the peptide bond formation in organisms.

1.2 Formation of peptide bonds by biocatalysts

A peptide bond is a covalent bond formed between two molecules when the carboxyl group of one molecule reacts with the amino group of the other molecule, leading to the release of a molecule of water (H_2O). It usually occurs between amino acids (**Figure 13**).

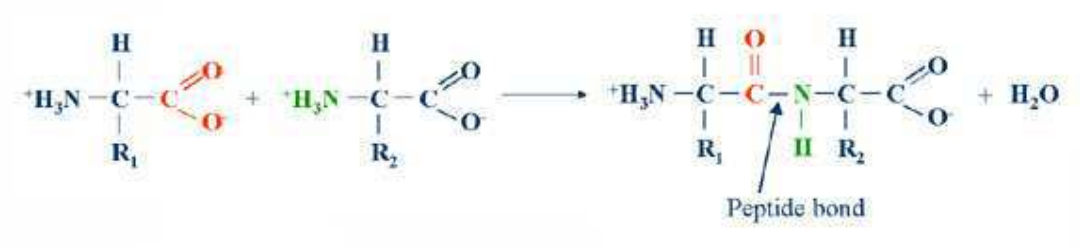


Figure 13: Formation of a peptide bond between two amino acids.

Peptide bonds are ubiquitous and important in organisms for the structures of many biological molecules such as proteins and numerous peptide derivatives. In organisms, peptide bond formation is catalyzed by certain biocatalysts in order to

accelerate the reaction since uncatalyzed reaction is very long. There exist in cells different synthetic systems to form peptide bonds, which can be divided into two types: ribosomal pathway and nonribosomal pathways. The ribosomal pathway involving the macromolecule ribosome is the main pathway to synthesize proteins and polypeptides by forming multi-peptide bonds. It is one of the most important activities in cells to ensure the normal cell metabolism. Formation of peptide bonds through nonribosomal pathways can be accomplished by several different enzyme families such as NRPSs, CDPSs, Fem-transferases, aminoacyl-tRNA transferases, and glutathione synthetase (**Figure 14**). No matter which pathway is used to form peptide bonds, activation of the carboxyl group is indispensable to bring about the reaction. In some cases, such as NRPS pathways, the carboxyl group of amino acid involved in the peptide bond formation is activated by ATP thereby forming an intermediate aminoacyl adenylate. In other cases, activated amino acid is present in the form of aminoacyl-tRNA (aa-tRNA), the formation of which is catalyzed by the aminoacyl-tRNA synthetase (aaRS).

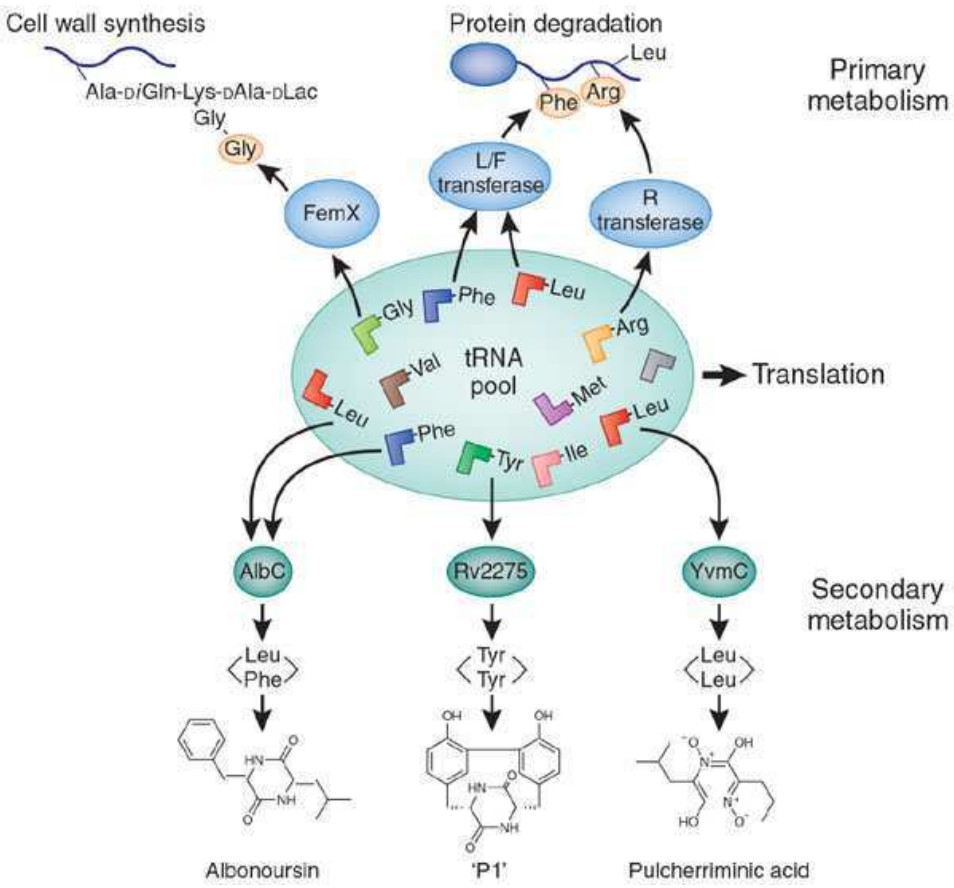


Figure 14: Some tRNAs are involved in primary processes such as cell wall formation, protein labeling for degradation, biosynthesis of amino acids and modification of porphyrin and lipids. Now tRNAs have been found in secondary metabolism in cyclic peptide formation (von Dohren 2009). 'P1' is a recently discovered product created by the P450-mediated oxidation of cyclo(Tyr-Tyr) (Belin et al. 2009).

In this part, I will firstly describe the biocatalysts of peptide bond formation which involve aa-tRNAs as substrates: ribosome, Fem-transferases, aminoacyl-tRNA transferases, CDPS, and a recently discovered transferase named PacB (Zhang et al. 2011). Before the description of these biocatalysts, aaRSs, enzymes catalyzing the aminoacylation of tRNA, will be particularly discussed because of their structural homology with our research object CDPSs. This part will then be ended with the description of some others biocatalysts of peptide bond formation which function in a tRNA-independent manner, in particular the NRPSs.

1.2.1 Peptide bond formation involving aa-tRNA

Above all, we will briefly evoke tRNA and aaRS. Then we will present the ribosomal pathway of peptide bond formation, the essential peptide and protein synthesis pathway in cells. Finally we will talk about some enzyme families, which catalyze the peptide bond formation by using the aa-tRNA as substrate.

1.2.1.1 Transfer RNA (tRNA)

Representing the single largest, best-understood class of non-protein coding RNA genes found in all living organisms, tRNAs are adaptor molecules composed of RNA, typically 73 to 93 nucleotides in length, that are used in protein biosynthesis to link the codons in a messenger RNA (mRNA) to the amino acids that they specify (Crick 1968).

tRNAs have *cloverleaf* secondary structure due to four base-paired stems (**Figure 15A**). The secondary structure folds into a compact L-shaped three-dimensional structure *via* a set of tertiary and triple base-pair interactions, as well as a number of magnesium ions that are crucial for folding and stability (Jovine et al. 2000) (**Figure 15B**). Generally, tRNA structure contains an acceptor arm, an anticodon arm, a D arm, and a TΨC arm (**Figure 15**). The acceptor arm is made by the base pairing of the 5'-terminal nucleotide with the 3'-terminal nucleotide, which contains the CCA 3'-terminal group used to attach the amino acid. The loop of anticodon arm contains the anticodon, which can match with the triplet codon on mRNA. The D arm often contains hydrouridines where as the TΨC arm contains the sequence TΨC where Ψ is a pseudouridine.

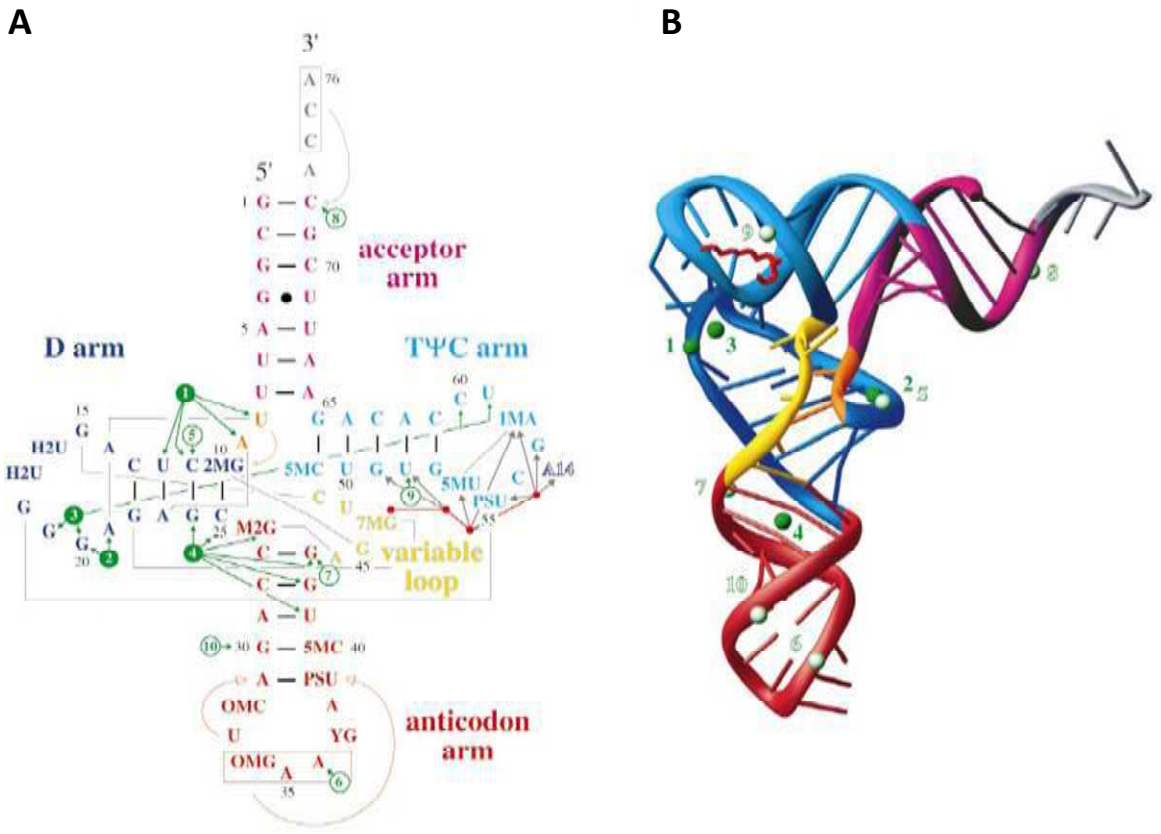


Figure 15: Secondary and tertiary structure of yeast $tRNA^{Phe}$ (Jovine et al. 2000). (A) "Cloverleaf" representation of yeast $tRNA^{Phe}$ secondary structure. Different regions of the molecule are indicated and colour-coded accordingly. Watson-Crick base-pairs are represented by short black lines, the single G4·U69 wobble pair is marked by a black dot. Tertiary interactions are shown as grey lines, while arrows indicate stacking of the enclosed nucleotides onto 5′ or 3′ helices. The anticodon and the terminal CCA sequence are boxed in red and grey, respectively. Tightly and weakly bound Mg^{2+} are shown as green and white circles, respectively, with green arrows indicating the nucleotides with which they interact. The spermine molecule is represented by a red stick, with four dots corresponding to its nitrogen atoms (not drawn to scale); grey arrows indicate interactions with the RNA. The A14 nucleotide of a symmetry-related molecule is labelled in reverse colours. Modified nucleotide name abbreviations are; 2MG, *N*2-methylguanosine; H2U, dihydrouridine; M2G, *N*2,*N*2-dimethylguanosine; OMC, 2′-*O*-methylcytidine; OMG, 2′-*O*-methylguanosine; YG, wybutosine; PSU, pseudouridine; 5MC, 5-methylcytidine; 7MG, 7-methylguanosine; 5MU, ribosylthymine; 1MA, 1-methyladenosine. (B) Ribbon representation of the three-dimensional structure. The structure is colour coded as in (A).

In general, tRNA biogenesis involves the synthesis of the initial transcript, followed by processing to remove the 5′ leader, trim the 3′ trailer, add CCA, splice introns that may be present, modify multiple nucleoside residues (**Figure 15A**), and, for eukaryotes, export the tRNA to the cytoplasm, before its use in translation (Phizicky and Hopper 2010). tRNA genes are highly transcribed to meet the needs of

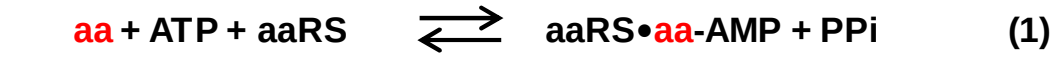
cells to tRNA. tRNA transcription is a complicated process involving multiple layers of regulation (Phizicky and Hopper 2010). Freshly transcribed pre-tRNAs with extensions 5′ and 3′ as well as introns need to be processed to generate functional tRNA molecules. Maturation on 5′ involves an endoribonuclease RNase P consisting of one RNA and a variable number of protein subunits: one in bacteria, four in archaea, nine in yeast and ten in mammalian (human, nuclear) RNase P (Kirsebom 2007). RNase P RNA is the catalytic component in bacteria (Guerrier-Takada et al. 1983; Torres-Larios et al. 2005) whereas the protein subunit increases the catalytic efficiency of the RNA component (Crary et al. 1998; Sun et al. 2006). Maturation of the 3′ end of tRNA always requires removal of the 3′ trailer from the original transcript, and also often requires the subsequent addition of CCA after the nucleotide N_{73}, although, in some bacteria and archaea, some or all tRNA genes have encoded CCA (Vogel et al. 2005; Hartmann et al. 2009). 3′ processing involves numerous endonucleases and exonucleases, and the mechanism is different among the three phyla. Furthermore, although introns occur in only a minority of tRNA genes, they are found in all sequenced eukaryotes and archaea, and splicing is essential (or nearly essential) in all of these organisms. In this process, introns are excised by certain endonucleases, then dissected tRNAsare repaired by special ligases (Phizicky and Hopper 2010). Finally, before definitely getting matured and functional, tRNAs have to suffer some post-transcriptional modifications on their nucleotides (methylated nucleotides, pseudouridines, dihydrouridines and so on) (**Figure 15A**). A total of 92 different tRNA modifications are listed in the RNA Modification Database (http://biochem.ncsu.edu/RNAmods). In tRNAs, which are among the most strongly modified RNAs, most nucleotide modification sites are clustered in two sites, one in the structural core of the L-shaped 3D structure and the other in the anticodon domain (Grosjean et al. 1996). These modifications have different affects on tRNA activity. Some modifications in or around the anticodon loop affect translation or growth; many modifications in the main body of the tRNA affect tRNA folding or stability; some other modifications at various positions specifically affect tRNA identity (Phizicky and Hopper 2010).

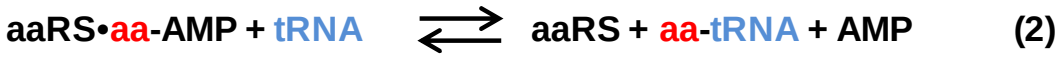
After those complicated post-transcriptional modifications, matured tRNAs are ready to perform their duties as transporter of amino acids by forming aa-tRNAs. Loading of amino acids to tRNAs, that is aminoacylation of tRNAs, is performed by aminoacyl-tRNA synthetases.

1.2.1.2 Aminoacyl-tRNA synthetase (aaRS)

Aminoacyl-tRNA synthetases were first discovered to ligate an amino acid to tRNA in 1958 (Hoagland et al. 1958; Martinis et al. 1999). They are the linchpins of translation, the link between the worlds of protein and nucleic acid by covalently linking an amino acid to a tRNA that contains the triplet anticodon for that amino acid. The charged tRNAs are then available for use in polypeptide chain elongation on the ribosome or other tRNA-dependent peptide bond formation reactions. In addition to their housekeeping role in peptide synthesis, recently, aaRSs have been paid attention for their potential pathophysiological implications in tumorigenesis (Kim et al. 2011).

Generally, the ligation of a substrate amino acid to its cognate tRNA proceeds in two steps:





The first step involves the activation of an amino acid to aminoacyl-adenylate in consumption of one ATP, and the second delivers the activated amino acid to the acceptor end of its congate tRNA (Ibba and Söll 2000). This reaction is catalyzed by a specific aaRS. There exist 20 aaRSs responsible for the synthesis of the 20 canonical aa-tRNA families (Woese et al. 2000). In organisms, an amino acid could be charged to several different cognate tRNAs which are called isoacceptors (Saks et al. 1998). All the isoacceptors are charged by a single aaRS (Woese et al. 2000). For example, transfer of the leucine to its 6 different isoaccepting tRNAs in *E. coli* is catalyzed by

the same LeuRS. In a word, each aaRS is specific for one amino acid and one or more isoaccepting tRNAs. A set of tRNA isoacceptors can be distinguished by their main points of contact for aaRS, that is the acceptor arm, the anticodon loop and a few base pairs in the T and D stems (Arnez and Moras 1997).

1.2.1.2.1 Classification of aaRSs

Although all aaRSs share the same catalytic mechanism, they appear to be different from each other in terms of molecular size, quaternary structure and amino acid sequence. Overall, aaRSs are divided into two unrelated classes (I and II) based on mutually exclusive sequence motifs that reflect distinct active site topologies. In addition, each class is divided into subclasses based on additional common features among some members of each class (Arnez and Moras 1997; Ibba and Söll 2000). Principal features of the two classes of aaRSs and their classification are summarized in **Table 2**. Some exceptions should be noted. LysRS exists in two unrelated forms thereby is found as both a class I and a class II aaRS (Arnez and Moras 1997; Ibba and Söll 2000; Ambrogelly et al. 2002). Most prokaryotes and all eukaryotes contain a class II LysRS, whereas most archaea and a few bacteria contain a less common class I LysRS (Ambrogelly et al. 2002). GlyRS exists also in two forms and is classed in two different subclasses of class II (Freist et al. 1996).

	Class I	Class II
Principal features of the two classes of aaRSs		
Active sites	Rossmann fold (parallel β–sheet nucleotide-binding fold)	Antiparallel β-sheet
Conserved motifs	HIGH and KMSKS	Motif 1: ...**P**... Motif 2: ...**FR**xE... Motif 3: ...(Gx)3E**R**...
ATP-binding conformation	Extended	Bent
tRNA binding : - **Acceptor stem** - **Variable loop**	- Minor groove side - Faces solvent	- Major groove side - Faces protein
Aminoacylation site	2'-OH	3'-OH
Subclasses		
	Ia	***IIa***
	ArgRS α	GlyRS α_2
	CysRS α	HisRS α_2
	IleRS α	ProRS α_2
	LeuRS α	ThrRS α_2
	MetRS α, α_2	SerRS α_2
	ValRS α	
	LysRS I α	
	Ib	***IIb***
	GlnRS α	AsnRS α_2
	GluRS α	AspRS α_2
		LysRS II α_2
	Ic	***IIc***
	TyrRS α_2	AlaRS α, α_4
	TrpRS α_2	GlyRS $(\alpha\beta)_2$
		PheRS $(\alpha\beta)_2$, α

Table 2: Classes of aaRSs and their principal features (Arnez and Moras 1997; Ibba and Söll 2000). Both LysRS and GlyRS exist in two unrelated forms, so they are both found in two different subclasses.

1.2.1.2.1.1 Class I aaRSs

As shown in **Table 2**, class I aaRSs share sequence motifs HIGH and KMSKS, and have active sites based on the Rossmann fold. This fold is characterized by a central β–sheet composed of several β-strands linked by α–helix in the topological order β-α-β.

During the first step of aminoacylation, for example catalyzed by GlnRS (**Figure 16A**), ATPs adopts an extended conformation reminiscent of that found in other proteins containing the Rossmann fold. In this catalytic site, the two signature motifs, HIGH and KMSKS, interact with the ATP and stabilize the transition state of the reaction for amino acid activation. Whereas, in contrast to other aaRS, GlnRS (together with ArgRS, GluRS, and LysRS I) requires bound tRNA for amino acid activation. One significant consequence of this requirement is that the sequence of the tRNA strongly influences the affinity of GlnRS for glutamine (Hong et al. 1996; Ibba et al. 1996). The second step involves binding and aminoacylation of tRNA. In tRNA binding mode, class I aaRSs approach the acceptor stem of tRNA from the minor groove side with the variable loop facing the solvent. In addition, class I aaRSs attach amino acids to the 2'-OH of the terminal adenosine of the tRNA.

Class I aaRSs are furthermore subgrouped to three subclasses. Members of each subclass share common enzymatic modules and generally charge the same type amino acids onto cognate tRNAs. Thus, the subclass Ia aaRSs process mostly aliphatic or sulfphur-containing amino acids except ArgRS and LysRS I; the subclass Ib aaRSs require tRNA for activation of charged amino acids, Glu and Gln; whereas the assembly of aromatic amino acids (Tyr and Trp) to their cognate tRNAs are taken charge by the subclass Ic aaRSs (**Table 2**). In the class I aaRSs, amino acids are bound in relatively open and relaxed pockets.

With regard to the quaternary structure, class I aaRSs are mostly monomeric although TyrRS and TrpRS of the subclass Ic are obligate dimmers (Arnez and Moras 1997).

1.2.1.2.1.2 Class II aaRSs

The Rossmann fold is absent in the class II aaRSs. Instead, their active sites are built on a novel antiparallel β–sheet surrounded by α–helix (**Figure 16B**). Class II aaRSs share three conserved sequence motifs (**Table 2**). Motif 1 consists of a long α–helix linked to a β–strand. Motif 2 comprises two antiparallel β–strands connected by a long loop, and motif 3 consists of a β–strand followed immediately by an α–helix. Motifs 2 and 3 are part of the active site (**Figure 16B**).

In the first step of aminoacylation, in contrast to class I aaRSs, the ATP exhibits a compact conformation: the γ–phosphate folds back over the adenine base. The ATP is cradled by the antiparallel β–sheet, which is formed in part by the conserved motifs 2 and 3. Amino acids are bound by specific rigid templates in class II aaRSs. Motif 2 is involved in the binding of amino acids. In the second step of aminoacylation, class II aaRSs, such as AspRS (**Figure 16B**), bind tRNA on the major groove side. Thus, the variable loop of the tRNA faces the aaRS. Class II aaRSs attach amino acids to the 3'-OH of tRNA except PheRS which attaches Phe to the 2'-OH of $tRNA^{Phe}$ (Arnez and Moras 1997).

Class II aaRSs are divided into 3 subclasses (**Table 2**). The class IIa contains five aaRSs, which charge small and polar amino acids onto their cognate tRNAs; the class IIb aaRSs process charged amino acids; whereas the class IIc aaRSs take charge of small amino acids (Ala and Gly) and an aromatic amino acid (Phe).

Most of class II aaRSs are dimmers (Arnez and Moras 1997). The conserved motif 1 participates in the dimer interface. The latter participates in the formation of the active site and cooperative coupling upon tRNA binding. Class IIc aaRSs are obligate or could be tetramers. For example, PheRS is an $(\alpha\beta)_2$ tetramer.

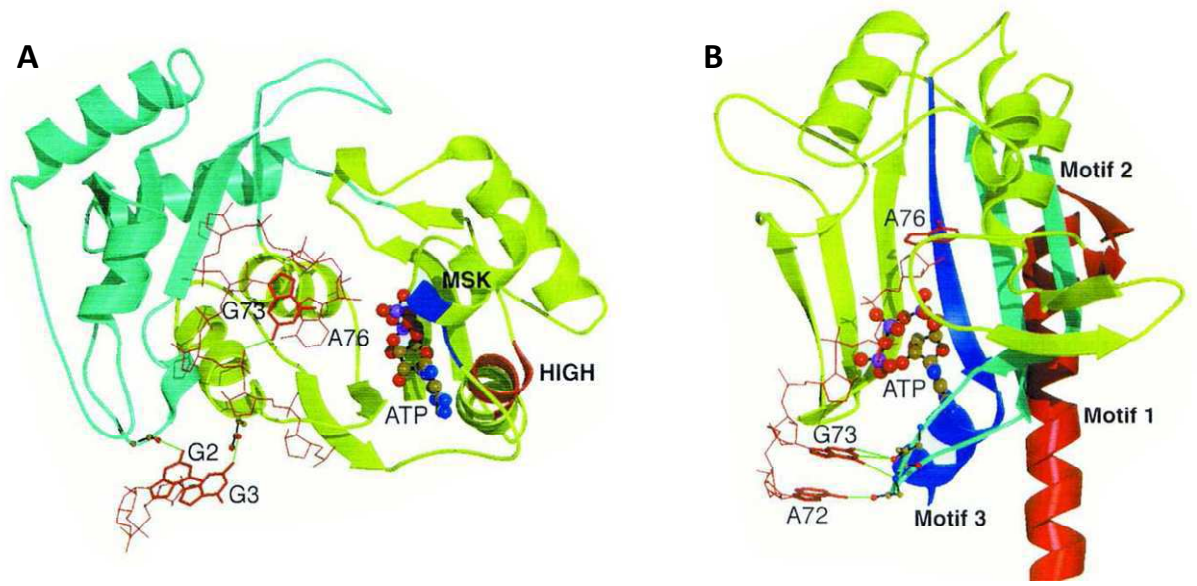


Figure 16: Active-site domains of (A) class I aaRS, e.g. GlnRS, and (B) class II aaRS, e.g. AspRS. Shown are ATP and the acceptor ends of cognate tRNAs (red). The locations of the characteristic motifs are indicated: in (A), MSK (dark blue), HIGH (red); in (B), motif 1 (red), motif 2 (light blue) and motif 3 (dark blue).

1.2.1.2.2 Evolution of aaRSs

The universal distribution across the phylogenetic tree of the aaRSs suggests that they are among the oldest polypeptide families (Ribas de Pouplana and Schimmel 2001).

Aminoacylation of all 20 universal amino acids is partitioned between two major aaRS folds (class I and class II). Within each class, the aaRSs with different amino acid specificity show distant shared ancestry as revealed by structural, sequence, and enzymatic similarity. Within each class, only the catalytic domain is conserved among all the members. Each class has its own specific structural fold forming the catalytic domain (**Table 2**). No evidence for a common ancestor to the two classes has been uncovered. Although the aaRSs of two different classes are not related by divergent evolution, they are clearly the result of a functional evolutionary convergence, as they carry out the same basic biochemical function, that is linking specific amino acids to their cognate tRNAs (O'Donoghue and Luthey-Schulten 2003).

It was found that two aaRSs from different class can be docked simultaneously onto the tRNA acceptor stem by binding opposite sides of the acceptor stem (Ribas de Pouplana and Schimmel 2001; Ribas de Pouplana and Schimmel 2001). Remarkably, the docking pairs are not arbitrary. These specific pairings linked two members of class Ia (IleRS and ValRS) respectively with two members of class IIa (ThrRS and SerRS), a member of class Ib (GlnRS) with one of class IIb (AspRS), and one from class Ic (TyrRS) with a member of class IIc (PheRS) (Ribas de Pouplana and Schimmel 2001). This complementation phenomenon of a pair aaRSs in the tRNA binding manner is supposed to be related to a special role of the early aaRSs, which is acting as "chaperones" to cover and protect the acceptor end of tRNA therefore to preserve aminoacylation probably in a high-temperature environment. Thus, a specific primordial chaperone/synthetase pair may have been involved in the discrimination between sterically similar amino acids like as tyrosine and phenylalanine. The individual members of each pair may have acquired exact amino acid specificity. In this way, the primordial chaperone/synthetase pairs would have contributed to the final development of the genetic code. As the members of each pair gained exact amino acid specificity, and the pairs split apart, new codons would have to be assigned to differentiate between similar amino acids (Ribas de Pouplana and Schimmel 2001).

In addition to their aminoacylation role, catalytic domains of aaRSs were found in a class of HisRS-like proteins named HisZ which lack aminoacylation activity but are involved in histidine biosynthesis (Sissler et al. 1999). Moreover, the overall structure of *E. coli* asparagine synthetase (AsnA) which generates asparagine from aspartate is remarkably similar to that of the catalytic domain of yeast AspRS (Ruff et al. 1991). The active site architecture and most of the catalytic residues are also conserved in both enzymes. They were thereby supposed to have evolved from a common ancestor and share a similar chemistry (Nakatsu et al. 1998). In addition, the newly discovered cyclodipeptide synthases show structural similarity to the catalytic domains of class I aaRSs (Sauguet et al. 2011), which will be detailed in the **chapter 2**. Recently, a new class of SerRS-like proteins was found structurally related to the

catalytic core of atypical SerRS (aSerRS) found only in methanogenic archaea. These truncated SerRS homologs acylate carrier proteins by selected amino acids. It is speculated that single-domain SerRS-like proteins are ancestors of both thio-acylating enzymes needed for nonribosomal peptide synthesis and of class II synthetases required for ribosomal protein synthesis. Therefore, they provide a link between ribosomal and nonribosomal peptide synthesis (Mocibob et al. 2010). These observations show that the evolution of aaRSs is probably closely related to the evolution of the genetic code, amino acids and their metabolism.

After presenting the tRNA and the enzymes aaRSs responsible for its aminoacylation, we will describe the ribosome with an emphasis on the structure of its active site and the catalytic mechanism of peptide bond formation currently accepted.

1.2.1.3 Ribosome

The ribosome is the protein synthesis factory in the cell. It decodes the genetic information carried by mRNA. Considering its important role in protein biosynthesis, the ribosome is extensively studied, no matter in structural or in functional aspect, in order to understand the catalytic mechanism of the peptide bond formation thereby to decrypt how the genetic information is translated to a protein form.

1.2.1.3.1 Ribosome structure

The ribosome is the large ribonucleoprotein particle using mRNA as the template and tRNAs as substrates to perform the protein synthesis. Ribosomes are composed of two subunits, a large one and a small one. Prokaryotic ribosomes and eukaryotic ribosomes have significantly different structures and RNA sequences. Ribosomes from bacteria consist of a large subunit which has a sedimentation coefficient of 50S and a small (30S) subunit, which together compose the 70S ribosome; their eukaryotic counterparts are the 60S and 40S subunits and the 80S ribosome (Schmeing and Ramakrishnan 2009). Here, we will mainly present prokaryotic ribosomes. In prokaryote, the ribosomal 50S subunit consists of 23S RNA

(~2900 nucleotides), 5S RNA (~120 nucleotides) and about 34 proteins; the 30S subunit consists of 16S RNA (~1500 nucleotides) and about 20 proteins. The 50S subunit contains the site of catalysis, the peptidyl transferase center (PTC), which is responsible for making peptide bonds and for the hydrolysis of peptidyl-tRNA. The 30S subunit contains the decoding site, where base-pairing interactions between the mRNA codon and the tRNA anticodon determine the selection of the cognate aa-tRNA. In addition, several protein factors act on the ribosome at various stages of translation (Schmeing and Ramakrishnan 2009). The resolution of the ribosome structure is the basis for the phasing and molecular interpretation of every subsequent structure of the ribosome. During the last ten years, progress has been made from low-resolution structures of the 70S ribosome by crystallography (Yusupov et al. 2001) to recent high-resolution structures of the 70S ribosome (Schuwirth et al. 2005; Selmer et al. 2006).

The basic architecture of the ribosome is shown in **Figure 17** (Schmeing and Ramakrishnan 2009). The interface between the two subunits consists mainly of RNA. The mRNA binds in a cleft of the 30S subunit, where its codons interact with the anticodons of the tRNA. The ribosome has three tRNA binding sites: the A site that binds the incoming aminoacyl-tRNA, the P site that holds the peptidyl-tRNA attached to the nascent polypeptide chain, and the E (exit) site to which the deacylated P-site tRNA moves after peptide bond formation before its ejection from the ribosome (Schmeing and Ramakrishnan 2009). The whole peptide synthesis process occurs on these sites.

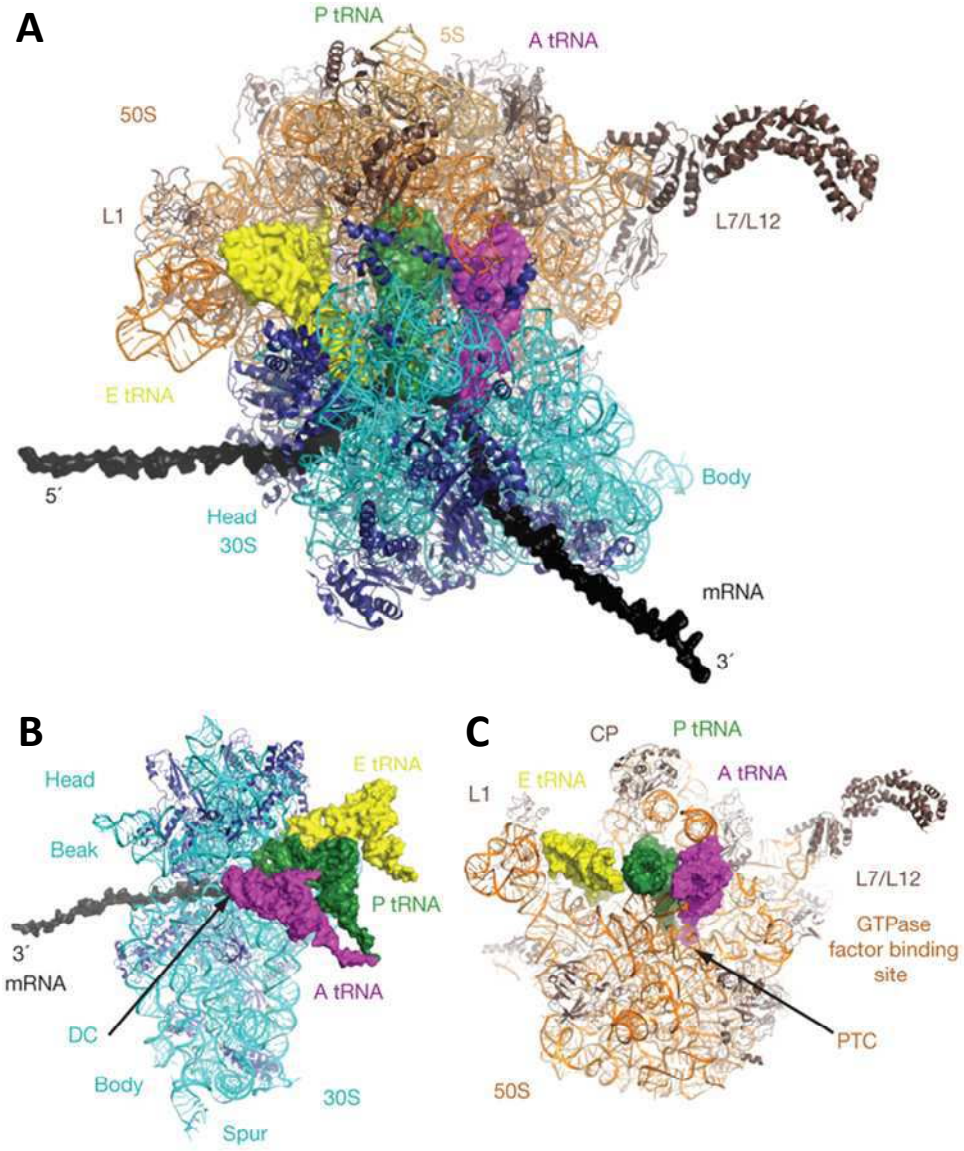


Figure 17: Structure of the ribosome (Schmeing and Ramakrishnan 2009). A, Top view of the 70S ribosome with mRNA and A-P- and E-site tRNAs. B and C, Exploded view of the 30S subunit (B) and 50S subunit (C).

1.2.1.3.2 General mechanism of the peptide synthesis by the ribosome

Bacterial translation can be roughly divided into three main stages: initiation, elongation and termination. The overview of bacterial translation is described in **Figure 13** (Schmeing and Ramakrishnan 2009) although not every detail is clear up to now.

Initiation requires the ribosome to position by base pairing the initiator, formylmethionine (fMet)-tRNA^{fMet} in bacteria, over the start codon of mRNA in the P site. The precise positioning requires three initiation factors named IF1, IF2, and IF3 (**Figure 18, Initiation**). However, exactly how the correct tRNA is selected and the roles of these factors still remain unclear.

Once the translation is successfully initiated, the elongation cycle begins to sequentially add amino acids to the polypeptide chain. An elongation cycle can be divided into three steps: decoding, peptide bond formation, and translocation (**Figure 13, Elongation**). Decoding ensures that the correct aa-tRNA, as dictated by the mRNA codon, is selected in the A site. During decoding, aa-tRNA is delivered to the A site in a ternary complex with elongation factor Tu (EF-Tu) and GTP. Following the GTP hydrolysis by EF-Tu and the dissociation of the factor from the ribosome, aa-tRNA accommodates in the A site to prepare for the next step concerning the peptide bond formation. During this step, the aminoacyl part of aa-tRNA reacts with peptidyl-tRNA bound to the P site to form a new peptide bond, yielding deacylated tRNA in the P site and A site peptidyl-tRNA extended by one amino acid residue. The mechanism of peptide bond formation will be discussed in the next part (section **1.2.1.3.3**). Before the next round of elongation, the tRNAs and mRNA need to move relative to the ribosome. This process is termed translocation. During translocation, catalyzed by another elongation factor named EF-G, the mRNA normally shifts by precisely one codon. The deacylated tRNA translocates from the P site to the E site, whereas the nascent peptidyl-tRNA moves from the A site to the P site. In consequence, a new codon is exposed in the A site for the interaction with the next aa-tRNA. The deacylated tRNA bound to the E site is released during the beginning of the next elongation cycle.

The elongation cycle continues until an mRNA stop codon moves into the A site, signaling the end of the coding sequence. A class I release factor (RF1 or RF2 in bacteria) recognizes the stop codon and cleaves the nascent polypeptide chain from the P site tRNA, resulting in the release of the newly synthesized protein from the ribosome. Then the deacylated tRNAs and mRNA are subsequently released from

the ribosome which is then recycled into subunits for a new round of protein synthesis to begin in the help of various factors including RF3, EF-G, and the ribosome recycling factor (RRF) (**Figure 18, Release and Recycling**).

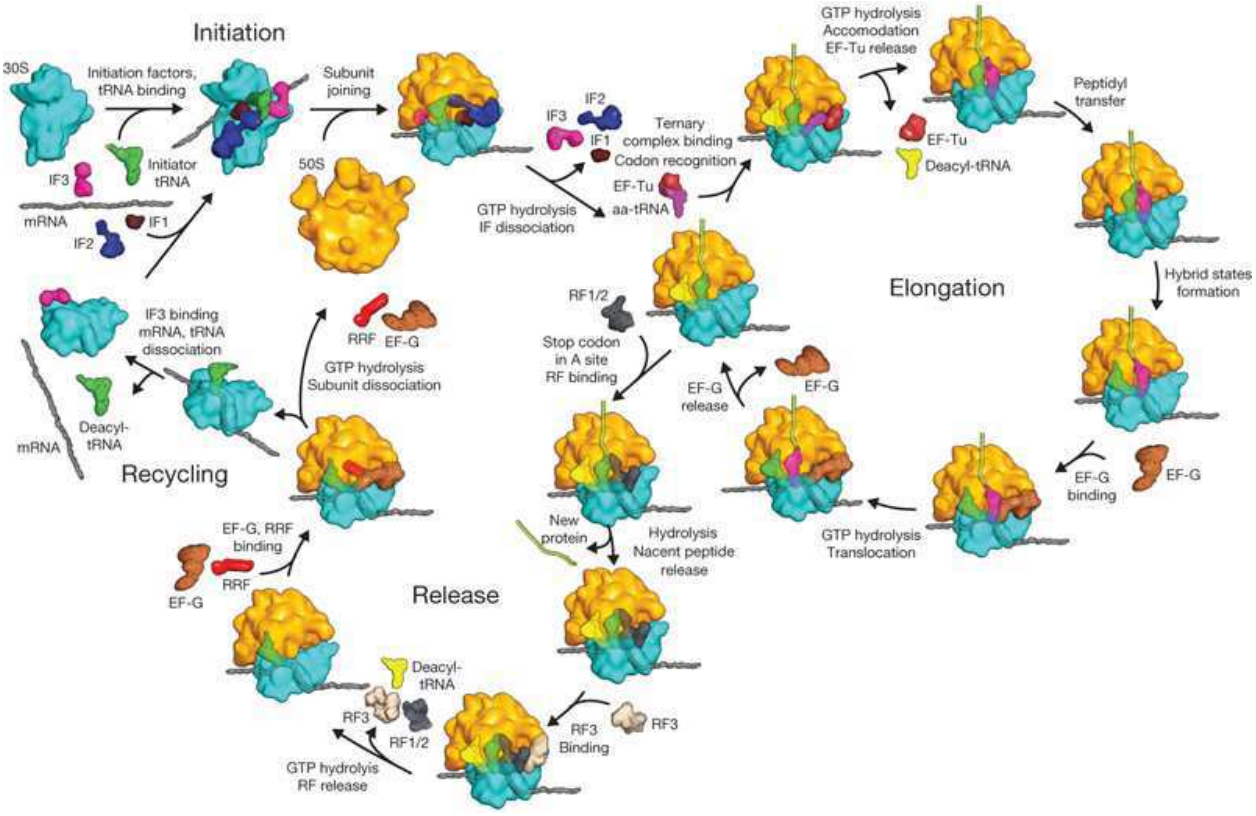


Figure 18: Overview of bacterial translation (Schmeing and Ramakrishnan 2009). EF, elongation factor; IF, initiation factor; RF, release factor.

1.2.1.3.3 Mechanism of peptide bond formation

In this part, we will specially discuss the mechanism of peptide bond formation during the elongation cycle of peptide synthesis in the ribosome. As previously presented, the peptidyl transferase center (PTC) of the 50S ribosomal subunit is the site of catalysis responsible for the peptide bond formation and the hydrolysis of peptidyl-tRNA. In fact, the central chemical event in protein synthesis is the peptidyl-transferase reaction, in which the α–amino group of the aa-tRNA bound to the A site nucleophillically attacks the ester carbon of the peptidyl-tRNA attached to the P site to form a new peptide bond (Beringer and Rodnina 2007; Schmeing and

Ramakrishnan 2009) (**Figure 19**). Compared to the uncatalyzed reaction, the ribosome increases the rate of this reaction by at least ~10^5-fold (Sievers et al. 2004).

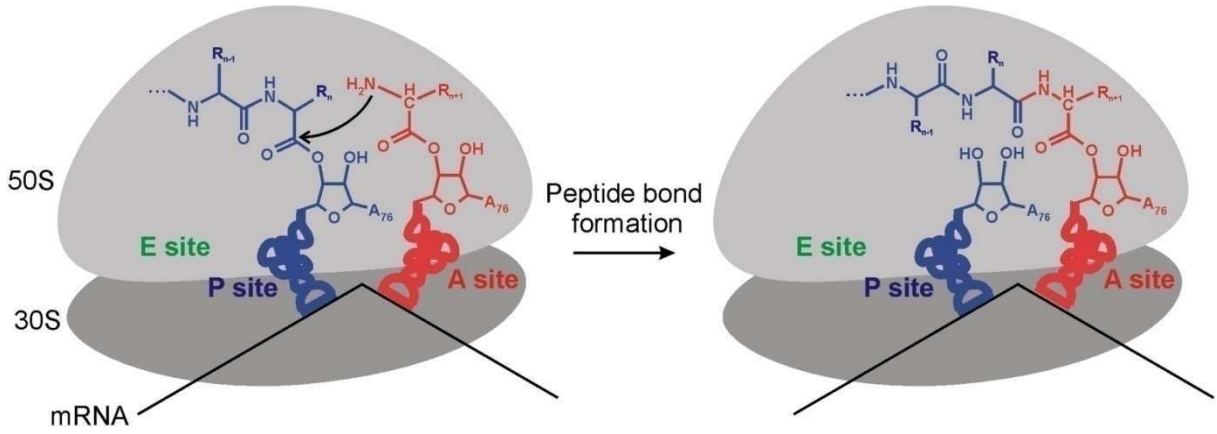


Figure 19: Schematic of peptide bond formation on the ribosome (Beringer and Rodnina 2007).

How does the ribosome so efficiently catalyze the peptide formation? Now the catalytic mechanism of peptide bond formation on the ribosome is becoming more and more clear thanks to the progress in ribosome crystallography, as well as kinetic, biochemical, genetic, and computational approaches. It has been proved that no general acid/base catalysis involving a group with near-neutral pKa because when the reactive α–amine was substituted with a much less reactive hydroxyl, making chemistry rate limiting, a pH-independent reaction rate was observed (Bieling et al. 2006). A comparison of the rate of peptide bond formation by the ribosome and by a ribosome free model system suggested that the ribosome accelerated the reaction solely by entropic effects (Sievers et al. 2004). In addition, the combined evidence of all the studies supports the idea that peptide bond formation on the ribosome is driven by a favorable entropy change. The most favorable catalytic mechanism, named concerted proton shuttle mechanism, has been proposed (Beringer and Rodnina 2007). It is substrate-assisted catalysis of peptide bond formation involving a six-membered transition state in which proton shuttling occurs *via* the 2′-OH of the P site peptidyl-tRNA (**Figure 15**). In other words, the 2′-OH group of the P site tRNA A76 ribose moiety donates its proton to the adjacent leaving 3′-O of A76 and simultaneously receives a proton from the nucleophilic $-NH_2$ group of the A site aa-

tRNA. The presence of this 2′-OH is crucial for the reaction. Substitution of the 2′-OH by hydrogen (2′-H) or fluor (2′-F) results in at least a 10^6-fold reduction in the rate of peptide bond formation by the ribosome (Weinger et al. 2004). The ribosome plays an important role in the process although it is not involved in chemical catalysis. Certain ribosomal residues (C2063, A2451, and U2584) participate in the construction of the hydrogen bond network that stabilizes the transition state (**Figure 20**). In addition to placing the reactive groups into close proximity and precise orientation to each other, the ribosome appears to work by providing an electrostatic environment that reduces the free energy for forming the highly polar transition state, shielding the reaction against bulk water, helping the proton shuttle forming the leaving group, or a combination of these effects. With this preorganized network, the ribosome avoids the extensive solvent reorganization that is inevitable in the corresponding reaction in solution, resulting in significantly more favorable entropy of activation of the reaction on the ribosome (Beringer and Rodnina 2007).

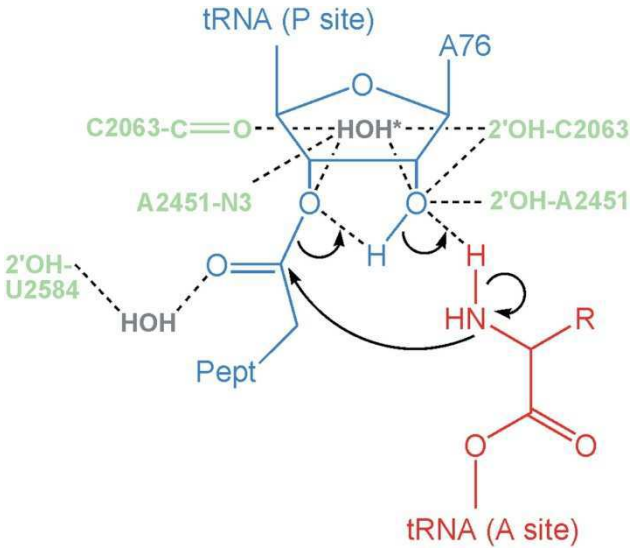


Figure 20: Concerted proton shuttle mechanism of peptide bond formation by the ribosome (Beringer and Rodnina 2007). A six-membered transition state is formed between the P site peptidyl-tRNA (in blue) and the A site aa-tRNA (in red). Ribosomal residues (in green) stabilize the transition state by forming a hydrogen bond network *via* H_2O (in gray).

Although the critical role of aa-tRNAs is playing a central role in protein biosynthesis, aa-tRNAs are also involved in other cellular processes (**Figure 14**)

(RajBhandary and Söll 2008). For example, aa-tRNAs are used as substrates for two enzyme families which catalyze one peptide bond formation: the Fem transferases and the aminoacyl-tRNA-protein transferases. The two enzyme families are respectively involved in cell wall formation and protein labeling for degradation. Next, we will briefly present the Fem transferases then the aminoacyl-tRNA-protein transferases with an emphasis on the L/F transferase.

1.2.1.4 Fem transferases

1.2.1.4.1 Biological implication and function

Cell wall integrity is critical for bacterial survival. Peptidoglycan (murein) is an essential and specific structural element of the bacterial cell wall surrounding the cytoplasmic membrane. It serves both as a barrier to osmotic pressure by providing a mechanical protection against the turgor and a scaffold for attachment of various proteins (Ton-That et al. 1998; Dramsi et al. 2008; Vollmer et al. 2008). It is particularly abundant in Gram-positive bacteria, which possess a very thick cell wall containing around twenty layers of peptidoglycan. The main structural features of peptidoglycan are linear glycan strand cross-linked by short peptides. The glycan strands are made up of alternating *N*-acetylglucosamine (Glc*N*Ac) and *N*-acetylmuramic acid (Mur*N*Ac) residues linked by β-1→4 bonds. The D-lactoyl group of each Mur*N*Ac residue is substituted by a peptide stem whose composition is most often (L-Ala)-(γ-D-Glu)-X-(D-Ala)-(D-Ala) in nascent peptidoglycan, where X is often meso-diaminopimelic acid (DAP), Lys, or ornithine (Orn) depending on the bacterial species (Vollmer et al. 2008; Dare and Ibba 2012). The last D-Ala residue is lost in the mature macromolecule. Cross-linking of the glycan strands generally occurs between the carboxyl group of D-Ala at position 4 and the amino group of the diamino acidat position 3, either directly (most Gram-negative bacteria) or through a short peptide bridge (most Gram-positive bacteria) (Vollmer et al. 2008). The peptide bridge varies from one to seven amino acid residues, which could be the glycine, or certain L- or D-amino acids. Generally, transfer of the glycine and L-amino acids to peptidoglycan precursors are ensured by Fem transferases, a family of nonribosomal peptide bond

forming enzymes utilizing aa-tRNAs as substrates. In a word, Fem transferases transfer the amino acid moiety from aa-tRNA to the pentapeptide of the peptidogylycan precursor to form branched peptide chains that link these precursors to from the peptidoglycan layer (**Figure 21**) (Dare and Ibba 2012). The completion of these peptide bridges is paramount to the structural integrity of the cell wall. Some antibiotics are conceived to damage the integrity of these peptide bridges resulting in inadequately linked peptidoglycan and weak cell walls. Therefore, Fem transferases have became the subject of studies to develop novel antibiotics (Hegde and Shrader 2001). The peptide bridges formed by Fem transferases differ among bacterial species (Hegde and Blanchard 2003; Dare and Ibba 2012) (**Figure 22**). For example, in *Streptomyces aureus*, the peptide bridge is composed of five glycines, which are added by the proteins FemA, FemB, and FemX (Ehlert et al. 1997; Rohrer et al. 1999). FemX catalyzes the addition of the first glycine to the ε–amino group of L-Lys of the peptidoglycan precursor lipid II, followed by the addition of two glycines by FemA, and two more glycines by FemB (**Figure 21** and **Figure 22**). All the three Fem transferases use Gly-tRNA^{Gly} as substrate to donate the glycine. The last step of peptidoglycan biosynthesis is performed by transpeptidase (TP) (**Figure 21** and **Figure 22**).

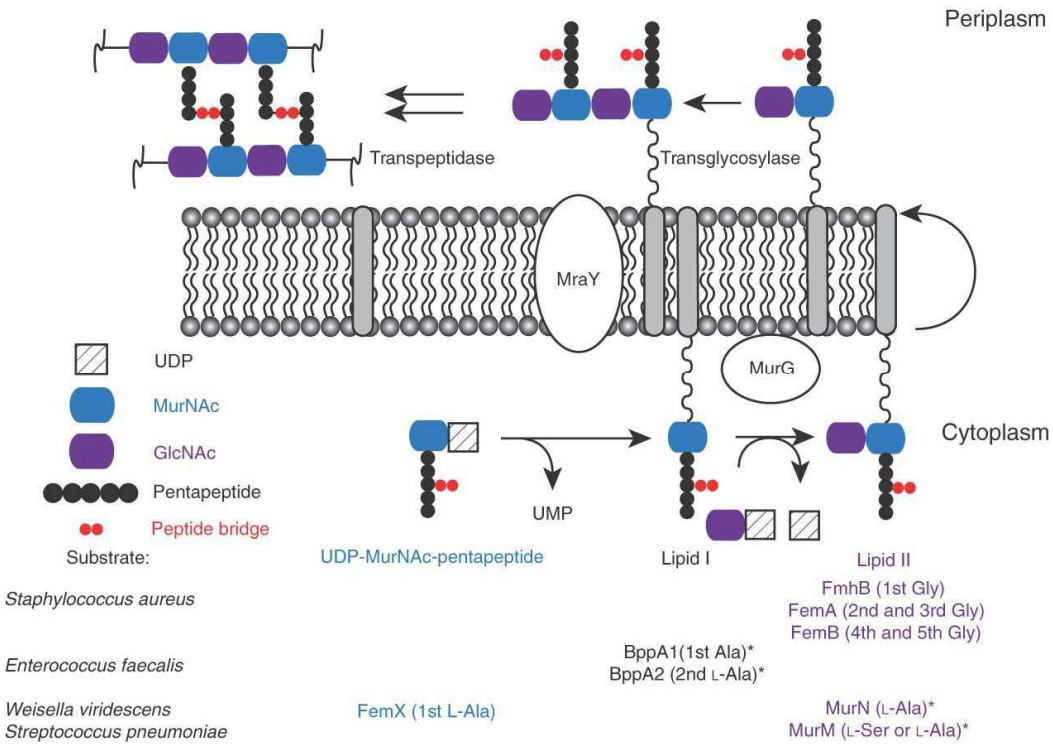
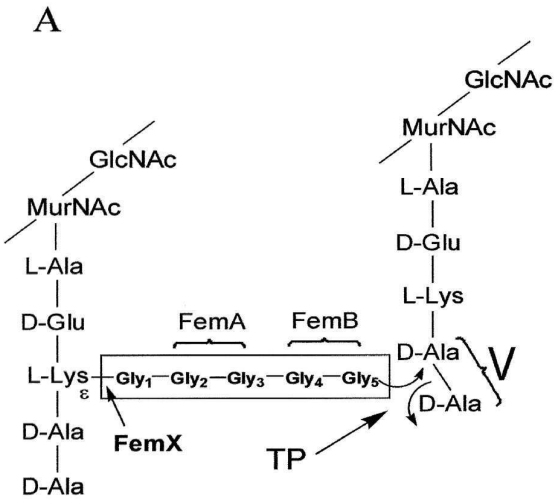


Figure 21: Formation of the peptidoglycan layer in some bacterial species involves peptide bridges formed by Fem tansferases (Dare and Ibba 2012). Shown is a schematic of cell wall biogenesis, below each precursor the amino acid specificity of the Fem enzyme(s) found in the organism listed to the left is indicated in the same color text as the precursor used by the enzyme for amino acid transfer.



B

Species	Diamino acid	Cross-bridge
S. aureus	L-Lys	$(Gly)_5$
S. pneumoniae	L-Lys	$(Ala/Ser)_1$—Ala_2
S. pyogenes	L-Lys	Ala_1—Ala_2
E. faecalis	L-Lys	Ala_1—Ala_2
L. viridescens	L-Lys	Ala_1—Ser_2
L. lactis	L-Lys	D-Asp
E. faecium	L-Lys	D-Asx
B. burgdorferi	L-Orn	Gly
T. palladium	L-Orn	Gly
E. coli	*m*-Dap	None
B. subtilis	*m*-Dap	None

Figure 22: Primary structure of cross-linked peptidoglycan(Hegde and Blanchard 2003). (A) Primary structure of a fragment of *S. aureus* peptidoglycan. *S. aureus* peptidoglycan contains L-Lys in the stem peptide (diamino acid) and a pentaglycine interchain peptide. *S. aureus* FemX transfers Gly_1 from glycyl-tRNA to the ϵ-NH_2 group of Lys followed by the addition of Gly_{2-3} and Gly_{4-5} by FemA and FemB, respectively. The α-NH_2 group of the terminal Gly_5 is cross-linked to the penultimate d-Ala residue (d-Ala_4) of a neighboring pentapeptide releasing the terminal d-Ala residue (*curved arrows*). The cross-linking step is a transpeptidase (*TP*) reaction. (B) Sequence of interchain peptide bridges in different species.

1.2.1.4.2 Structures of Fem transferases

The Fem transferases constitute a family of proteins of 330-450 amino acids that have been divided into two classes depending on the presence or absence of a coiled-coil region of about 60 amino acids. Structures of one member of each class have been determined, FemA from *S. aureus*, and FemX from *W. viridescens* (Benson et al. 2002; Biarrotte-Sorin et al. 2004). *S. aureus* FemA is the first protein structure solved from this family. It consists of an antiparallel coiled-coil domain formed by the helical domain and a globular domain that can be separated into two subdomains (domains 1A and 1B) (**Figure 23A**). *W. viridescens* FemX lacking the antiparallel coiled-coil region consists of two structurally equivalent domains, separated by a cleft. So the two Fem enzymes display a similar structure except the coiled-coil region (**Figure 23**). The fold of each subdomain of the two Fem enzymes displays a structural similarity to that of a large N-acetyltransferase superfamily, the GCN5-related N-acetyltransferase (GNAT) superfamily, which catalyzes the transfer of the acetyl group from acetyl coenzyme A to a primary amine (Neuwald and Landsman 1997) (**Figure 23**). In addition, the complex structure of *W. viridescens* FemX with its substrate UDP-MurNAc-pentapeptide shows that the binding site of the substrate is situated in the clef that separates the two domains (Biarrotte-Sorin et al. 2004).

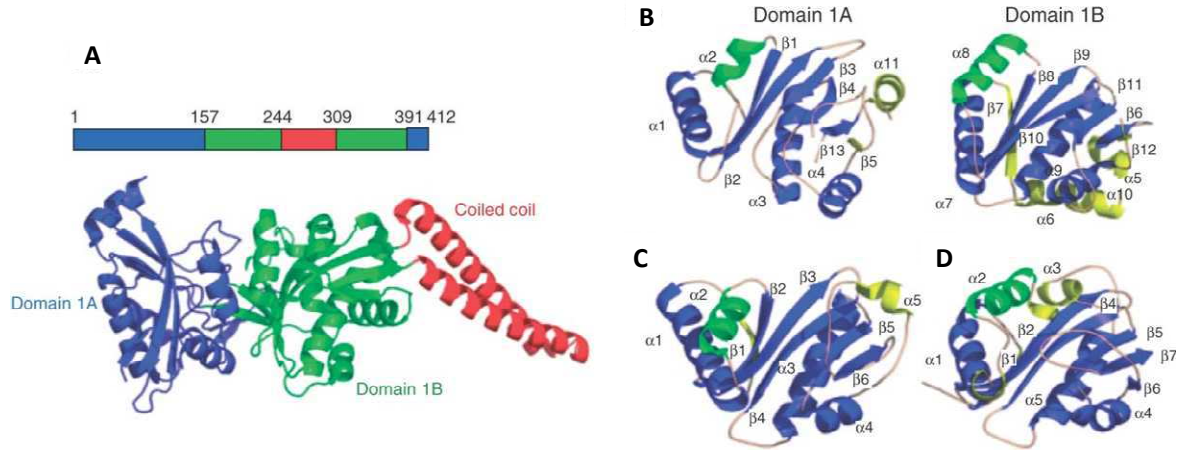


Figure 23: Structural comparisons of *Staphylococcus aureus* FemA and *Weissella viridescens* FemX (Mainardi et al. 2008; Dare and Ibba 2012). (A) Sequence and structure of FemA with an additional coiled coil absent in FemX. (B) Domains 1A and 1B of *W. viridescens* FemX. The two proteins can be structurally compared to the catalytic domains of the histone acetyltransferase of *Tetrahymena thermophila* (C), and the serotonin acetyltransferase of *Ovis ovaries* (D).

1.2.1.4.3 Catalytic mechanism of Fem transferases

Since 2003, there is little progress on kinetic and mechanistic characterization of Fem transferases. The first related work was effectuated on recombinant *Lactobacillus viridescens* (also known as *W. viridescens*) FemX that catalyzes the transfer of L-Ala from Ala-tRNA^{Ala} to UDP-MurNacpentapeptide (UDP-MPP) (Hegde and Shrader 2001; Hegde and Blanchard 2003). In this work, it was suggested that FemX use a sequential kinetic mechanism where both substrates (UDP-MPP and Ala-tRNA) must be bound to the enzyme for catalysis to occur (**Figure 24**) (Hegde and Blanchard 2003). In addition, the authors proposed a model for the reaction catalyzed by FemX that includes the acid/base assistance (**Figure 24**) (Hegde and Blanchard 2003). UDP-MPP binds to the free enzyme followed by Ala-tRNA. A ternary complex is consequently formed among UDP-MPP, Ala-tRNA and FemX. In the ternary complex, nucleophilic attack on the carbonyl of Ala-tRNA generates the zwitterionic, tetrahedral intermediate. A general acid participates in the catalysis by assisting in protonating the 3′-OH of the tRNA ribose moiety and a general base deprotonates the positively charged amine. Ala is thereby transferred to the ε–amino group of L-Lys of UDP-MPP with the help of the general acid and base (Hegde and Blanchard 2003).

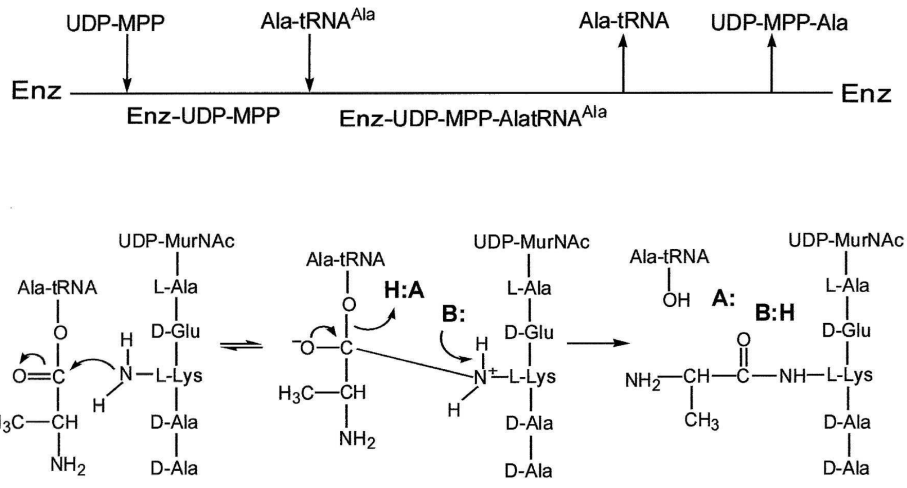


Figure 24: Proposed kinetic and chemical mechanisms for Ala transfer activity by *W. viridescens* FemX (Hegde and Blanchard 2003).

1.2.1.5 Aminoacyl-tRNA protein transferases

Aminoacyl-tRNA protein tranferases are ubiquitous proteins involved in an essential cellular process: regulated protein degradation.

1.2.1.5.1 Biological implication and function

Intracellular protein degradation plays an essential role in many physiological processes by removing either damaged polypeptides or proteins that harbor specific destruction tags. No matter in eukaryotes or in prokaryotes, the protein degradation is regulated by the N-end rule pathway. The N-end rule defines the stability of proteins according to the nature of their N-terminal residues (Mogk et al. 2007). Unstable proteins harbor specific degradation signals, termed degrons, which are recognized by components of the proteolytic systems and subsequently delivered to hydrolyzing proteases. In different organisms, the generation of N-degrons and the composed residues show distinct differences (Mogk et al. 2007). One of the strategies to create functional N-degrons concerns the aminoacyl-tRNA protein transferases, which catalyze the transfer of destabilizing residues to the N-terminal of the proteins to be degraded. These destabilizing residues are recognized as destruction tags and thereby provoke the protein degradation. For example, in the majority of prokaryotes, such as *E. coli*, the primary destabilizing residues (Phe, Leu) are conjugated to the second destabilizing N-terminal residue Arg (or Lys) of unstable proteins by a leucyl/phenylalanyl-tRNA protein transferase (L/F transferase) which uses Leu-$tRNA^{Leu}$ or Phe-$tRNA^{Phe}$ as substrate (Tobias et al. 1991). L/F tranferase is widely studied in the last decades, including its tridimensional structure.

1.2.1.5.2 Structure of L/F transferase

The crystal structure of *E. coli* L/F transferase was separately resolved in 2006 (Suto et al. 2006; Dong et al. 2007). It is the first resolved structure of the aminoacyl-tRNA protein transferase family. L/F transferase forms a compact structure and consists of two domains: an NH_2-terminal domain of 60 residues and a COOH-terminal domain of 170 residues (**Figure 25**) (Suto et al. 2006). The COOH-terminal

domain consists of the GCN5-related N-acetyltransferase fold thus presents a structural similarity to domain 2 of *W. viridescens* FemX (Biarrotte-Sorin et al. 2004) and that of *S. aureus* FemA (Benson et al. 2002) (**Figure 25C**). The topological similarity may reflect the similarity of their chemical reactions, where an amino acid is transferred from an aminoacyl-tRNA to the amino group of a protein (or peptide). Nevertheless, there is no significant amino acid similarity between them, suggesting that they might have arisen from a common ancestor, but have divergently evolved (Suto et al. 2006). In contrast, no significant topological similarity is identified between the NH_2-terminal domain of L/F transferase and the domain 1 of FemX and FemA (**Figure 25C**) (Suto et al. 2006).

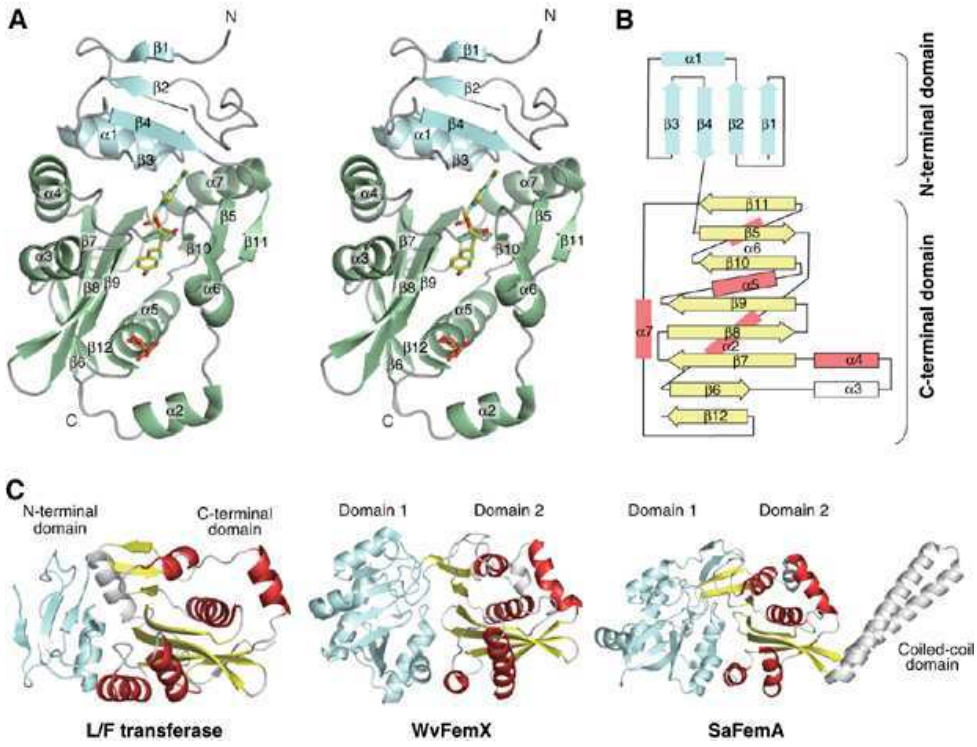


Figure 25: Overall architecture of *E. coli* L/F-transferase (Suto et al. 2006). (A) Stereo view of the *E. coli* L/F-transferase structure. The NH_2-terminal domain (residues 2–62) and the COOH-terminal domain (residues 63–232) are colored blue and green, respectively. The puromycin (analog of the 3'-terminus of an aa-tRNA) bound to the hydrophobic pocket is colored yellow. (B) Topology diagram of L/F-transferase. The rimmed elements in the COOH-terminal domain (α3–α5) and (β5–β12) are common to the GNAT superfamily fold. The α-helices and β-strands in the COOH-terminal domains are colored red and yellow, respectively. (C) Comparison of the structures of *E. coli* L/F-transferase (left), *W. viridescens* FemX (*wv*FemX; middle) and *S. aureus* FemA (*sa*FemA; right). The COOH-terminal domain of L/F-transferase is topologically similar to the domain 2's of *wv*FemX and *sa*FemA. The conserved α-helices and β-strands in L/F-transferase, *wv*FemX and *sa*FemA, are colored red and yellow, respectively.

1.2.1.5.3 Recognition of its substrate aminoacyl-tRNA by L/F transferase

As what is previously described, L/F transferase catalyzes peptide bond formation by using Leu-tRNA^{Leu} (or Phe-tRNA^{Phe}) and a protein bearing an N-terminal Arg (or Lys) as donor and acceptor substrates, respectively. But how does L/F transferase recognize its substrate aa-tRNA? The structure of its complex with a substrate aa-tRNA, combined with site-directed mutagenesis study, could help us unveil the recognition mechanism.

The first complex structure of L/F transferase with an analog of the 3'-terminus of an aa-tRNA, puromycin, was published in 2006 (**Figure 25A** and **Figure 26**) (Suto et al. 2006). The chemical structure of puromycin is similar to that of the 3'-teminus of an aa-tRNA; the carboxyl group of *p*-methoxyphenylalanine is linked to the 3'-amino group of 3'-amino-6-*N,N*-dimethyladenosine by an amide bond (**Figure 26A**). The *p*-methoxybenzyl group and the 6-*N,N*-dimethyladenosine correspond to the side chain of an amino acid and the adenosine at the CCA end of an aa-tRNA, respectively. Puromycin reportedly inhibits the activity of L/F transferase by preventing the aa-tRNA from binding to the enzyme suggesting that puromycin binds to the same position in L/F transferase as the 3'-end of an aa-tRNA (Horinishi et al. 1975; Abramochkin and Shrader 1996). The complex structure shows that the *p*-methoxybenzyl group of puromycin is docked within a deep pocket of L/F transferase composed of the side chains of several hydrophobic residues. Thus, the *p*-methoxybenzyl group of puromycin should be recognized by L/F transferase through a hydrophobic interaction with these hydrophobic residues. The reported inhibitory effect by puromycin on L/F transferase activity might reflect the similar hydrophobic properties of the *p*-methoxybenzyl group of puromycin to the side chains of phenylalanine and leucine (**Figure 26A**). In addition to a series of structure-based site-directed mutations, it was assumed that the hydrophobic leucyl and phenylalanyl moieties of Leu-tRNA^{Leu} and Phe-tRNA^{Phe}, respectively, are recognized by the highly hydrophobic pocket of L/F transferase (Suto et al. 2006). Furthermore, this hydrophobic pocket is well conserved among L/F transferases of different organisms. All these results explain why L/F transferase recognizes aa-tRNAs

attached to hydrophobic amino acids, and excludes those coupled to hydrophilic or charged amino acids. Besides the hydrophobicity, the size and the shape of the pocket also restrict the nature of bound amino acid. Taken all these together, only Leu and Phe can be perfectly accommodated in the pocket, although Trp and Met can barely be docked in the pocket. This is consistent with that Met and Trp can be transferred to proteins bearing an N-terminal Arg or Lys by L/F transferase but much less efficiently than Leu and Phe (Abramochkin and Shrader 1996). In addition to the hydrophobic aminoacyl moiety recognizing pocket, there is a 3′-nucleotide recognition site. The present structural and biochemical studies of L/F transferase strongly suggest that the 3′-terminus of the aa-tRNA is recognized by the combination of the hydrophobic aminoacyl moiety recognizing pocket and the 3′-nucleotide recognition sitem although the 3′-nucleotide binding site is not specific for the adenosine (Suto et al. 2006). Furthermore, the electrostatic potential surface of L/F transferase reveals the highly biased distribution of charged residues. A cluster of positively charged residues are present and protrude toward the solvent. These basic residues are conserved among the eubacterial L/F transferase. Mutations of these residues remarkably reduced the L/F transferase activity. Thus, according to a tRNA docking model on L/F transferase, the tRNA moiety of the substrate aa-tRNA was supposed to interact with the basic patch (**Figure 27**) (Suto et al. 2006). Concerning the aa-tRNA specificity of L/F transferase, studies suggest that it is determined by the aminoacyl moiety attached to the 3′-terminus of the tRNA, rather than by the tRNA itself (Suto et al. 2006).

Afterwards, in 2007, the crystal structure of *E. coli* L/F transferase complex with phenylalanyl adenosine (rA-Phe) was published (Watanabe et al. 2007). rA-Phe was found located at almost the same position in the hydrophobic pocket of L/F transferase as puromycin, an aa-tRNA analog described previously (Horinishi et al. 1975; Abramochkin and Shrader 1996; Watanabe et al. 2007). The new complex structure and a series of site-directed mutagenesis study confirmed the recognition mode of aa-tRNA by L/F transferase previously proposed by Suto et al (Suto et al.

2006). Besides the new complex structure, Watanabe et al. also proposed a catalytic mechanism of L/F transferase in this work.

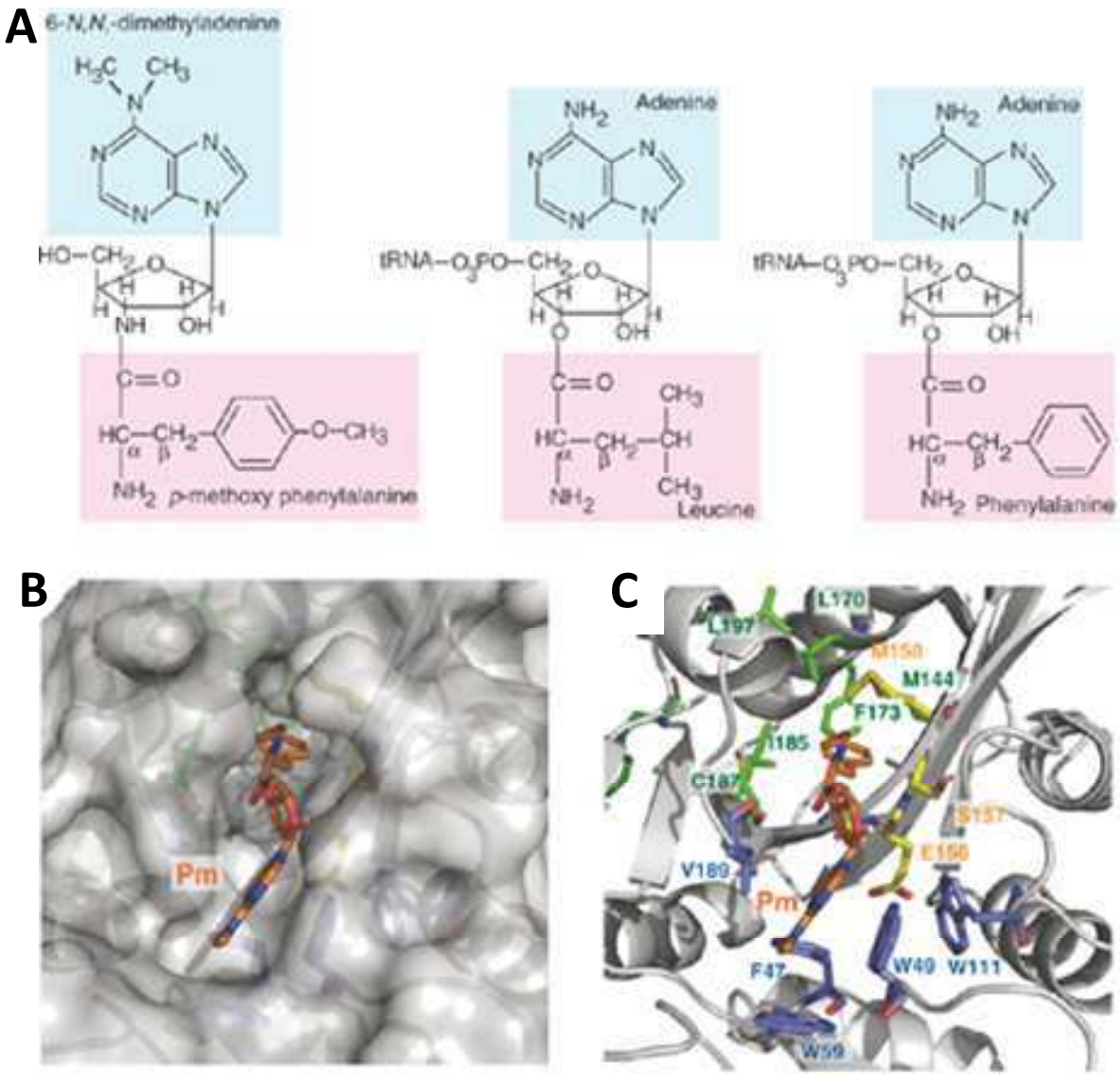


Figure 26: Recognition of the puromycin by *E. coli* L/F transferase (Suto et al. 2006). (A) Chemical structure of puromycin (left) and that of the 3'-ends of Leu-tRNA^{Leu} and Phe-tRNA^{Phe} (middle and right, respectively). The amino-acid moiety and the base moiety are colored pink and blue, respectively. (B) Recognition of the *p*-methoxybenzyl group and the puromycin base by the hydrophobic pocket, as shown by a surface model. (C) Ribbon model of (B).

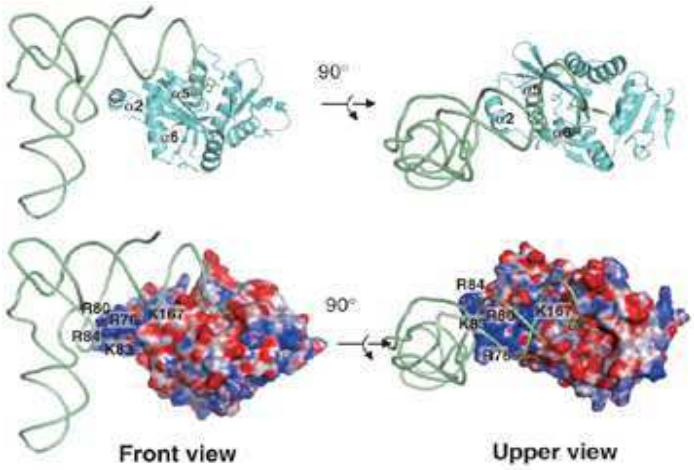


Figure 27: Model of aminoacyl-tRNA binding to L/F transferase (Suto et al. 2006). Two views of a ribbon diagram of the docking model of L/F transferase and tRNA. The tRNA backbone is shown as a green line (upper two panels). Two views of the L/F transferase-tRNA complex model, showing the surface colored according to its calculated electrostatic potential (lower panel; blue, positively charged +8KT; red, negatively charged –8KT). tRNA interacts with a positively charged cluster composed of basic residues (blue) of helice α2.

1.2.1.5.4 Catalytic mechanism of L/F transferase

When the crystals of L/F transferase were soaked in a solution containing its minimal substrates: rA-Phe and a short peptide bearing an N-terminal Arg, which respectively mimic the aa-tRNA and the target protein bearing an N-termianal Arg, the reaction took place in the crystal and the product peptide bearing an N-terminal Phe was observed in the active site of L/F transferase (Watanabe et al. 2007). Comparing the structure before the reaction, that is L/F transferase in complex with the minimal substrates, and the structure after the reaction, that is L/F transferase in complex with the product, acatalytic mechanism of L/F transferase was proposed (**Figure 28A**) (Watanabe et al. 2007). Before the reaction, two hydrogen bonds respectively form between the α–amino group of the N-terminal Arg of the target protein and the Oε atom of Gln 188, and between the Nε atom of Gln 188 and the Oε atom of Asp 186. Therefore, Asp 186 helps Gln 188 attract a proton from the adjacent α–amino group of the N-terminal Arg. This hydrogen bond network consequently facilitates the nucleophilic attack of the α–amino group of the N-terminal Arg on the carbonyl carbon of the aa-tRNA. Meanwhile, another hydrogen bond forms between Nδ atom of Asn 191 and the carbonyl oxygen of aa-tRNA, which may increase the

polarity of the carbonyl oxygen thus enhance the electrophilicity of the carbonyl carbon of aa-tRNA for attack by the nucleophile. Moreover, this hydrogen bond stabilizes the tetrahedral intermediate of the oxyanion during the peptide bond formation by L/F transferase. Finally, the proton abstracted from the α–amino group by Gln188 is transferred to the 3'-O of the tRNA, which is liberated with the completion of peptide bond formation.In addition, biochemical studies agree with the proposed catalytic mechanism of peptide bond formation by L/F transferase (Watanabe et al. 2007).

The proposed catalytic mechanism of peptide bond formation by L/F transferase is analogous to the reverse reaction of the acylation step observed in the peptide hydrolysis reaction by serine proteases, such as chymotrypsin (Blow et al. 1969; Steitz and Shulman 1982) (**Figure 28B**). In the model for the catalytic mechanism of L/F transferase, the aa-tRNA corresponds to the acyl-Ser 195 of chymotrypsin, and Gln 188 and Asp 186 of L/F transferase correspond to His 57 and Asp 102 of chymotrypsin. The role of the Nδ atom of Asn 191 of L/F transferase corresponds to that of the main chain of the amide of Ser 195 and Gly 193 of chymotrypsin, which stabilizes the tetrahedral intermediate (Watanabe et al. 2007).

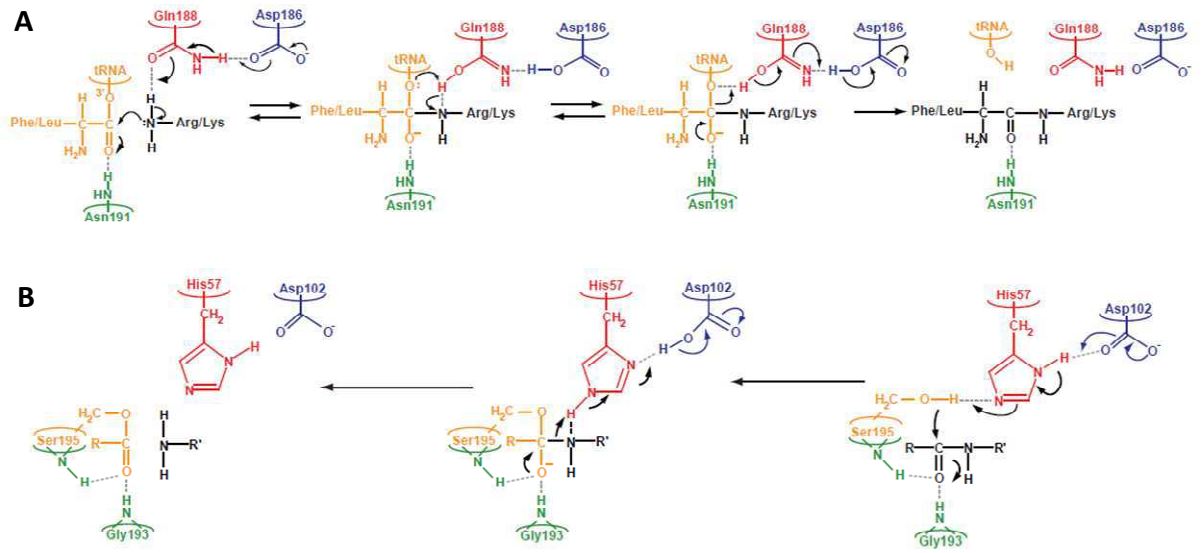


Figure 28: Comparison of the catalytic mechanism for peptide bond formation by L/F-transferase (A) with the reverse reaction of the acylation step during peptide hydrolysis by chymotrypsin (B) (Watanabe et al. 2007).

However, a recent study on catalytic mechanism of L/F transferase does not agree with the mechanism previously described. The results demonstrate that the roles of Asp 186 and Gln 188 are just for proper substrate binding and orientation, and are not directly involved in the chemistry of the reaction as previously proposed by Watanabe et al. (Fung et al. 2011).

We just illustrate, through examples of the ribosomal machinery, of Fem transferases and aa-tRNA protein transferases, the implication of aa-tRNAs in cellular processes. Besides these primary processes, aa-tRNAs are also involved in secondary metabolism (von Dohren 2009). This can be illustrated by the CDPS family (Gondry et al. 2009) and also a new group of tRNA-dependent peptide bond-forming enzymes, represented by the transferase PacB, involved in the biosynthesis of the antibiotics "pacidamycins" (Zhang et al. 2011). VlmA, identified from *Streptomyces viridifaciens*, is another tRNA-dependent peptide bond forming enzyme involved in the biosynthesis of the antibiotic valanimycin (Garg et al. 2008).

1.2.1.6 Cyclodipeptide synthases (CDPSs)

In the introduction of the manuscript, we have described the primary characterization of CDPSs. My thesis work consists of deep studies on CDPSs including the molecular bases of the substrate specificity, the catalytic mechanism and the kinetics of CDPSs. They will be described in the result chapters (**Chapters 2-5**).

1.2.1.7 Other tRNA-dependent peptide bond-forming enzymes involved in the secondary metabolism: example of PacB

Besides in the biosynthesis of cyclodipeptides by CDPSs, aa-tRNAs are also involved in other biosynthetic pathways of secondary metabolites, such as some antibiotics. Taking for example, we will briefly introduce a new group of tRNA-depending peptide bond-forming enzymes, named PacB, in biosynthesis of pacidamycins.

Pacidiamycins are a family of uridyl tetra/pentapeptide antibiotics produced by *streptomyces coeruleorubidus* with antipseudomonal activities through inhibition of the translocase Mra Y in bacterial cell wall assembly (Chen et al. 1989). They are reported to exert antibiotic activities by virtue of functioning as a substrate analog of the UDP-MurNAc-pendapeptide of Mra Y in bacterial cell wall assembly of the pentapeptidyl-bactoprenaol intermediate (Boojamra et al. 2001; Winn et al. 2010). More than 10 related family members have been reported (Chen et al. 1989; Fronko et al. 2000). Their common scaffold contains a central N_{β}-methyl *2S,3S*-diaminoburytic acid (DABA) moiety which is α-amino-capped by a ureido dipeptide (C-terminal), β-amino-capped by a single amino acid or a dipeptide (N-terminal), and carboxy-linked to a 3′-deoxy-4′,5′-enaminouridine (**Figure 29**). Pacidamycins are of interest due to their unusual peptidyl-nucleoside structure features, and for the development of next generation antibacterial drugs inhibiting the clinically underexplored cell wall enzyme target MraY (Zhang et al. 2011). Studies revealed that the 5′-aminouridyl-tetrapeptide framework of the pacidamycin family is synthesized by nine NRPS enzymes (Zhang et al. 2011) (**Figure 29**). But what is the biosynthetic route to pentapeptides from that set of tetrapeptide scaffolds? In other words, how is the alanine added to the N-terminal of the uridyl tetrapeptide? Recently, a NRPS-unrelated enzyme, PacB, coded by the gene *pacB* was demonstrated to be responsible of this reaction because the deletion of *pacB* abolished the production of all of the pentapeptide compounds *in vivo*, while uridyl tetrapeptides were produced (Zhang et al. 2011). In addition, *in vitro* experiments showed that PacB catalyzed transfer of the alanyl residue to the N-terminal of the uridyl tetrapeptide on NRPS assembly line, as the free uridyl tetrapeptide was not converted to the uridyl pentapeptide by PacB (Zhang et al. 2011). Moreover, PacB was found prefer Ala-tRNA as a donor over other activated alanyl species in transferring the alanyl residue to the NRPS assembly line. Thus, PacB is a tRNA-dependent aminoacyltransferase involved in peptide bond formation in secondary metabolise, specifically in the addition of an N-terminal alanyl residue to a tetrapeptidyl intermediate in the highly dissociated NRPS assembly line in pacidamycin biosynthesis (**Figure 30**) (Zhang et al. 2011).

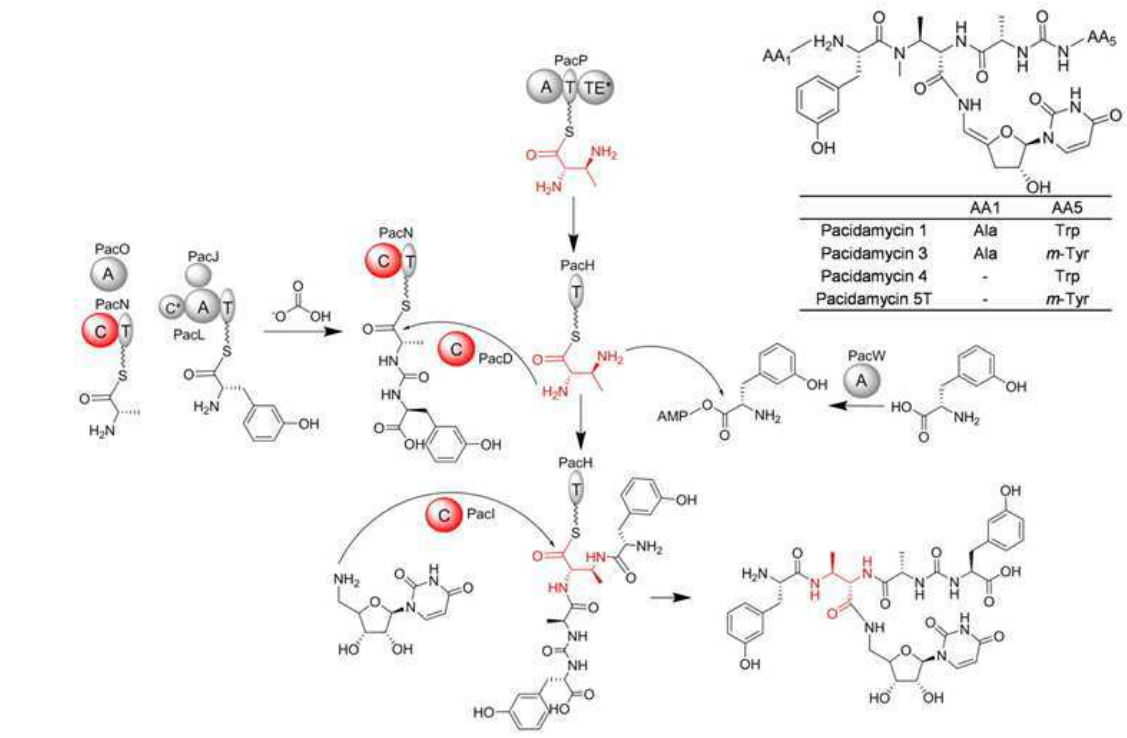


Figure 29: Structures of selected pacidamycins, and biosynthetic pathway for uridyl tetrapeptide using nine NRPS enzymes (Zhang et al. 2011). Domain notation: T, thiolation; A, adenylation; C, condensation; TE, thioesterase. The diamine spacer DABA is shown in red.

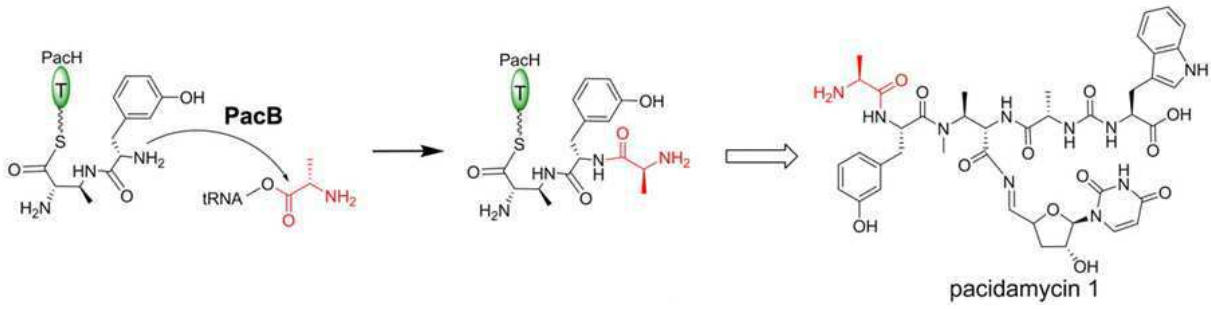


Figure 30: peptide bond-forming reaction catalyzed by tRNA-dependent enzyme PacB in the biosynthesis of pacidamycin 1 (Zhang et al. 2011). The alanyl residue transferred from Ala-tRNA is shown in red. PacH is a NRPS enzyme involved in the catalysis.

Regarding the structure of PacB, bioinformatic analysis using the program HHpred revealed structural homology to *W. viridescens* FemX (Biarrotte-Sorin et al. 2004) and *S. aureus* FemA (Benson et al. 2002). As previously described in this thesis, these Fem transferases catalyze the addition of amino acid (FemX: Ala; FemA: Gly) on the peptidoglycan precursor using an aa-tRNA as a substrate for subsequent

peptide bridge cross-linking to strengthen the cell wall (Hegde and Shrader 2001). Thus, in term of amino acid addition to a peptidyl chain, PacB also shows functional homologies to the Fem transferases. According to the structural model of PacB generated based on the highest scoring template FemX, PacB is presumed to have two domains separated by a cleft. One domain is predicted to have a three-dimensional fold similar to that of the GCN5-like *N*-acetyltransferase superfamily (Zhang et al. 2011) (**Figure 31**).

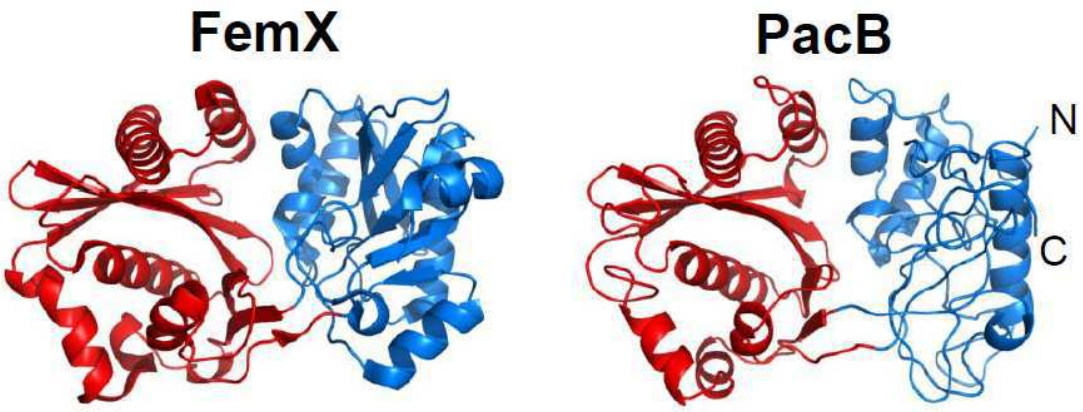


Figure 31: Three-dimensional structural model of PacB in comparison to the reported structure of FemX (PDB ID 3GKR) (Zhang et al. 2011). Domain 1 is represented in blue and domain 2 in red.

1.2.2 tRNA-independent peptide bond-forming enzymes

tRNA-independent peptide bond-forming enzymes have been identified to catalyze the nonribosomal peptide formation. These peptides generally show interesting biological and pharmacologic activities. Therefore, more and more attentions are being paid to their synthetic pathways. Among these enzymes, NRPS family is the most important class. This part will be dedicated to the description of this enzyme family as well as other relevant enzymes by taking glutathione synthetase as an example.

1.2.2.1 Nonribosomal peptide synthetases (NRPSs)

Nonribosomally synthesized microbial peptides show remarkable structural diversity and constitute a widespread class of the most potent antibiotics and other

important pharmaceuticals that range from penicillin to the immunosuppressant cyclosporine (Walsh 2007). They are generally produced in the secondary metabolism of bacteria and fungi by the consecutive condensation of amino acids, which is achieved by a group of multimodular megaenzymes called nonribosomal peptide synthetases (NRPSs) (Finking and Marahiel 2004; Walsh 2007). We previously presented that several DKPs were synthesized by NRPSs (section **1.1.2.2.1**). However, nonribosomal peptides synthesized by NRPSs are far more than this. They are generally polypeptides with more than two peptide bonds. In addition, NRPSs relay not only on the 20 canonical amino acids, but also use several different building blocks, including nonproteinogenic amino acids (Caboche et al. 2008). The nonproteinogenic building blocks contribute to structural versatility of nonribosomal peptides and are likely to contribute substantially to the observed biological activity (**Figure 32**) (Walsh 2007).

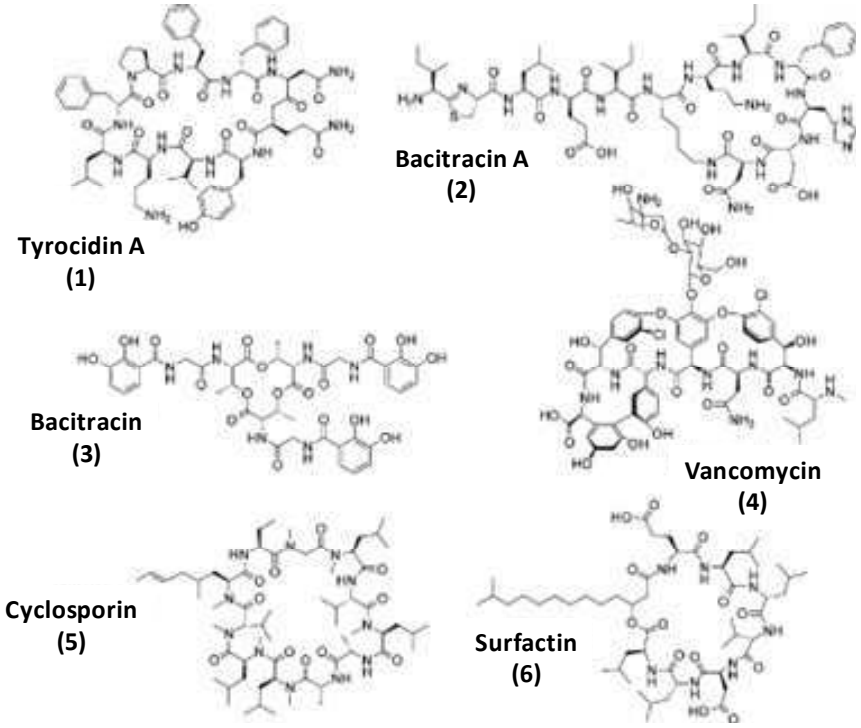


Figure 32: Examples of non-ribosomally synthesized bioactive macrocyclic peptides that comprise unusual structures (D-configured and *N*-methylated amino acids, heterocyclic rings, fatty acids, sugars and non-proteinogenic amino acids) and exhibit antibiotic activity (1, 2, 4), act as siderophore (3), biosurfactant (6), or exhibit immunosuppressive activity (5) (Marahiel 2009).

Next, we will describe NRPS assembly line, their structure, and their catalytic mechanism.

1.2.2.1.1 Assembly logic of NRPSs

In principle, NRPSs constitute an array of distinct modular sections, each of which is responsible for incorporation and if necessary modification of one defined monomer into the final product (Konz and Marahiel 1999; Sieber and Marahiel 2005; Caboche et al. 2008). As a consequence, in linear NRPS assembly lines, the number of the catalytic modules and their order exactly matches the number and order of amino acids incorporated in the backbone of the final peptide product (Marahiel 2009). Precisely, the identity and order of a module in an assembly line specifies: first, the sequence of monomer units activated and incorporated; second, the chemistry that occurs at each way station in the assembly line; third, the length and functionality of the product released from the distal end of the assembly line (Fischbach and Walsh 2006) (**Figure 33**). The modules can be divided into catalytic domains, each responsible for a specific synthetic step during peptide synthesis (Mootz et al. 2002; Schwarzer et al. 2003; Felnagle et al. 2008). In each so-called elongation module, three domains are ubiquitous in nonribosomal peptide synthesis and essential for peptide elongation. The domains are responsible for the activation of the amino acid (adenylation (A) domains, ~500 amino acids), the propagation of the growing peptide chain (thiolation or peptidyl carrier protein (PCP) domains, 80-100 amino acids), and the condensation of the amino acids (condensation (C) domains, ~450 amino acids). A fourth essential NRPS catalytic unit associated with product release is the thioesterase (TE) domain (~280 amino acids). The TE domain is located in the termination module and catalyzes peptide release by either hydrolysis or macrocyclization. In addition, some modification domains could be integrated into NRPS modules at different locations in order to modify amino acids incorporated, such as epimerization (E) domains and *N*-methyltransferase domains (Mt) catalyzing respectively the generation of D-configurated and methylated amino acids (**Figure 33**) (Kopp and Marahiel 2007; Marahiel 2009; Strieker et al. 2010).

However, during the past few years, it has become clear that modules from acterial megasynthetases can in fact operate iteratively, resulting in elongated products or, in contrast to this, can be skipped during the biosynthetic process to yield shortened products. Reports on multimodular assembly lines not following the co-linearity rules are emerging, as well as descriptions of deviations from the standard modular architecture. This reports illustrate that NRPSs have a much greater diversity in their biosynthetic potential than originally expected (Wenzel and Müller 2005).

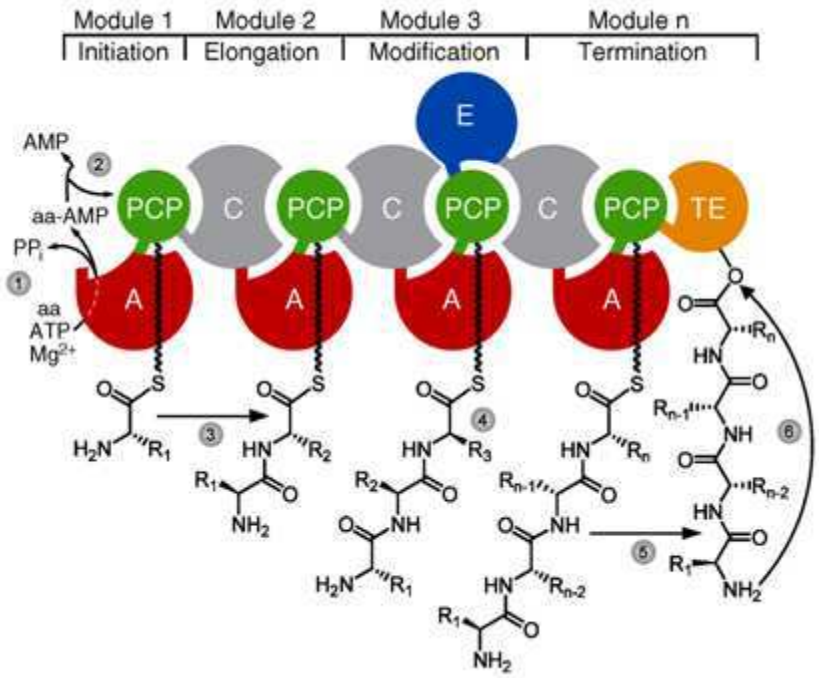


Figure 33: Simplified mechanism of nonribosomal peptide synthesis (Strieker et al. 2010). (1) The amino acid is activated as aminoacyl-AMP by the adenylation domain. (2) Transfer of the amino acid onto the PCP domain. (3) Condensation of PCP-bound amino acids. (4) Possibility of amino acid modifications, for example by epimerization domains. (5) Transesterification of the peptide chain from the terminal PCP onto the TE domain. (6) TE catalyzed product release by either hydrolysis or macrocyclization. The number of modification domains and modules is very variable.

Structures of the NRPS domains are necessary to be described in order to clarify the catalytic mechanism.

1.2.2.1.2 NRPS structures

So far, the structures of the four essential NRPS domains described above had been subsequently determined. The structural information provides profound insights into the catalytic mechanisms and dynamics of these multidomain enzymes.

1.2.2.1.2.1 The adenylation (A) domain

The A domains catalyze reactions concerning ATP-dependent amino acid activation and subsequent transfer to the PCP domain. The first A domain structure, Phe A, was published in 1997 (Conti et al. 1997). Each A domain comprises a large N-terminal core domain (~450 amino acids) and a small C-terminal subdomain (~100 amino acids), which are connected by a small 5-10 residues hinge, with a hydrophobic active site in between (**Figure 34**) (Marahiel 2009; Strieker et al. 2010). The A domains show a significant structural similarity to several members of the adenylate-forming-superfamily, which include in addition to NRPS A domains the firefly luciferase, acetyl CoA synthetase, and yeast 4-chlorobenzoate: CoA ligase (Marahiel 2009). The core and subdomain organization of A domains is highly conserved even in other adenylate generating enzymes just mentioned. It was reported that an A domain adopted two different states, open and closed, during different catalytic stage. During the catalytic cycle (substrate and ATP binding followed by acyl-AMP formation) the small C-terminal subdomain can adopt different orientations relative to the large N-terminal subdomain (Yonus et al. 2008). This motion ranges from 140° rotation in the open state during substrate loading (ATP and amino acid) to a closed state that promotes the adenylation reactions (**Figure 34**). Also, extensive biochemical studies combined with structural and phylogenetic studies on A domains revealed the so-called "specificity-conferring code" of A domains (Stachelhaus et al. 1999). This code is composed of 10 amino acid residues, which are responsible for substrate binding within the active site of A domains. This code can be easily identified and extracted out of the sequence of any A domain and used for *in silico* studies to make a good prediction on the possible

substrate without prior biochemical studies. Different online tools for A domain specificity prediction are available (Rausch et al. ; Weber et al. 2009).

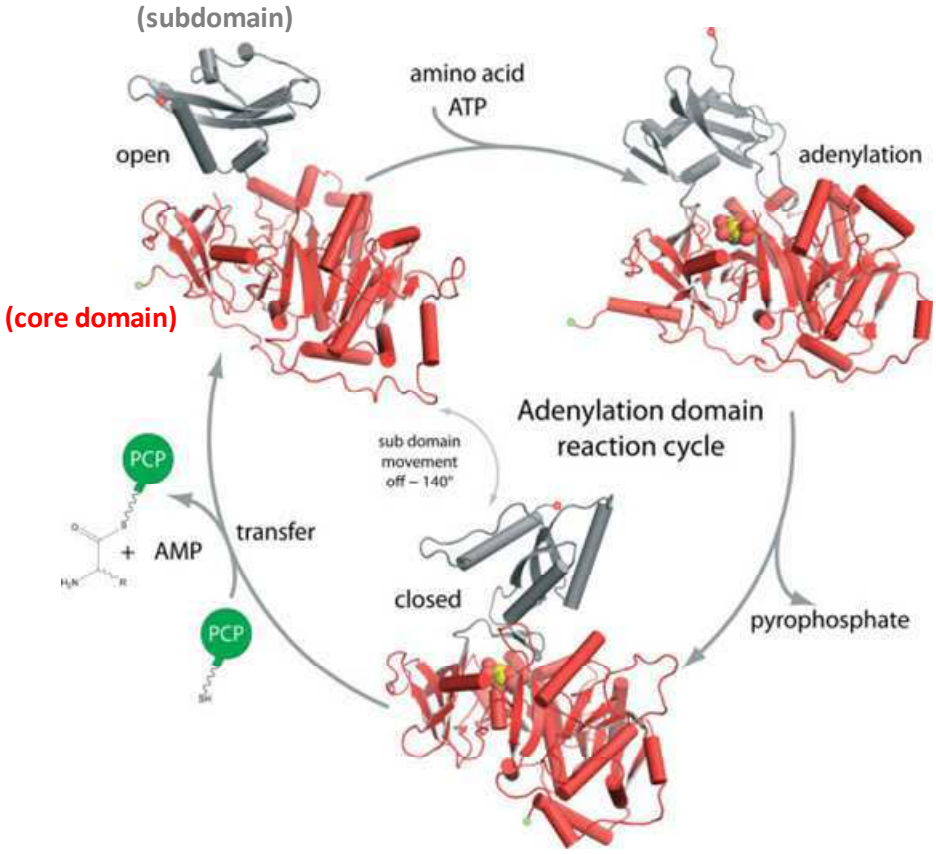


Figure 34: An proposed reaction cycle of an A domain showing the conformational movement of the small *C*-terminal subdomain (gray) compared to the large *N*-terminal core domain (red) upon ATP and substrate amino acid binding (Marahiel 2009). During this initial reaction of the A domain, aminoacyl-AMP formation and pyrophosphate release, the *C*-terminal subdomain rotates by some 140° from an open into a close conformation. In the second half reaction, through a specific PCP/A interaction, the activated substrate is then transferred to the reactive thiol group of the cofactor bound to the carrier domain PCP (green).

1.2.2.1.2.2 The peptidyl carrier protein (PCP) domain

The PCP domain is a small protein containing the binding site of the cofactor 4'-phospopantetheine (ppant), derived from CoA. Its highly conserved serine residue (GGXS-motif) is post-translationnally modified by the 20 Å "swinging arm" ppant, which is carried out by ppant-transferases such as the promiscuous *Bacillus subtilis* ppant-transferase Sfp (Reuter et al. 1999; Strieker et al. 2010). The ppant offers a thiol group to covalently bind its substrate and synthetic intermediates in the form of acyl-

S-PCP (**Figure 33**). Like other members of the carrier protein superfamily, PCP shows a simple four helix bundle structure. A large loop region (10-15 amino acids) located between helix 1 and helix 2 contains the conserved serine residue to which the ppant-cofactor is attached (**Figure 35**) (Weber et al. 2000; Marahiel 2009). Recent studies revealed that PCP forms three different and slowly interconverting conformations, that is the A (apo), H (holo), and A/H states, depending on its chemical modification states (**Figure 35**) (Koglin et al. 2006). It was shown that ppant-bound PCP domain could be generated only with the A-state PCP.

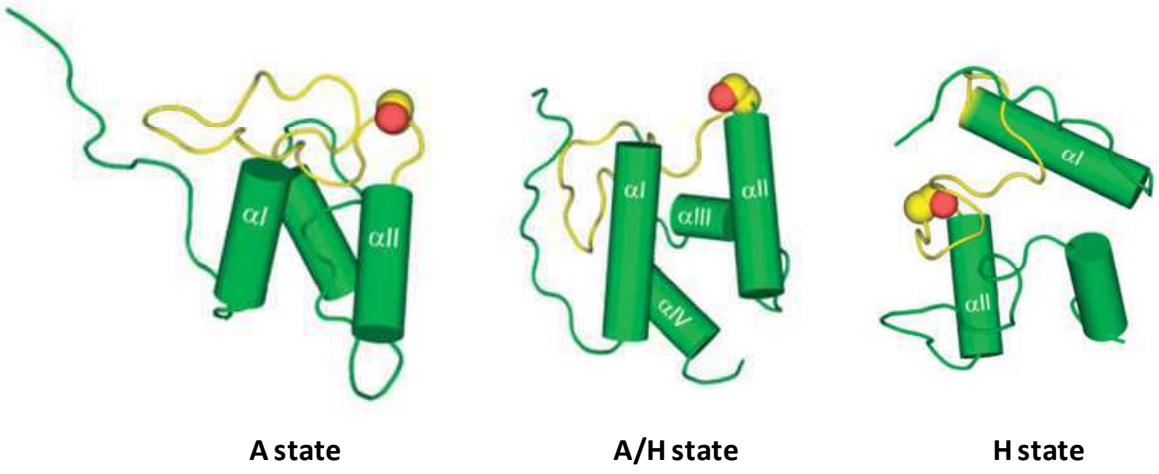


Figure 35: NMR structures of the PCP conformers (A, A/H, and H states) (Marahiel 2009).

1.2.2.1.2.3 The condensation (C) domain

The C domains are the control entities of NRPS elongation (C-A-PCP) modules as they catalyze the peptide bond formation between two adjacent PCP molecules loaded with their respective amino acids (Linne and Marahiel 2000). An aminoacyl-S-PCP activated amino acid is bound to the acceptor site of the C domain, whereas its donor site accommodates the incoming peptidyl-S-PCP substrate. The overall architecture of the C domain revealed a V-shaped form with a canyon-like structure in which the two PCP-bound substrates can be positioned from both sides (acceptor and donor sites) in close proximity to the active site histidine located in the conserved HHxxxDG core motif (**Figure 36A**). The V-shape C domain structure

shows two similar N- and C-terminal subdomain folds that both belong to the well-known chloramphenicol acetyltransferase (CAT) fold. In the C domain structure the two CAT-folds build the canyon-like active site with the HHxxxDG-motif located on the middle of its floor. In the PCP-C didomain structure (**Figure 36A**), the PCP domain is attached to the donor site of the C domain and has an A/H conformation. The flexible linker region between PCP and C is 18 residues in length and show little interaction with both folds (Marahiel 2009).

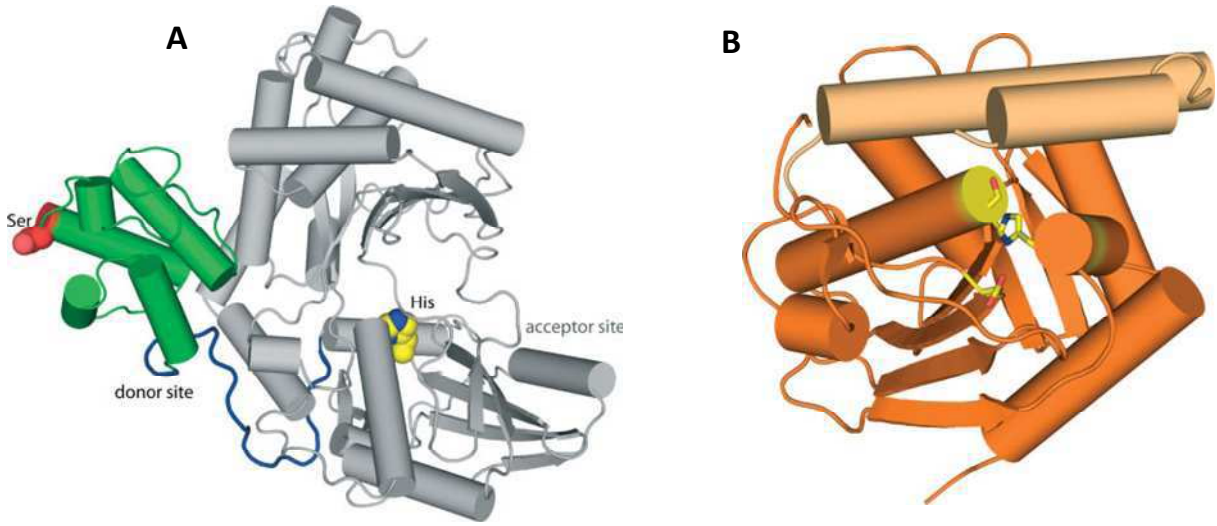


Figure 36: Structures of the PCP-C didomain(A) and of the dissected TE domain (B) from the surfactin synthetase (Marahiel 2009). PCP, C and TE domains are respectively shown in green, gray and orange.

1.2.2.1.2.4 The thioesterase (TE) domain

TE domains are only found in termination modules. Several structures of dissected TE domains have been solved, showing that all have the common fold of α/β-hydrolases (**Figure 36B**) (Bruner et al. 2002; Samel et al. 2006), such as esterases and lipases, and all harbor in the active center a catalytic triad of serine, histidine, and aspartate. The serine within this triad plays an important catalytic role, which will be detailed later in the catalytic mechanism part.

1.2.2.1.2.5 An entire termination module

Recently, the crystal structure of the surfactin NRPS (SrfA-C) termination module composed of the four domains (A-PCP-C-TE, 1274 residues) was determined (**Figure 37**) (Tanovic et al. 2008). The folds of the individual domains in SrfA-C module were found to be exactly the same as those solved before for the individually dissected A, C, PCP, and TE domains. The four domains of SrfA-C termination module catalyze the activation and incorporation of the terminal amino acid leucine and also the release of the final product as a cyclic lipopeptide lactone. Although the structures of these domains had been individually resolved, the entire module structure of SrfA-C provided unique information on domain-domain interactions within an NRPS module as well as on the nature of the linker regions and on how the essential catalytic domains are oriented and connected in space. Overall, the structural core of the module is a compact rectangular catalytic platform mainly built by the intimate association of the C domain and the large N-terminal part of the A domain. The two domains are "glued" together by extensive interactions at their interfaces and by the intervening linker region (32 residues) sandwiched in between. Both active sites of the A and C domains are arrayed on the same side of the platform. The C-terminal lid-region of the A domain (~100 residues of the C-terminal part) and the PCP domains are tethered to each other on top of the platform and connected to the large N-terminal region of A by a flexible linker of 15 residues. In this arrangement, PCP and A-lid region can easily move relative to the static C–A-platform. Considering the distances unsuitable for the catalysis between the active sites of the module, large structural rearrangements are mandatory during a full catalytic cycle (Marahiel 2009; Strieker et al. 2010).

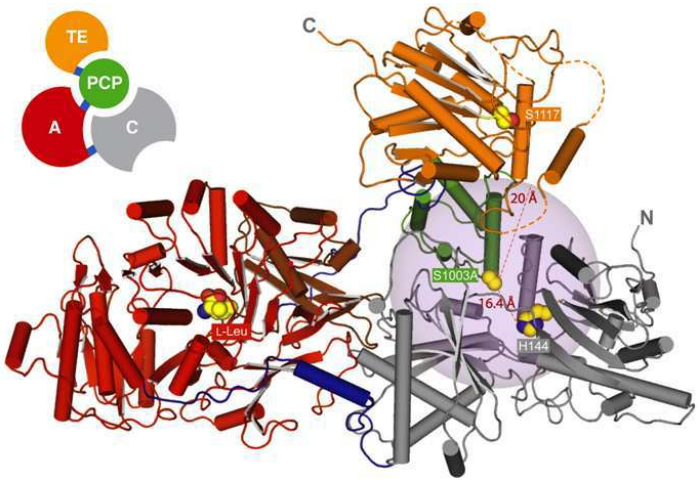


Figure 37: Back view or acceptor site view of the termination module of the surfactin NRPS subunit SrfA–C (Strieker et al. 2010). The 20 Å radius, which is the reachable region of the carrier-arm, around the ppant-attachment site of the PCP domain (S1003A) is depicted as a gray sphere. The PCP domain is well positioned to interact with the H144 active site of the C domain. The active sites of the TE (S1117) and the A domain substrate leucine cannot be reached. The active site of each domain is shown in spheres. Linker regions are shown in blue, the domains are arranged and colored corresponding to the schematic illustrations in upper left corner.

1.2.2.1.3 Catalytic mechanism of NRPSs

As described previously, an A domain is responsible for activation of its cognate amino acid by ATP. The activated amino acid is present in the form of aminoacyl-adenylate (**Figure 38A**). Then the activated amino acid is transferred onto the PCP by a yet unknown mechanism, forming a thioester via ppant bound to PCP (**Figure 38B**). Next, the C domain catalyzes the peptide bond formation between two adjacent PCP domains (upstream PCP1 and downstream PCP2) loaded with their respective amino acids. The aminoacyl-S-PCP2 activated amino acid is bound to the acceptor site of the C domain, whereas the incoming aminoacyl (or peptidyl)-S-PCP1 substrate is bound to the donor site. The C domain catalyzes the nucleophilic attack of the downstream amino acid on the carbonyl group of the upstream amino acid (or nascent peptide), thus forming a new peptide bond (**Figure 38C**). Then a new peptide bond-forming cycle begins and is catalyzed by the downstream NRPS module. In this manner, the peptide chain is elongated until the last module containing the TE domain. The TE domains catalyzes product release by a two step reaction mechanism: transfer of the full length peptide chain attached to the terminal PCP to a highly

conserved serine residue in the active site of TE by the formation of an acyl-O-TE intermediate that is subsequently cleaved by a region- and stereoselective intramolecular macrocyclization using a peptide internal nucleophile (Trauger et al. 2000; Kopp and Marahiel 2007) (**Figure 38D**). Some TE domains catalyze product release either by the generation of cyclic or branched-cyclic products (lactones and lactams), whereas others catalyze product release just by hydrolysis. Anyways, macrocyclization by TE seems to be the predominant mechanism for product release and regeneration of the NRPS (Marahiel 2009). We have previous mentioned that the TE domain harbors in the active center a catalytic triad of serine, histidine, and aspartate (**Figure 36B** and **Figure 37**). The serine within this triad is the site of tetrahedral intermediate formation that is stabilized by an oxyanion hole on the way to the acyl-O-TE intermediate. This intermediate breaks down by the nucleophilic attack of a peptide internal nucleophile. In this way, a nonribosomal peptide is formed by NRPS.

More and more medically relevant nonribosomal peptides and their synthetic mechanisms by NRPSs are being discovered. Nevertheless, although remarkable progress regarding the revelation of the mechanisms underlying nonribosomal peptide synthesis has been made during the last decade, there are still a lot of puzzles. We are still far away from understanding the whole picture.

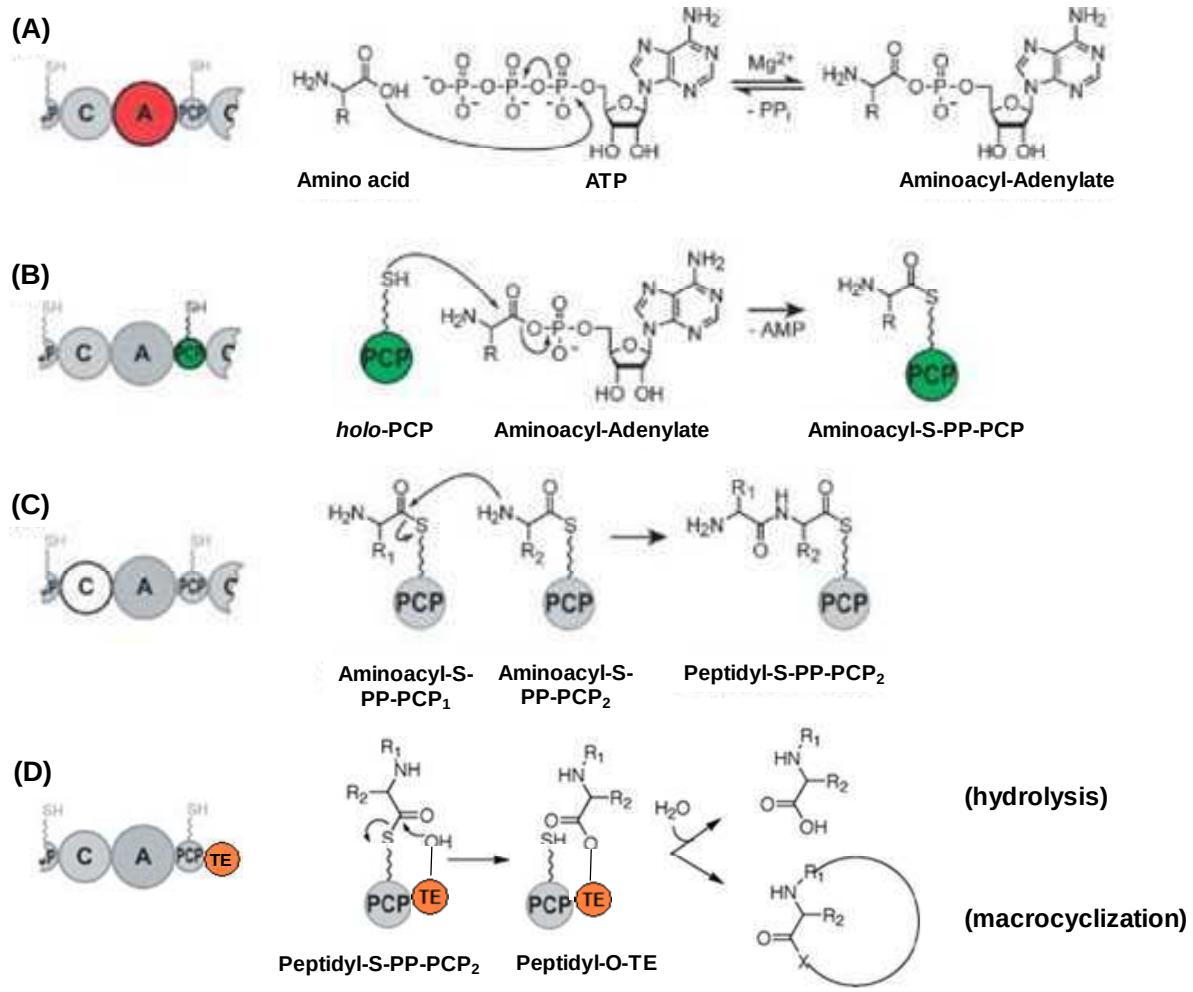


Figure 38: Reactions catalyzed by the four essential NRPS domains A, PCP, C and TE (Felnagle et al. 2008; Marahiel 2009). The cofactor ppant attached to PCP domain is shown as a wavy line.

1.2.2.2 Other peptide synthetases: example of the glutathione synthetase

Glutathione (GSH) is the most abundant and important low molecular weight antioxidant synthesized in cells. GSH play critical roles in protecting cells from oxidative damage and the toxicity of xenobiotic electrophiles, and maintaining redox homeostasis. Moreover, GSH is also involved in regulation of the cell cycle (Forman et al. 2009). Thus, it is of interest to understand the biosynthetic pathway of the GSH.

1.2.2.2.1 General introduction to the glutathione biosynthesis

GSH is a tripeptide, γ-glutamyl-cysteinyl-glycine. The GSH biosynthesis consists of two steps of reaction (**Figure 39**) (Forman et al. 2009). The first step

involves the combination of cysteine with glutamate to produce γ-glutamylcysteine. This reaction is catalyzed by the enzyme glutamate cysteine ligase, also called γ-glutamylcysteine synthetase. This enzyme requires coupled ATP hydrolysis to form an amide bond between the γ-carboxyl group of glutamate and the amino group of cysteine (Huang et al. 1993). The next step involves the enzyme glutathione synthetase (GSH synthetase), responsible for adding glycine to the dipeptide to produce GSH. This step also requires coupled ATP hydrolysis (Meister 1974). Here, we will just briefly describe the structure and the catalytic mechanism of the GSH synthetase.

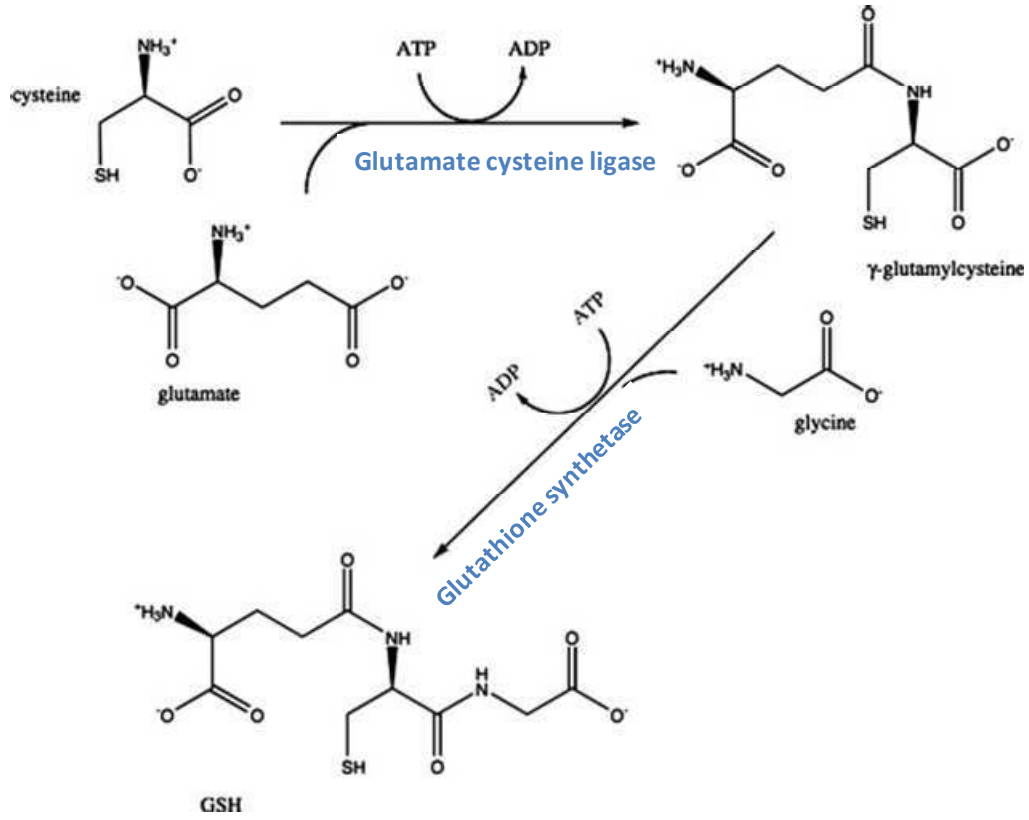


Figure 39: Two-step glutathione synthetic pathway (Forman et al. 2009).

1.2.2.2.2 Structure of the GSH synthetase

Up to now, GSH synthetase structures from three different species have been determined: *E. coli* (Yamaguchi et al. 1993), human (Polekhina et al. 1999), and yeast (Gogos and Shapiro 2002). Taking for example, we will highlight the human GSH

synthetase, the crystal structure of which was determined at 2.1 Å resolution in complex with ADP, two Mg^{2+}, a sulfate ion and GSH (**Figure 40**) (Polekhina et al. 1999). Human GSH synthetase is composed of about 470 amino acid residues. It is a compact molecule with the shape of a flat and a central cavity. The ligands mentioned above are bound in the central cavity, with ADP stacked between two antiparallel β-sheets. The cavity is covered by three loops (designated gray in **Figure 40**) projecting from three of the lain structural units of the structure. The first loop, named substrate-binding loop, forms interactions with the GSH; the other two loops are named the Gly-rich loop and Ala-rich loop. The main structural units are an antiparallel β-sheet together with helices packing on either side of the sheet (designated blue in **Figure 40**), a parallel β-sheet together with helices on both sides (designated red in **Figure 40**), and a domain called the lid because of its role in providing access to the ATP-binding site (designated green in **Figure 40**). The lid domain consists of an antiparallel β-sheet with helices packed on one side. In addition to the three main structural units, residues 3-48 form a dimerization unit (designated purple in **Figure 40**) and residues 295-335 make up the helical connection (designated orange in **Figure 40**) between the parallel β-sheet unit and the lid domain (Polekhina et al. 1999).

The structure indicates that the human GSH synthetase belongs to the ATP-grasp superfamily (Polekhina et al. 1999), the structures of which are characterized by an ATP-binding cleft formed by two antiparallel β-sheets and a phosphate-binding loop (Galperin and Koonin 1997).

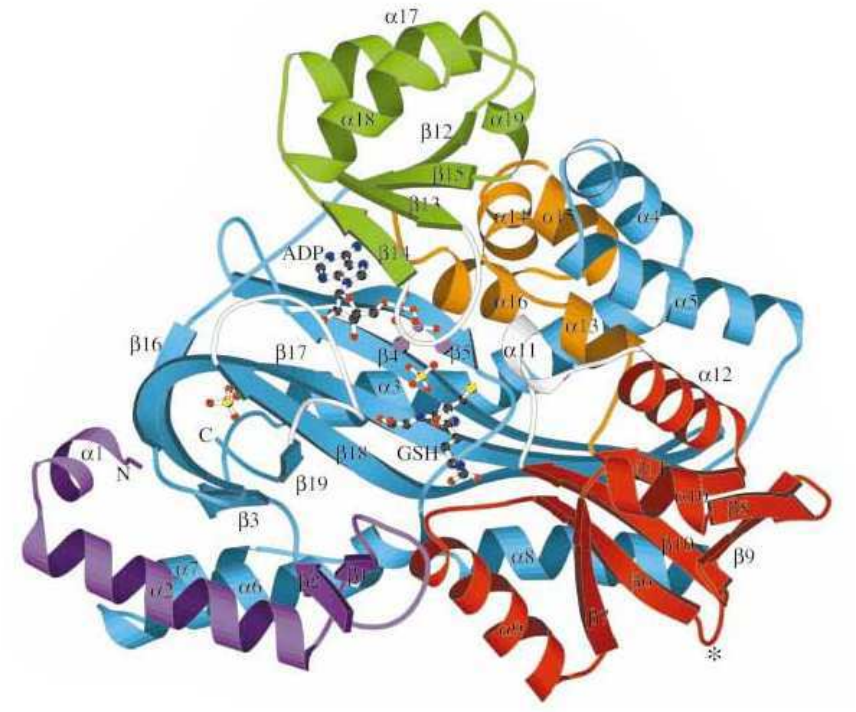


Figure 40: Structure of human GSH synthetase (Polekhina et al. 1999). The elements are shown insecondary structures. GSH and sulfate ions are shown as ball-and-stick and magnesium ions as mauve spheres. Major structural units (see text) are shown in different colors.

1.2.2.2.3 Catalytic mechanism of the GSH synthetase

We evoke that the GSH synthetase is involved in the second catalytic step of the GSH by catalyzing the addition of glycine to the first step-formed dipeptide, γ-glutamylcysteine.

Recently, combining the structures of the GSH synthetases and biochemical studies, the highly conserved residues that form polar interaction with ATP, Mg^{2+} and GSH were found important even essential to the activity. A catalytic mechanism for GSH synthesis in eukaryotes is thus suggested (**Figure 41**) (Herrera et al. 2007). Several amino acid residues of the enzyme were involved in the catalysis. In the first part of the reaction (**Figure 41A**), formation of an electrophilic acylphosphate intermediate occurs by transfer of the γ-phosphate of ATP to γ-glutamylcysteine. During this step, the residues coordinating the Mg^{2+} to ATP are required for nucleotide binding and for orienting the γ-phosphate group in the active site. Arg

132 plays a critical role in catalysis. The guanidyl group of the side chain likely stabilizes formation of the pentavalent transition state that yields the phosphorylated γ-glutamylcysteine intermediate and ADP. In the second half of the reaction (**Figure 41B**), the amino group of glycine acts as a nucleophile to attack the acylphosphate intermediate. Arg 132 and Mg^{2+} bound by Glu 148 and Asn 150 are positioned to stabilize the tetrahedral transition state that decomposes to yield glutathione and inorganic phosphate. The side chain guanidyl group of Arg 454 interacts with the carboxylate moiety of glycine to orient the substrate for attack on the acylphosphate intermediate (Herrera et al. 2007).

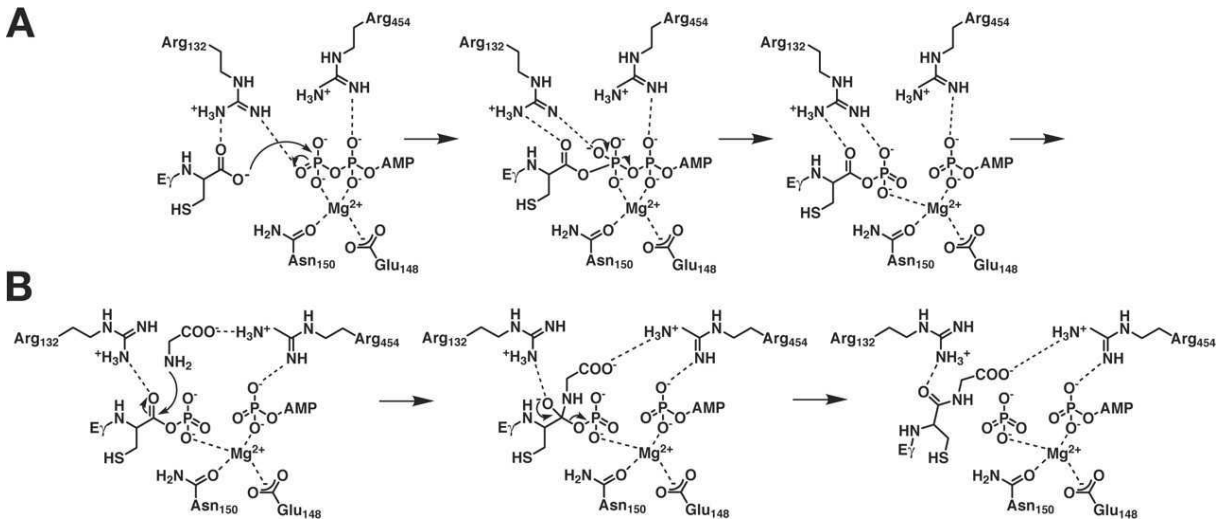


Figure 41: Proposed reaction mechanism for the GSH synthetase (Herrera et al. 2007). A, formation of the acylphosphate intermediate. B, addition of glycine to theγ-glutamylcysteine phosphate intermediate.

We have presented, in **Chapter 1**, the DKP family and the biologic peptide bond-forming systems. These bibliographic studies will help us understand the results of my thesis work, which will now be exposed.

2 STRUCTURAL AND MECHANISTIC CHARACTERIZATION OF THE CDPSs

2.1 Introduction

In the section **1.1** and the introduction of the manuscript, we have illustrated numerous interests to study the DKP family and presented the recently discovered CDPS family which is dedicated to catalyzing the formation of cyclodipeptides, precursors of DKPs. Up to now, eight CDPSs of different bacterial phyla have been identified and characterized, including the first characterized CDPS, AlbC from *S. noursei*. AlbC catalyzes the formation of cFL, which is the precursor of the DKP albonoursin (Lautru et al. 2002; Gondry et al. 2009).

This chapter consists of an article describing the structure and site-directed mutagenesis studies of AlbC, which provides insight into the catalytic mechanism and the interaction between AlbC and aa-tRNA substrates. In the end of the chapter, we will present related studies on another CDPS YvmC-Blic published shortly after those on AlbC (Sauguet et al. 2011).

The crystallographic structure of AlbC was determined by Ludovic Sauguet during his Ph.D work in our laboratory. Concerning my contribution to the work of this article, I participated in the site-directed mutagenesis studies of AlbC in order to understand the role of some residues considered relevant to the catalytic activity of AlbC. Firstly, I participated in the selection of these residues. Secondly, I participated in the construction of the plasmids encoding different variants of AlbC. Then I tested the *in vivo* activities of each of the variants. The variants were expressed in *E. coli* and

the culture supernatants were analyzed by LC-MS/MS in order to identify and quantify the various cyclodipeptides produced then secreted in the culture medium. I also analyzed the expression levels of the variants by SDS-PAGE and western blot. Concretely describing this work, as the crystal structure of AlbC was found to be similar to that of the catalytic domain of the class-Ic aaRSs, we compared the sequences of the six identified CDPSs with those of three TyrRSs from different kingdoms. Among the five conserved CDPS residues located in a surface-accessible pocket, three are well superimposed with their corresponding residues in TyrRS. Considering the possible positioning of the aminoacyl moiety of an aa-tRNA substrate in AlbC pocket, six pocket-forming residues suspected to interact with the aminoacyl moiety of the substrate were independently substituted by alanine and, where relevant, other amino acids. It included the five conserved CDPS residues and another relevant residue located near to the bottom of the pocket. This study allowed understanding the role of these residues in the interaction with the aminoacyl moiety of the substrate. In addition to AlbC pocket, we also independently substituted with alanine the seven residues conserved among CDPSs but located outside the active site. They showed different effect on the activity or production amounts of the corresponding variants. On the basis of their structural positioning, we gave possible explications to their various effects. Moreover, we also studied two patches of basic residues suspected to interact with the tRNA moiety of the substrate. All of these basic residues were substituted with alanine. The site-directed mutagenesis studies on these relevant residues described above allowed understanding the interaction mode of aa-tRNA substrates with CDPSs.

2.2 Article

Nucleic Acids Research Advance Access published February 3, 2011

Nucleic Acids Research, 2011, 1–15
doi:10.1093/nar/gkr027

Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis

Ludovic Sauguet[1], Mireille Moutiez[1], Yan Li[1], Pascal Belin[1], Jérôme Seguin[1], Marie-Hélène Le Du[2], Robert Thai[1], Cédric Masson[1], Matthieu Fonvielle[1], Jean-Luc Pernodet[3], Jean-Baptiste Charbonnier[2,*] and Muriel Gondry[1,*]

[1]CEA, IBITECS, Service d'Ingénierie Moléculaire des Protéines (SIMOPRO), F-91191 Gif-sur-Yvette, [2]CEA, IBITECS, Service de Bioénergétique, Biologie Structurale et Mécanismes (SB[2]SM), CNRS, URA2096, F-91191 Gif-sur-Yvette and [3]Institut de Génétique et Microbiologie, Univ. Paris-Sud 11, CNRS, UMR8621, F-91405 Orsay, France

Received November 5, 2010; Revised January 6, 2011; Accepted January 12, 2011

ABSTRACT

Cyclodipeptide synthases (CDPSs) belong to a newly defined family of enzymes that use aminoacyl-tRNAs (aa-tRNAs) as substrates to synthesize the two peptide bonds of various cyclodipeptides, which are the precursors of many natural products with noteworthy biological activities. Here, we describe the crystal structure of AlbC, a CDPS from *Streptomyces noursei*. The AlbC structure consists of a monomer containing a Rossmann-fold domain. Strikingly, it is highly similar to the catalytic domain of class-I aminoacyl-tRNA synthetases (aaRSs), especially class-Ic TyrRSs and TrpRSs. AlbC contains a deep pocket, highly conserved among CDPSs. Site-directed mutagenesis studies indicate that this pocket accommodates the aminoacyl moiety of the aa-tRNA substrate in a way similar to that used by TyrRSs to recognize their tyrosine substrates. These studies also suggest that the tRNA moiety of the aa-tRNA interacts with AlbC *via* at least one patch of basic residues, which is conserved among CDPSs but not present in class-Ic aaRSs. AlbC catalyses its two-substrate reaction *via* a ping-pong mechanism with a covalent intermediate in which L-Phe is shown to be transferred from Phe-tRNA^{Phe} to an active serine. These findings provide insight into the molecular bases of the interactions between CDPSs and their aa-tRNAs substrates, and the catalytic mechanism used by CDPSs to achieve the non-ribosomal synthesis of cyclodipeptides.

INTRODUCTION

Cyclodipeptide synthases (CDPSs) constitute a family of peptide-bond forming enzymes that use aminoacyl-tRNAs (aa-tRNAs) as substrates to form various cyclodipeptides (1). They are associated with cyclodipeptide-tailoring enzymes in biosynthetic pathways dedicated to the synthesis of diketopiperazines (DKPs) (2,3), a large class of natural products mainly produced by microorganisms. The CDPS family includes at least eight identified members found in various bacterial phyla. All members are similar in size (239–289 residues). They display a moderate sequence similarity, but conserved residues are found in 13 positions.

CDPSs use previously activated amino acids in the form of aa-tRNAs and catalyse the formation of the peptide bonds of the DKP scaffold (1) (Figure 1). Only two other enzymes families also use loaded tRNA as substrates to form peptide bonds (4–7): Fem aminoacyl transferases form cross-bridges in peptidoglycan biosynthesis (8), and

*To whom correspondence should be addressed. Tel: +33 169 08 76 47; Fax: +33 169 08 90 71; Email: muriel.gondry@cea.fr
Correspondence may also be addressed to Jean-Baptiste Charbonnier. Tel: +33 169 08 76 77; Fax: +33 169 08 47 12; Email: jb.charbonnier@cea.fr
Present address:
Ludovic Sauguet, Institut Pasteur, CNRS, URA2185, Unité de Dynamique Structurale des Macromolécules, 25, rue du Dr Roux, F-75015 Paris, France.

The authors wish it to be known that, in their opinion, the second and third authors should be regarded as joint First Authors.

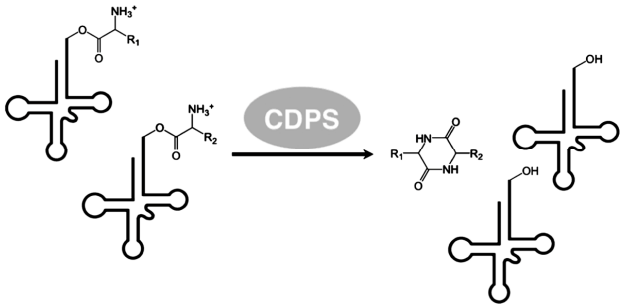


Figure 1. The reaction catalysed by CDPSs. CDPSs use activated amino acids, in the form of aa-tRNAs, as substrates to catalyse the formation of cyclodipeptides, also known as DKPs.

aa-tRNA protein transferases modify the N-terminal extremity of proteins for subsequent degradation (9). Fem aa-transferases and aa-tRNA protein transferases thus catalyse similar chemical reactions, in which the amino acid loaded on an aa-tRNA donor is transferred to the N-terminal amino group of a peptide or a protein acceptor. They both also belong to the same structural GCN5-related *N*-acetyltransferase (GNAT) protein superfamily (9–11). CDPSs differ from Fem aa-transferases and aa-tRNA protein transferases in that they use two aa-tRNAs to form two peptide bonds, whereas transferases use only one aa-tRNA to form only one peptide bond. CDPSs and transferases have also different structural organizations. Indeed, the crystal structure of a CDPS, namely Rv2275 of *Mycobacterium tuberculosis*, was recently determined (12), revealing that Rv2275 is a homodimer, each monomer being structurally related to class-Ic aminoacyl-tRNA synthetases (aaRSs) (13–16). Rv2275 uses two tyrosyl-tRNA^{Tyr} (Tyr-tRNA^{Tyr}) as substrates to produce cyclo(L-Tyr–L-Tyr) (cYY) (1).

Here, we solved at 1.9 Å resolution the crystal structure of *Streptomyces noursei* AlbC, a CDPS that uses mainly phenylalanyl-tRNA^{Phe} (Phe-tRNA^{Phe}) and leucyl-tRNA^{Leu} (Leu-tRNA^{Leu}) as substrates to synthesize cyclo(L-Phe–L-Leu) (cFL). The AlbC structure consists of a monomer. It contains a Rossmann-fold domain and shares similarity with the catalytic domain of class-Ic aaRSs, as previously reported for Rv2275 (12). The reaction catalysed by Rv2275 was suggested to proceed *via* a ping-pong kinetic mechanism with a covalent intermediate in which L-Tyr is transferred from Tyr-tRNA^{Tyr} to an active serine, but such a transfer was not conclusively demonstrated (12). Here, we showed that the serine residue is covalently bound to the aminoacyl transferred during the catalytic cycle, demonstrating that CDPSs use a mechanism involving a covalent aminoacyl-enzyme intermediate. Moreover, we also showed that the interaction between AlbC and its aa-tRNA substrates is driven by the binding of the substrate aminoacyl moiety in a deep pocket—similar to the tyrosine-binding pocket of TyrRSs—and the interaction of the tRNA moiety with at least one patch of basic residues. This provides clues about the molecular bases of the interactions between CDPSs and their aa-tRNAs substrates.

MATERIALS AND METHODS

Mutagenesis and purification of AlbC and variants

The expression plasmid encoding AlbC was constructed from the vector pQE60 (Qiagen). The resulting plasmid encodes the full-length AlbC (residues 1–239) with the C-terminal addition of two residues (RS) followed by a His_6 tag (1). The desired variants of AlbC were created *via* PCR mutagenesis according to the QuikChange method (Stratagene). Thirty-two variants were constructed (28 variants with a single mutation, and four variants with a double mutation). For 12 variants, the mutations are localized in the putative catalytic pocket (L33Y/L185D, G35A, S37A, S37C, Y178A, Y178F, E182A, E182Q, E182D, L200N, Y202A, Y202F), and for the remaining variants, the mutations are outside this pocket (H31A, W54A, F59A, G79A, R87A, K90A, R91A, K94A, R97A, R98A, R99A, R98A/R99A, R101A/R102A, S120A, Y128A, C149A, P184A, R214A/R215A, R220A, R231A). The C-terminal His_6-tagged proteins were produced and purified as described previously (1,17). Protein production was followed by SDS–PAGE analyses or western blotting according to standard protocols, and His-tagged proteins were visualized using alkaline phosphatase-conjugated anti-His antibody and BCIP/NBT solution (Sigma).

Crystallization and structure determination

Small crystals of AlbC ($80 \times 10 \times 10\,\mu m^3$) were grown by vapour diffusion in sitting drops at 17°C by mixing 2 µl of protein sample at 4 mg/ml with 2 µl reservoir solution. First crystalline precipitates were obtained by a reverse screening approach (18). By screening different pHs, crystals were obtained with reservoir containing 1.7 M Na/K phosphate at pH 6.0. A screening of different additives showed that the presence of reducing agents improves crystals size and the addition of a large excess of dithiothreitol (DTT) (100 mM) gave the best results. These crystals were cryocooled with different cryoprotectants, but they did not diffract better than 3.5 Å. Larger

crystals ($300 \times 40 \times 40\,\mu m^3$) were obtained by a batch method under silicon oil and microseeding from solutions of crushed AlbC crystals (19). The native crystals were cryocooled with 35% glycerol and diffracted up to 1.9 Å on Proxima1 (Soleil, France) and ID29 (ESRF, France) beamlines. The crystal structure was determined with selenomethionine-substituted AlbC by MAD (multi-wavelength anomalous dispersion) using phase information derived from anomalous scattering data collected at the Se K-edge. The crystals belong to the space group I4 with a single copy of the protein in the asymmetric unit. The positions of four selenium atoms were determined by SHELXD (20). Experimental phases were calculated in SHARP (21), and the tracing of an initial model in COOT that was subsequently completed by automatic building in ArpwARP (22) using the AlbC native data set.

The final model was refined in REFMAC5 (23) to R/R_{free} values of 0.186/0.236 at 1.9 Å resolution with one TLS parameter for the protein chain. In the final model, residues 1–27, 217, the eight C-terminal residues [RS$(His)_6$] introduced during cloning, and the side chain of residue R214 were not visible in the electron density map, they were not included in the crystallographic model. The MolProbity score (24) for the refined model is 1.91, in the 75th percentile of structures refined at comparable resolution. Almost all residues (97.6%) were in favoured regions of the Ramachandran plot, and 100.0% were in allowed regions.

In the final electron density, remaining peaks that could not be attributed to water were modelled with three molecules of phosphate and a molecule of cyclo-dithiothreitol (cDTT), which were present in the crystallization solution as a precipitant agent and an additive, respectively (see above, cDTT comes from the oxidized fraction of the DTT used; Supplementary Figure S1). This observation suggests that the large quantity of DTT used during crystal optimisation served both to maintain AlbC under its reduced state, and to feed crystals with cDTT present in equilibrium with reduced DTT. Figures were prepared with PyMOL (PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC). The programme Superpose (25) was used for structural alignment and Areaimol (26) for accessible surface area measurements. The coordinates and structure factors of *S. noursei* AlbC have been deposited in the Protein Data Bank (PDB) (accession code 3OQV).

Cyclodipeptide-synthesizing activity assays

Assays with supernatants of cultures of *Escherichia coli* cells expressing the genetic constructs encoding AlbC wild-type or an AlbC variant were analysed by LC-MS/MS as described previously (1). The *in vivo* activities are reported as percentage of the wild-type activity. Error bars represent the standard deviation for at least two independent experiments.

Detection and identification of the covalent aminoacyl-enzyme intermediate

To detect the covalent aminoacyl-enzyme intermediate, purified AlbC or relevant variants were incubated with [^{3}H]Phe-tRNA^{Phe}, the proteins were separated by SDS–PAGE before being transferred on a polyvinylidene fluoride (PVDF) membrane, which was analysed with a Beta-Imager^{TM} 2000 from Biospace (Paris, France). Protein concentrations were quantified by UV spectroscopy. The tritiated substrate was obtained as follows: a 100 µl reaction mixture containing 3.2 µM *E. coli* tRNA^{Phe} (Sigma), 0.32 mM Phe, 0.9 µM [^{3}H]Phe (Phenylalanine, L [2,3,4,5,6-^{3}H], Perkin-Elmer, 110 Ci/mmol), 2 mM ATP, 1 mM DTT, 20 mM $MgCl_2$, 50 mM KCl, 1 µM *E. coli* PheRS (PheRS$_{Ec}$) in 50 mM Hepes–KOH, pH 7.0 was incubated at 30°C for 30 min. Then, AlbC or AlbC variant was added at a final concentration of 1 µM. After 20 s incubation, the reaction was quenched by the addition of 5 µl Laemmli loading buffer [final concentration, 0.1% (w/v) bromophenol blue, 2% (w/v) SDS, 10% (v/v) glycerol, 50 mM Tris–HCl, pH 6.8 and 100 mM DTT] followed by heating the samples 2 min at 50°C. These samples were immediately resolved by SDS–PAGE electrophoresis followed by blotting. Blotting of proteins onto a PVDF membrane was achieved using a semi-dry transfer blot apparatus (Bio-Rad). After transfer (400 mA, 30 min), the membrane was dried before radioactivity analysis with the Beta-Imager^{TM} 2000. This apparatus allows an absolute counting of the tritium β particles, with a detection threshold of 0.007 cpm/mm^2 for tritium.

To identify the residue acylated and the nature of the acylation, the aminoacyl-enzyme intermediate was produced by a direct assay using purified Phe-tRNA^{Phe}. The aminoacylation reaction was performed as described above. Charged Phe-tRNA^{Phe} was recovered by phenol/chloroform extraction and ethanol precipitation [the loading of tRNA^{Phe} with Phe was estimated to be 70%, as previously described (1)]. Two micrograms of Phe-tRNA^{Phe} were mixed with 3 µg of AlbC (or variant) in 20 µl of the reaction buffer containing 50 mM Hepes–KOH pH 7.0 and 50 mM KCl. After 20 s of incubation, a 5 µl-aliquot was removed and submitted to trypsin digestion: the withdrawn aliquots were diluted up to 10 µl in 50 mM NH_4AcO_2, pH 8.0 and 1 µl of trypsin (400 ng/µl porcine trypsin, Promega) was added. Digestion was performed at 50°C up to 10 min. Sample digests were then acidified by 1.25% trifluoroacetic acid (TFA) in water, desalted and concentrated by C18 ziptip and spotted with 4-HCCA matrix solution (10 mg/ml in 49.8/49.8/0.4: CH_3CN/H_2O/TFA: v/v/v) on MALDI plate. MS and MS/MS spectra were registered using a 4800 MALDI-TOF/TOF mass spectrometer (Applied Biosystems, Foster City, USA) in a positive reflectron mode. Each MS spectrum was the result of 1000 laser shots and calibration was applied by using trypsin autolysis fragments as internal standards. MS/MS analyses were performed under post-source decay (PSD) mode. Each MS/MS spectrum was the result of 10 000 shots. The sequence of the trypsic fragments was identified by using Data Explorer® processing software (Version 4.9, Applied Biosystems) to assign the potential chemical modifications of the serine or cysteine residues.

RESULTS

Crystal structure of *S. noursei* AlbC

We produced AlbC in *E. coli*, purified it to homogeneity (1) and crystallized the purified AlbC. We solved the crystal structure of AlbC by the multi-wavelength anomalous dispersion (MAD) method using selenomethionine-labelled protein. The native structure was refined to a final R-factor/R-free of 18.6/23.6% at a resolution of 1.9 Å (Table 1). The crystal structure of AlbC is a monomer that presents a compact α/β structure (Figure 2). It is composed of a central β-sheet with five parallel β strands (β3–β7) surrounded by ten α helices (α2–α9) (Figure 2A). The AlbC structure contains a Rossmann-fold domain (27) formed by strands β3–β5, helix α2 and the C-terminal part of helix α4. This domain is followed by a helical region composed of three main helices (α5–α7) that pack against the Rossmann fold. The structure ends with two β strands (β6 and β7) that extend the Rossmann β-sheet and two small connecting helices (α8 and α9). AlbC residues 1–27 were not visible in the electron density map. Crystal packing analysis showed the presence of large cavities near the residue 28 that are compatible with the presence of a flexible N-terminal extension (Supplementary Figure S2). Some regions of AlbC were visible in the electron density map, but they present higher B-factors than the rest of the protein (Supplementary Figure S3). These regions contain the residues 38–46 in loop β3–α2, the residues 105–111 in loop α4–β5, the residues 215–220 in loop α8–β7, and the C-terminal residues 236–239.

A striking feature of the structure of AlbC is the presence of a deep (14 Å long) and narrow (diameter ranging from 4 to 7 Å) surface-accessible pocket, which is delineated by helices α6, α7, α8 and β-strands β3, β4 and β6. The buried part of the pocket in AlbC is bordered mostly by side chains of nine hydrophobic residues: L33, V65, L119, L185, F186 and L200 at its deepest point and V67, M152 and I208 on the edges (Figure 2B). Four additional residues contribute to define the buried part of the pocket, G35, E182 and the main chains of Y202 and H203. Near the residue G35, the pocket presents a small bulge bordered by the hydrophobic residues I34, F43, L51, C201 and V230. Residues P207 and V156, the carboxylate moiety of E182 and the side chains of four polar amino acids—S37, N40, Y178, T206—define the entrance of the pocket (Figure 2B). Additional electron density was attributed to a molecule of phosphate and to a molecule of cDTT, which are located at the entrance and at the bottom, respectively, of the pocket (Figure 2B; 'Materials and Methods' section). The phosphate molecule, probably present as a $H_2PO_4^-$ ion at the pH used for crystallization (pH = 6), is within hydrogen bond distances of residues Y178 and E182, and to two water molecules. The cDTT is surrounded by hydrophobic residues in the most buried part of the pocket and its oxygens are within hydrogen bond distances of phosphate and water molecules.

Table 1. Data collection, phasing and refinement statistics

	Native	SeMet Crystal 1			SeMet Crystal 2
Data collection					
Space group	I4	I4			I4
Cell dimensions					
a, *b*, *c* (Å)	97.2, 97.2, 45.5	97.9, 97.9, 45.5			98.0, 98.0, 45.5
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0			90.0, 90.0, 90.0
		Peak	Inflection	Remote	Peak
Wavelength	0.9793	0.9792	0.9794	0.9757	0.9792
Resolution (Å)	48.6–1.90 (2.00–1.90)	69.0–2.70 (2.70–2.85)	69.0–3.00 (3.16–3.00)	69.0–3.50 (3.69–3.50)	69.0–3.00 (3.16–3.00)
R_{sym} or R_{merge}	8.70 (61.5)	8.10 (28.6)	9.80 (36.6)	9.80 (24.2)	9.10 (40.6)
$I/\sigma I$	13.0 (3.10)	13.8 (4.40)	11.1 (3.50)	11.4 (5.50)	12.9 (3.80)
Completeness (%)	99.7 (100.0)	99.3 (98.9)	99.1 (99.8)	98.5 (99.7)	99.9 (100.0)
Redundancy	7.4 (6.9)	4.0 (3.8)	3.7 (3.9)	3.4 (3.6)	5.6 (5.7)
Refinement					
Resolution (Å)	1.90				
No. reflections	15963				
R_{work}/R_{free}	18.6/23.6				
No. atoms					
Protein	1621				
Ligand/ion	23				
Water	107				
B-factors (Å^2)					
Protein	29.3				
Ligand/ion	55.8				
Water	36.2				
Rmsd					
Bond lengths (Å)	0.006				
Bond angles (°)	0.793				

Values in parentheses are for highest resolution shell.

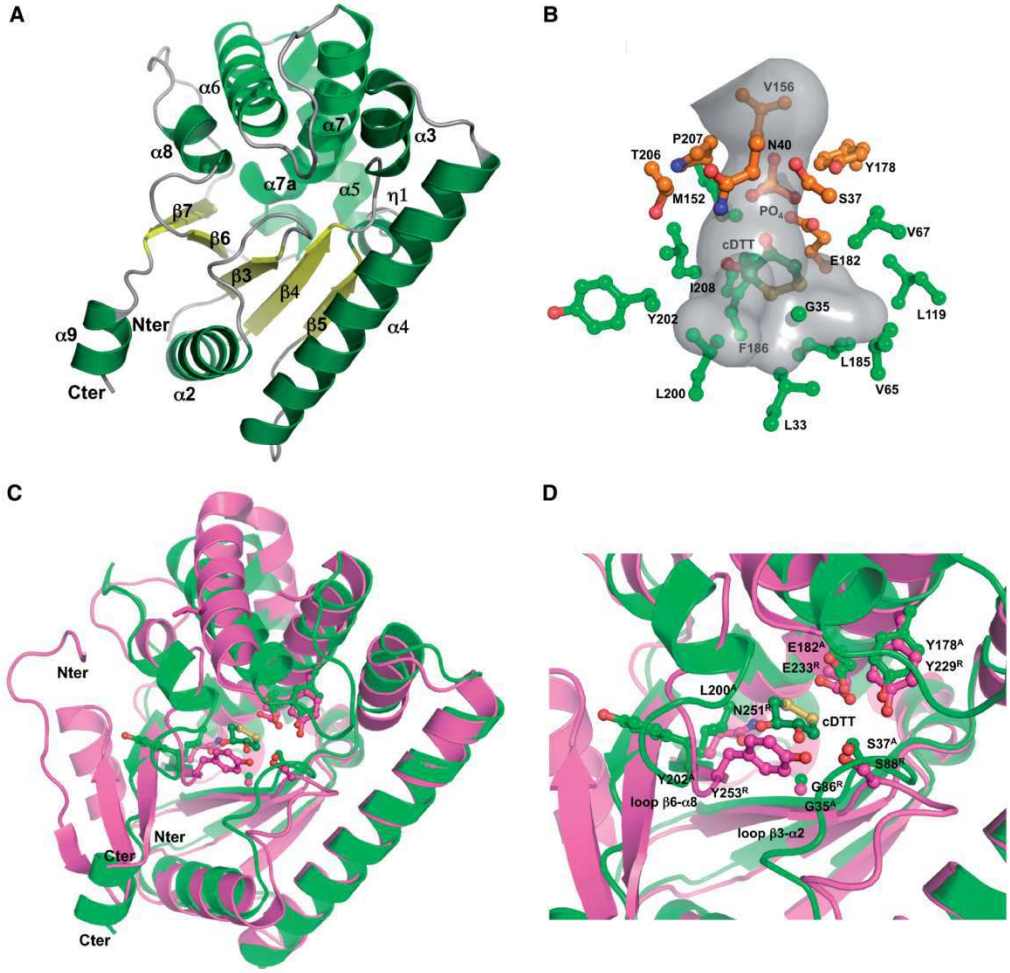


Figure 2. Crystal structure of *S. noursei* AlbC and comparison with that of *M. tuberculosis* Rv2275. (**A**) Overall structure of AlbC (region 28–236) in cartoon mode, with α helices, β-strands, and loops coloured in green, yellow and grey, respectively. AlbC secondary-structure elements are numbered according to Rv2275 numbering (12). The Rossmann-fold domain is constituted by helix α2, the C-terminal part of helix 4 and the β-sheet (β3–β5) that is extended by two additional strands β6 and β7. A helical region (in light green), composed of helices α5, α6 and α7, packs against the Rossmann fold. (**B**) Structure of the AlbC surface-accessible pocket. Residues that delineate the AlbC pocket are shown in ball and stick representation. The most buried part of the pocket is mainly constituted of hydrophobic residues (coloured in green). The entrance of the pocket is mainly delineated by polar residues (coloured in orange). The boundaries of the pocket, calculated using the program CAVER (58), are defined by a semi-transparent grey surface. The molecules of phosphate and cDTT modelled in the electron density are shown in orange and dark green, respectively. (**C**) Superimposition of the structures of AlbC and Rv2275 in cartoon representation. AlbC and Rv2275 are coloured in green and magenta, respectively. Both proteins superpose well. Most significant structural differences concern N- and C-terminus regions, and also relative positions of helices α5, α6 and α8. (**D**) Superimposition of the surface-accessible pockets of AlbC and Rv2275, with an orientation similar to that shown in Figure 2C. Residues represented in ball and stick are essential for CDPS activity. AlbC and Rv2275 residues are indicated by A and R letters, respectively. The superimposition highlights the structural deviations between loops β3–α2 and β6–α2, and residues $Y202^A$ and $Y253^R$ located at the entrance of the active sites.

The three-dimensional (3D) superimposition of AlbC and Rv2275 (PDB id: 1j1u) [root mean square deviation (rmsd) value of 2.27 Å over 192 Cα] revealed that the two enzymes superimpose well despite sequence divergence (AlbC shares 25.9 % identity and 40.6 % similarity with Rv2275; Figure 2C). Almost all secondary-structure elements superimpose well, in particular the Rossmann-fold domain and the two additional strands β6 and β7. However, there are some major differences between the two enzymes. First, the region 50–77 of Rv2275 is structured (strands β1, β2 and helix α1) while the corresponding region of AlbC (1–27) is not observed in the

electron density maps. At opposite, two regions of AlbC, the loop α6–α7 and the C-terminal helix α9 are structured in AlbC, and the corresponding regions are not observed for Rv2275. Second, the axes of helices α5, α6 and α8 are tilted by some degrees, so that residues located at these helices extremities superimpose poorly. Finally, important deviations are observed between loops, in particular loops β3–α2, α4–β5, β6–α8 and α8–β7.

Both enzymes possess a similar surface-accessible pocket that contains six residues conserved in the CDPS family. Four of these residues, G35 (AlbC numbering) belonging to the strand β3, and Y178, E182 and P184 belonging to the helix α7, are remarkably well superimposed. The two other conserved residues S37 and Y202, respectively, belong to loops β3–α2 and β6–α8 that have different conformations as mentioned above. In particular, the side chain of Y202 adopts a conformation directed towards solvent in AlbC whereas it is directed towards the active site entrance and residue S37 in Rv2275 (Figure 2D).

Structural similarity of AlbC with the catalytic domain of the class-I aminoacyl-tRNA synthetases

We used the Dali server (28) to compare the structure of AlbC to protein structures in the PDB. Due to the Rossmann-fold domain, AlbC was found to share structural similarity with more than 300 proteins with a high *Z*-score ($Z > 4$) (Supplementary Table S1). Most of these proteins are class-I aaRSs, enzymes that catalyse the activation of amino acids and their transfer to cognate tRNAs to form aa-tRNAs. The 120 solutions with the highest *Z*-scores, from 10.7 to 7.3, were all obtained with the closely related class-Ic TyrRSs and TrpRSs, as previously reported for Rv2275 (12). The greatest similarity is with the archaeal *Archaeglobus fulgidus* (PDB id: 2cyb) and *Methanococcus jannaschii* TyrRSs ($TyrRS_{Mj}$) (PDB id: 1zh6, 1j1u), and with the eukaryotic *Entamoeba histolytica* TrpRS ($TrpRS_{Eh}$) (PDB id: 3hzr) (Supplementary Table S1). The 3D superimposition of AlbC and $TyrRS_{Mj}$ (rmsd value of 3.5 Å over 160 Cα) (Figure 3) or $TrpRS_{Eh}$ (rmsd value of 3.0 Å over 163 Cα) (Supplementary Figure S4) revealed that AlbC superimposes well with the N-terminal catalytic domains of the two aaRSs, despite low-sequence similarity (AlbC shares 15.5 and 15.8% similarity with the catalytic domains of $TyrRS_{Mj}$ and $TrpRS_{Eh}$, respectively). Note that these enzymes possess C-terminal tRNA-binding domains whereas AlbC does not. The catalytic domains of each of the two aaRSs are composed of a Rossmann-fold domain and a connective-polypeptide 1 (CP1) domain that are both found in AlbC. The overall position of the two domains is conserved between AlbC and the two aaRSs and is mediated by substantial conservation of residues at the interface between the two domains. Most secondary-structure elements are well conserved, and the central β-sheets are well superimposed, but the helices that pack against these sheets present slight differences in length and positioning (Figure 3; Supplementary Figure S4). Moreover, the pocket in AlbC and the aminoacyl binding pockets in the two aaRSs are positioned similarly relative to the

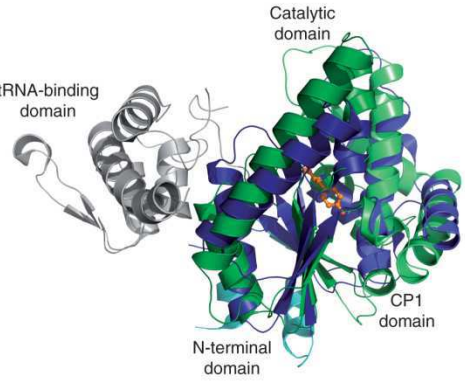


Figure 3. Superimposition of the structures of AlbC and $TyrRS_{Mj}$ in complex with its L-tyrosine substrate (PDB id: 1j1u). The two enzymes are shown in cartoon mode, and L-tyrosine is in ball and stick coloured in orange. The Rossmann-fold and the CP1 domains of $TyrRS_{Mj}$ are coloured in dark and light blue, respectively, and the corresponding domains of AlbC are coloured in dark and light green, respectively. The Rossmann-fold and CP1 domains of the two proteins have similar rmsd values (3.25 Å over 96 Cα and 3.26 Å over 53 Cα, respectively). $TyrRS_{Mj}$ possesses two additional regions not present in AlbC: a C-terminal domain involved in tRNA-binding and anticodon recognition coloured in grey, and a short N-terminal region coloured in light blue.

β-sheet forming part of the Rossmann-fold domain (29–33).

However, there are some major differences between AlbC and the two aaRSs. The hydrophobic regions of the CP1 domains of TyrRS or TrpRS involved in the homodimerization of these enzymes (34,35) are not found in AlbC. In addition, when we mimicked this association for two AlbC monomers by superimposing them on the crystal structure of $TyrRS_{Mj}$, we did not find any complementary dimer interface, and we observed major steric hindrances between helices α5, α6 and α7 of each AlbC monomer (Supplementary Figure S5A). However, the crystal structure of Rv2275 was obtained as a homodimer. But, the Rv2275 residues that make up the interface belong to other secondary-structural elements than those involved in the homodimerization of the two aaRSs (12). Noted that these residues are not conserved in CDPSs (Figure 4 and Supplementary Figure S5B). All these observations are consistent with AlbC being found as a monomer both in solution (1) and in crystal forms. Another major difference between CDPSs and class-I aaRSs is that CDPSs do not have the two consensus motifs that are conserved in class-I aaRSs and are involved in ATP binding (36–38) (Figure 4).

The pocket of AlbC, a binding site for the aminoacyl moiety of the substrate

Using 3D superimposition, we compared the pocket in AlbC to the conserved amino acid-binding pockets in $TyrRS_{Mj}$ and $TrpRS_{Eh}$ (29–33). For example, the AlbC pocket was compared to that of the $TyrRS_{Mj}$ complexed with its tyrosine substrate (Figure 5), which forms a

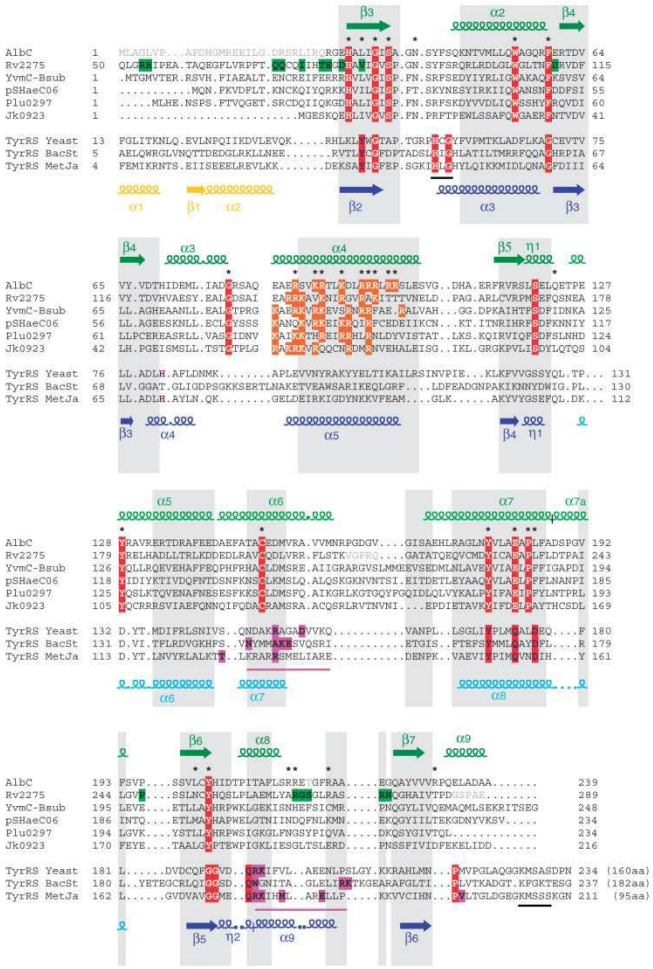


Figure 4. Structure-based sequence alignment of six CDPSs and three TyrRSs from the three primary kingdoms. The secondary-structural elements of AlbC and the $TyrRS_{Mj}$ are indicated above and below the alignment, respectively. The Rossmann-fold and the CP1 domains of AlbC are shown in dark and light green, respectively. The N-terminal region, the Rossmann-fold domain, and the CP1 domain of the $TyrRS_{Mj}$ are shown in yellow, dark and light blue, respectively. The core regions of AlbC and the $TyrRS_{Mj}$ displaying greatest structural similarity (rmsd value of 2.36 Å over 116 Cα) are shown with a gray background. The residues conserved among CDPSs, and those conserved among TyrRSs are indicated by a red background. Black stars above the AlbC sequence indicate the residues that were modified by site-directed mutagenesis in this study. The basic residues of CDPSs, which define a patch of positively charged residues belonging to the helix α4, are indicated by an orange background. The residues that are disordered in the crystal structure of AlbC and Rv2275 are shown in grey. The residues of TyrRSs previously shown to be involved in tyrosine recognition are indicated in dark blue; the position of the two clusters involved in acceptor stem recognition (41) are indicated by pink lines below the TyrRSs sequences, and the residues involved in this recognition are indicated by a pink background. Black lines below the TyrRSs sequences indicate the class I signature motifs, HIGH and KMSKS, which are involved in ATP binding. Rv2275 residues that are involved in the interface of the dimer are shown with a green background. Protein sequence alignments were done according to structural alignments for AlbC and TyrRSs and were corrected manually for CDPS based on previously described alignments (1). Two CDPSs are not mentioned, namely YvmC from *Bacillus thuringiensis* and YvmC from *B. licheniformis*, because they are very close to YvmC from *B. subtilis*. Abbreviations: Bsub, *B. subtilis*; BacSt, *Geobacillus stearothermophilus*; MetJa, *M. jannaschii*.

hydrogen bond network with the conserved residues Y32, Y151, Q155, D158 and Q173, and interacts with the conserved residue G34 (29–31) (PDB id: 1j1u) (Figure 5A). The two pockets are positioned similarly relative to the Rossmann-fold. Each pocket contains some residues that are highly conserved in the corresponding family and that are remarkably well superimposed: the AlbC residues G35, Y178 and E182 superimpose with the $TyrRS_{Mj}$ residues G34, Y151 and Q155, the rmsd values on all atoms being 1.30, 1.29 and 1.95 Å, respectively (Figures 4 and 5B). These structural similarities suggest that the pocket of AlbC could accommodate the aminoacyl

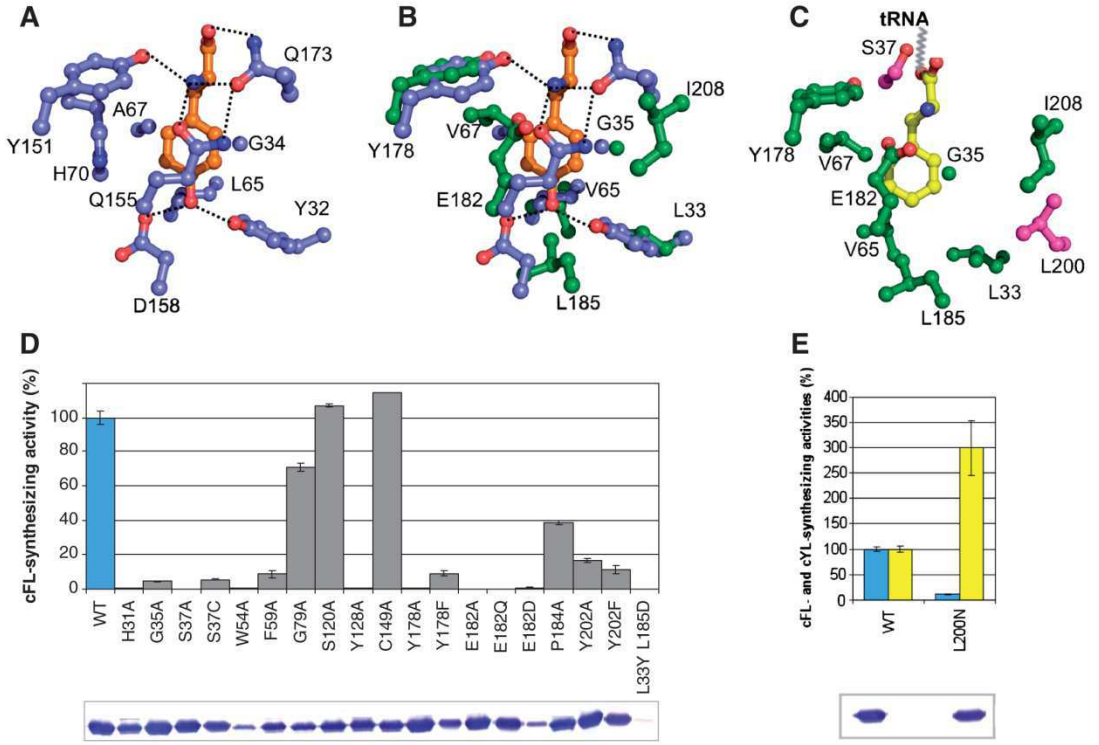


Figure 5. Comparison of the AlbC and $TyrRS_{Mj}$ pockets. (**A**) The tyrosine-binding pocket of $TyrRS_{Mj}$ in complex with its L-tyrosine substrate. TyrRS residues and L-tyrosine are represented in ball and stick, and coloured in blue and orange, respectively, with oxygen and nitrogen atoms coloured in red and dark blue, respectively. (**B**) Superimposition of the TyrRS residues (A) with the corresponding AlbC residues. AlbC residues are represented in ball and stick, and coloured in green, with oxygen and nitrogen atoms coloured in red and dark blue, respectively. Residues lining the two pockets superimpose with an rmsd value of 1.49 Å over 32 main chain atoms. (**C**) The AlbC pocket in the presence of the phenylalanyl moiety of a Phe-$tRNA^{Phe}$ substrate. The phenylalanyl moiety position is mimicked on that of the L-tyrosine in the pocket of $TyrRS_{Mj}$, as shown in (A) and (B), except that the hydroxyl group has been removed. The conserved residue S37 and the residue L200, which have no corresponding residues in TyrRS, are shown in pink, with oxygen and nitrogen atoms coloured in red and dark blue, respectively. (**D** and **E**) Site-directed mutagenesis study of AlbC. Cyclodipeptide-synthesizing activities are shown with error bars. The corresponding western blots indicating amounts of the proteins are also shown. (D) cFL-synthesizing activities of the wild-type AlbC (in blue) and of each of the variants in which a residue conserved among CDPSs is substituted (in grey). (E) cFL- and cYL-synthesizing activities (in blue and yellow, respectively) of the wild-type AlbC and the variant L200N. The wild-type AlbC synthesizes $33.5 \pm 2.5\,mg\,l^{-1}$ of cFL and $4.9 \pm 0.5\,mg\,l^{-1}$ of cYL (Supplementary Figure S6).

moiety of an aa-tRNA substrate in the same way that TyrRSs bind their tyrosine substrates. To test this possibility, we independently substituted each residue in AlbC suspected to interact with the aminoacyl moiety of the substrate by alanine and, where relevant, other amino acids. We produced the resulting variants in *E. coli* and compared their *in vivo* synthesis of cFL, the major cyclodipeptide produced by AlbC, to that of the wild-type enzyme (1). The substitution of Y178 with alanine resulted in an almost inactive variant (Y178A), and its substitution with phenylalanine gave a poorly active variant (Y178F), despite being produced in amounts similar to that of the wild-type enzyme (Figure 5D). The variants E182A and E182Q were inactive, and the variant E182D had a weak but detectable activity although its production was strongly affected. The variant G35A was poorly active despite being abundantly produced (Figure 5D). Thus, the hydroxyl group of Y178 and the carboxylate group of E182 are essential for AlbC activity, suggesting that these groups may mediate hydrogen bonds with the main chain amino group of the substrate, in a way similar to what is described for $TyrRS_{Mj}$. The conserved residue G35 is also essential for the accommodation of the CDPS substrate. The conserved residues Y32 and D158 in $TyrRS_{Mj}$ form hydrogen bonds with the hydroxyl group of the tyrosine substrate (Figure 5A), and superimpose with the hydrophobic residues L33 and L185, respectively, in AlbC (Figure 5B). This is consistent with hydrophobic phenylalanyl or leucyl moieties in the AlbC substrates. To try to accommodate a tyrosyl-$tRNA^{Tyr}$ substrate in the AlbC pocket, we attempted to construct an L33Y/L185D AlbC variant, but the genetic construct was not expressed (Figure 5D).

All these data are consistent with the pocket of AlbC being the binding site of the aminoacyl moiety of an aa-tRNA, but do not prove it. We therefore replaced a residue in the AlbC pocket suspected to interact with the aminoacyl moiety of the substrate to see if this may modify the enzyme's specificity. The sequence alignment of CDPSs (Figure 4), and the specificity determined for each characterized CDPS (1), led us to focus on Rv2275 from *M. tuberculosis* that mainly synthesizes cyclo(L-Tyr–L-Tyr) (cYY) (1,3). The residue L200 located near the bottom of the AlbC pocket (Figure 5C) corresponds to an asparagine in Rv2275 (Figure 4). This asparagine residue may form a hydrogen bond with the hydroxyl group of the tyrosyl moiety of a tyrosyl-tRNA^{Tyr} substrate. We constructed the variant L200N, and tested its *in vivo* specificity: it synthesized mainly cYL instead of cFL and thus showed some of the specificity of Rv2275 (Figure 5E, Supplementary Figure S6). This indicates that the substitution of one residue located in the AlbC pocket is sufficient to change the nature of one of the aminoacyl residue incorporated into the cyclodipeptides it produces. This also demonstrates that the AlbC pocket is the binding site of the phenylalanyl moiety of the Phe-tRNA^{Phe} substrate. But, this result does not provide data on the binding of the leucyl moiety of the other substrate, which is required to synthesize a molecule of cFL.

Formation of a covalent intermediate during the catalytic cycle of AlbC

The superimposition of the pockets of AlbC and TyrRS$_{Mj}$ revealed significant similarities (see above), but also intriguing differences. The strictly conserved residue S37 of AlbC has no corresponding conserved residue in class Ic-aaRSs (Figures 4 and 5B). The position of the phenylalanyl moiety of a Phe-tRNA^{Phe} in the AlbC pocket, as predicted by analogy with the tyrosine substrate in the TyrRS$_{Mj}$ pocket (Figure 5C), indicates that the Oγ atom of S37 would be located in the proximity of the carbonyl carbon of the esterified Phe-tRNA^{Phe} substrate. This organization together with the catalysed reaction—the formation of a peptide bond from aminoacyl-tRNA—suggests that S37 may be a catalytic residue serving as a nucleophile. This is consistent with the fact that substitution of S37 with alanine resulted in an inactive variant (S37A), but its substitution with the nucleophilic cysteine (S37C) did not completely abolish the enzymatic activity (Figure 5D). So, S37 may attack the carbonyl carbon of the esterified Phe-tRNA^{Phe} substrate, leading to the covalent phenylalanylation of S37. The formation of a covalent intermediate in which L-Tyr is transferred from Tyr-tRNA^{Tyr} to an active serine (S88) was also suggested for Rv2275, using radioactive labelling (12). Moreover, the S88 residue of Rv2275 was suggested to be activated by the conserved Y253 residue, which corresponds to Y202 in AlbC (12). To verify the existence of a covalent intermediate, we further analysed AlbC and the five variants S37A, S37C, Y178F, E182Q and Y202F. The enzymes were incubated in the presence of [^{3}H]Phe-tRNA^{Phe}, and they were separated by SDS–PAGE before being transferred on a PVDF membrane, which was analysed with a radioimager. As observed for Rv2275, wild-type AlbC retained a high amount of the radiolabel, whereas the variant S37A did not retain any radiolabel. The variant S37C presented behaviour similar to that of the wild-type enzyme, confirming the formation of a covalent intermediate and the probable status of S37 as the point of covalent attachment (Figure 6). The variants Y178F and E182Q retained significant amounts of the radiolabel although much less than the wild-type enzyme (Figure 6). This result, combined with the substantial and complete loss of enzyme activity for the variants Y178F and E182Q, respectively (Figure 5D), argues that Y178 and E182 may play a crucial role in the good positioning of the aminoacyl moiety of the substrate in the active site—essential for an efficient formation of the covalent intermediate—concerning the variant Y202F, it retained ~11% of the enzymatic activity (Figure 5D), and at least as much radiolabel as the wild-type enzyme (Figure 6). As the removal of the hydroxyl group in the variant Y202F does not prevent the formation of the covalent intermediate, this demonstrated that Y202 is not responsible for the S37 activation in AlbC.

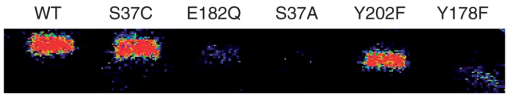


Figure 6. Covalent labelling of AlbC and variants by tritiated Phe transferred from [^{3}H]Phe-tRNA^{Phe}. Enzymes were incubated with [^{3}H]Phe-tRNA^{Phe}, as described in 'Materials and Methods' section, separated on SDS–PAGE, then transferred onto a PVDF membrane that was analysed with a radioimager.

In order to unambiguously identify the residue acylated and the nature of the acylation, we performed peptide mass fingerprint (PMF) analyses on trypsin-digested AlbC, previously incubated or not with Phe-tRNA^{Phe}. AlbC without substrate incubation gave two identified fragments containing the residue S37 (fragments [29–46] and [27–46]) (Supplementary Figure S7A). The corresponding *m*/*z* of these fragments (calculated *m*/*z* 1878.94 and 2163.09, respectively) were selected and isolated as precursor ion for PSD MS/MS sequence characterization (Supplementary Figure S7B and C). The same experiment performed on AlbC incubated with Phe-tRNA^{Phe} gave two additional fragments (Figure 7A). The *m*/*z* of these two fragments were 2026.01 (observed *m*/*z* 2026.00) and 2310.16 (observed *m*/*z* 2310.17), and corresponded to the values expected for the addition of a phenylalanyl moiety (addition of 147.07) on fragments [29–46] and [27–46], respectively. These fragments, sequenced unambiguously by both b- and y-ions series, contained a phenylalanyl moiety on S37 (Figure 7B and Supplementary Figure S8). This demonstrated the formation of a covalent intermediate, in which L-Phe is transferred from Phe-tRNA^{Phe} to S37 during the catalytic cycle. Moreover, the same experiment was performed on relevant AlbC variants. No modified fragments were observed for the variant S37A (Supplementary Figure S9). For the variants E182Q and

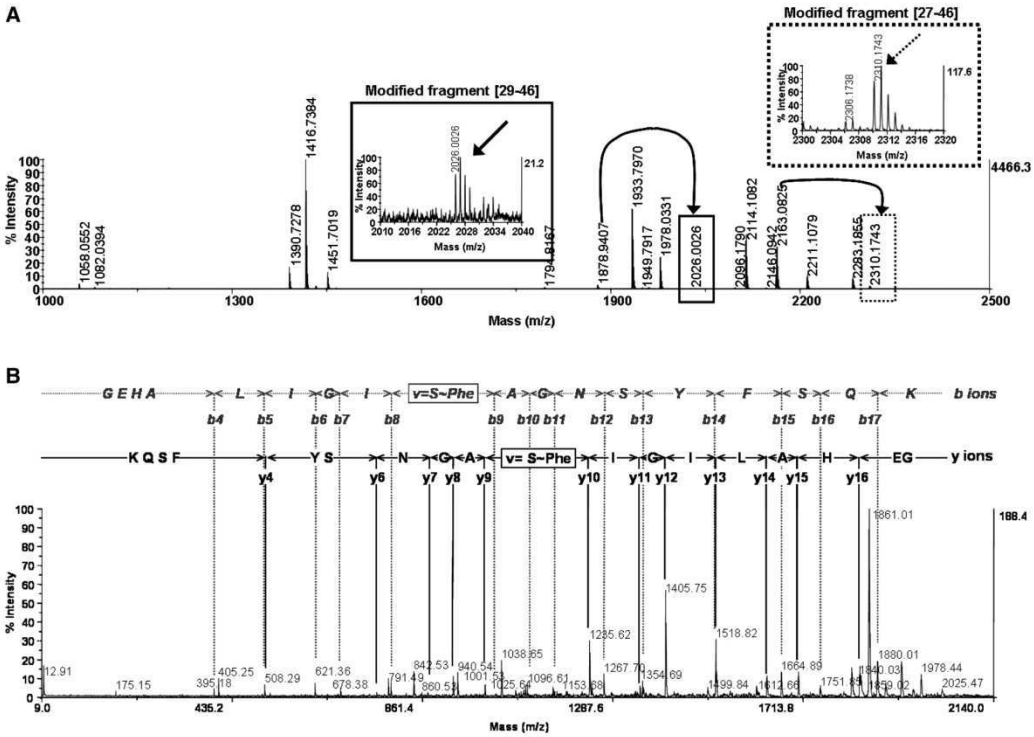


Figure 7. Identification of the covalent aminoacyl-enzyme intermediate. (**A**) Peptide mass fingerprint (PMF) analysis of trypsin-digested AlbC pre-incubated with Phe-tRNA^{Phe}. Arrows indicate the formation of two additional fragments compared to PMF analysis of trypsin-digested AlbC not incubated with Phe-tRNA^{Phe} (Supplementary Figure S7): one at m/z 2026.00 (continuous line frame) corresponding to phenylalanylated fragment [29–46] at m/z 1878.94, and the other at m/z 2310.17 (dashed line frame) corresponding to phenylalanylated fragment [27–46] at m/z 2163.08. (**B**) PSD-MS/MS spectrum of m/z 2026.00. Both b- and y-ions series identify the amino acids sequence of the modified AlbC fragment [29–46], and the residue 37 as the phenylalanylated residue. PSD-MS/MS spectrum of m/z 2310.17 is shown in Supplementary Figure S8.

Y178F, phenylalanylated fragments could be significantly detected, but in much less amounts than those obtained with the wild-type enzyme (Supplementary Figures S10 and S11), in agreement with radiolabelling experiments (Figure 6). These fragments were formed with amounts insufficient for PSD-MS/MS analyses (Supplementary Figures S10 and S11). For the variant Y202F, the phenylalanylated fragments were observed in proportions similar to those obtained with the wild-type enzyme, and were unambiguously sequenced (Supplementary Figure S12). The results confirm that Y202 is not responsible for S37 activation. For the variant S37C, the phenylalanylation of the fragments [27–49] and [29–46] (respectively, calculated m/z of 2040.97 and 2326.13 instead of 1894.90 and 2179.06) was detected (Supplementary Figure S13A). MS/MS sequencing of both m/z demonstrated the formation of a thioester link between the phenylalanyl moiety and the thiol group of C37 (Supplementary Figure S13B and C). All these data allow the proposal of a mechanism for the formation of a covalent intermediate during the catalytic cycle of the CDPS AlbC (see Discussion section).

Distribution and role of the residues conserved among CDPSs but located outside the active site

Seven of the 13 residues conserved among CDPSs are outside the active site: H31, W54, F59, G79, S120, Y128 and C149 (Figure 4 and Supplementary Figure S14). Each of G79, S120 and C149 was substituted with alanine without any significant effect on the activity or production of the corresponding variants amounts (Figure 5D). Variants H31A, W54A and F59A expressed very substantially lower activities than the wild-type enzyme, and were produced in significantly (H31A and F59A) or substantially (W54A) smaller amounts (Figure 5D). The residues H31, W54 and F59 are located in strand β3, helix α2 and loop α2–β4, respectively. The residue H31 is fully buried, and its side chain makes two hydrogen bonds with the residues D63 and S198 that are located in neighbouring strands. The residues W54 and F59 are highly buried and their side chains interact with hydrophobic residues from strand β6 (Supplementary Figure S14A and B). The substitution of each of these three residues with alanine should lead to a significant structural destabilization,

which is consistent with our results showing that the production of the corresponding variants is strongly affected. The substitution of Y128 by alanine did not impair the production of the corresponding variant, but abolished cFL synthesis (Figure 5D). The residue Y128, which is located at the beginning of helix α5, is fully buried and its side chain makes two hydrogen bonds with Q123 from helix η1 and R132 from helix α5 (Supplementary Figure S14A and C). The substitution of Y128 by alanine should alter the relative positions of these helices and neighbouring interactions, resulting in a loss of AlbC activity.

Recognition of the tRNA moiety of the aa-tRNA substrates by AlbC

The electrostatic potential surface of AlbC presents a highly biased distribution of charged residues, which could interact with the tRNA substrate by forming salt bridges with the phosphates of the tRNA backbone or hydrogen bonds with the nucleotide bases. In particular, there is a large patch of positively charged residues covering a surface of 1040 $Å^2$ (Figure 8A). This patch is mostly composed of residues belonging to the helix α4, which contains nine basic residues. These residues, R87, K90, R91, K94, R97, R98, R99, R101 and R102, protrude toward the solvent, and a similar pattern of basic residues is found in all CDPSs (Figures 4 and 8B). We substituted with alanine each of the basic residues in helix α4. All the resulting variants were produced with yields similar to that of the wild-type enzyme but their cFL-synthesizing activities were diverse (Figure 8C). The variants R98A, R99A and R98A/R99A displayed only 19, 8 and <2% of the wild-type activity, respectively. Five other variants (K90A, K94A, R97A, R101A and R102A) had activities that were 53–72% of the wild-type enzyme. The activities of the two remaining variants, R87A and R91A, were similar to that of the wild type. These results showed that the basic residues located in the C-terminal half of the helix α4, especially R98 and R99, are important for cFL production, probably because they interact with the tRNA moiety of the substrate. A second patch of basic residues, consisting of R214, R215 and R220 from loop α8–β7 and R231 from loop β7–α9, is present on AlbC (Figure 8A). We substituted these residues and showed that the corresponding variants displayed properties similar to the wild-type enzyme (Figure 8C), indicating that this patch is not involved in tRNA binding. This is consistent with the fact that this patch is not conserved among CDPSs (Figure 4). Our results suggest that the interaction between AlbC and its aa-tRNA is driven by the binding of the substrate aminoacyl moiety in its deep pocket and the interaction of the tRNA moiety with at least one patch of basic residues on helix α4.

DISCUSSION

We report the 3D structure of AlbC, a member of the recently identified CDPS family. AlbC presents a compact α/β structure with a Rossmann-fold domain, as it was also recently reported for the CDPS Rv2275 (12). We show that, although having many structural similarities, AlbC and Rv2275 have significant differences that raise intriguing questions regarding the catalytic mechanism used by the CDPSs to synthesize cyclodipeptides.

AlbC, and Rv2275, share a high structural similarity with the catalytic domain of class-Ic aaRSs. aaRSs catalyse specific esterifications of particular amino acids to their cognate tRNAs in a two-step reaction. First, amino acids are activated with ATP to form enzyme-bound aminoacyl-adenylate intermediates. Second, the aminoacyl moieties are transferred to tRNAs to form aa-tRNA products. The structural similarities concern only the catalytic domains of class-I aaRSs, which are responsible for the binding of amino acid and ATP substrates. The amino acid-binding pockets of class-I aaRSs superimpose well with the AlbC pocket and we show that this pocket interacts with the aminoacyl moiety of the aminoacyl-tRNA substrate. However, AlbC does not possess the HIGH motif involved in ATP binding that is conserved among all catalytic domains of class-I aa-RSs (36–38). Also, AlbC does not have the loop connecting the catalytic domain to the C-terminal tRNA-binding domain, which contains the second ATP-binding motif —KMSKS—in class-I aa-RSs. This is consistent with the fact that CDPSs do not activate amino acids using ATP, but use already activated amino acids in the forms of aa-tRNAs (1). Another significant difference is that AlbC does not possess the C-terminal domain of aaRSs, which is crucial for their interaction with tRNAs (39).

The AlbC specificity seems to be directed mainly at the aminoacyl moiety, and not at the sequence of the tRNA moiety of the substrate. Indeed, the substitution of one residue of the aminoacyl-binding pocket of AlbC by the corresponding residue of Rv2275 gave a variant that, in *E. coli*, mainly bound tyrosyl-$tRNA^{Tyr}$ rather than phenylalanyl-$tRNA^{Phe}$. Comparing the acceptor stems of the two tRNA moieties revealed that they share the sequences of the single-strand extremity ($A^{76}C^{75}C^{74}A^{73}$) and the first base pair (C^{72}–G^{1}), suggesting that some specificity determinants could lay in these conserved sequences of the tRNA moiety of the substrate. We previously showed that AlbC is able to efficiently use various *E. coli* aa-tRNAs to synthesize various cyclodipeptides (1). We compared the sequences of all *E. coli* aa-tRNAs that can serve as substrates for AlbC (predominantly phenylalanyl- and leucyl-tRNAs, and to a lesser extent methionyl-, tyrosyl-, alanyl- and valyl-tRNAs). For all these tRNA families, we found molecules that also share the same sequence: $A^{76}C^{75}C^{74}A^{73}C^{72}$-$G^{1}$. This might suggest that sequences of the single-strand extremity and the first base pair, but not those of the double-stranded acceptor stem, are specifically recognized by AlbC. Noted that aa-tRNA protein transferases recognize the aa-tRNA in a tRNA sequence-independent manner, the double-stranded acceptor stem being fully dispensable (40).

An interesting question concerns the tRNA-binding mode of AlbC. Could it be similar to that described for class-Ic aaRSs? Recognition of tRNAs by TyrRSs and TrpRSs mainly involves their C-terminal tRNA-binding domains, essential for tRNA binding (39). However, several residues belonging to helices α7 and α9 of their

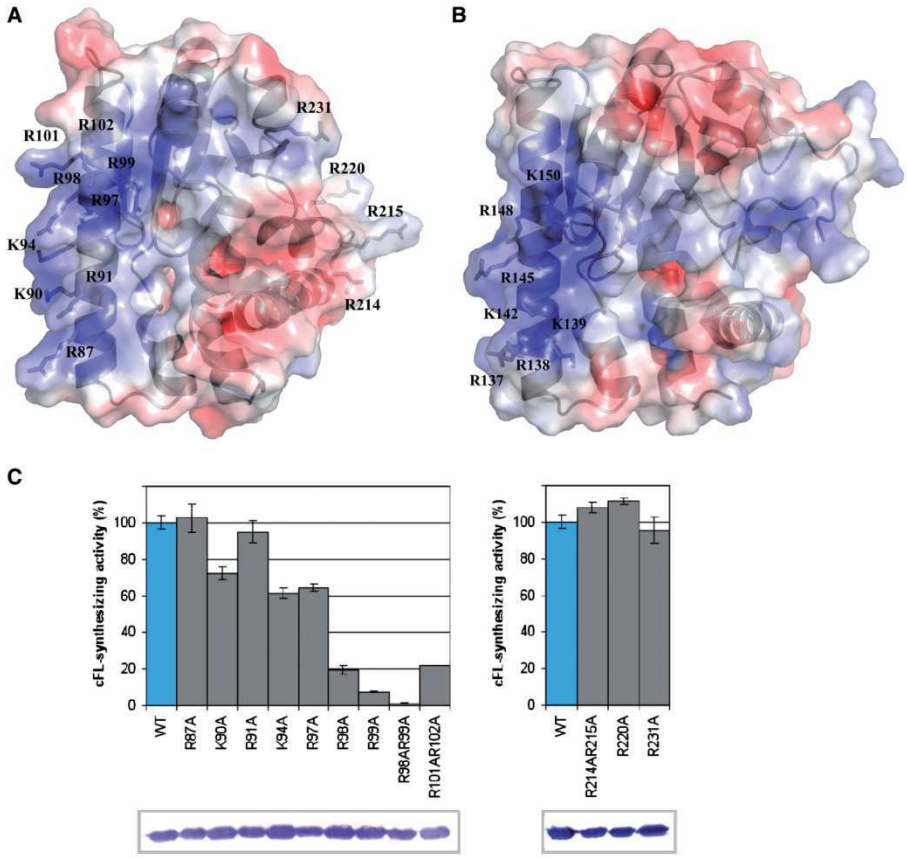


Figure 8. Basic patch residues contributing to CDPS activity. (**A**) Electrostatic surface potential of AlbC, mapped on its solvent-accessible surface at contouring levels of ± 5 kTe–1. Positive charge is in blue, negative charge in red. The potential was calculated using APBS in PyMol (59) (http://www.pymol.org/) from a derived model of AlbC where the missing side chain of residue R215 and the missing residue T217 were added. AlbC structure is shown in grey and the side chains of all mutated basic residues are identified as sticks. (**B**) Electrostatic surface potential of Rv2275 generated like that of AlbC. (**C**) Site-directed mutagenesis study of AlbC. cFL-synthesizing activity of the wild-type AlbC (in blue) and of each of the variants in which a basic residue is substituted (in grey) are shown with error bars. The corresponding western blots indicating amounts of the proteins are also shown. Left, the residues belonging to the helix α4; right, other basic residues.

N-terminal catalytic domains interact with the acceptor stems of tRNAs (Figure 4) (29,32,41–46). AlbC does not possess a C-terminal tRNA-binding domain like those in aaRSs nor any residue identical to those of the N-terminal catalytic domain involved in the tRNA interaction (Figure 4). Otherwise, we showed that a patch of basic residues, which is conserved among CDPSs but not present in class-I aa-RSs, is important for AlbC activity. This patch may interact with the phosphate backbone of the tRNA body, as proposed for aa-tRNA-protein transferases and Fem aa-transferases (47). Comparing the position of the basic patch in AlbC with that of the acceptor stem-binding region in $TyrRS_{Mj}$ revealed that they are different (Supplementary Figure S15). All these data might suggest that the tRNA-binding mode of AlbC is different from that of class-I aa-RSs, but this issue remains to be addressed.

We wondered how AlbC uses two tRNA substrates to form cyclodipeptides. One possibility is that AlbC acts as a homodimer, each monomer interacting with one aa-tRNA. Although AlbC was obtained as a monomer both in solution (1) and in crystal forms, we cannot exclude that aa-tRNA binding induces dimerization. Noted that class-Ic aaRSs act as functional dimers, but they are also obtained as dimers in crystals (44). Class-Ic aaRSs dimers are mainly formed by the association of a hydrophobic surface involving residues of the CP1 domain of each catalytic domain (34,35). The positions of these residues are not conserved in AlbC. In addition, when we mimicked this association for two AlbC monomers by superimposing them on different crystal structures of TyrRSs or TrpRSs, we observed major steric hindrances between helices α5, α6 and α7 of each AlbC monomer. Therefore, it is unlikely that AlbC dimerizes in a manner

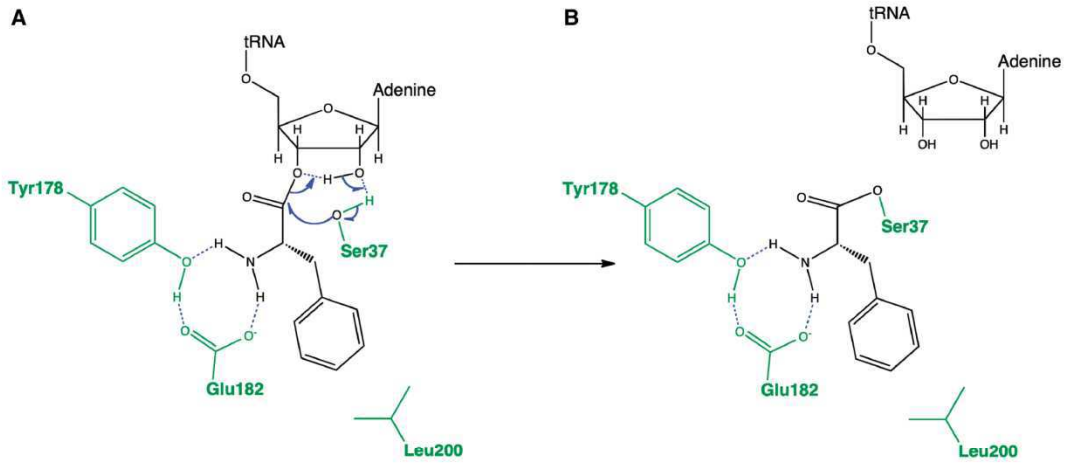


Figure 9. Proposed mechanism of covalent phenylalanyl-enzyme formation for AlbC. (**A**) Proposed anchoring of the Phe-tRNA^{Phe} substrate at the AlbC active site. The residues S37, Y178 and E182, which are essential for the covalent intermediate formation (Figures 6 and 7), are represented in green. The residue L200, shown to interact with the phenylalanyl moiety of the substrate (Figure 5E), is also represented in green. As no residue in the AlbC active site is likely to deprotonate the hydroxyl group of S37, the S37 activation is proposed to be achieved by a concerted proton shuttling mechanism involving the two adjacent vicinal hydroxyls of the nucleotide A^{76} of the tRNA moiety. (**B**) The resulting covalent phenylalanyl-enzyme.

similar to class-Ic aaRSs, and furthermore, any such dimerization would result in a physical separation of the two AlbC catalytic pockets, inconsistent with the formation of two peptide bonds. This is indeed what was observed for Rv2275, which is found as a homodimer: the residues that make up the interface belong to other secondary structural elements than those involved in the dimerization of class-Ic aaRSs, and the two Rv2275 catalytic pockets are not at the dimer interface, but are physically separated (12). Moreover, it is unlikely that AlbC dimerizes in a manner similar to Rv2275 since the residues that are involved in the Rv2275 interface are not conserved in AlbC and other CDPSs. As it seems more probable that AlbC is active as a monomer, we considered whether AlbC catalyses its two-substrate reaction using a ternary complex or a ping-pong mechanism. The size of the AlbC pocket does not appear appropriate to accommodate the aminoacyl moieties of two aa-tRNA substrates concomitantly. Thus, one aa-tRNA may bind to AlbC and its aminoacyl moiety would be stored as a covalent intermediate, i.e. as an aminoacyl-enzyme intermediate. The formation of a covalent intermediate resulting from the transfer of L-Tyr from Tyr-tRNA^{Tyr} to an active serine was suggested for Rv2275, but such a transfer was not conclusively demonstrated (12). We demonstrated that this transfer occurs in AlbC because we showed that the active serine residue is covalently bound to the phenylalanyl transferred from Phe-tRNA^{Phe} during the catalytic cycle. An intriguing point that remains to be determined concerns the activation of the catalytic serine, i.e. the deprotonation of its hydroxyl group serving as the nucleophile. For Rv2275, it was suggested that the catalytic serine (S88) residue is activated by a conserved tyrosine residue (Y253). This is consistent with the hydrogen bonding of Y253 to the side chain hydroxyl group of S88, and the much lower (by a factor of 200) catalytic activity of the Y253F variant (12). For AlbC, the corresponding tyrosine residue (Y202) adopts a different conformation in the AlbC structure, and the Y202F variant displays only a 10-fold decrease in catalytic activity. We also unambiguously showed that the formation of the covalent aminoacyl intermediate is not affected by the substitution of Y202 with phenylalanine, demonstrating that Y202 is not responsible for the activation of S37 in AlbC. As no other residue in the vicinity of S37 can account for its activation, an attractive hypothesis is that the serine activation results from a concerted proton shuttling mechanism involving the two adjacent vicinal hydroxyls of the nucleotide A^{76} of the tRNA moiety (Figure 9), as it has been proposed for the ribosomal peptididyl transferase (48) and the FemX aa-transferase of *Weissella viridescens* (49). At this stage, we cannot exclude some rearrangements in AlbC active site during the different steps of the reaction. The analysis of B-factors around AlbC active site suggests that the loop β3–α2, which precedes the position S37, may present some flexibility. The crystal structure of AlbC complexed with different intermediates of the reaction should clarify this point. But, whatever how the serine activation is performed, we demonstrated that CDPSs use a ping-pong mechanism involving a covalent aminoacyl-enzyme intermediate. The aminoacyl-enzyme may then react with the aminoacyl moiety of the second aa-tRNA to form a dipeptidyl-enzyme or dipeptidyl-tRNA intermediate, which may then undergo intramolecular cyclization leading to the formation of the cyclodipeptide product.

In conclusion, we report the crystal structure of AlbC, a CDPS from *S. noursei*. It confirms that CDPSs are structurally similar to the catalytic domain of class-I aaRSs. This structural similarity has provided clues about the

molecular bases of the interactions between CDPSs and their aa-tRNAs substrates. Moreover, we demonstrate that CDPSs catalyse their two-substrate reactions via a ping-pong mechanism involving a covalent aminoacyl-enzyme intermediate. CDPSs provide a novel example of the exaptation of aaRSs in processes not related to ribosomal protein synthesis. Some atypical aaRSs and aaRS-like proteins have already been shown to be involved in amino acid biosynthesis, RNA modification (50,51), and in non-ribosomal peptide synthesis (52,53). These proteins include enzymes having the catalytic domain of an aaRS and lacking its tRNA-binding domain. Thus, the archaeal Asn synthetase, a paralog of AsnRS, does not form Asn-$tRNA^{Asn}$, but synthesizes Asn by amidation of the β-carboxyl group of aspartate with ammonia (54). YadB (renamed glutamyl-queuosine $tRNA^{Asp}$ synthetase), a paralog of GluRS, catalyses glutamylation of queuosine in the wobble position of $tRNA^{Asp}$ (55,56). Amino acid:[carrier protein] ligases, which are SerRS homologues, catalyse the ATP-dependent activation of particular amino acids and their transfer to the phosphopantetheine prosthetic group of carrier proteins (57). Finally, CDPSs differ from these other aaRS-like proteins in that they have not only diverged from conventional aaRSs, but they have also acquired new active site residues, converting them into self-contained cyclodipeptide-forming enzymes, which can use aa-tRNAs as substrates.

ACCESSION NUMBER

PDB, 3OQV.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank P. Cuniasse, J. Baillon, and S. Lautru for helpful discussions. We thank E. Stura, A. Ducruix, and M. Ries-Kaut for helpful advice on crystallogenesis, and R. Guerois for help in bioinformatic. We thank S. Bregant and B. Czarny for advice on the use of the Beta-imager. We thank M. Delarue for the critical reading of the article. We are grateful to J. Marquez for access to the P-CUBE HTX platform (EMBL, Grenoble) and to F. Fenaille for access to mass spectrometry facilities (IBITECS/SPI).

FUNDING

Commissariat à l'Energie Atomique; Centre National de la Recherche Scientifique, Université Paris-Sud 11; Agence Nationale de la Recherche (ANR-10-1501; CyDiPepS); Commissariat à l'Energie Atomique, doctoral fellowships (to L.S. and Y.L.). Funding for open access charge: Commissariat à l'Energie Atomique, IBITECS/SIMOPRO.

Conflict of interest statement. None declared.

REFERENCES

1. Gondry,M., Sauguet,L., Belin,P., Thai,R., Amouroux,R., Tellier,C., Tuphile,K., Jacquet,M., Braud,S., Courçon,M. *et al.* (2009) Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. *Nat. Chem. Biol.*, **5**, 414–420.
2. Gondry,M., Lautru,S., Fusai,G., Meunier,G., Ménez,A. and Genet,R. (2001) Cyclic dipeptide oxidase from *Streptomyces noursei*. Isolation, purification and partial characterization of a novel, amino acyl alpha,beta-dehydrogenase. *Eur. J. Biochem.*, **268**, 1712–1721.
3. Belin,P., Le Du,M.H., Fielding,A., Lequin,O., Jacquet,M., Charbonnier,J.B., Lecoq,A., Thai,R., Courçon,M., Masson,C. *et al.* (2009) Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA*, **106**, 7426–7431.
4. von Döhren,H. (2009) Charged tRNAs charge into secondary metabolism. *Nat. Chem. Biol.*, **5**, 374–375.
5. RajBhandary,U.L. and Söll,D. (2008) Aminoacyl-tRNAs, the bacterial cell envelope, and antibiotics. *Proc. Natl Acad. Sci. USA*, **105**, 5285–5286.
6. Banerjee,R., Chen,S., Dare,K., Gilreath,M., Praetorius-Ibba,M., Raina,M., Reynolds,N.M., Rogers,T., Roy,H., Yadavalli,S.S. *et al.* (2010) tRNAs: cellular barcodes for amino acids. *FEBS Lett.*, **584**, 387–395.
7. Francklyn,C.S. and Minajigi,A. (2010) tRNA as an active chemical scaffold for diverse chemical transformations. *FEBS Lett.*, **584**, 366–375.
8. Mainardi,J.L., Villet,R., Bugg,T.D., Mayer,C. and Arthur,M. (2008) Evolution of peptidoglycan biosynthesis under the selective pressure of antibiotics in Gram-positive bacteria. *FEMS Microbiol. Rev.*, **32**, 386–408.
9. Watanabe,K., Toh,Y., Suto,K., Shimizu,Y., Oka,N., Wada,T. and Tomita,K. (2007) Protein-based peptide-bond formation by aminoacyl-tRNA protein transferase. *Nature*, **449**, 867–871.
10. Benson,T.E., Prince,D.B., Mutchler,V.T., Curry,K.A., Ho,A.M., Sarver,R.W., Hagadorn,J.C., Choi,G.H. and Garlick,R.L. (2002) X-ray crystal structure of *Staphylococcus aureus* FemA. *Structure*, **10**, 1107–1115.
11. Biarrotte-Sorin,S., Maillard,A.P., Delettré,J., Sougakoff,W., Arthur,M. and Mayer,C. (2004) Crystal structures of *Weissella viridescens* FemX and its complex with UDP-MurNAc-pentapeptide: insights into FemABX family substrates recognition. *Structure*, **12**, 257–267.
12. Vetting,M.W., Hegde,S.S. and Blanchard,J.S. (2010) The structure and mechanism of the *Mycobacterium tuberculosis* cyclodityrosine synthetase. *Nat. Chem. Biol.*, **6**, 797–799.
13. Eriani,G., Delarue,M., Poch,O., Gangloff,J. and Moras,D. (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*, **347**, 203–206.
14. Cusack,S., Berthet-Colominas,C., Härtlein,M., Nassar,N. and Leberman,R. (1990) A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature*, **347**, 249–255.
15. Arnez,J.G. and Moras,D. (1997) Structural and functional considerations of the aminoacylation reaction. *Trends Biochem. Sci.*, **22**, 211–216.
16. Ibba,M. and Söll,D. (2000) Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.*, **69**, 617–650.
17. Braud,S., Moutiez,M., Belin,P., Abello,N., Drevet,P., Zinn-Justin,S., Courçon,M., Masson,C., Dassa,J., Charbonnier,J.B. *et al.* (2005) Dual expression system suitable for high-throughput fluorescence-based screening and production of soluble proteins. *J. Proteome Res.*, **4**, 2137–2147.
18. Stura,E.A., Satterthwait,A.C., Calvo,J.C., Kaslow,D.C. and Wilson,I.A. (1994) Reverse screening. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 448–455.
19. Ducruix,A. and Giegé,R. (1999) *Crystallization of Nucleic Acids and Proteins: A Practical Approach*. Oxford University Press, Oxford.
20. Sheldrick,G.M. (2008) A short history of SHELX. *Acta Crystallogr. A*, **64**, 112–122.
21. Vonrhein,C., Blanc,E., Roversi,P. and Bricogne,G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.*, **364**, 215–230.

22. Langer,G., Cohen,S.X., Lamzin,V.S. and Perrakis,A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.*, **3**, 1171–1179.
23. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
24. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B. 3rd, Snoeyink,J., Richardson,J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–W383.
25. Krissinel,E. and Henrick,K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2256–2268.
26. Lee,B. and Richards,F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**, 379–400.
27. Rao,S.T. and Rossmann,M.G. (1973) Comparison of super-secondary structures in proteins. *J. Mol. Biol.*, **76**, 241–256.
28. Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.
29. Kobayashi,T., Nureki,O., Ishitani,R., Yaremchuk,A., Tukalo,M., Cusack,S., Sakamoto,K. and Yokoyama,S. (2003) Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol.*, **10**, 425–432.
30. Zhang,Y., Wang,L., Schultz,P.G. and Wilson,I.A. (2005) Crystal structures of apo wild-type *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O-methyl-L-tyrosine. *Protein Sci.*, **14**, 1340–1349.
31. Kuratani,M., Sakai,H., Takahashi,M., Yanagisawa,T., Kobayashi,T., Murayama,K., Chen,L., Liu,Z.J., Wang,B.C., Kuroishi,C. *et al.* (2006) Crystal structures of tyrosyl-tRNA synthetases from Archaea. *J. Mol. Biol.*, **355**, 395–408.
32. Shen,N., Guo,L., Yang,B., Jin,Y. and Ding,J. (2006) Structure of human tryptophanyl-tRNA synthetase in complex with $tRNA^{Trp}$ reveals the molecular basis of tRNA recognition and specificity. *Nucleic Acids Res.*, **34**, 3246–3258.
33. Zhou,M., Dong,X., Shen,N., Zhong,C. and Ding,J. (2010) Crystal structures of *Saccharomyces cerevisiae* tryptophanyl-tRNA synthetase: new insights into the mechanism of tryptophan activation and implications for anti-fungal drug design. *Nucleic Acids Res.*, **38**, 3399–3413.
34. Brick,P. and Blow,D.M. (1987) Crystal structure of a deletion mutant of a tyrosyl-tRNA synthetase complexed with tyrosine. *J. Mol. Biol.*, **194**, 287–297.
35. Doublié,S., Bricogne,G., Gilmore,C. and Carter,C.W. Jr (1995) Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. *Structure*, **3**, 17–31.
36. Webster,T., Tsai,H., Kula,M., Mackie,G.A. and Schimmel,P. (1984) Specific sequence homology and three-dimensional structure of an aminoacyl transfer RNA synthetase. *Science*, **226**, 1315–1317.
37. Ludmerer,S.W. and Schimmel,P. (1987) Gene for yeast glutamine tRNA synthetase encodes a large amino-terminal extension and provides a strong confirmation of the signature sequence for a group of the aminoacyl-tRNA synthetases. *J. Biol. Chem.*, **262**, 10801–10806.
38. Rould,M.A., Perona,J.J., Söll,D. and Steitz,T.A. (1989) Structure of *E. coli* glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution. *Science*, **246**, 1135–1142.
39. Waye,M.M., Winter,G., Wilkinson,A.J. and Fersht,A.R. (1983) Deletion mutagenesis using an 'M13 splint': the N-terminal structural domain of tyrosyl-tRNA synthetase (*B. stearothermophilus*) catalyses the formation of tyrosyl adenylate. *EMBO J.*, **2**, 1827–1829.
40. Abramochkin,G. and Shrader,T.E. (1996) Aminoacyl-tRNA recognition by the leucyl/phenylalanyl-tRNA-protein transferase. *J. Biol. Chem.*, **271**, 22901–22907.
41. Nair,S., Ribas de Pouplana,L., Houman,F., Avruch,A., Shen,X. and Schimmel,P. (1997) Species-specific tRNA recognition in relation to tRNA synthetase contact residues. *J. Mol. Biol.*, **269**, 1–9.
42. Bedouelle,H. and Winter,G. (1986) A model of synthetase/transfer RNA interaction as deduced by protein engineering. *Nature*, **320**, 371–373.
43. Labouze,E. and Bedouelle,H. (1989) Structural and kinetic bases for the recognition of $tRNA^{Tyr}$ by tyrosyl-tRNA synthetase. *J. Mol. Biol.*, **205**, 729–735.
44. Yaremchuk,A., Kriklivyi,I., Tukalo,M. and Cusack,S. (2002) Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. *EMBO J.*, **21**, 3829–3840.
45. Tsunoda,M., Kusakabe,Y., Tanaka,N., Ohno,S., Nakamura,M., Senda,T., Moriguchi,T., Asai,N., Sekine,M., Yokogawa,T. *et al.* (2007) Structural basis for recognition of cognate tRNA by tyrosyl-tRNA synthetase from three kingdoms. *Nucleic Acids Res.*, **35**, 4289–4300.
46. Yang,X.L., Otero,F.J., Ewalt,K.L., Liu,J., Swairjo,M.A., Köhrer,C., RajBhandary,U.L., Skene,R.J., McRee,D.E. and Schimmel,P. (2006) Two conformations of a crystalline human tRNA synthetase-tRNA complex: implications for protein synthesis. *EMBO J.*, **25**, 2919–2929.
47. Suto,K., Shimizu,Y., Watanabe,K., Ueda,T., Fukai,S., Nureki,O. and Tomita,K. (2006) Crystal structures of leucyl/phenylalanyl-tRNA-protein transferase and its complex with an aminoacyl-tRNA analog. *Embo J.*, **25**, 5942–5950.
48. Beringer,M. and Rodnina,M.V. (2007) The ribosomal peptidyl transferase. *Mol. Cell*, **26**, 311–321.
49. Fonvielle,M., Chemama,M., Lecerf,M., Villet,R., Busca,P., Bouhss,A., Ethève-Quelquejeu,M. and Arthur,M. (2010) Decoding the logic of the tRNA regiospecificity of nonribosomal FemX(Wv) aminoacyl transferase. *Angew. Chem. Int. Ed. Engl.*, **49**, 5115–5119.
50. Schimmel,P. and Ribas De Pouplana,L. (2000) Footprints of aminoacyl-tRNA synthetases are everywhere. *Trends Biochem. Sci.*, **25**, 207–209.
51. Francklyn,C., Perona,J.J., Puetz,J. and Hou,Y.M. (2002) Aminoacyl-tRNA synthetases: versatile players in the changing theater of translation. *RNA*, **8**, 1363–1372.
52. Sareen,D., Steffek,M., Newton,G.L. and Fahey,R.C. (2002) ATP-dependent L-cysteine:1D-myo-inosityl 2-amino-2-deoxy-alpha-D-glucopyranoside ligase, mycothiol biosynthesis enzyme MshC, is related to class I cysteinyl-tRNA synthetases. *Biochemistry*, **41**, 6885–6890.
53. Aravind,L., de Souza,R.F. and Iyer,L.M. (2010) Predicted class-I aminoacyl tRNA synthetase-like proteins in non-ribosomal peptide synthesis. *Biol. Direct*, **5**, 48.
54. Roy,H., Becker,H.D., Reinbolt,J. and Kern,D. (2003) When contemporary aminoacyl-tRNA synthetases invent their cognate amino acid metabolism. *Proc. Natl Acad. Sci. USA*, **100**, 9837–9842.
55. Dubois,D.Y., Blaise,M., Becker,H.D., Campanacci,V., Keith,G., Giegé,R., Cambillau,C., Lapointe,J. and Kern,D. (2004) An aminoacyl-tRNA synthetase-like protein encoded by the *Escherichia coli yadB* gene glutamylates specifically $tRNA^{Asp}$. *Proc. Natl Acad. Sci. USA*, **101**, 7530–7535.
56. Blaise,M., Olieric,V., Sauter,C., Lorber,B., Roy,B., Karmakar,S., Banerjee,R., Becker,H.D. and Kern,D. (2008) Crystal structure of glutamyl-queuosine $tRNA^{Asp}$ synthetase complexed with L-glutamate: structural elements mediating tRNA-independent activation of glutamate and glutamylation of $tRNA^{Asp}$ anticodon. *J. Mol. Biol.*, **381**, 1224–1237.
57. Mocibob,M., Ivic,N., Bilokapic,S., Maier,T., Luic,M., Ban,N. and Weygand-Durasevic,I. (2010) Homologs of aminoacyl-tRNA synthetases acylate carrier proteins and provide a link between ribosomal and nonribosomal peptide synthesis. *Proc. Natl Acad. Sci. USA*, **107**, 14585–14590.
58. Petrek,M., Otyepka,M., Banás,P., Kosinová,P., Koca,J. and Damborský,J. (2006) CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics*, **7**, 316.
59. Baker,N.A., Sept,D., Joseph,S., Holst,M.J. and McCammon,J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.

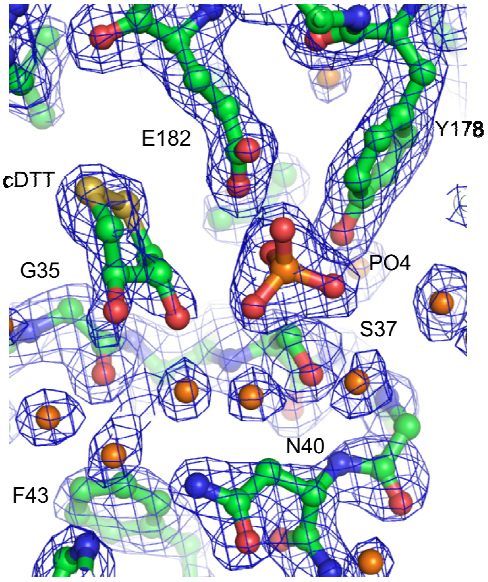
SUPPLEMENTARY INFORMATIONS

Supplementary Table 1. First hits of the Dali search of the AlbC structure against the Protein Data Bank.

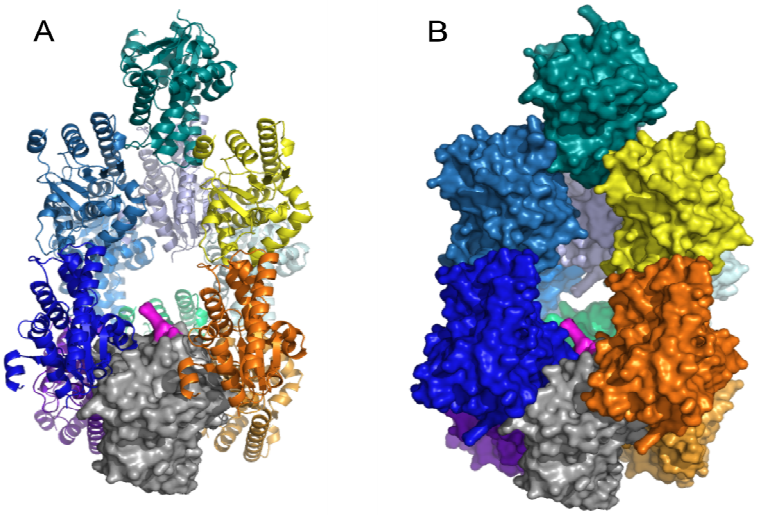
Dali number	Protein number	PDB id	Z	rmsd	lali	nres	%id	Protein	Organism
1	1	2cyb-B	10.7	3.3	162	317	12	TyrRS	*Archaeglobus fulgidus*
3	2	3hzr-E	10.5	3.0	161	374	6	TrpRS	*Entamoeba histolytica*
6	*3*	*1zh6-A*	*10.5*	*3.5*	*160*	*307*	*9*	*engineered TyrRS*	*Methanococcus jannaschii*
8	4	2cyc-A	10.4	3.3	164	371	12	TyrRS	*Pyrococcus horikoshii*
20	5	3foc-A	9.7	3.1	165	404	5	TrpRS	*Giardia lamblia*
23	6	1q11-A	9.6	3.3	166	329	11	fragment TyrRS	*Homo sapiens*
24	7	1i6l-A	9.6	3.1	159	326	10	TrpRS	*Geobacillus stearothermophilus*
28	8	2ake-A	9.5	3.3	162	373	6	TrpRS	*Homo sapiens*
37	9	2j5b-B	9.3	3.1	147	321	12	TyrRS	*Acanthomoeba polyphaga mimivirus*
39	10	3hv0-A	9.2	3.3	160	353	6	TrpRS	*Cryptosporidium parvum*
44	11	3i05-B	9.1	3.2	161	331	8	TrpRS	*Trypanosoma brucei*
45	12	2yy5-B	9.1	3.1	158	346	16	TrpRS	*Mycoplasma pneumoniae*
47	13	2ip1-A	9.1	3.2	162	371	8	TrpRS	*Saccharomyces cerevisiae*
50	14	2el7-A	9.1	3.1	157	318	11	TrpRS	*Thermus thermophilus*
55	15	2dlc-X	9.0	3.4	167	339	14	TyrRS	*Saccharomyces cerevisiae*
58	16	2g36-A	8.9	3.1	153	323	11	TrpRS	*Thermotoga maritima*
85	17	1yi8-A	8.4	3.5	160	331	9	TrpRS	*Deinococcus radiodurans*
92	18	1jii-A	8.3	4.0	161	319	6	TyrRS	*Staphylococcus aureus*
104	19	1tya-E	7.9	3.6	155	317	8	TyrRS	*Geobacillus stearothermophilus*
106	20	3jxe-A	7.9	3.3	159	361	8	TrpRS	*Pyrococcus horikoshii*
116	21	2jan-A	7.6	4.1	159	419	9	TyrRS	*Mycobacterium tuberculosis*
118	22	1h3f-B	7.5	4.1	152	407	16	TyrRS	*Thermus thermophilus*
120	23	1y42-X	7.3	3.8	152	370	11	TyrRS	*Neurospora crassa*
122	**24**	**1q77-A**	**7.2**	**3.3**	**122**	**138**	**11**	**Putative universal stress protein**	***Aquifex aeolicus***

Redundancies, *i.e.* hits that correspond to the different structures of the same protein, have been removed. Hits that correspond to the structures of TyrRSs and TrpRSs are indicated by an orange or yellow background, respectively. TyrRS from *M. jannaschii* (PDB id: 1j1u) has the Dali number 13 and the protein number 3.

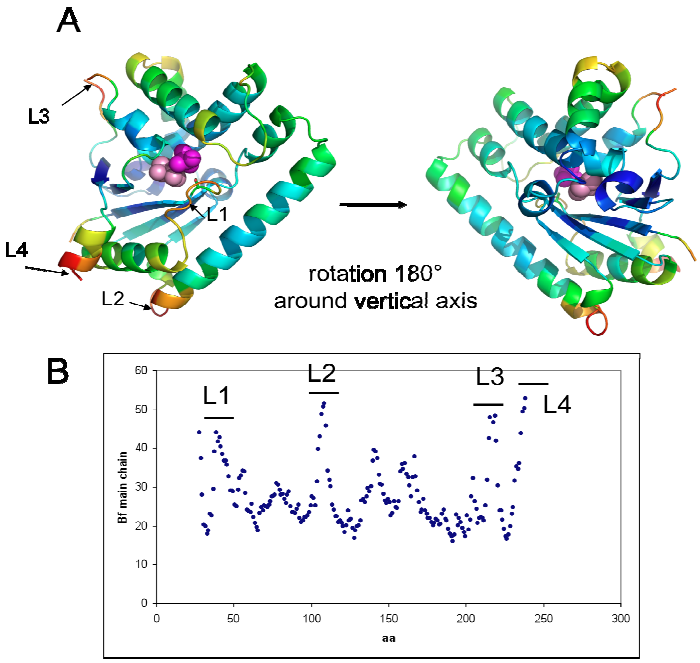
SUPPLEMENTAL FIGURES



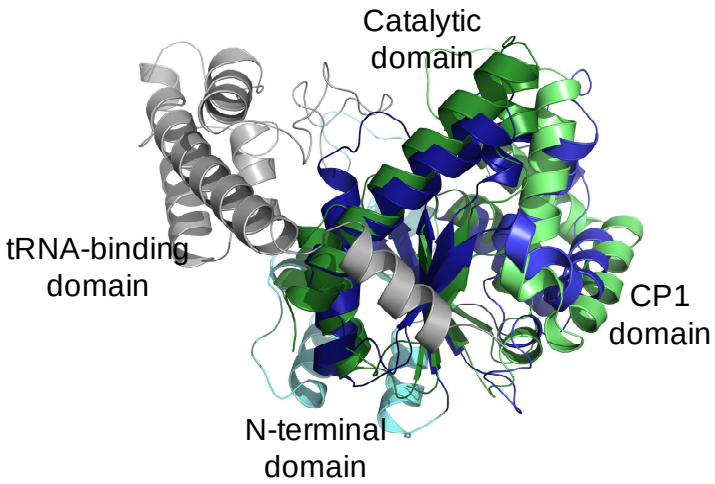
Supplementary Figure S1. Electron density map in AlbC pocket. Residues from AlbC pocket are represented in ball and stick and coloured in green. A molecule of cyclodithiothreitol and a molecule of phosphate have been modelled in the residual electron density observed in the pocket. Water molecules presented in the pocket are represented in orange sphere. The electron density corresponds to a 2Fo-Fc map contoured at one sigma.



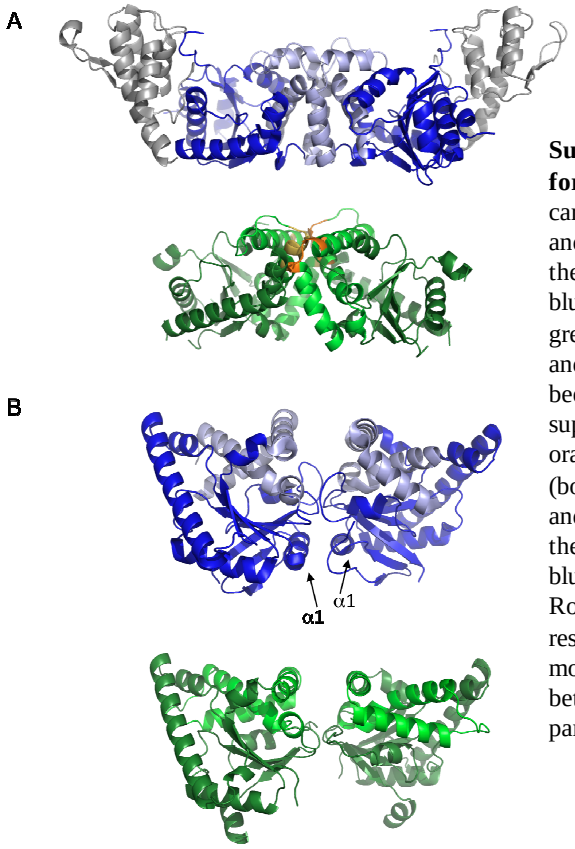
Supplementary Figure S2. Positioning of the flexible 27 amino acids extension in the crystal packing. An AlbC monomer is represented in grey surface with the first four N-terminal residues visible in the electron density map shown in magenta. Twelve symmetric molecules surrounding the N-terminal region of AlbC are presented either in cartoon (**A**) or in surface (**B**) representation. The AlbC N-terminus is located in a large cavity that is about 45 Å large in each direction. Three other symmetric molecules have their N-terminus directed towards this cavity. The dimension of this cavity about 4 10^5 $Å^3$ is fully compatible with the presence of four flexible 27 amino acids extensions.



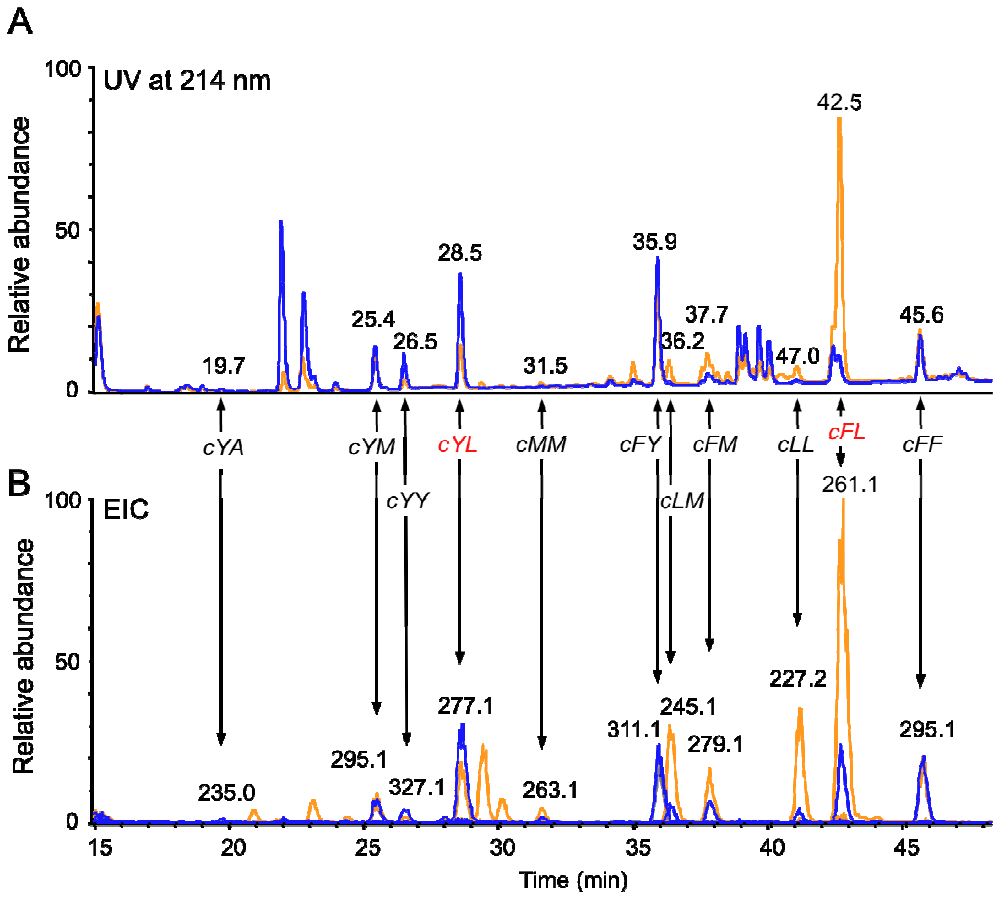
Supplementary Figure S3. Analysis of AlbC B-factors. (**A**) Cartoon representation of AlbC with residues coloured from high B-factors (red) to low B-factors (blue). The four loops with highest B-factors (L1: A38-K46; L2: A105-E111; L3: R215-R220; L4: A236-A239) are indicated by arrows. The ligands observed in the active site are represented in sphere mode and coloured in magenta (phosphate) and pink (cDTT). (**B**) Graph representing the B-factors along the AlbC sequence. The four loops with highest B-factors are indicated.



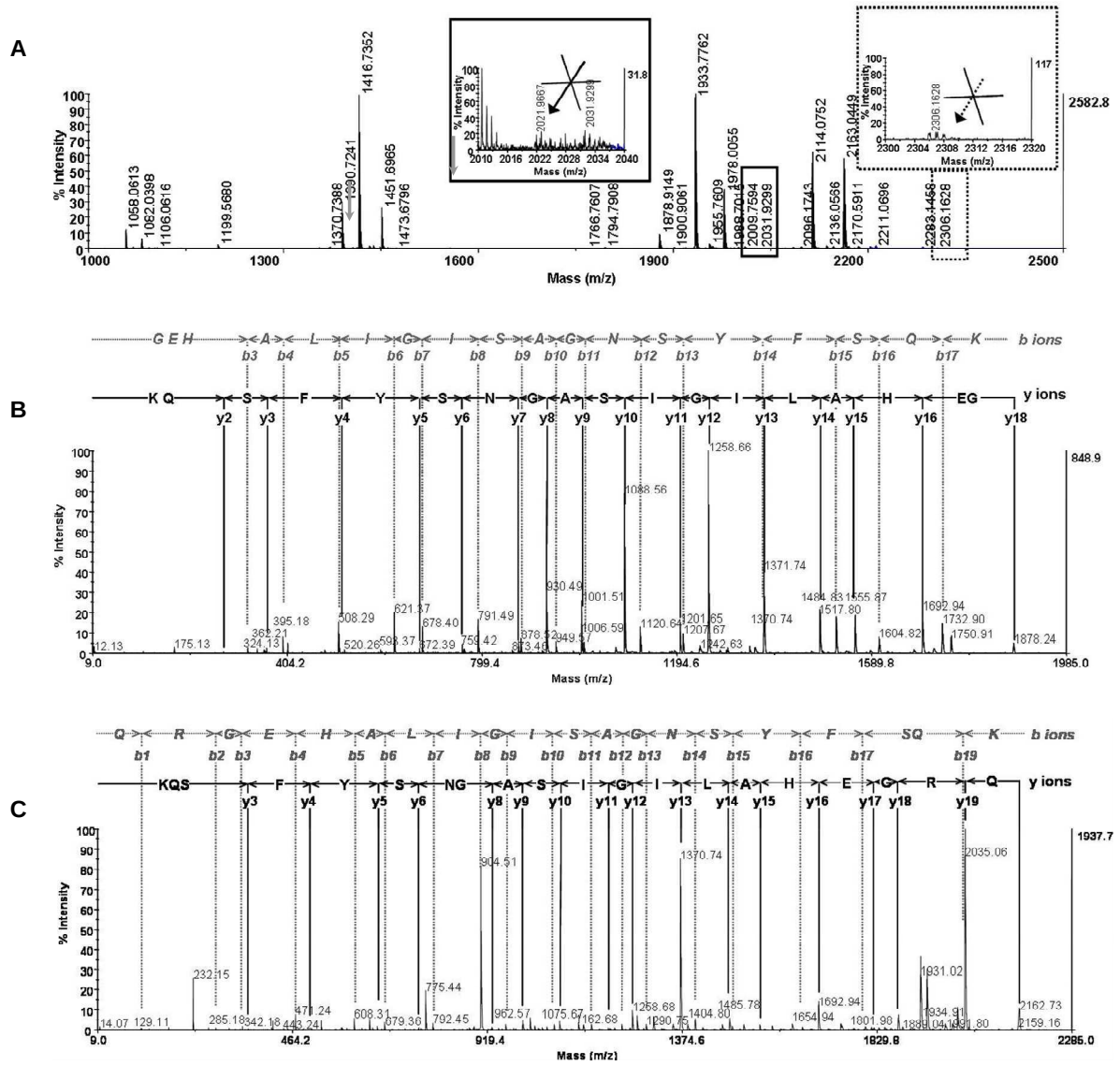
Supplementary Figure S4. Superimposition of the structures of AlbC and *E. histolytica* TrpRS (pdb 3hzr). The two enzymes are shown in cartoon mode. The Rossmann-fold and the CP1 domains of *E. histolytica* TrpRS are coloured in dark blue and blue, respectively, and the corresponding domains of AlbC are coloured in dark green and green, respectively. Rossmann-fold and CP1 domains of the two proteins have similar rmsd (respectively 3.07 Å over 91 Ca and 3.17 Å over 55 Ca). *E. histolytica* TrpRS possesses two additional regions compared to AlbC: a C-terminal domain involved in tRNA-binding and anticodon recognition coloured in grey, and a short N-terminal region coloured in light blue.



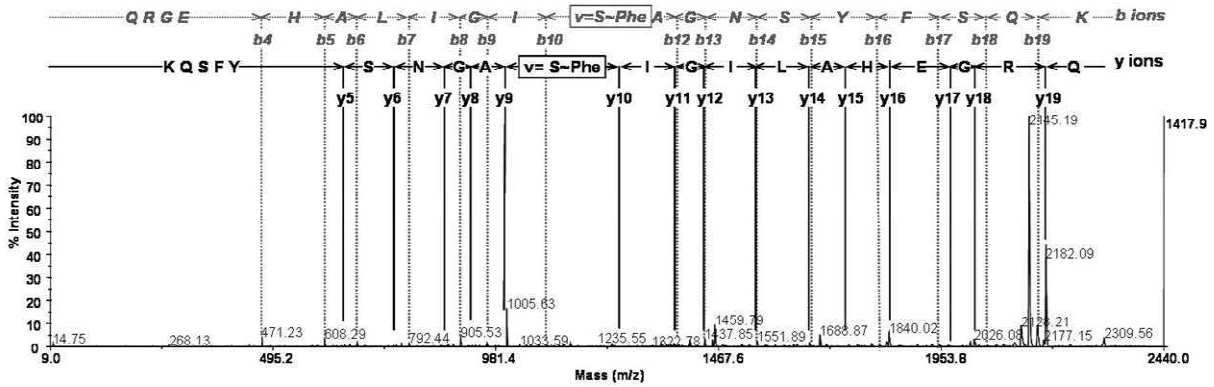
Supplementary Figure S5. Comparison of oligomeric states for AlbC, Rv2275, and TyrRS$_{Mj}$. The enzymes are shown in cartoon mode. (**A**) Comparison for TyrRS$_{MJ}$ (PDB id: 1j1u) and AlbC (PDB id: 3oqv). The TyrRS$_{Mj}$ dimer is shown with the Rossmann-fold and the CP1 domains in dark and light blue, respectively. The tRNA-binding domain is represented in grey (top view). The AlbC monomer, with the Rossmann-fold and the CP1 domains in dark and light green respectively, has been superimposed on each of TyrRS$_{Mj}$ monomer. The superimposition shows major steric hindrance, coloured in orange, between helices α5, α6 and α7 of each AlbC monomer (bottom view). (**B**) Comparison for Rv2275 (PDB id: 2x9q) and AlbC (PDB id: 3oqv). The Rv2275 dimer is shown with the Rossmann-fold and the CP1 domains in dark and light blue, respectively (top view). The AlbC monomer, with the Rossmann-fold and the CP1 domains in dark and light green respectively, has been superimposed on each of Rv2275 monomer. The superimposition shows smaller interface between AlbC monomer that for that of Rv2275, due in particular to the absence of helix α1 (bottom view).



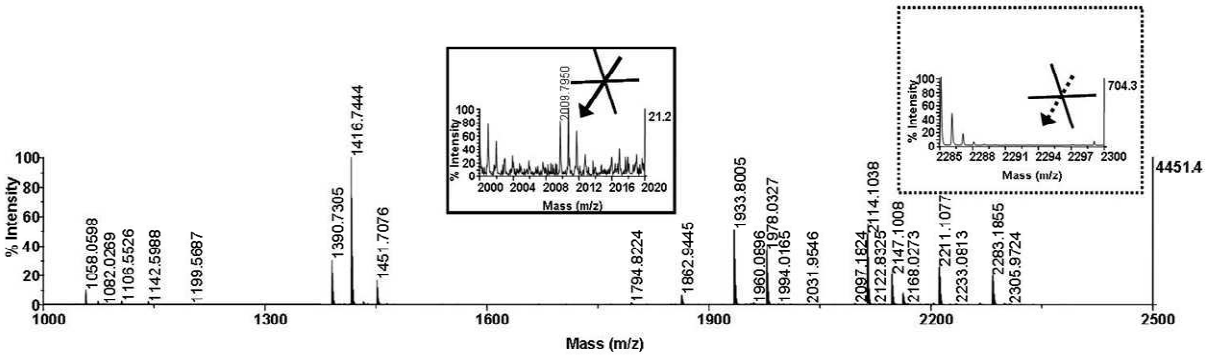
Supplementary Figure S6. Cyclodipeptides synthesized by the wild-type AlbC and the variant L200N expressed in *E. coli*. (**A**) LC-MS analyses of the culture medium of *E. coli* cells expressing the wild-type AlbC and the variant L200N; the UV traces (λ = 214 nm) are shown in yellow and blue, respectively. (**B**) The EIC peaks corresponding to the wild-type AlbC and the variant L200N are shown in yellow and blue, respectively. The cyclodipeptides were identified as previously described in Gondry et al[1].



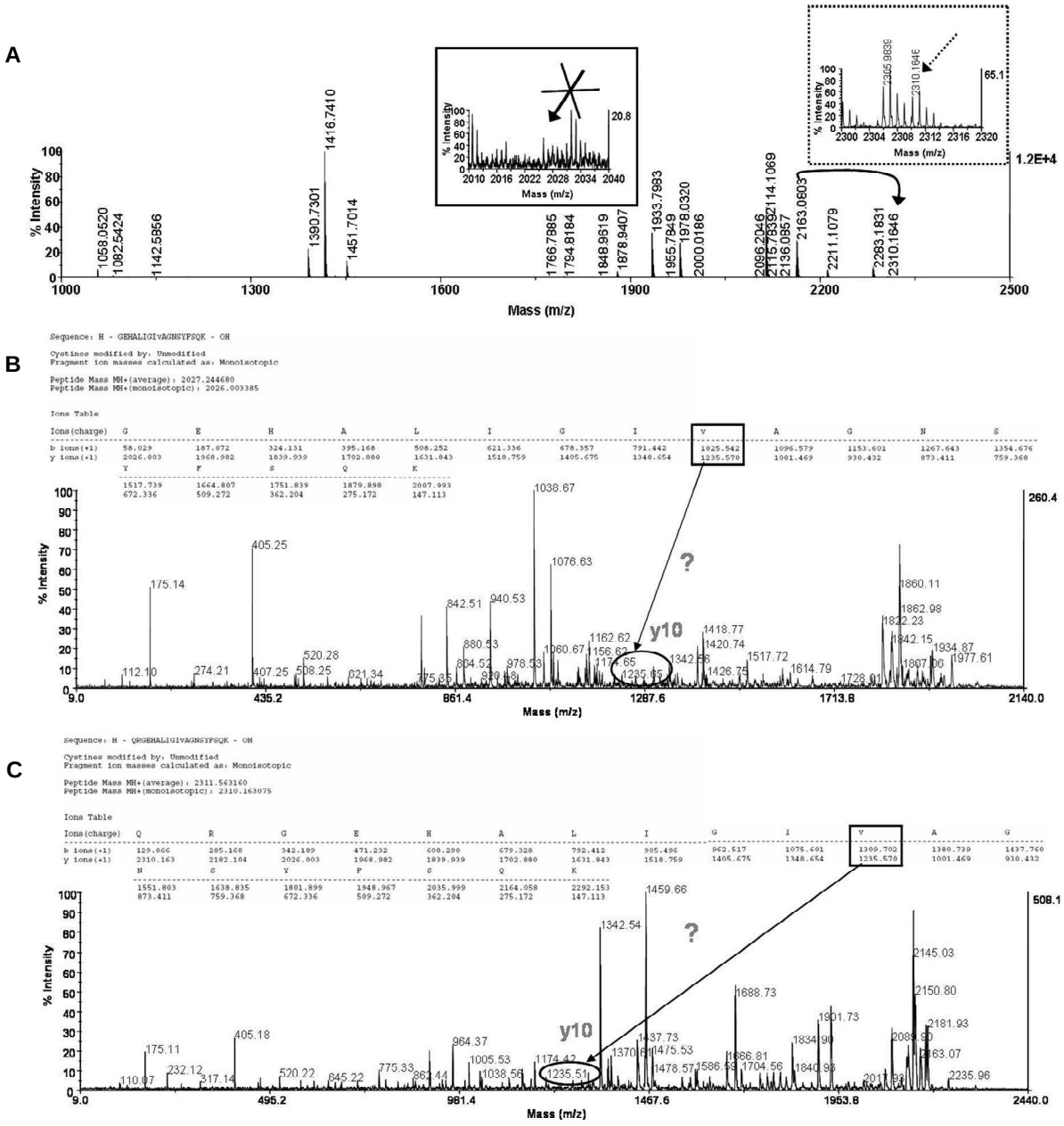
Supplementary Figure S7. Analysis of trypsin-digested AlbC not incubated with Phe-tRNA^{Phe}. (**A**) Peptide mass fingerprint analysis. Thick grey arrows indicate the observed m/z of AlbC peptide fragments containing residue S37 (fragments [29-46] at calculated m/z 1878.94 and [27-46] at calculated m/z 2163.09). Continuous and dashed line frames are closed views of the m/z regions of the mass spectrum corresponding to calculated m/z of potential phenylalanylated [29-46] fragment at 2026.01 and phenylalanylated [27-46] fragment at 2310.16, respectively. (**B**) PSD-MS/MS spectrum of m/z 1878.91. Both b- and y-ions series identify the amino acids sequence of AlbC fragment [29-46]. (**C**) PSD-MS/MS spectrum of m/z 2163.04. Both b- and y-ions series identify the amino acids sequence of AlbC fragment [27-46].



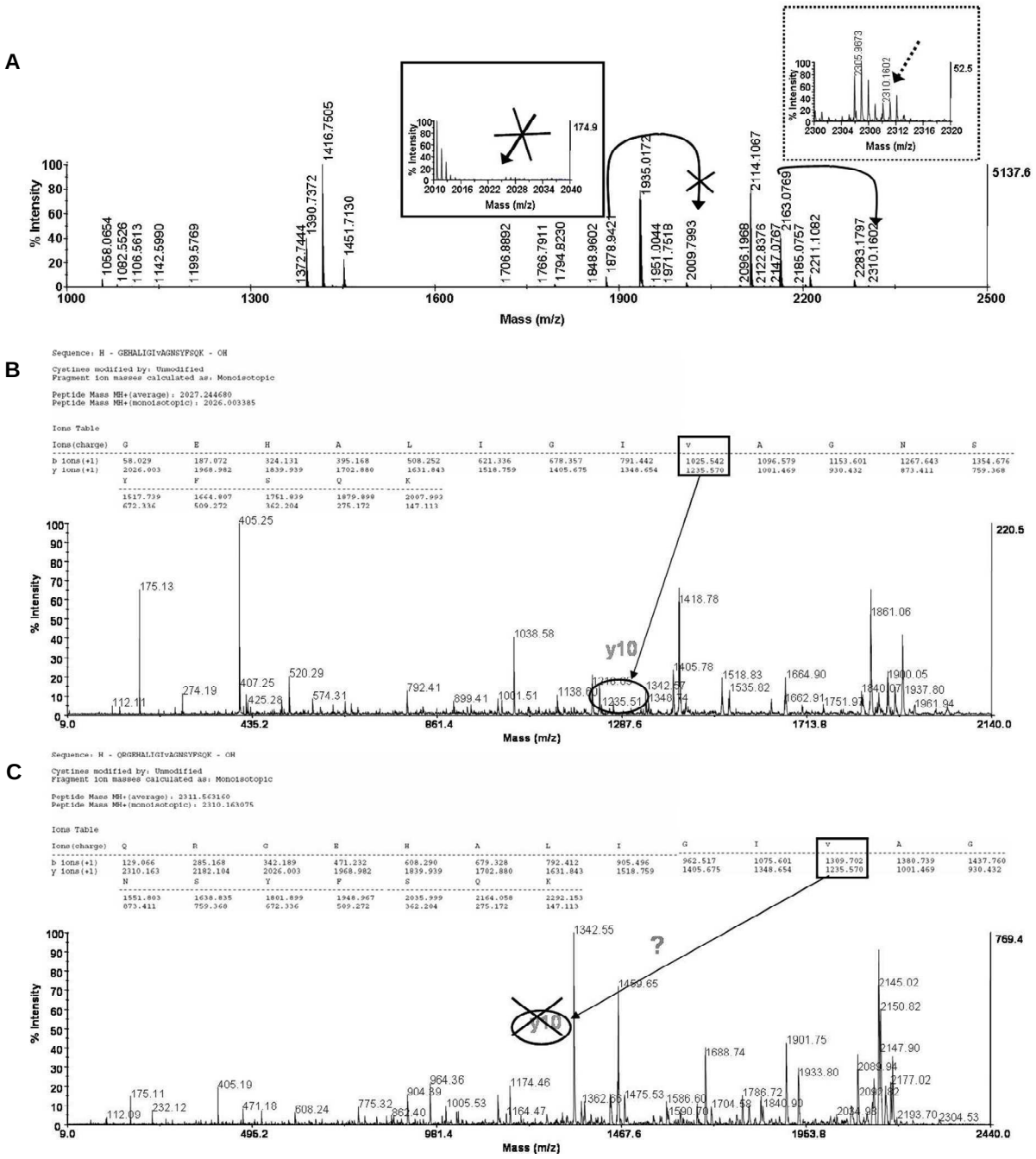
Supplementary Figure S8. Analysis of trypsin-digested AlbC pre-incubated with Phe-tRNA^{Phe}. Corresponding PMF analysis is shown in Figure 7. PSD-MS/MS spectrum of m/z 2310.17. Both b- and y-ions series identify the amino acids sequence of the modified AlbC fragment [27-46], and the residue 37 as the phenylalanylated residue.



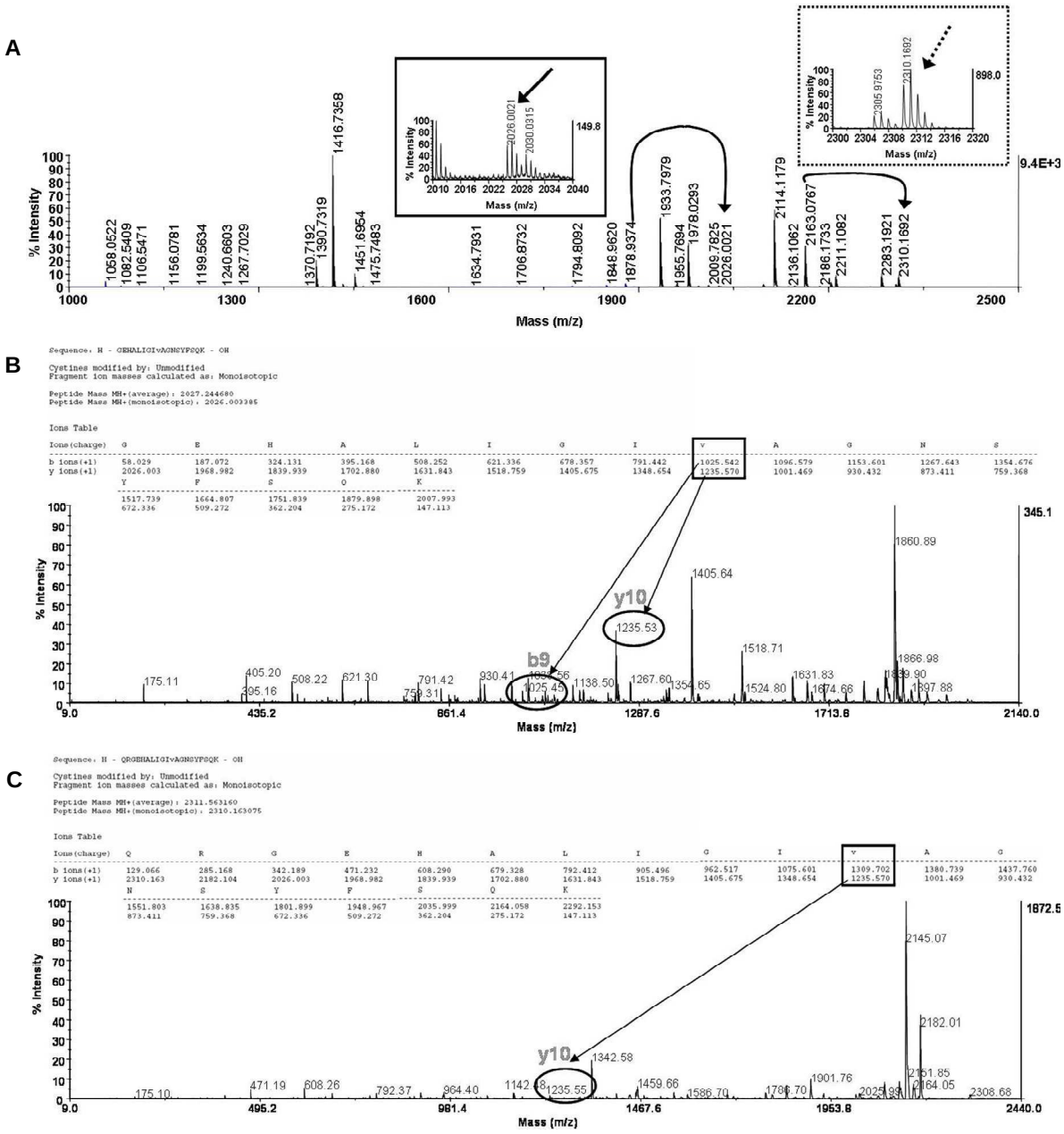
Supplementary Figure S9. Peptide mass fingerprint analysis of the trypsin-digested variant S37A pre-incubated with Phe-tRNA^{Phe}. Continuous and dashed line frames are closed views of the m/z regions of the mass spectrum corresponding to respectively calculated m/z of potential phenylalanylated [29-46] fragment at 2009.97 and phenylalanylated [27-46] fragment at 2294.16. None of these modified fragments was detected and identified by PSD-MS/MS analyses.



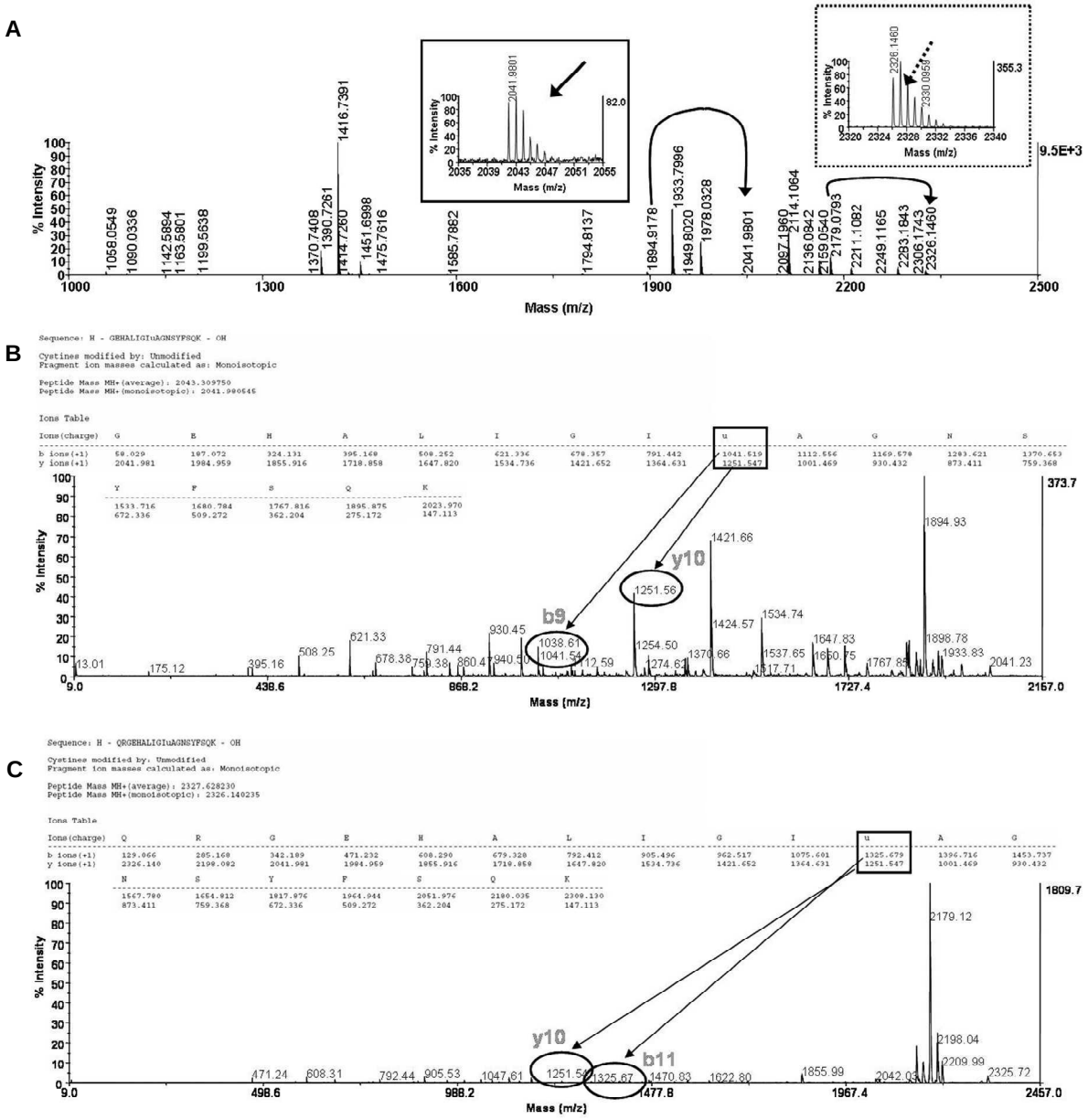
Supplementary Figure S10. Analysis of the trypsin-digested variant E182Q pre-incubated with Phe-tRNA^{Phe}. (**A**) Peptide mass fingerprint analysis. Continuous and dashed line frames are closed views of the m/z regions of the mass spectrum corresponding to calculated m/z of potential phenylalanylated [29-46] fragment at 2026.00, and phenylalanylated [27-46] fragment at 2310.17, respectively. The y-ions of the phenylalanylated serine residue were detected in both [29-46] fragment at 2026.00, and [27-46] fragment at 2310.17 (respectively **B** and **C**). However, none of these modified fragments was formed with a sufficient amount to be detected, isolated as precursor-ion, and sequenced unambiguously by PSD-MS/MS analyses.



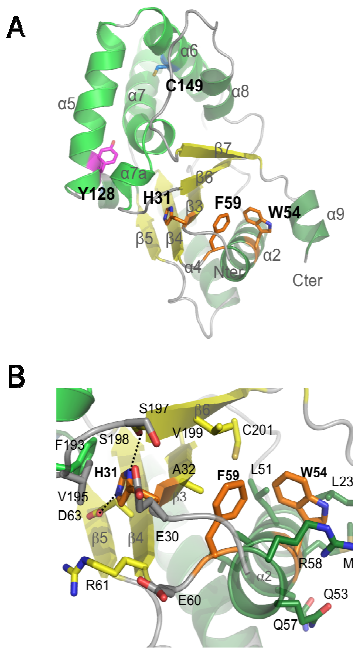
Supplementary Figure S11. Analysis of the trypsin-digested variant Y178F pre-incubated with Phe-tRNA^{Phe}. (**A**) Peptide mass fingerprint analysis. Continuous and dashed line frames are closed views of the m/z regions of the mass spectrum corresponding to calculated m/z of potential phenylalanylated [29-46] fragment at 2026.00, and phenylalanylated [27-46] fragment at 2310.17, respectively. The y-ions of the phenylalanylated serine residue were detected in the [29-46] fragment at 2026.00, but not in the [27-46] fragment at 2310.17 (respectively **B** and **C**). The modified [29-46] fragment was not formed with a sufficient amount to be detected, isolated as precursor-ion, and sequenced unambiguously by PSD-MS/MS analyses.



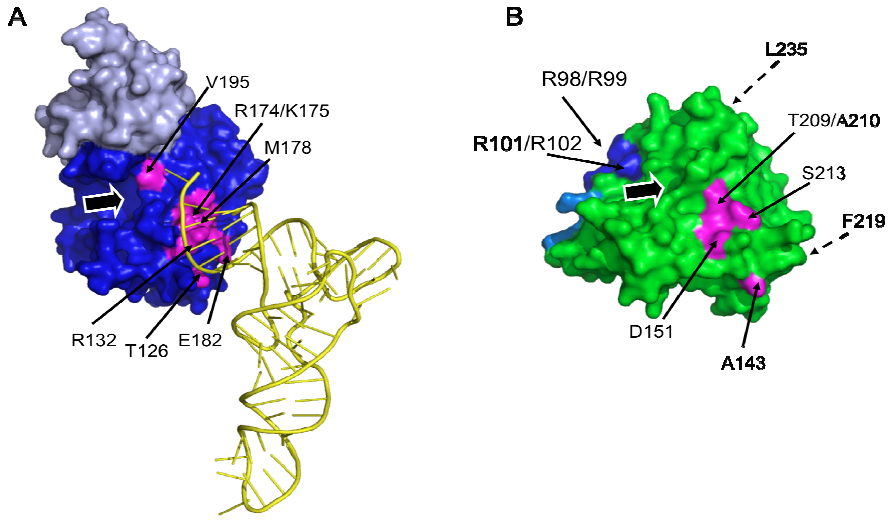
Supplementary Figure S12. Analysis of the trypsin-digested variant Y202F pre-incubated with Phe-tRNA^{Phe}. (**A**) Peptide mass fingerprint analysis. Continuous and dashed line frames are closed views of the m/z regions of the mass spectrum corresponding to calculated m/z of potential phenylalanylated [29-46] fragment at 2026.00, and phenylalanylated [27-46] fragment at 2310.17, respectively. (**B**) PSD-MS/MS spectrum of m/z 2026.00. Both b- and y-ions series identify the amino acids sequence of the modified AlbC fragment [29-46], and b9 and y10 daugther-ions confirm that S37 is the phenylalanylated residue. (**C**) PSD-MS/MS spectrum of m/z 2310.17. Both b- and y-ions series identify the amino acids sequence of the modified AlbC fragment [27-46], and y10 daugther-ion confirms that S37 is the phenylalanylated residue.



Supplementary Figure S13. Analysis of the trypsin-digested variant S37C pre-incubated with Phe-tRNA^{Phe}. (**A**) Peptide mass fingerprint analysis. Continuous and dashed line frames are closed views of the m/z regions of the mass spectrum corresponding to calculated m/z of potential phenylalanylated [29-46] fragment at 2041.99, and phenylalanylated [27-46] fragment at 2326.15, respectively. Both modified fragments were detected as indicated by the arrows. (**B**) PSD-MS/MS spectrum of m/z 2041.99. Both b- and y-ions series identify the amino acids sequence of the modified fragment [29-46], and b9 and y10 daugther-ions unambiguously confirm that C37 is the phenylalanylated residue. (**C**) PSD-MS/MS spectrum of m/z 2326.15. Both b- and y-ions series identify the amino acids sequence of the modified fragment [27-46], and b11 and y10 daugther-ion unambiguously confirm that C37 is the phenylalanylated residue.



Supplementary Figure S14. Description of the residues conserved among CDPSs but located outside the active site. (**A**) Overall view. The side chains of conserved residues are in stick representation. They are coloured in blue if their substitution with alanine does not affect enzyme activity and protein expression (G79, S120 and C149). They are coloured in orange if the substitution affects both enzyme activity and protein expression (H31, W54, and F59), and in magenta if the substitution affects enzyme activity but not protein expression (Y128). Residues G79 and S120 are not visible in this view. (**B**) Position of residues H31, W54 and F59. Residues that are within 4.0 Å of these three residues are represented. Hydrogen bonds made by H31 are represented in dashed lines. (**C**) Position of residues Y128 and S120. Residues that are within 4.0 Å of these two residues are represented. Hydrogen bonds are represented in dashed lines.



Supplementary Figure S15. Comparison of the tRNA-binding regions in AlbC and TyrRS$_{Mj}$. (**A**) Structure of the TyrRS$_{Mj}$-tRNA^{Tyr}-tyrosine complex (PDB id: 1j1u). TyrRS$_{Mj}$ and tRNA^{Tyr} are respectively shown in surface and cartoon representations. The catalytic and tRNA-binding domains are respectively in dark blue and grey. The TyrRS$_{Mj}$ residues involved in acceptor stem binding (see Figure 4) are coloured in magenta. The amino acid-binding pocket is indicated by a black arrow. (**B**) Structure of AlbC. AlbC is shown in surface representation, coloured in green, and oriented as TyRS$_{MJ}$ in (**A**). The residues forming the basic patch identified in AlbC are shown in blue; among them, the residues essential for enzyme activity are indicated.

2.3 Additional information

Now, the crystal structure of another CDPS named YvmC-Blic from *Bacillus licheniformis*is also available (Bonnefond et al. 2011). Here we will compare the three CDPSs: AlbC, Rv2275 and YvmC-Blic. Their structures are respectively determined at 1.9-Å, 2.0-Å and 1.7-2.4-Å resolutions. The 3D structures of the three CDPSs (**Figure 42A**) (Belin et al. 2012) superimposed well (rms deviation value of 2.27 Å, 2.2 Å and 2.46 Å over 196, 211 and 180 matched Cα positions for Rv2275 versus AlbC, for Rv2275 versus YvmC-Blic and AlbC *versus* YvmC-Blic, respectively) despite their sequence divergence (only about 27% sequence identity). Strikingly, AlbC and YvmC-Blic are monomers whereas Rv2275 is a homodimer. The common monomer has a compact α/β fold consisting essentially of a central β-sheet with five parallel β strands surrounded by α helices (**Figure 42A**). Just as class I aaRSs, the common monomer of the three CDPSs also contains a Rossmann-fold domain. This motif is formed by strands β3-β5, helix α2 and the C-terminal part of helix α4, followed by a helical region consisting of three helices (α5-α7) that is packed against this Rossmann fold. Structures of all the three CDPSs are highly similar to that of the catalytic domain of class Ic aaRS. They all possess a similar surface-accessible pocket containing five of the seven essential residues conserved among the eight characterized CDPSs. Four residues (AlbC numbered Gly35, Ser37, Tyr178, and Glu182) are well superposed whereas the fifth, Tyr202, adopts an opposed position for Rv2275 to AlbC and YvmC-Blic (**Figure 42B**). In addition, the cluster of basic residues of AlbC supposed to interact with the tRNA moiety of aa-tRNA substrate is conserved in all CDPSs characterized to date, including Rv2275 and YvmC-Blic. Thus, the basic patch of the CDPS may interact with the tRNA moiety of the substrate, potentially with the phosphate backbone of the tRNA body. However, there are some significant differences between the three CDPSs. First, the N-terminal region (amino acids 50-77) of Rv2275 is structured (strands β1, β2 and helix α1) whereas the corresponding regions of YvmC-Blic (1-11) and AlbC (1-27) are not observed. Conversely, the C-terminal helix α9 is structured in AlbC whereas the corresponding regions of Rv2275 (285-289) and YvmC-Blic (236-249) are not observed (**Figure 42A**).

Second, there are major deviations between the loops containing residues shown to be involved in substrate interaction or catalytic mechanism.

Studies on the three CDPSs all suggest that the surface-accessible pockets in CDPSs are the binding sites for the aminoacyl moiety of the aa-tRNA substrate. Furthermore, studies on the pocket of AlbC proved this hypothesis *via* the results obtained with the L200N variant of AlbC (Sauguet et al. 2011). These studies not only demonstrate the binding of the aminoacyl moiety of the substrate in the pocket, but also suggest that the specificity of CDPSs for at least one of the substrates is mainly directed at the aminoacyl moiety, and not at the sequence of the tRNA moiety of the substrate. Knowing that AlbC, Rv2275 and YvmC-Blic mainly synthesize cFL, cYY and cLL, respectively, in addition to the other characterized CDPSs mainly synthesizing cLL (Gondry et al. 2009), it is reasonable that the bottom of the pocket be found essentially hydrophobic. Nevertheless, the residues of this site are loosely conserved among the CDPS family, potentially accounting for their relaxed substrate specificity.

Studies performed on all the three CDPSs suggest that these enzymes use the two aa-tRNA substrates in a sequential ping-pong mechanism, with a similar first catalytic step: the binding of the first aa-tRNA and the transfer and storage of its aminoacyl moiety on the conserved serine residue of the enzyme pocket (*e.g.* Ser37, AlbC numbering), that is the covalent acyl-enzyme intermediate (**Figure 43**) (Belin et al. 2012). However, one divergence among the studies on the three CDPSs is about the activation of the catalytic serine. In Rv2275 and YvmC-Blic, the catalytic serine (Ser88, in Rv2275 numbering; or Ser37, in YvmC-Blic numbering) is proposed to be activated by a sterically adjacent tyrosine (Tyr253, in Rv2275 numbering; or Tyr204, in YvmC-Blic numbering). In AlbC, the corresponding tyrosine (Tyr202, in AlbC numbering) is proved not to be responsible for serine activation because the Y202F variant does not affect the formation of the covalent acyl-enzyme intermediate mentioned above. This divergence is also reflected to the conformation of the controversial tyrosine, which is located on the highly flexible loop β6 - α8 (**Figure 42B**). In the structures available for Rv2275 and YvmC-Blic, this loop adopts a

confirmation narrowing the entrance of the active site and bringing the tyrosine close to the active serine (**Figure 42B**). Nevertheless, in AlbC, this loop adopts a different conformation, making the active site appears more "open" and the Try202 is oriented towards the solvent rather than the active Ser37 (**Figure 42B**). Nureki and coworkers noticed the narrow access to the pocket in both YvmC-Blic and Rv2275 and suggested that the solved structures may represent a "closed" state of the enzymes, with tRNA binding required to open up the structure (Bonnefond et al. 2011). The impact of aminoacyl-tRNA binding on the enzyme structure remains unclear, but the structures of both YvmC-Blic and Rv2275 strongly suggested that some remodeling of the loops close to the active serine is required to allow access to the catalytic pocket. As no relevant residue for activation of the catalytic serine has yet been clearly identified, an attractive hypothesis is that the serine activation results from a concerted proton shuttling mechanism involving the two adjacent vicinal hydroxyls of the nucleotide A76 of the tRNA moiety (Sauguet et al. 2011).

The steps of the catalytic mechanism occurring after the formation of the covalent intermediate remain to be determined. The cyclodipeptides formed by a given CDPS enzyme almost invariably have one amino acid in common (Gondry et al. 2009). Based on the available data, the most likely hypothesis for this is that only the first aminoacyl group, probably the "common" aminoacyl group, binds and remains in the catalytic pocket, the second aminoacyl group being accommodated at a different site with less stringent recognition that has yet to be characterized. The aminoacyl moiety of the second aa-tRNA to form either a dipeptidyl-enzyme or a dipeptidyl-tRNA intermediate, which may undergo intramolecular cyclization to generate the final cyclodipeptide product (**Figure 43**) (Sauguet et al. 2011; Belin et al. 2012).

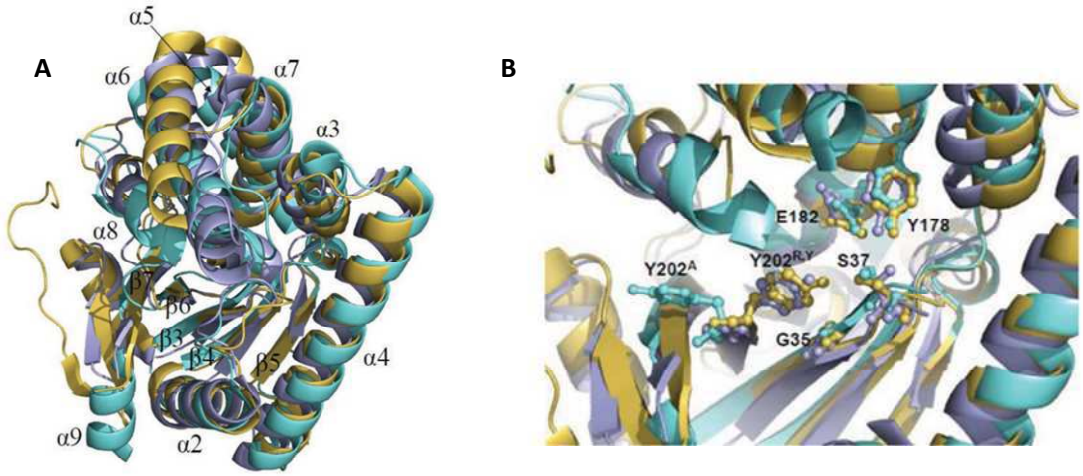


Figure 42: Structural comparison between AlbC (cyan), Rv2275 (yellow) and YvmC-Blic (salmon) (Belin et al. 2012). (A) Superimposition of the three CDPSs structures in cartoon mode. AlbC secondary-structure α-helices and β-strands are numbered. (B) Superimposition of conserved residues in the surface-accessible pockets of AlbC, Rv2275 and Yvmc-Blic. Residues are represented in ball and stick and are numbered as for AlbC except for Y204^{Y} (Yvmc-Blic) and Y253^{R} (Rv2275).

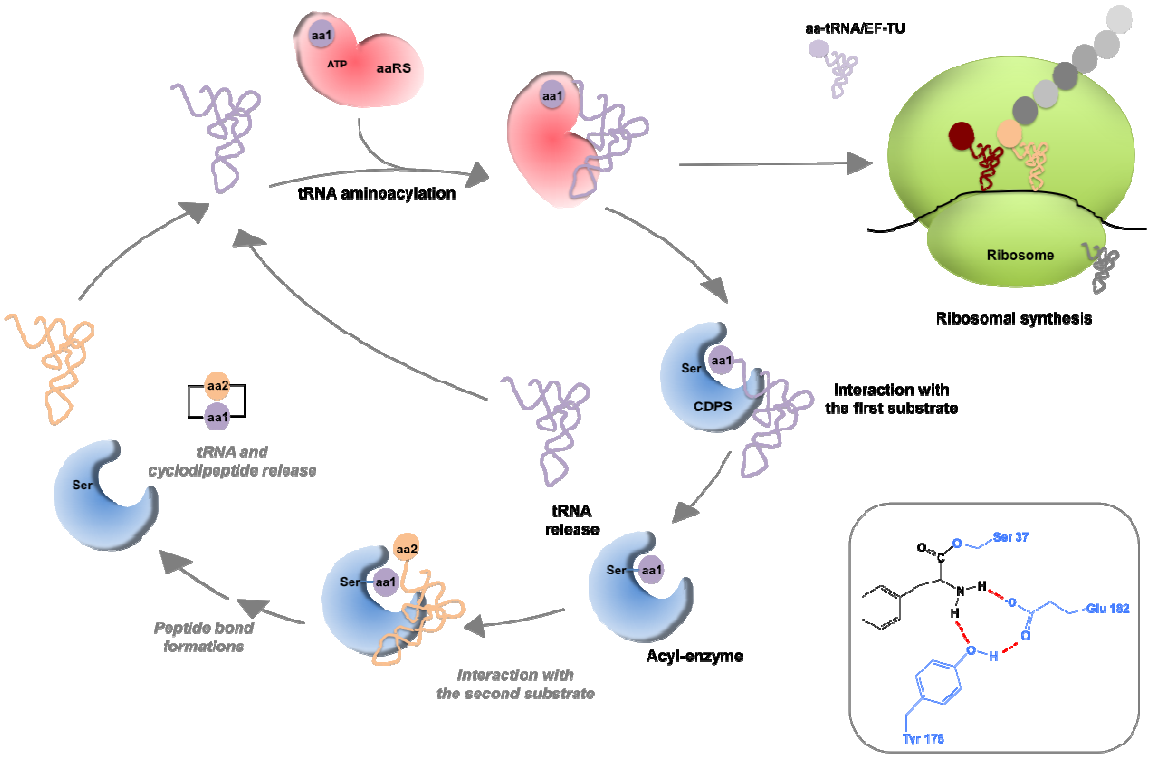


Figure 43: CDPSs hijack aa-tRNAs to produce cyclodipeptides (Belin et al. 2012). aa-tRNAs are generated from amino acid, ATP and tRNAs by aaRSs (in red). aa-tRNAs are typically delivered to the ribosome (in green) for the synthesis of peptide bonds in nascent polypeptides, but CDPSs (in blue) hijack some aa-tRNAs to make cyclodipeptides. Known and putative (in italics) steps of the CDPS mechanism are summarized. The inset describes the conserved residues (AlbC numbering) essential for formation of the covalent aminoacyl-enzyme (here, phenylalanyl-AlbC).

3 NDAS_1148 FROM *NOCARDIOPSIS DASSONVILLEI*, A NEW ACTIVE CDPS WITH A PROTEIN SEQUENCE CLOSE TO THAT OF ALBC

3.1 Introduction

In the databases, characterized CDPSs and putative CDPSs have been detected in approximately 50 different bacterial genomes (of the 2019 publicly available) (Belin et al. 2012). The evolutionary relationship of these CDPSs in the form of unrooted phylogenetic tree (Belin et al. 2012) attracted our attention to a putative CDPS Ndas_1148 from the actinobacterium *Nocardiopsis dassonvillei* because it lies on the same branch as the characterized CDPS AlbC from *S. noursei*. Ndas_1148 is the nearest identified putative CDPS to AlbC compared to the others. It should thus be of great interest to study this protein in order to have a better understanding of CDPSs.

This chapter consists of studies on Ndas_1148 by comparison with AlbC. It mainly concerns studies of grafting the Ndas_1148 catalytic pocket on AlbC in order to elucidate of the molecular bases of the substrate specificity of CDPSs. This work will be presented in the form of manuscript of an article.

3.2 Article manuscript

Changing the aminoacyl-tRNA substrate specificity of a cyclodipeptide synthase by swapping its aminoacyl-moiety binding pocket

Yan Li[1], Pascal Belin[1], Mireille Moutiez[1], Jérôme Seguin[1], Cécile Martel[2], Jean-Luc Pernodet[2] and Muriel Gondry[1*]

[1]CEA, Institut de Biologie et Technologies de Saclay, Service d'Ingénierie Moléculaire des Protéines, F-91191 Gif-sur-Yvette, France.

[2]Univ. Paris-Sud, CNRS, UMR8621, Institut de Génétique et Microbiologie,F-91405 Orsay, France.

Contact Information

[*] Correspondence: muriel.gondry@cea.fr; Tel: +33 169 08 76 47; Fax: +33 169 08 90 71

Running title

An active CDPS found in Nocardiopsis *dassonvillei*

SUMMARY

Cyclodipeptide synthases (CDPSs) hijack aminoacyl-tRNAs (aa-tRNAs) to produce various cyclodipeptides. They contain a surface-accessible pocket accommodating the aminoacyl moiety of one of the two aa-tRNA substrates. We show here that the putative CDPS Ndas_1148 identified in *Nocardiopsis dassonvillei* is an active CDPS. Ndas_1148 displays a high level of sequence identity to the CDPS AlbC but both enzymes do not have the same profile of cyclodipeptide production. We demonstrate that the chimera that corresponds to AlbC having the Ndas_1148 pocket has acquired the substrate specificity of Ndas_1148. This shows that the difference in residues constituting the two CDPS pockets is responsible for the difference in cyclodipeptide production observed for the two enzymes. This also indicates that CDPS pockets contain specificity determinants responsible for the incorporation of the aminoacyl moiety of the second aa-tRNA substrate into cyclodipeptides.

HIGHLIGHTS

- Identification of a new active CDPS
- CDPS engineering
- Molecular bases of substrate specificities in CDPSs
- Changing CDPS specificities

INTRODUCTION

Cyclodipeptide synthases (CDPSs) form a family of small enzymes that hijack aminoacyl-tRNAs (aa-tRNAs) from their canonical role in ribosomal protein synthesis for cyclodipeptide formation (Gondry et al., 2009). Nine CDPSs have been biochemically characterized, eight originating from bacteria and one from an animal (Gondry et al., 2009; Seguin et al., 2011). Genome mining revealed the presence of putative CDPSs in approximately 60 different genomes from bacteria, fungi or animal (Aravind et al. 2010; Belin et al. 2012). In bacteria and fungi, CDPSs are associated with cyclodipeptide-tailoring enzymes in biosynthetic pathways dedicated to the synthesis of diketopiperazines (DKPs) (Belin et al., 2012). Three CDPS-dependent pathways have been fully characterized in terms of both the DKPs produced and the enzymes involved. In *Streptomyces noursei*, the CDPS AlbC mostly catalyzes the synthesis of the cyclo(L-Phe-L-Leu) (cFL) intermediate, into which the cyclic dipeptide oxidase (CDO) introduces two α,β-dehydrogenations, yielding albonoursin that has been shown to have antibacterial activity (Fukushima et al., 1973; Gondry et al., 2001; Lautru et al., 2002). In *Bacillus subtilis*, the CDPS YvmC mostly produces the cyclo(L-Leu-L-Leu) (cLL), into which the cytochrome P450 CYP134A1 catalyzes a DKP ring oxidation, yielding pulcherriminic acid that chelates iron (Cryle et al., 2010; Gondry et al., 2009; Tang et al., 2006). In *Mycobacterium tuberculosis*, the CDPS Rv2275 mostly synthesizes the cyclo(L-Tyr-L-Tyr) (cYY), into which the cytochrome P450 CYP121A1 introduces C-C aryl coupling, yielding mycocyclosin that may be essential for *M. tuberculosis* viability (Belin et al., 2009; Gondry et al., 2009; McLean et al., 2008; Vetting et al., 2010).

The crystallographic structures of AlbC, Rv2275 and YvmC have been determined. The three CDPSs share a common architecture highly similar to that of the catalytic domain of class-Ic aminoacyl-tRNA synthetases, the enzymes catalyzing the activation of amino acids and their transfer to cognate tRNAs to produce aa-tRNAs. These structural data combined with extensive mutagenesis analyses have provided insight into the molecular bases of the interactions between CDPSs and their substrates and the catalytic mechanism used by these enzymes (Bonnefond et al., 2011; Sauguet et al., 2011; Vetting et al., 2010). CDPSs contain a patch of basic residues that may interact with the tRNA moiety of an aa-tRNA substrate. They also have a surface-accessible pocket that accommodates the aminoacyl moiety of the substrate. CDPSs catalyze their two-substrate reaction using a sequential ping-pong

mechanism. The first catalytic steps correspond to the binding of the first aa-tRNA then the transfer and the covalent attachment of its aminoacyl moiety on a conserved serine residue located in the pocket. The subsequent steps remain to be demonstrated but it is currently assumed that the aminoacyl-enzyme reacts with the aminoacyl moiety of the second tRNA to form a dipeptidyl-intermediate that undergoes intramolecular cyclization, yielding the cyclodipeptide product (Bonnefond et al., 2011; Sauguet et al., 2011; Vetting et al., 2010).

CDPSs are promiscuous enzymes as they predominantly produce one major cyclodipeptide but also other minor cyclodipeptides. For example, AlbC mostly synthesizes cFL but also other L-Phe-containing cyclodipeptides. The other CDPSs produce L-Tyr-, L-Leu- or L-Trp-containing cyclodipeptides (Gondry et al., 2009; Seguin et al., 2011). The molecular bases responsible for the difference in substrate specificities remain poorly documented. We show here that the CDPS Ndas_1148 identified in *Nocardiopsis dassonvillei* does not produce cFL as the major cyclodipeptide although this CDPS has a protein sequence and a genomic context close to those of AlbC. We also show that the difference in residues constituting the two CDPS pockets is responsible for the difference in profiles of cyclodipeptides production observed for these two enzymes. Finally, we demonstrate that the grafting of the Ndas_1148 pocket on AlbC results in an AlbC chimera having the substrate specificity of Ndas_1148.

RESULTS

The putative CDPS found in Nocardiopsis *dassonvillei*

Eight bacterial and one animal members of the CDPS family have been experimentally characterized so far (Gondry et al., 2009; Seguin et al., 2011). However, new putative CDPSs have been identified in diverse organisms by iterative PSI_BLAST searches (Aravind et al. 2010; Belin et al. 2012). One putative CDPS, Ndas_1148, has been detected in *Nocardiopsis dassonvillei*. We aligned sequences for Ndas_1148 and the nine characterized CDPSs (Figure 1). The protein Ndas_1148 (244 residues) is of a similar size to the CDPSs (216-289 residues) and displays the highest sequence similarity to the CDPS AlbC (40% identity and 53% similarity). The 13 residues conserved among bacterial CDPSs, including the seven residues (G35, S37, G79, Y128, Y178, E182, Y202, AlbC numbering) previously shown to be involved in CDPS activity (Sauguet et al., 2011), are also found in Ndas_1148. The

characterized CDPSs contain a patch of basic residues, suspected to interact with the tRNA moiety of an aa-tRNA substrate (Sauguet et al., 2011), which is also present in Ndas_1148 (Figure 1).

In *S. noursei*, AlbC catalyzes the synthesis of cFL, into which the cyclic dipeptide oxidase (CDO) introduces two α,β-dehydrogenations leading to the formation of albonoursin (cycloΔPhe-ΔLeu) (Gondry et al., 2001; Lautru et al., 2002). The genes encoding the two subunits of CDO, AlbA and AlbB, are located immediately upstream from that of AlbC. In *N. dassonvillei*, a similar operon-like structure is found (see Figure S1 available online). Proteins encoded by the two genes preceding the Ndas_1148 gene display a significant sequence similarity to AlbA (39% identity and 50% similarity) and AlbB (45% identity and 58% similarity). Two other genes encoding proteins annotated as methyl transferases are also present in the vicinity of the Ndas_1148 gene but absent in that of AlbC (Figure S1 available online).

The similar protein sequence and genomic context of Ndas_1148 and AlbC strongly suggest that Ndas_1148 may be member of the CDPS family.

Ndas_1148, a cyclodipeptide-synthesizing enzyme

We investigated whether Ndas_1148 catalyzes the formation of cyclodipeptides. We had previously shown that all characterized CDPSs produced in *Escherichia coli* synthesize cyclodipeptides that are further released into the culture medium (Gondry et al., 2009; Seguin et al., 2011). We then introduced a construct encoding Ndas_1148 in this host and analyzed the cyclodipeptide content of the culture supernatant of these Ndas_1148 producing cells. We compared the UV-chromatograms for the Ndas_1148 sample and the control – consisting of the culture supernatant of *E. coli* cells containing the empty cloning vector. Several peaks specific to Ndas_1148-producing cells were observed (Figure 2A). The corresponding extracted ion current (EIC) chromatograms (Figure 2B) and MS/MS fragmentation patterns (see Figure S2 available online) indicated that Ndas_1148 synthesizes eight cyclodipeptides: cFL, cFY, cFF, cFM, cFA, cYL, cYY and cYM. Thus, Ndas_1148 catalyzes the formation of various cyclodipeptides in *E. coli*. It is a new active member of the CDPS family.

Similarities and differences between Ndas_1148 and AlbC

The cyclodipeptides produced by Ndas_1148 are also produced by AlbC (Gondry et al., 2009). However, although the nature of the cyclodipeptides produced by the two CDPSs is almost identical, their quantities and their relative proportions appear different. Concerning Ndas_1148, the cyclodipeptides cFY and cFF are predominantly produced (3.5 ± 0.6 and 1.9 ± 0.1 mg l^{-1} of culture supernatant, respectively), the cyclodipeptides cYL and cYM are obtained in lower quantities (about 0.5 mg l^{-1} of culture supernatant) and the cyclodipeptides cFL, cFM, cFA and cYY are hardly detectable (Figure 3A). To accurately compare the Ndas_1148 profile of cyclodipeptide production to that of AlbC, we again analyzed the cyclodipeptide content of a culture supernatant of AlbC-producing cells. AlbC synthesizes cFL (43 ± 5 mg l^{-1} of culture supernatant), cFY, cFF, cFM, cYL, cYM, cLL, cLM (between 10 and 4 mg l^{-1} of culture supernatant), cYY, cYA, cMM (less than 1 mg l^{-1} of culture supernatant) (Figure 3A and Figure S3 available online) and trace amounts of cFA and cLA that had not been previously identified (Gondry et al., 2009). Thus, Ndas_1148 synthesizes less cyclodipeptides than AlbC, not only in variety but also in quantities, which is consistent with its much lower expression in soluble form (Figure 3B). Interestingly, Ndas_1148 does not synthesize significant quantities of cyclodipeptides containing a leucyl residue: cFL, the major cyclodipeptide produced by AlbC, is almost not synthesized (the decrease in cFL synthesis is 99.5% compared to AlbC) and cLL is not at all produced (Figure 3A). The quantity of L-Phe-containing cyclodipeptides is also reduced (the decrease in cFY and cFF is about 60-70%), but the relative proportion of each of the L-Phe-containing cyclodipeptides is almost similar for the two CDPSs (except for cFL). The same conclusion can be made for L-Tyr-containing cyclodipeptides although their synthesis by Ndas_1148 is further decreased.

To understand the difference in production profiles of the two CDPSs, we generated a homology-based structural model for Ndas_1148 with Modeler software (Sali and Blundell, 1993), using the crystal structure of AlbC (Protein Data Bank [PDB] id: 3OQV) as a template (Figure 4). Superimposition of the three-dimensional structures of AlbC and Ndas_1148 gave a root-mean-square deviation value of 0.29 Å over 152 Cα. Comparison of the distribution of electrostatic surface potential of the two proteins indicates that the patch of basic residues present in AlbC is also present in Ndas_1148 (Figures 4A and 4B). The residues constituting the two basic patches are particularly well conserved (see also Figure 1) and most of them protrude toward the solvent, consistent with the fact that they may interact with the tRNA moiety of an aa-tRNA substrate (Sauguet et al., 2011). For AlbC, the aminoacyl moiety of the substrate is accommodated in a surface-accessible pocket that consists of 19 residues (Figure

4C). Compared to the AlbC pocket, the Ndas_1148 pocket is constituted of 13 residues that are identical and similarly positioned, and six residues that are different (Figure 4D). The identical residues contain in particular the five residues (G35, S37, Y178, E182, Y202, AlbC numbering) conserved among all known CDPSs and essential for the cyclodipeptide-synthesizing activity (Sauguet et al., 2011). The different residues are L69, A71 and L186 – that respectively correspond in AlbC to the conservative residues V65, V67 and F186 – and also H156, C200 and M206 –that respectively correspond in AlbC to the non conservative residues M152, L200 and T206 – (Figures 4C and 4D). These residues are positioned at the entrance (M206), the side (A71 and H156) and the base (L69, L186 and C200) of the Ndas_1148 pocket (Figure 4D).

Exchange of the catalytic pockets of the two CDPSs

To investigate whether the difference in residues constituting the two CDPS pockets is responsible for the difference in profiles of cyclodipeptides production observed for these enzymes, we substituted in AlbC all the six residues different in the two CDPSs by their corresponding residues in Ndas_1148. Thus, we obtained the AlbC chimera that possesses the Ndas_1148 pocket (V65L/V67A/M152H/F186L/L200C/T206M), named AlbC-p(Ndas_1148). We produced the resulting chimera in *E. coli* and compared its properties to those of AlbC and Ndas_1148. Its expression profile in soluble form and its profile of cyclodipeptides production are both similar to those of Ndas_1148 (Figure 3). The chimera does not synthesize significant quantities of cFL (the decrease in cFL synthesis is 99.5% compared to AlbC, as observed for Ndas_1148) or other leucyl-containing cyclodipeptides. Thus, the grafting of the Ndas_1148 pocket on AlbC results in an AlbC chimera that has almost acquired the substrate specificity of Ndas_1148.

To better identify the specificity determinants located in the two CDPS pockets, we independently substituted in AlbC each of the six different residues by their corresponding residue in Ndas_1148. The six resulting variants were produced in soluble forms in amounts almost similar to that of the wild-type AlbC, with the exception of the variant V65L that is less soluble (Figure 5). None of these variants has a profile of cyclodipeptides production similar to that of AlbC-p(Ndas_1148), indicating that none of the mutated residues is responsible by itself for the change of the substrate specificity (Figure 5A and Figures S4-S9).

However, compared to AlbC, each of the mutations induces a decrease in cFL synthesis about 30 % for the variants V65L, L200C and T206M to about 50-60% for the variants V67A and F186L. The largest decrease is observed for the variant M152H (about 90%) but this variant is almost inactive. As the residue M152 is located at the entrance of the AlbC pocket (Figure 4), we hypothesized that its substitution with histidine could narrow the entrance of the pocket and prevent the anchoring of the aminoacyl moieties of substrates. We attempted to construct a variant corresponding to AlbC-p(Ndas_1148) without introducing the substitution of M152, but the genetic construct (named AlbC-p(Ndas_1148)/M152) was almost not expressed in a soluble form (Figure 5B). Concerning the variants F186L, L200C and T206M, besides a decrease in the synthesis of L-Leu-containing cyclodipeptides, they have also a significant increase in cFY synthesis (between 80 and 100%) that is not observed for the variants V65L and V67A (Figure 5). The substitution of residue L200 with asparagine in AlbC was previously shown to give a variant that produces mainly cYL instead of cFL (Sauguet et al., 2011). Here we show that its substitution with cysteine leads to a significant increase in cFY synthesis (Figure 5A). We constructed a variant corresponding to AlbC-p(Ndas_1148) containing the L200N substitution (named AlbC-p(Ndas_1148)/L200N) and the variant Ndas_1148 C200N. However, these two variantsare only poorly soluble (Figure 5B) and almost inactive.

DISCUSSION

We show here that Ndas_1148 is an active member of the CDPS family. It displays a high level of identity to the CDPS AlbC. These two CDPSs differ in specificity from other characterized CDPSs as they predominantly synthesize L-Phe-containing cyclodipeptides whereas the others produce L-Tyr-, L-Leu- and L-Trp-containing cyclodipeptides (Gondry et al., 2009; Seguin et al., 2011). Compared to AlbC, Ndas_1148 is much less soluble and it synthesizes less cyclodipeptides in terms of both variety and quantity. In particular, it synthesizes almost no cFL, which is the major cyclodipeptide produced by AlbC.

In *S. noursei*, the AlbC-dependent biosynthetic pathway has been fully characterized in terms of both the DKPs produced and the enzymes involved (Gondry et al., 2009; Lautru et al., 2002). The gene encoding AlbC is clustered to the two genes encoding the subunits AlbA and AlbB of the tailoring enzyme CDO. The cyclodipeptides produced by AlbC are substrates for

CDO that catalyzes the α,β-dehydrogenation of the two aminoacyl side chains of a wide range of cyclodipeptides (Gondry et al., 2001). The major DKP produced is albonoursin – cyclo(α,β-dehydroPhe-α,β-dehydroLeu) (cΔFΔL)–, which corresponds to the α,β-dehydrogenation of cFL, the major cyclodipeptide produced by AlbC. In *N. dassonvillei*, the protein encoded by the two genes preceding the Ndas_1148 gene isa CDO homologue. The cyclodipeptides produced by Ndas_1148 are likely to be dehydrogenated by the CDO homologue. However, as Ndas_1148 produces much more cFY and other cyclodipeptides (*e.g.* cFF) than cFL, cycloΔFΔY and other dehydrogenated cyclodipeptides (*e.g.* cΔFΔF) but not albonoursin are probably the major DKPs synthesized. Two other genes encoding proteins annotated as methyl transferases are also present in the vicinity of the Ndas_1148 gene but absent in that of AlbC. As methyl transferases are known to be involved in the transfer of methyl group on various atoms (C, N, S or O) of various molecules, the final products from the Ndas_1148-dependent biosynthetic pathway are likely to be both dehydrogenated and methylated.

Ndas_1148 and the other characterized CDPSs respectively display 40% and 19-27% identity with AlbC. Thus, Ndas_1148 has the protein sequence closest to that of AlbC. In particular, the residues constituting the pockets of the two CDPSs are well conserved, with only six of 19 residues that are different for the two enzymes. However, unlike AlbC, Ndas_1148 does not synthesize significant quantities of cyclodipeptides containing a leucyl residue. The sequence similarity and the difference in cyclodipeptide-synthesizing activity observed for the two CDPSs offer the opportunity to document the molecular bases of their specificity, in particular those located in their pockets. We previously demonstrated that the AlbC pocket is the binding site of the phenylalanine moiety of its Phe-tRNA^{Phe} substrate, but we had no data on the binding of the leucyl moiety of the Leu-tRNA^{Leu} substrate, which is required for cFL production (Sauguet et al., 2011). Here, we constructed the chimera enzyme AlbC-p(Ndas_1148) that corresponds to the CDPS AlbC having the Ndas_1148 pocket. AlbC-p(Ndas_1148) has acquired the poor protein solubility and the substrate specificity of Ndas_1148. This result indicates that the grafting of the Ndas-1148 pocket is sufficient to switch the substrate specificity of AlbC to that of Ndas_1148. In particular, it shows that specificity determinants responsible for the incorporation of a leucyl residue into cyclodipeptides are also located in the pocket.

Moreover, we showed that none of the six substituted residues is responsible by itself for the drastic decrease in the incorporation of a leucyl residue into the cyclodipeptides produced. This differs from what we previously obtained concerning the incorporation of a phenylalanyl residue.

Indeed, the substitution in AlbC of only one residue, namely L200, with asparagine - its equivalent residue in the CDPS Rv2275 that mainly produces cYY - was sufficient to induce the synthesis of cYL instead of cFL, probably because the asparagine may form a hydrogen bond with the hydroxyl group of the tyrosyl moiety of a tyrosyl-tRNA^{Tyr} substrate (Sauguet et al., 2011). We here substituted L200 with cysteine, which is not favorable for such hydrogen bonding, and the resulting variant does not synthesize more cYL. However, this variant displays a significant increase in cFY (about 100%). To try to better understand this result, we constructed the variant corresponding to AlbC-p(Ndas_1148) containing the L200N substitution and the variant Ndas_1148 C200N, butthese two variants are almost inactive, which is consistent with their poor expression in soluble form.

AlbC specifically recognizes the aminoacyl moieties of its aa-tRNA substrates *via* its pocket (this work and (Sauguet et al., 2011)). We previously hypothesized that AlbC also recognizes the sequences $A^{76}C^{75}C^{74}A^{73}$ of the single-strand extremity and C^{72}-G^{1} of the first base pair of the tRNA moieties of its substrates (Sauguet et al., 2011). AlbC and Ndas_1148 use as substrates the same *E. coli* tRNA families, namely phenylalanyl-, leucyl-, tyrosyl-, methionyl- and alanyl-tRNAs, that possess members sharing these sequences of the acceptor stems. This could explain why the change in AlbC specificity is achieved by modifying only its aminoacyl moiety-binding pocket.

Due to their small size, CDPSs are good candidates for protein engineering, and they will rapidly receive a large interest to generate a variety of new cyclodipeptide derivatives. However, reprogramming CDPSs by altering their substrate specificity still remains a challenge. We show here that the grafting of the whole pocket of a given CDPS on another CDPS is an interesting option that can be successful.

SIGNIFICANCE

Cyclodipeptide synthases (CDPSs) are a family of small enzymes that hijack aminoacyl-tRNAs (aa-tRNAs) to produce various cyclodipeptides, biosynthetic precursors of many natural products exhibiting noteworthy biological activities. CDPSs are structurally similar to class-I aminoacyl-tRNA synthetases (aaRSs) and contain a surface-accessible pocket accommodating the aminoacyl moiety of one of the two aa-tRNA substrates in a way similar to that used by class-Ic aaRSs to recognize their amino acid substrates. We show here that the putative CDPS Ndas_1148 identified in *Nocardiopsis dassonvillei* is an active CDPS. Ndas_1148 displays a high level of sequence identity to the CDPS AlbC from *Streptomyces noursei* but both enzymes do not have the same profile of cyclodipeptide production. We demonstrate that the chimera that corresponds to AlbC having the Ndas_1148 pocket has acquired the substrate specificity of Ndas_1148. This shows that the difference in residues constituting the two CDPS pockets is responsible for the difference in cyclodipeptide production observed for the two enzymes. This also indicates that CDPS pockets contain specificity determinants responsible for the incorporation of the aminoacyl moiety of the second aa-tRNA substrate into cyclodipeptides. CDPSs are promising enzymes to generate a variety of new cyclodipeptide derivatives but reprogramming CDPSs by altering their substrate specificity still remains a challenge. However, our findings show that reprogramming CDPS specificity can be achieved by swapping the whole pocket of a given CDPS by that of another CDPS.

EXPERIMENTAL PROCEDURES

Structural model of Ndas_1148

Alignment of protein sequences, obtained with MUSCLE and HHPred (Soding et al., 2005), was used for model building with Modeler (Sali and Blundell, 1993). The template used for the homology modeling was the crystal structure of AlbC (PDB id: 3OQV). The quality of the Ndas_1148 model was checked with the Qmean server (Benkert et al., 2009).

Cloning and Production of Ndas_1148

Synthetic gene encoding Ndas_1148 was purchased from GENEART, inserted into the pQE60 vector (QIAGEN), and the His_6-tagged Ndas_1148 protein was then expressed in *E. coli* as previously described (Gondry et al., 2009).

Construction of expression plasmids encoding AlbC variants

The primers used in this study are listed in Table S1. The expression plasmid encoding AlbC was previously constructed (Gondry et al., 2009). It encodes the full-length AlbC (residues 1-239) with the C-terminal addition of two residues (RS) followed by a His_6 tag. The desired variants with a single mutation were created *via* PCR mutagenesis according to the QuikChange method (Stratagene). The gene encoding the variant AlbC(p-Ndas_1148), which contains six mutations, was obtained by the particular PCR method "gene Splicing by Overlap Extension (gene SOEing)" (Ho et al., 1989; Horton, 1995). The desired mutations were introduced into the designed primers (Sigma-Aldrich). The gene encoding AlbC(p-Ndas_1148) was synthesized by two PCR steps. In the first PCR step, four DNA fragments containing different mutations (F1, F2, F3 and F4) were generated and amplified by four separated PCR reactions by using pQE60-AlbC or pQE60-AlbC-F186L as template. To the four DNA fragments, two adjacent fragments shared a segment of identical sequence called the overlap region. F1 had 352 bp containing a NcoI restriction site and two mutations V65L, V67A; F2 was composed of 297 bp and contained three mutations V65L, V67A and M152H; F3 was a DNA fragment of 195 bp containing four mutations M152H, F186L, L200C and T206M; F4 was composed of 254 bp containing two mutations L200C and T206M and also a HindIII restriction site. The four DNA fragments were purified by agarose gel electrophoresis then extracted by using QIAquick Gel Extraction Kit. In the second PCR step, the four DNA fragments obtained previously were mixed together, melted and re-annealed. The top strand of one DNA fragment could anneal to the bottom strand of another DNA fragment sharing an overlap region in such a way that the two strands acted as primers on one another. Extension of this overlap by DNA polymerase generated a new longer DNA product connected by the two mother DNA fragments. After several cycles of reaction, the four DNA fragments F1, F2, F3 and F4 were joined together, so was obtained a full gene of AlbC(p-Ndas_1148) containing the six desired mutations and two restriction sites (NcoI and HindIII) to extremities.

The latter was then inserted into the pQE60 vector *via* the two restriction sites described above. The DNA sequence was checked by DNA sequencing using Sanger method. One unit of Phusion® High-Fidelity DNA Polymerase was used in each PCR reaction. The amplification of genes by PCR was realized by the following cycle profile: 35 main cycles; 98°C, 10 sec (denaturation), 60°C, 10 sec (annealing), 72°C, 15 sec (extension).

Analysis of cyclodipeptide synthesis by Ndas_1148, AlbC or AlbC variants

In vivo assays for cyclodipeptide-synthesizing activity were performed as described by ((Gondry et al., 2009). Briefly, *E. coli* cells expressing the selected CDPSs were grown in M9 liquid medium supplemented with a solution of vitamins and each culture supernatant was analyzed for cyclodipeptide content by LC-MS/MS. Cyclodipeptides were detected and identified from both their m/z value (MS) and their daughter ion spectra (MS/MS), as a result of their common fragmentation patterns. The nature of the detected cyclodipeptides was unambiguously confirmed by comparison with authentic standards. Cyclodipeptides were quantified on the basis of their peak area at 214 nm using calibration curves performed with the standards.

ACKNOWLEDGMENTS

We thank Robert Thai and Steven Dubois for helpful advice on mass spectrometry. We thank Alain Lecoq for the synthesis of cyclodipeptide standards. This work was supported by the CEA and the ANR (ANR 2010 BLAN 1501 01). Y.L. holds a doctoral fellowship from the CEA. The authors have no conflict of interest to declare.

REFERENCES

Aravind, L., de Souza, R.F., and Iyer, L.M. (2010). Predicted class-I aminoacyl tRNA synthetase-like proteins in non-ribosomal peptide synthesis. Biol Direct *5*, 48.

Belin, P., Le Du, M.H., Fielding, A., Lequin, O., Jacquet, M., Charbonnier, J.B., Lecoq, A., Thai, R., Courcon, M., Masson, C., *et al.*(2009). Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A *106*, 7426-7431.

Belin, P., Moutiez, M., Lautru, S., Seguin, J., Pernodet, J.L., and Gondry, M. (2012). The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways. Nat Prod Rep doi:10.1039/C2NP20010D

Benkert, P., Kunzli, M., and Schwede, T. (2009). QMEAN server for protein model quality estimation. Nucleic Acids Res *37*, W510-514.

Bonnefond, L., Arai, T., Sakaguchi, Y., Suzuki, T., Ishitani, R., and Nureki, O. (2011). Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog. Proc Natl Acad Sci U S A *108*, 3912-3917.

Braud, S., Moutiez, M., Belin, P., Abello, N., Drevet, P., Zinn-Justin, S., Courcon, M., Masson, C., Dassa, J., Charbonnier, J.B., *et al.*(2005). Dual expression system suitable for high-throughput fluorescence-based screening and production of soluble proteins. J Proteome Res *4*, 2137-2147.

Cryle, M.J., Bell, S.G., and Schlichting, I. (2010). Structural and biochemical characterization of the cytochrome P450 CypX (CYP134A1) from *Bacillus subtilis*: a cyclo-L-leucyl-L-leucyl dipeptide oxidase. Biochemistry *49*, 7282-7296.

Fukushima, K., Yazawa, K., and Arai, T. (1973). Biological activities of albonoursin. J Antibiot (Tokyo) *26*, 175-176.

Gondry, M., Lautru, S., Fusai, G., Meunier, G., Menez, A., and Genet, R. (2001). Cyclic dipeptide oxidase from *Streptomyces noursei*. Isolation, purification and partial characterization of a novel, amino acyl alpha,beta-dehydrogenase. Eur J Biochem *268*, 1712-1721.

Gondry, M., Sauguet, L., Belin, P., Thai, R., Amouroux, R., Tellier, C., Tuphile, K., Jacquet, M., Braud, S., Courcon, M., *et al.*(2009). Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. Nat Chem Biol *5*, 414-420.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene *77*, 51-59.

Horton, R.M. (1995). PCR-mediated recombination and mutagenesis. SOEing together tailor-made genes. Mol Biotechnol *3*, 93-99.

Lautru, S., Gondry, M., Genet, R., and Pernodet, J.L. (2002). The albonoursin gene Cluster of *S. noursei:* biosynthesis of diketopiperazine metabolites independent of nonribosomal peptide synthetases. Chem Biol *9*, 1355-1364.

McLean, K.J., Carroll, P., Lewis, D.G., Dunford, A.J., Seward, H.E., Neeli, R., Cheesman, M.R., Marsollier, L., Douglas, P., Smith, W.E., *et al.* (2008). Characterization of Active Site Structure in CYP121: a cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. J Biol Chem *283*, 33406-33416.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol *234*, 779-815.

Sauguet, L., Moutiez, M., Li, Y., Belin, P., Seguin, J., Le Du, M.H., Thai, R., Masson, C., Fonvielle, M., Pernodet, J.L., *et al.*(2011). Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis. Nucleic Acids Res *39*, 4475-4489.

Seguin, J., Moutiez, M., Li, Y., Belin, P., Lecoq, A., Fonvielle, M., Charbonnier, J.B., Pernodet, J.L., and Gondry, M. (2011). Nonribosomal Peptide synthesis in animals: the cyclodipeptide synthase of nematostella. Chem Biol *18*, 1362-1368.

Soding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res *33*, W244-248.

Tang, M.R., Sternberg, D., Behr, R.K., Sloma, A., and Berka, R.M. (2006). Use of transcriptional profiling & bioinformatics to solve production problems. Ind Biotechnol *2*, 66-74.

Vetting, M.W., Hegde, S.S., and Blanchard, J.S. (2010). The structure and mechanism of the *Mycobacterium tuberculosis* cyclodityrosine synthetase. Nat Chem Biol *6*, 797-799.

FIGURE LEGENDS

Figure 1. Structure-based alignment of the nine characterized CDPSs and the Ndas_1148 protein from *N. dassonvillei*

The secondary-structural elements of AlbC of *S. noursei* (entry 3OQV), Rv2275 of *M. tuberculosis* (entry 2X9Q) and YvmC of *B. licheniformis*(YvmC-Blic) (entry 3OQJ) are indicated above the alignment. The residues constituting the aminoacyl moiety-binding pocket are colored in pink. The residues that are not visible in X-ray structures are indicated in gray. The regions of the three CDPSs that do not superimpose well are shown with a gray background. The other CDPSs used in the alignment are YvmC from *B. subtilis* (YvmC-Bsub), YvmC from *B. thuringiensis* (YvmC-Bthu), pSHaeC06 from *S. haemolyticus*, Plu0297 from *P. luminescens*, Jk0923 from *C. jeikeium* and Nvec-CDPS2 from *N. vectensis*. The strictly conserved residues are indicated by a red (for the ten CDPSs) or orange (for the nine bacterial CDPSs) background. The basic residues suspected to interact with the tRNA moiety of the substrate are indicated by a blue background. Black stars above the AlbC sequence indicate the residues that were modified by site-directed mutagenesis in this study.

Figure 2. LC-MS/MS analyses of the cyclodipeptides secreted into the culture supernatant of *E. coli* cells producing Ndas_1148

(A) UV traces (λ = 214 nm) of the culture medium of Ndas_1148-producing cells (blue) and cells containing the empty cloning vector (gray). (B) Corresponding EIC chromatograms with the same color codes. The identity of the cyclodipeptides produced is indicated. Cyclodipeptides were identified by MS/MS (see Figure S2 for an example).

Figure 3. Comparison of cyclodipeptide-synthesizing activities and protein expression profiles for Ndas_1148, AlbC and AlbC-p(Ndas_1148).

(A) Histogram of the amounts of the various cyclodipeptides synthesized by Ndas_1148 (blue), AlbC (red), and AlbC-p(Ndas_1148) (green). Cyclodipeptide amounts are shown with errors bars because the values are obtained from at least two independent experiments, except for cLL for which only one experiment allowed its quantification. (B) Western blot analysis

of soluble (S) and insoluble (I) protein fractions after bacterial lysis with an Eaton press (Braud et al., 2005).

Figure 4. Comparison of the Ndas_1148 model with the crystal structure of AlbC

Electrostatic surface potential of AlbC (A) and Ndas_1148 (B), mapped on their solvent-accessible surfaces at contouring levels of 5 kTe-1. Positive charge is in blue and negative charge in red. Surface-accessible pockets of AlbC with the phenylalanyl moiety of a Phe-$tRNA^{Phe}$ (C) and Ndas_1148 (D). Residues at the base, the side and the entrance of the pockets are colored with green, orange and yellow carbons, respectively. The name of the corresponding residues that are different in the two pockets is written in pink in the Ndas_1148 pocket.

Figure 5. Comparison of cyclodipeptide-synthesizing activities and protein expression profiles for AlbC and its variants.

(A) Histogram of the amounts of the various cyclodipeptides synthesized by AlbC (red), AlbC T206M (gray), AlbC V65L (green), AlbC L200C (purple), AlbC F186L (blue), AlbC V67A (orange) and AlbC M152H (light blue). Cyclodipeptide amounts are shown with errors bars because the values are obtained from at least two independent experiments, except for cLL for which only one experiment allowed its quantification. (B) Western blot analysis of soluble (S) and insoluble (I) protein fractions after bacterial lysis with an Eaton press (Braud et al., 2005). P1 and P2 represent AlbC-p(Ndas_1148)/M152 and AlbC-p(Ndas_1148)/L200N, respectively.

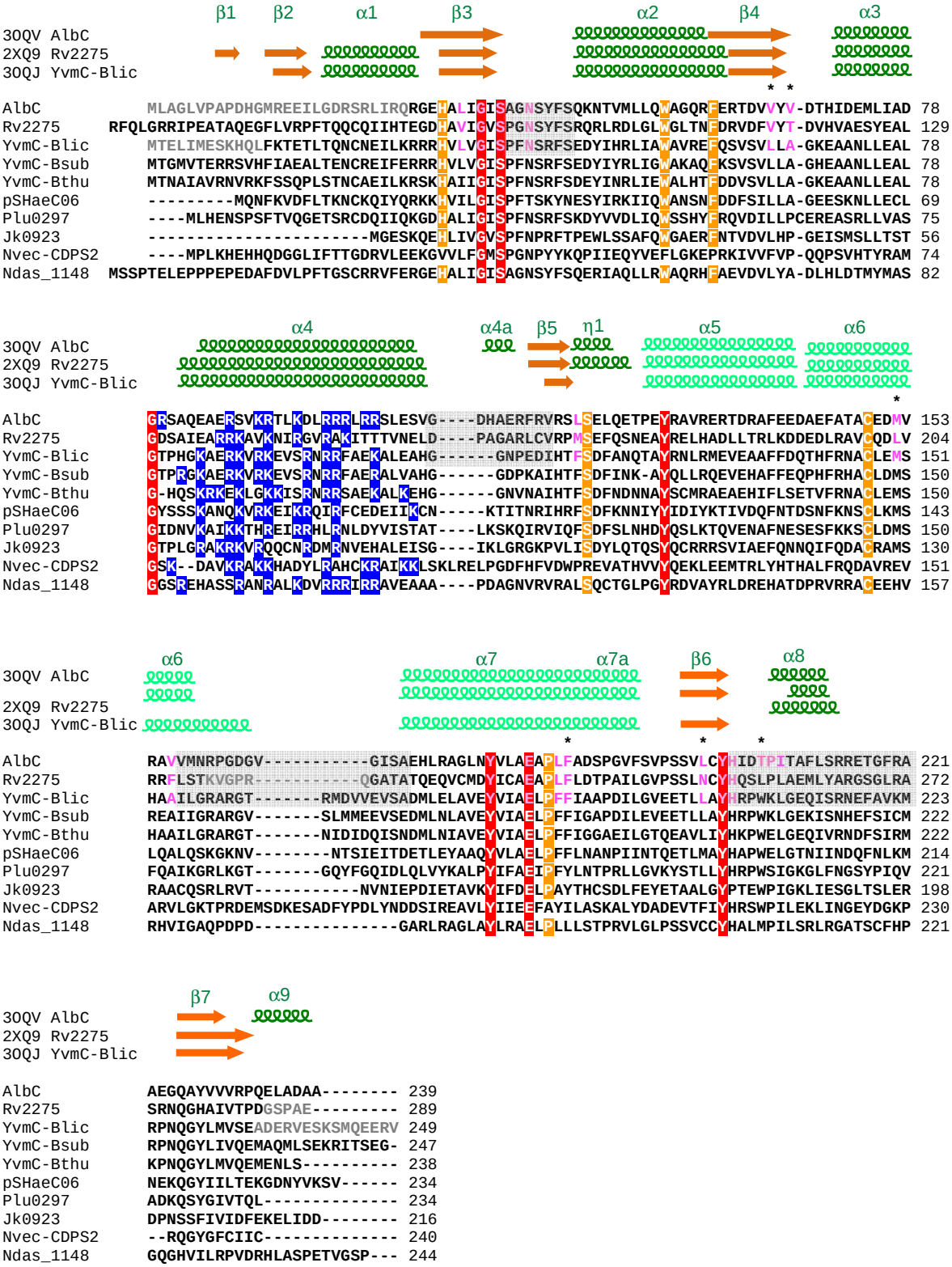

β1 β2 α1 β3 α2 β4 α3
3OQV AlbC
2XQ9 Rv2275
3OQJ YvmC-Blic
AlbC MLAGLVPAPDHGMREEILGDRSRLIRQRGEHALIGISAGNSYFSQKNTVMLLQWAGQRFERTDVYV-DTHIDEMLIAD 78
Rv2275 RFQLGRRIPEATAQEGFLVRPFTQQCQIIHTEGDHAVIGVSPGNSYFSRQRLRDLGLWGLTNFDRVDFVYT-DVHVAESYEAL 129
YvmC-Blic MTELIMESKHQLFKTETLTQNCNEILKRRRHVLVGISPFNSRFSEDYIHRLIAWAVREFQSVSVLLA-GKEAANLLEAL 78
YvmC-Bsub MTGMVTERRSVHFIAEALTENCREIFERRRHVLVGISPFNSRFSEDYIYRLIGWAKAQFKSVSVLLA-GHEAANLLEAL 78
YvmC-Bthu MTNAIAVRNVRKFSSQPLSTNCAEILKRSKHAIIGISPFNSRFSDEYINRLIEWALHTFDDVSVLLA-GKEAANLLEAL 78
pSHaeC06 ---------MQNFKVDFLTKNCKQIYQRKKHVILGISPFTSKYNESYIRKIIQWANSNFDDFSILLA-GEESKNLLECL 69
Plu0297 ----MLHENSPSFTVQGETSRCDQIIQKGDHALIGISPFNSRFSKDYVVDLIQWSSHYFRQVDILLPCEREASRLLVAS 75
Jk0923 ----------------------MGESKQEHLIVGVSPFNPRFTPEWLSSAFQWGAERFNTVDVLHP-GEISMSLLTST 56
Nvec-CDPS2 ----MPLKHEHHQDGGLIFTTGDRVLEEKGVVLFGMSPGNPYYKQPIIEQYVEFLGKEPRKIVVFVP-QQPSVHTYRAM 74
Ndas_1148 MSSPTELEPPPEPEDAFDVLPFTGSCRRVFERGEHALIGISAGNSYFSQERIAQLLRWAQRHFAEVDVLYA-DLHLDTMYMAS 82
α4 α4a β5 η1 α5 α6
3OQV AlbC
2XQ9 Rv2275
3OQJ YvmC-Blic
AlbC GRSAQEAERSVKRTLKDLRRRLRRSLESVG----DHAERFRVRSLSELQETPEYRAVRERTDRAFEEDAEFATAGEDMV 153
Rv2275 GDSAIEARRKAVKNIRGVRAKITTTVNELD----PAGARLCVRPMSEFQSNEAYRELHADLLTRLKDDEDLRAVCQDLV 204
YvmC-Blic GTPHGKAERKVRKEVSRNRRFAEKALEAHG------GNPEDIHTFSDFANQTAYRNLRMEVEAAFFDQTHFRNACLEMS 151
YvmC-Bsub GTPRGKAERKVRKEVSRNRRFAERALVAHG------GDPKAIHTFSDFINK-AYQLLRQEVEHAFFEQPHFRHACLDMS 150
YvmC-Bthu G-HQSKRKEKLGKKISRNRRSAEKALKEHG------GNVNAIHTFSDFNDNNAYSCMRAEAEHIFLSETVFRNACLEMS 150
pSHaeC06 GYSSSKANQKVRKEIKRQIRFCEDEIIKCN-----KTITNRIHRFSDFKNNIYYIDIYKTIVDQFNTDSNFKNSCLKMS 143
Plu0297 GIDNVKAIKKTHREIRRHLRNLDYVISTAT----LKSKQIRVIQFSDFSLNHDYQSLKTQVENAFNESESFKKSCLDMS 150
Jk0923 GTPLGRAKRKVRQQCNRDMRNVEHALEISG----IKLGRGKPVLISDYLQTQSYQCRRRSVIAEFQNNQIFQDACRAMS 130
Nvec-CDPS2 GSK--DAVKRAKKHADYLRAHCKRAIKKLSKLRELPGDFHFVDWPREVATHVVYQEKLEEMTRLYHTHALFRQDAVREV 151
Ndas_1148 GGSREHASSRANRALKDVRRRIRRAVEAAA----PDAGNVRVRALSQCTGLPGYRDVAYRLDREHATDPRVRRACEEHV 157
α6 α7 α7a β6 α8
3OQV AlbC
2XQ9 Rv2275
3OQJ YvmC-Blic
AlbC RAVVMNRPGDGV----------GISAEHLRAGLNYVLAEAPLFADSPGVFSVPSSVLCYHIDTPITAFLSRRETGFRA 221
Rv2275 RRFLSTKVGPR----------QGATATQEQVCMDYICAEAPLFLDTPAILGVPSSLNCYHQSLPLAEMLYARGSGLRA 272
YvmC-Blic HAAILGRARGT-------RMDVVEVSADMLELAVEYVIAELPFFIAAPDILGVEETLLAYHRPWKLGEQISRNEFAVKM 223
YvmC-Bsub REAIIGRARGV-------SLMMEEVSEDMLNLAVEYVIAELPFFIGAPDILEVEETLLAYHRPWKLGEKISNHEFSICM 222
YvmC-Bthu HAAILGRARGT-------NIDIDQISNDMLNIAVEYVIAELPFFIGGAEILGTQEAVLIYHKPWELGEQIVRNDFSIRM 222
pSHaeC06 LQALQSKGKNV--------NTSIEITDETLEYAAQYVLAELPFFLNANPIINTQETLMAYHAPWELGTNIINDQFNLKM 214
Plu0297 FQAIKGRLKGT-------GQYFGQIDLQLVYKALPYIFAEIPFYLNTPRLLGVKYSTLLYHRPWSIGKGLFNGSYPIQV 221
Jk0923 RAACQSRLRVT----------NVNIEPDIETAVKYIFDELPAYTHCSDLFEYETAALGYPTEWPIGKLIESGLTSLER 198
Nvec-CDPS2 ARVLGKTPRDEMSDKESADFYPDLYNDDSIREAVLYIIEEFAYILASKALYDADEVTFIYHRSWPILEKLINGEYDGKP 230
Ndas_1148 RHVIGAQPDPD--------------GARLRAGLAYLRAELPLLLSTPRVLGLPSSVCCYHALMPILSRLRGATSCFHP 221
β7 α9
3OQV AlbC
2XQ9 Rv2275
3OQJ YvmC-Blic
AlbC AEGQAYVVVRPQELADAA-------- 239
Rv2275 SRNQGHAIVTPDGSPAE--------- 289
YvmC-Blic RPNQGYLMVSEADERVESKSMQEERV 249
YvmC-Bsub RPNQGYLIVQEMAQMLSEKRITSEG- 247
YvmC-Bthu KPNQGYLMVQEMENLS---------- 238
pSHaeC06 NEKQGYIILTEKGDNYVKSV------ 234
Plu0297 ADKQSYGIVTQL-------------- 234
Jk0923 DPNSSFIVIDFEKELIDD-------- 216
Nvec-CDPS2 --RQGYGFCIIC---------- 240
Ndas_1148 GQGHVILRPVDRHLASPETVGSP--- 244

Figure 1

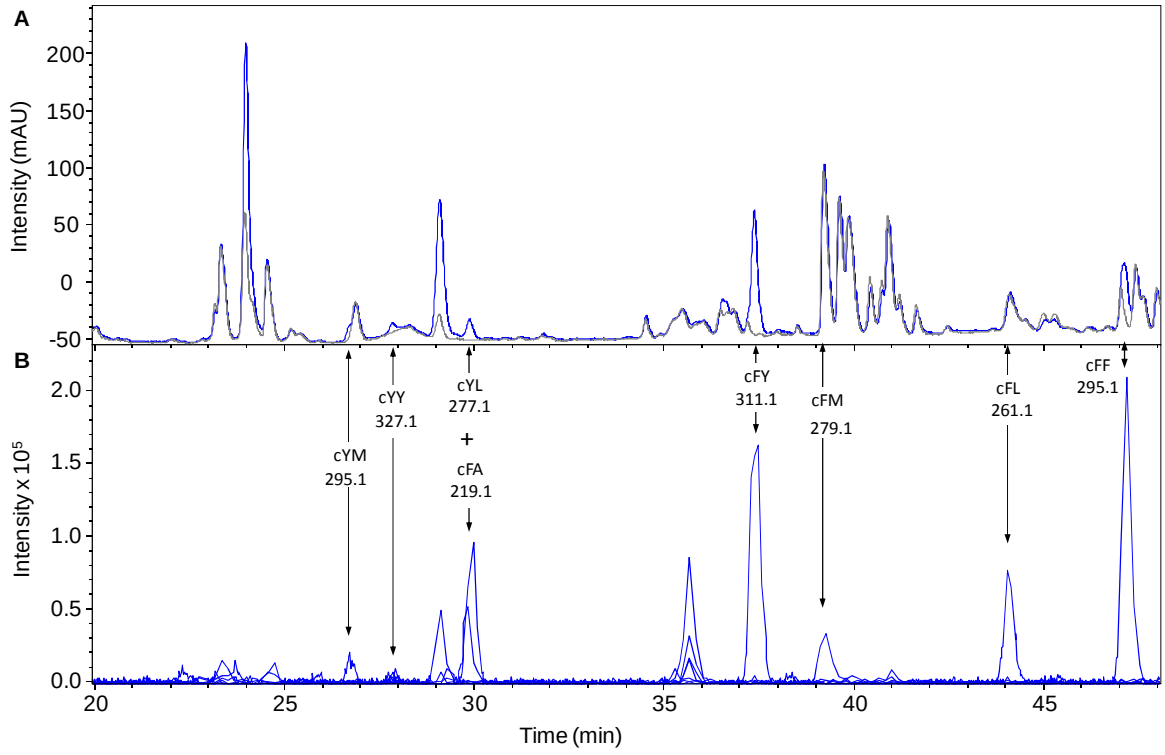

A
Intensity (mAU)
200
150
100
50
0
-50
B
Intensity x 10^5
2.0
1.5
1.0
0.5
0.0
cYM
295.1
cYY
327.1
cYL
277.1
+
cFA
219.1
cFY
311.1
cFM
279.1
cFL
261.1
cFF
295.1
20
25
30
35
40
45
Time (min)

Figure 2

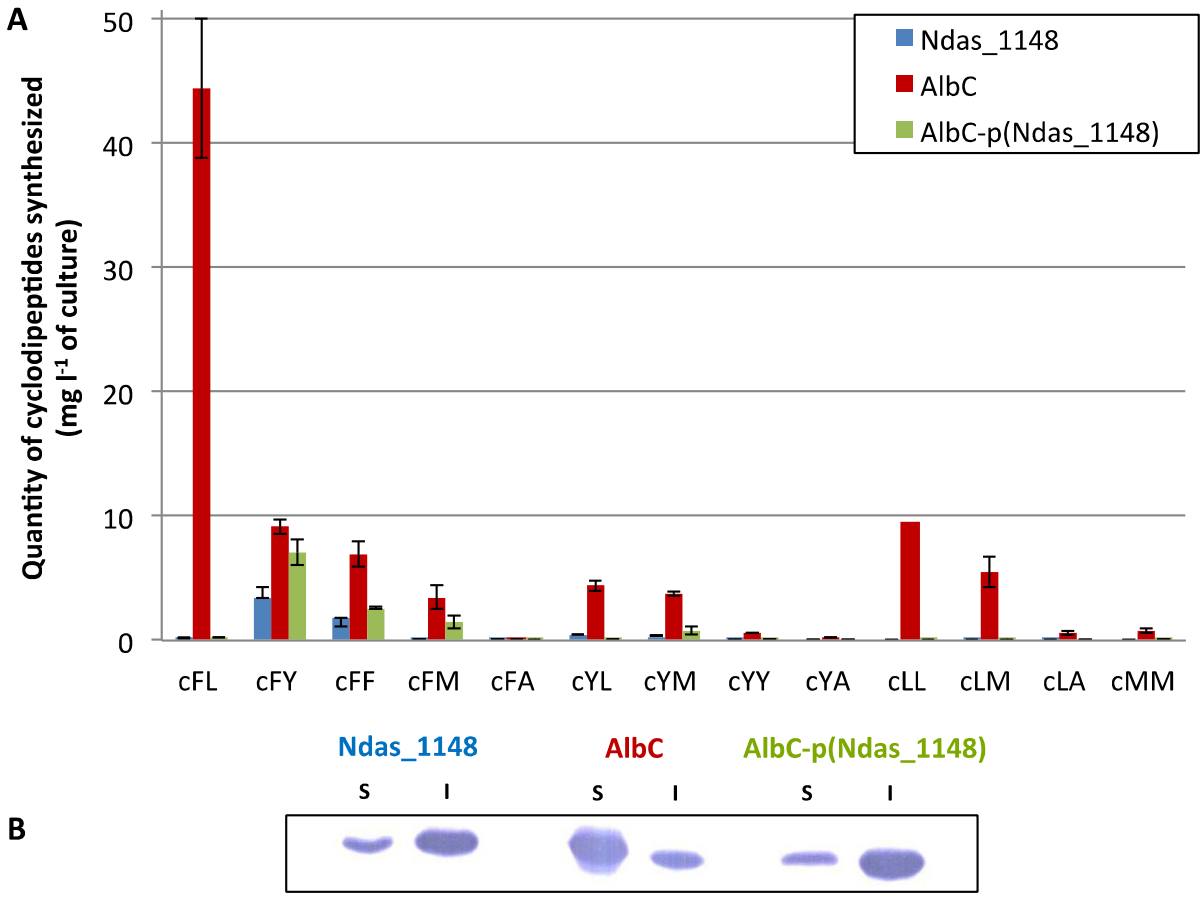

A
Quantity of cyclodipeptides synthesized (mg l-1 of culture)
50
40
30
20
10
0
Ndas_1148
AlbC
AlbC-p(Ndas_1148)
cFL
cFY
cFF
cFM
cFA
cYL
cYM
cYY
cYA
cLL
cLM
cLA
cMM
B
Ndas_1148
AlbC
AlbC-p(Ndas_1148)
S
I
S
I
S
I

Figure 3

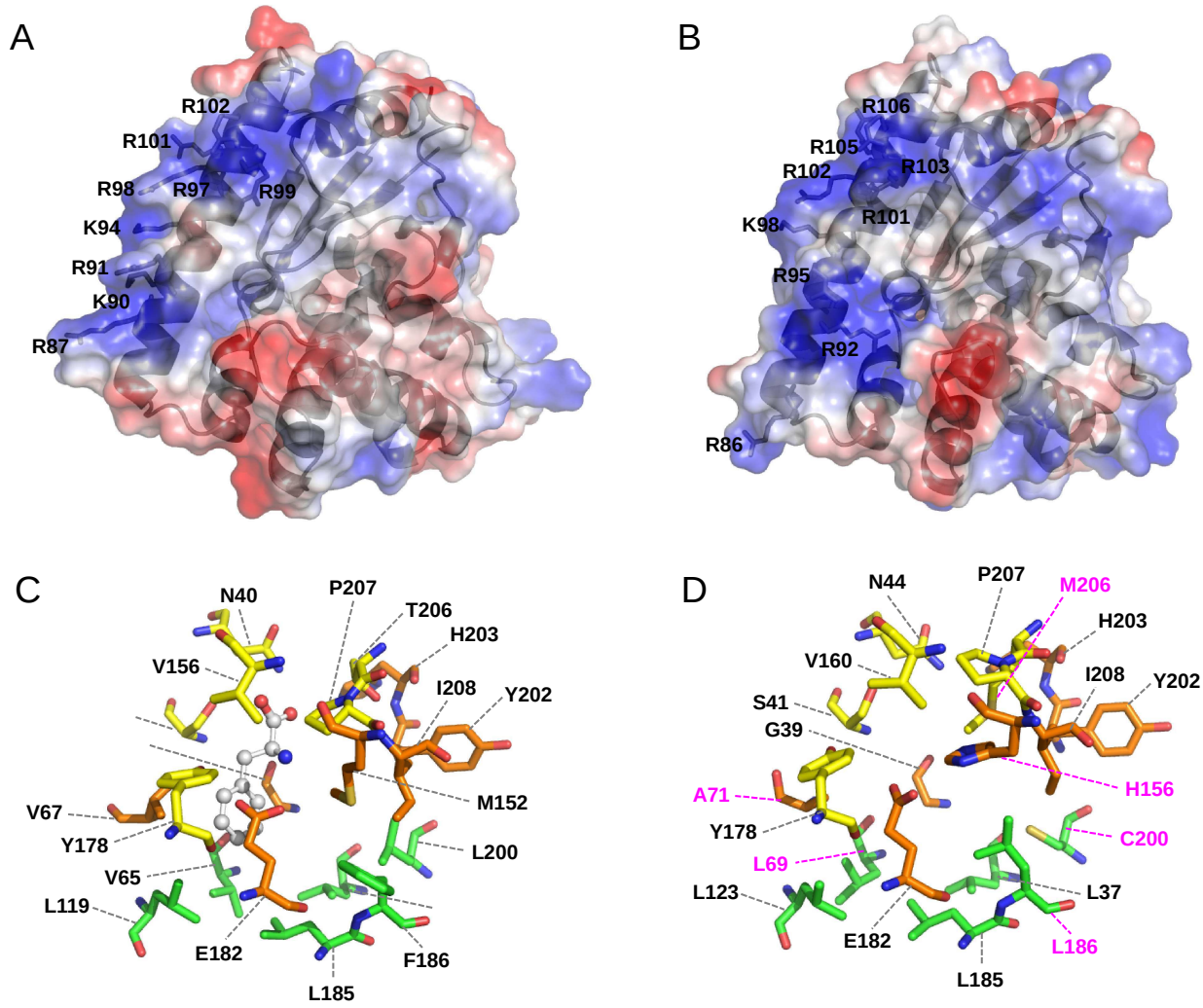

A
R102
R101
R98
R97
R99
K94
R91
K90
R87
B
R106
R105
R102
R103
K98
R101
R95
R92
R86
C
N40
P207
T206
H203
V156
I208
Y202
M152
V67
Y178
L200
V65
L119
E182
F186
L185
D
N44
P207
M206
H203
V160
I208
Y202
S41
G39
H156
A71
Y178
C200
L69
L123
L37
E182
L186
L185

Figure 4

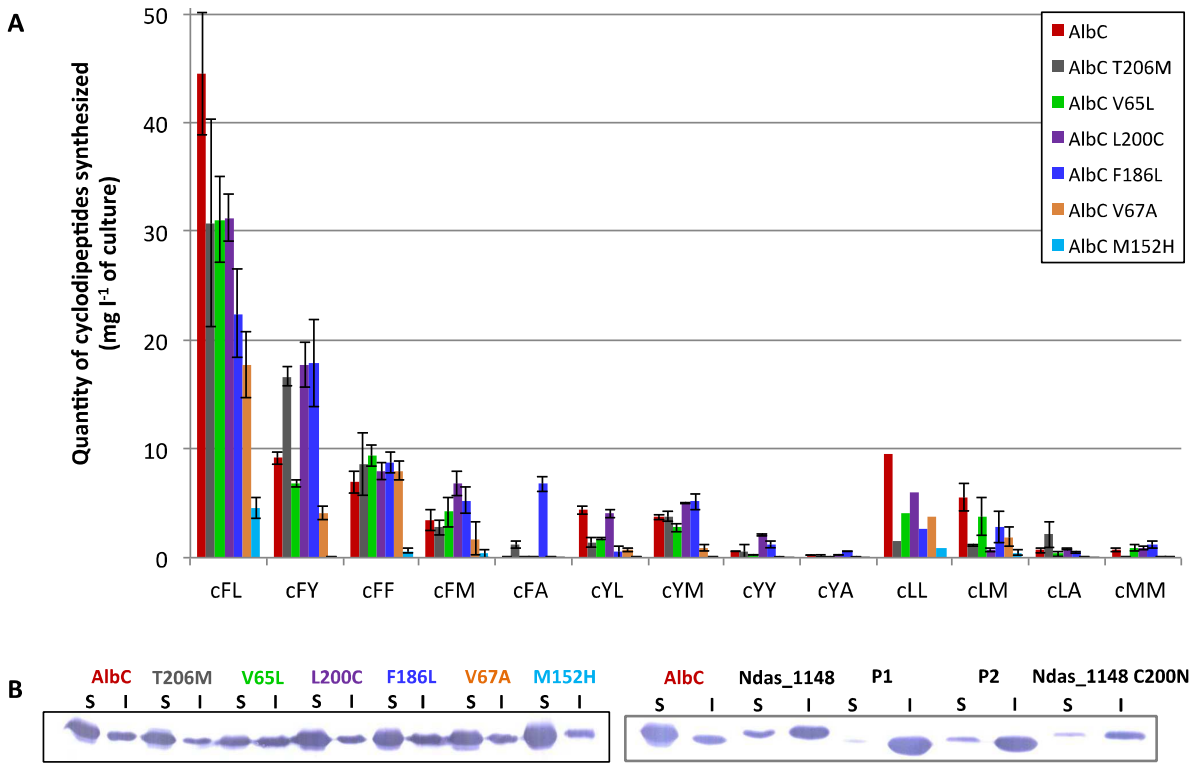

A
Quantity of cyclodipeptides synthesized (mg l-1 of culture)
50
40
30
20
10
0
cFL
cFY
cFF
cFM
cFA
cYL
cYM
cYY
cYA
cLL
cLM
cLA
cMM
AlbC
AlbC T206M
AlbC V65L
AlbC L200C
AlbC F186L
AlbC V67A
AlbC M152H
B
AlbC
T206M
V65L
L200C
F186L
V67A
M152H
S I S I S I S I S I S I S I
AlbC
Ndas_1148
P1
P2
Ndas_1148 C200N
S I S I S I S I S I

Figure 5

SUPPLEMENTARY FIGURES

Table S1. Primers used in this study.

For construction of the mutant AlbC-p(Ndas_1148):	
AlbC V65L V67A	5'-TTCGAGCGCACCGATGTCCTCTATGCCGACACCCACATCGAC
AlbC M152H	5'-ACCGCCTGCGAGGACCATGTGCGGGCCGTGGTG
AlbC F186L	5'-AGGCCCCGCTCCTCGCTGACTCGCCCGGA
AlbC L200C T206M	5'-GTCCCCTCCTCGGTGTGCTGCTACCACATCGACATGCCGATCACGGCGTTC
For construction of other AlbC mutants and the mutant Ndas_1148:	
AlbC V65L	5'-GAGCGCACCGATGTCCTCTATGTCGACACCCAC
AlbC V67A	5'-ACCGATGTCGTCTATGCTGACACCCACATCGAC
AlbC L200C	5'-GTCCCCTCCTCGGTGTGCTGCTACCACATCGAC
AlbC T206M	5'-CTCTGCTACCACATCGACATGCCAATCACGGCGTTC
AlbC-p(Ndas_1148)/M152	5'-CACCGCCTGCGAGGACATGGTGCGGGCCGTGGTG
AlbC-p(Ndas_1148)/L200N	5'-GTCCCCTCCTCGGTGAACTGCTACCACATCGAC
Ndas_1148 C200N	5'-GGTCTGCCGAGCAGCGTTAATTGTTATCATGCACTGATG

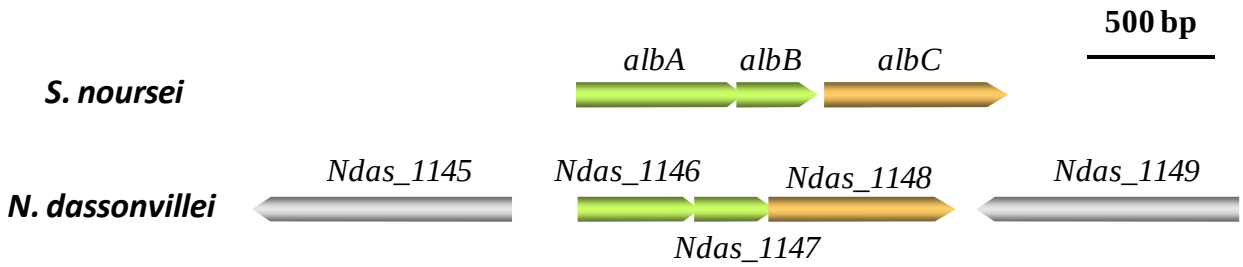


Figure S1. Genomic contexts showing the neighboring genes of the genes encoding AlbC and Ndas_1148. The CDPS genes, the genes encoding the two subunits of CDO and the genes encoding proteins annotated as methyl transferases are shown in brown, green and gray, respectively.

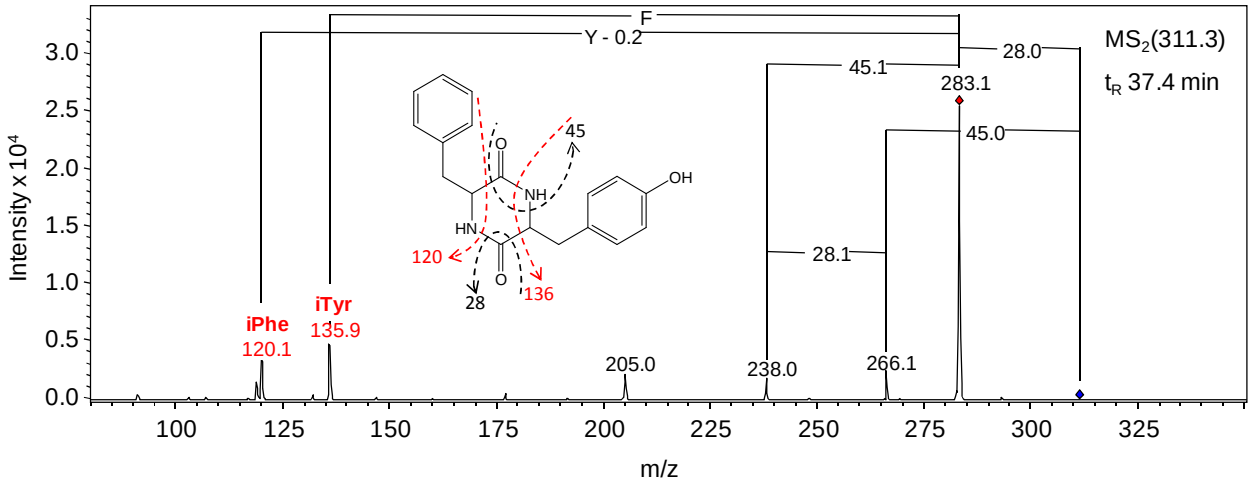


Figure S2. MS/MS of the major compound produced by Ndas_1148, identified as cFY. The fragmentation pattern is characteristic of a cyclodipeptide and red labels at 120.0 ± 0.1 and 136 ± 0.1 match the immonium ions of Phe (iPhe) and Tyr (iTyr), respectively.

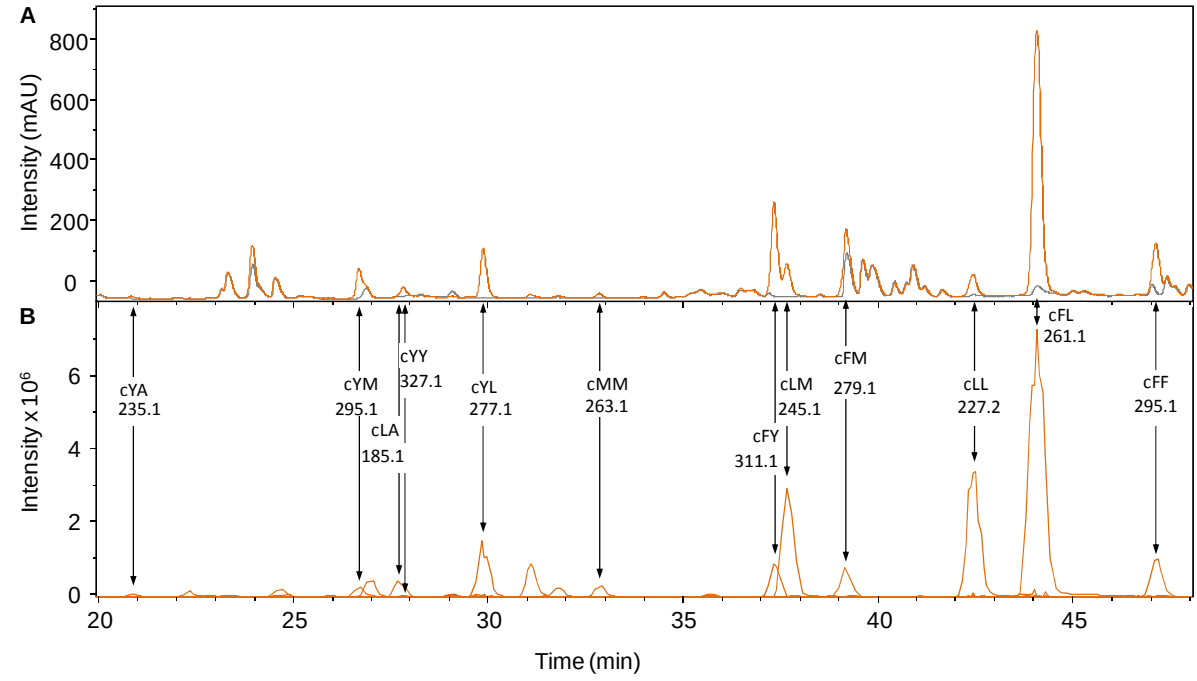


Figure S3. LC-MS/MS analysis of the cyclodipeptides secreted into the culture supernatant of *E. coli* cells producing AlbC. (A) UV traces (λ = 214 nm) of the culture medium of AlbC-producing cells (orange) and cells containing the empty cloning vector (gray). (B) Corresponding EIC chromatograms with the same color codes. This experiment was duplicated and gave similar results.

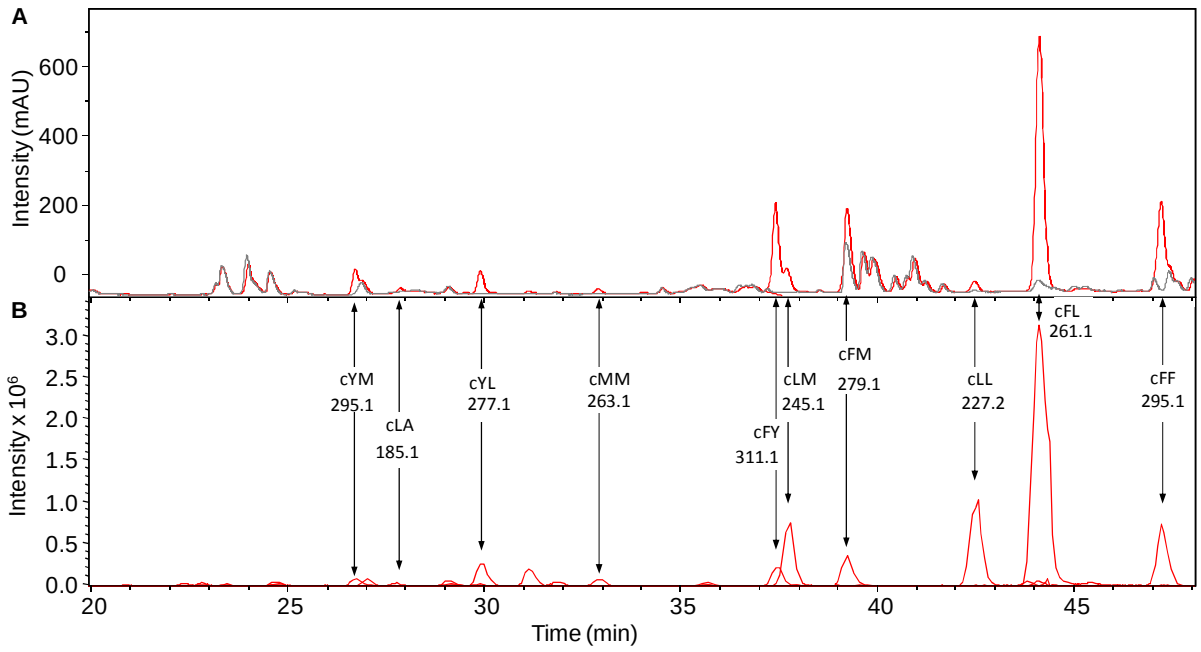


Figure S4. LC-MS/MS analysis of the cyclodipeptides secreted into the culture supernatant of *E. coli* cells producing AlbC V65L. (A) UV traces (λ = 214 nm) of the culture medium of AlbC V65L-producing cells (red) and cells containing the empty cloning vector (gray). (B) Corresponding EIC chromatograms with the same color codes. This experiment was duplicated and gave similar results.

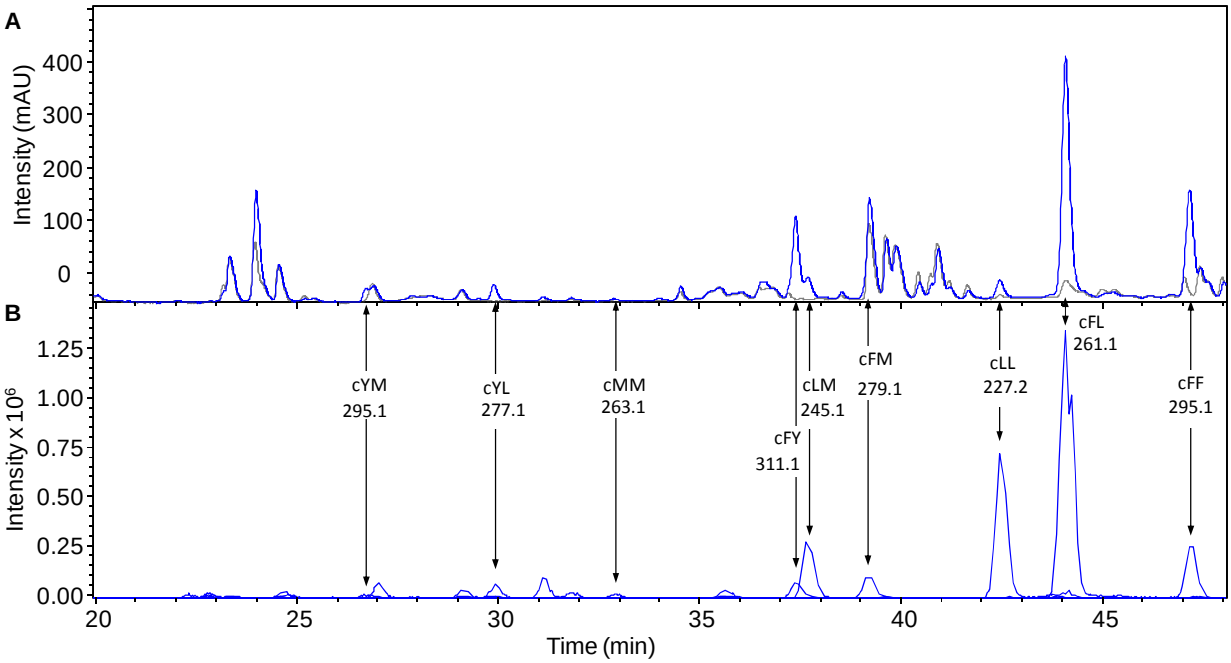


Figure S5. LC-MS/MS analysis of the cyclodipeptides secreted into the culture supernatant of *E. coli* cells producing AlbC V67A. (A) UV traces (λ = 214 nm) of the culture medium of AlbC V67A-producing cells (blue) and cells containing the empty cloning vector (gray). (B) Corresponding EIC chromatograms with the same color codes. This experiment was duplicated and gave similar results.

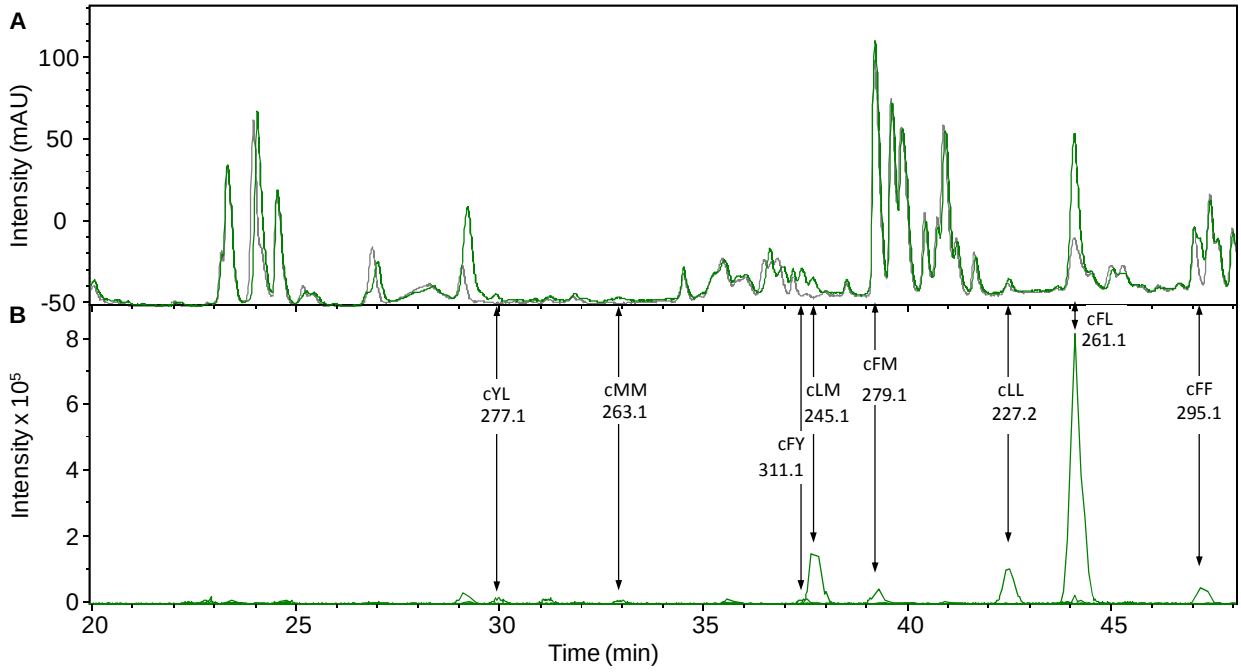


Figure S6. LC-MS/MS analysis of the cyclodipeptides secreted into the culture supernatant of *E. coli* cells producing AlbC M152H. (A) UV traces (λ = 214 nm) of the culture medium of AlbC M152H-producing cells (green) and cells containing the empty cloning vector (gray). (B) Corresponding EIC chromatograms with the same color codes. This experiment was duplicated and gave similar results.

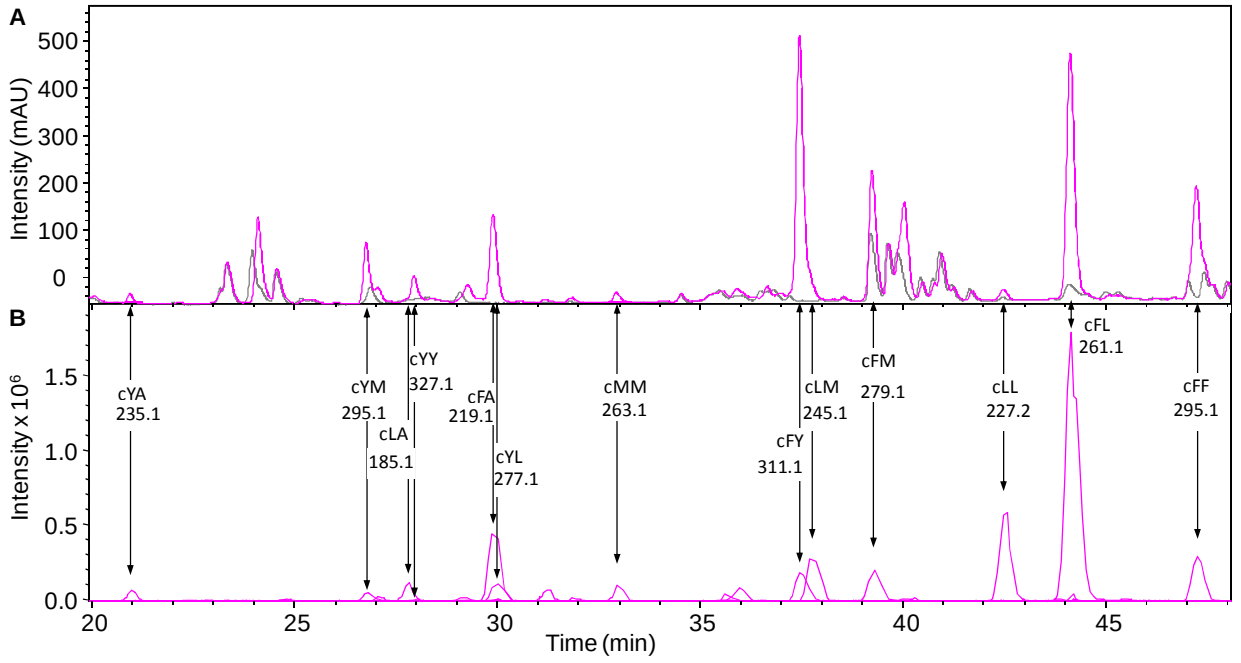


Figure S7. LC-MS/MS analysis of the cyclodipeptides secreted into the culture supernatant of *E. coli* cells producing AlbC F186L. (A) UV traces (λ = 214 nm) of the culture medium of AlbC F186L-producing cells (pink) and cells containing the empty cloning vector (gray). (B) Corresponding EIC chromatograms with the same color codes. This experiment was duplicated and gave similar results.

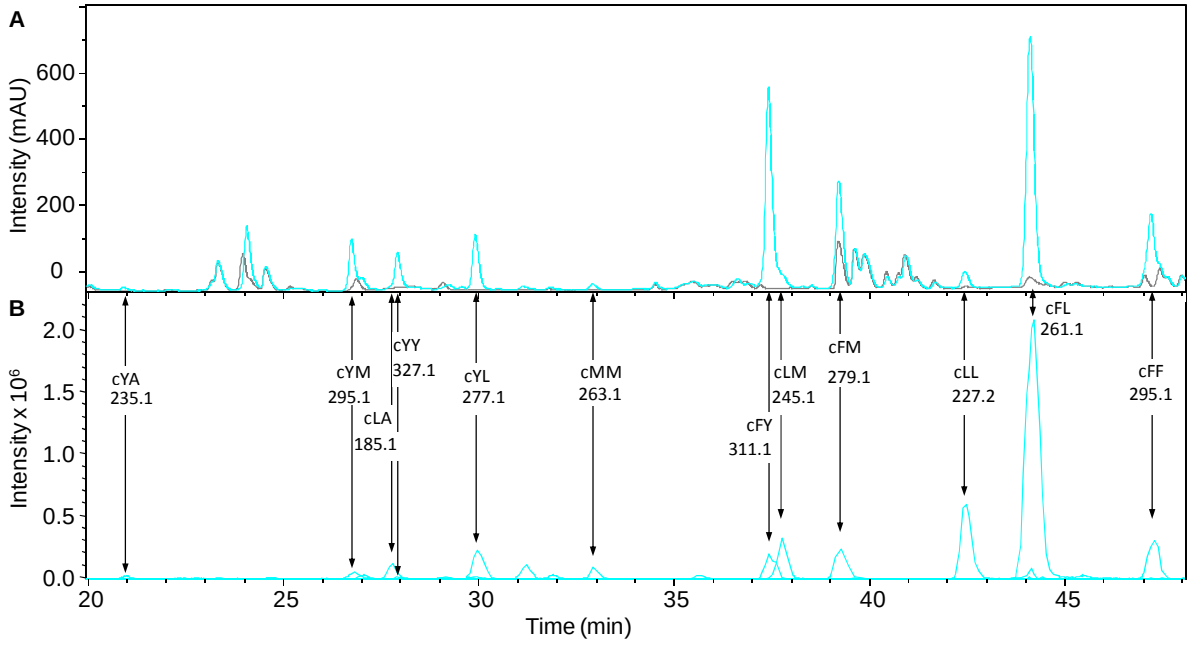


Figure S8. LC-MS/MS analysis of the cyclodipeptides secreted into the culture supernatant of *E. coli* cells producing AlbC L200C. (A) UV traces (λ = 214 nm) of the culture medium of AlbC L200C-producing cells (light blue) and cells containing the empty cloning vector (gray). (B) Corresponding EIC chromatograms with the same color codes. This experiment was duplicated and gave similar results.

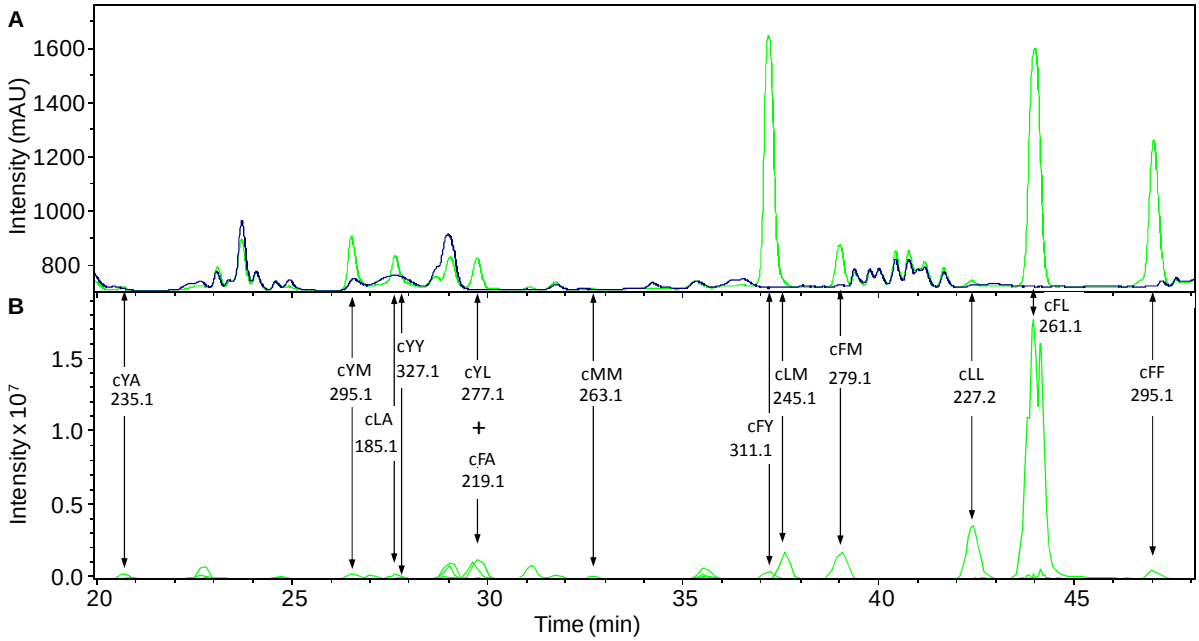


Figure S9. LC-MS/MS analysis of the cyclodipeptides secreted into the culture supernatant of *E. coli* cells producing AlbC T206M. (A) UV traces (λ = 214 nm) of the culture medium of AlbC T206M-producing cells (light green) and cells containing the empty cloning vector (gray). (B) Corresponding EIC chromatograms with the same color codes. This experiment was duplicated and gave similar results.

3.3 Additional information

In the manuscript of the article, we have described the cyclodipeptides produced by the recombinant Ndas_1148 expressed in *E. coli*. Although we supposed that the final products from the Ndas_1148-dependent biosynthetic pathway could be dehydrogenated and methylated cyclodipeptides because in *N. dassonvillei*, *Ndas_1148* is clustered with genes encoding a CDO homologue and methyl transferases (**Figure S1** of the section **3.2**), it should still be of great interest to experimentally determine the DKPs produced by Ndas_1148 and the putative cyclodipeptide-tailoring enzymes in *N. dassonvillei*. However, *N. dassonvillei* is a pathogen and can induce infections in humans (González-López et al. 2011). The order of the strain is thus a complicated procedure because of the strict eximination. Now, our collaborative team directed by Jean-Luc Pernodet (IGM, CNRS UMR 8621, Université Paris-Sud 11) has received the strain *N. dassonvillei* and has carried out the cultivation. We can thus analyze the culture supernatants by LC-MS/MS very soon so as to identify the final DKPs produced through the Ndas_1148-dependent biosynthetic pathway. These results will be added in the article manuscript before submission.

4 AlbC-IMI from *Streptomyces* sp. IMI 351155, a new CDPS to further explore the interactions between CDPSs and their substrates and the mechanism used by CDPSs

4.1 Introduction

CDPSs have been proved to use a sequential ping-pong mechanism to catalyze the formation of cyclodipeptides (Vetting et al. 2010; Bonnefond et al. 2011; Sauguet et al. 2011). Vetting and his coworkers have reported the kinetic parameters of Rv2275 for *E. coli* substrate (Vetting et al. 2010). However, they are the apparent kinetic parameters for Tyr-tRNA^{Tyr} which is not only the first substrate, but also the second one because Rv2275 produces cYY (Gondry et al. 2009). Thus, it is not possible to discriminate the binding of the two ordered substrates and in particular to determine the kinetic parameters for the first and the second substrates. In order to solve this problem, such a CDPS is needed: a protein which does not use the same aa-tRNA as the two ordered substrates in the reaction, and consequently does not produce the cyclodipeptide composed of the same amino acid residue. Unfortunately, no CDPS characterized to date meets the condition. Only AlbC from *S. nousei* mainly synthesizes a cyclodipeptide composed of two different amino acids, that is cFL, but it also produces a lot of cFF and significant amounts of cLL (Gondry et al. 2009).

However, such CDPS with the characteristic listed above was recently identified by the Jean-Luc Pernodet's team. They were particularly interested by the strain *Streptomyces* sp. IMI 351155 described to produce 1-*N*-methylalbonoursin (Gurney and Mantle 1993) (**Figure 44A**). By using a PCR approach with highly degenerated primer, they cloned a fragment of an *albC* homologue from *Streptomyces* sp. IMI 351155. By probing a cosmid genomic library of strain *Streptomyces* sp. IMI 351155, they isolated DNA fragments susceptible to contain the complete methylalbonoursin biosynthetic gene cluster from *Streptomyces* sp. IMI 351155. These fragments have been sequenced and their sequence analyzed and compared to the one of *Streptomyces noursei*. Comparisons of cluster organization and percentage of sequence identity are given in **Figure 44B** (non published data).

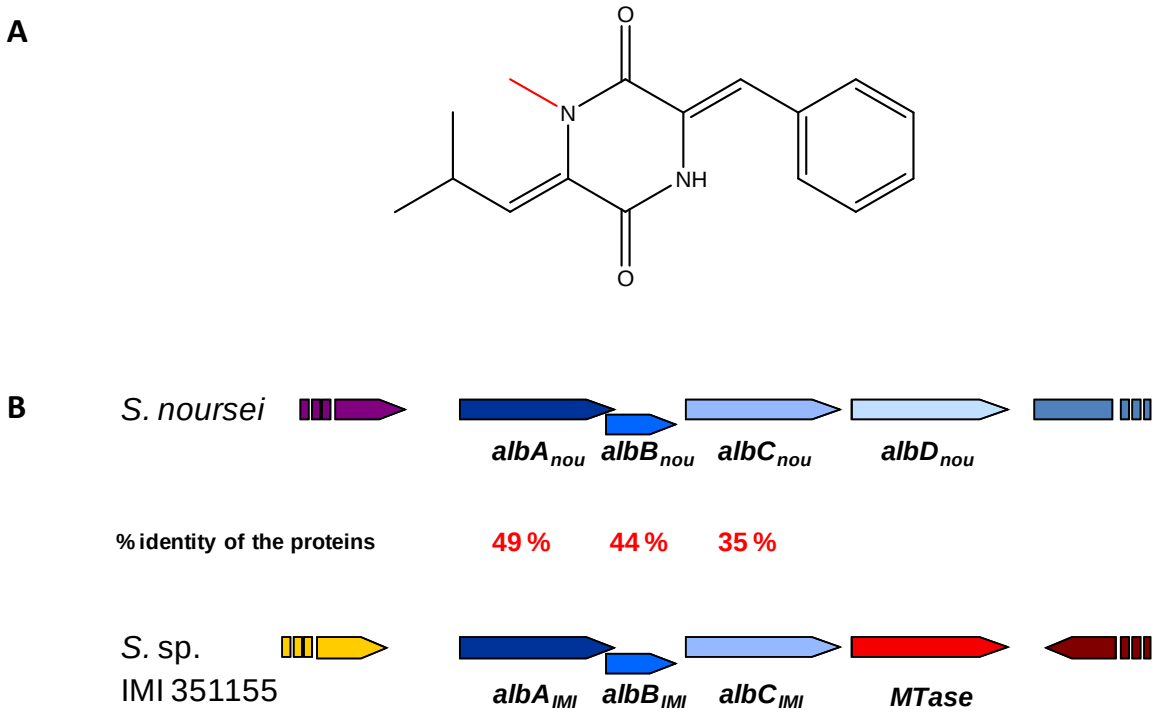


Figure 44: Structure of 1-*N*-methylalbonoursin (Gurney and Mantle 1993) with the methyl group shown in red (A), and comparison of cluster organization between *S. noursei* and *S.* sp. IMI 351155 with percentages of sequence identity indicated.

Besides, our laboratory analyzed by LC-MS/MS the culture supernatant of the strain *Streptomyces* sp. IMI 351155 and we have detected the presence of 1-*N*-methylalbonoursin. We have shown that the strain was also producing albonoursin

but cyclo(ΔPhe-ΔPhe) or its methylated derivative were not detected. This observation made us suppose that the CDPS AlbC-IMI produce cFL but not cFF. We thus decided to study this CDPS and by using its characteristics, to better understand its interaction with each of the two substrates as well as the mechanism of CDPSs.

In this chapter, we will present our preliminary results obtained so far, which concern the characterization of the *in vivo* activity of AlbC-IMI, the purification of the enzyme, as well as the preliminary work on substrate order used by AlbC-IMI. These results are significant because they allow performing a complete study on AlbC-IMI in the future.

4.2 Materials and methods

4.2.1 Expression of recombinant AlbC-IMI in *E. coli* and analysis of its *in vivo* activity

The pQE60 vector (Qiagen) was used to construct the plasmid pQE60-AlbC-IMI, as previously described in (Gondry et al. 2009). AlbC-IMI was expressed in *E. coli* M15 [pREP4] cells (Invitrogen) transformed with the plasmid pQE60-AlbC-IMI. The culture supernatant was then analyzed by LC-MS/MS. Cells were disrupted and expression of the protein AlbC-IMI was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and western blot. Proteins on SDS-PAGE gel were revealed by Coomassie brilliant blue whereas protein transferred to the Polyvinylidene fluoride (PVDF) membrane is revealed by an immuno-reaction. Meanwhile, the same experiments were carried out to the pQE60 vector without AlbC-IMI gene, which was taken as a control. The experimental detail was described by Gondry et al (Gondry et al. 2009).

4.2.2 Purification of AlbC-IMI

4.2.2.1 Expression of AlbC-IMI in *E. coli* BL21-AI™

The competent strain *E. coli* BL21-AI™ is transformed with the plasmid pQE60-AlbC-IMI by mixing 50 ng of the plasmid with 100 µl of competent *E. coli* BL21-AI™ cells. The mixture is incubated for 30 minutes in the ice. A thermal choc of 45 seconds at 42 °C is performed in order to make the plasmid penetrate into the bacterium. After that, the mixture is incubated for another 2 minutes in the ice. Then the transformation product is incubated in 250 µl of SOC medium for one hour at 37 °C with stir (200 rpm).

Preculture of AlbC-IMI is then performed by adding 150 µl of the transformed product in 50 ml of LB medium containing 100 µg l^{-1} ampicillin (Amp). The preculture lasts over night (about 15 hours) at 37 °C/200 rpm. The next morning, we inoculate 2 liters of LB/Amp so as to obtainan initial OD_{600} of 0.1. The culture mixture is incubated at 37 °C/200 rpm until OD_{600} reaches 0.6 - 0.7. Then, expression of AlbC-IMI is induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). The culture is incubated for 20 hours at 20 °C. Then the culture product is centrifuged for 40 minutes at 3500 g. The cell pellets are stored at -80 °C.

4.2.2.2 Extraction of soluble AlbC-IMI protein from *E. coli* BL21-AI™ cells

The cell pellets are resuspended in the lysis buffer consisting of Tris-HCl 100 mM (pH 8), 150 mM NaCl, 5% glycerol and a reducing agent (10 mM β-mercaptoethanol). Meanwhile, 1 mM PMSF and 10 µM phosphoramidon are added to inhibit activities of proteases. The bacterial cells were broken in the Eaton press to recover the recombinant AlbC-IMI protein. Eaton press allows to shatter the frozen cells by passing in a hole of small diameter by the application of a pressure up to six tons. Add again 1 mM PMSF and 10 µM phosphoramidon in the homogenate obtained. The homogenate was then treated with Benzonase® (a phosphodiesterase) of 250 U/µl in the presence of 10 mM $MgCl_2$ to degrade cellular DNA and RNA. The mixture is incubated in the cold room (4 °C) until the homogenate becomes fluid,

which indicates total degradation of nucleic acids. This process normally lasts 30 minutes to 1 hour. After that, the fluid homogenate is centrifuged for 30 minutes (4000 g, 4°C) to separate the soluble and insoluble fractions. The supernatant, that is the soluble fraction, is used to the purification procedure described below. The purification system is ÄKTA design (GE Healthcare). The purification process is monitored by UV detection at 280 nm and 260 nm.

4.2.2.3 Step one: immobilized metal ion affinity chromatography (IMAC)

His-tagged recombinant AlbC-IMI is firstly purified by IMAC with a precharged nickel column (HisTrap HP 5 ml, GE Healthcare). The column is equilibrated with 50 ml of **buffer A** (40 mM imidazole, 150 mM NaCl, 100 mM Tris-HCl (pH 8), 5% glycerol, 10 mM β-mercaptoethanol, and 1mM PMSF). The supernatant previously obtained is loaded onto the column at a flow rate of 1 ml/min. The column is then washed with **buffer A** at a flow rate of 3 ml/min until the UV detector shows a path back to its baseline. Then the elution is carried out with a linear gradient from 40 to 1000 mM imidazole at a flow rate of 1 ml/min. The fraction showing UV absorption is collected and analyzed with SDS-PAGE, then passed to the second step of purification.

4.2.2.4 Step two: heparin affinity chromatography

The UV spectrum of the fraction collected from the nickel column indicates a significant contamination by nucleic acids. In order to eliminate this contamination, the protein is further purified with a column charged with heparin (HiTrap Heparin HP 5ml, GE Healthcare). The column is equilibrated with 50 ml of **buffer C** (50 mM NaCl, 100 mM Tris-HCl (pH 8), 5% glycerol, 10 mM β-mercaptoethanol, and 1 mM PMSF). The protein obtained from the first step of purification is loaded onto the heparin column at a flow rate of 1 ml/min. The column is then washed with **buffer C** at a flow rate of 3 ml/min until the UV detector shows a path back to its baseline. The elution is carried out with a linear gradient from 50 to 1000 mM NaCl at a flow rate of 1 ml/min. The protein eluted is collected and analyzed with SDS-PAGE. The UV spectrum of the protein indicates the contamination of nucleic acids is removed.

4.2.2.5 Step three: size exclusion chromatography (SEC)

SEC is also called gel filtration chromatography or gel permeation chromatography. In order to obtain pure protein, the protein eluted from the second step is further purified by SEC with the column Sephacryl S-200 Hiprep 26/60 (GE Healthcare). In order to determine the apparent molecular weight of proteins eluted, the column is regularly calibrated from a calibration curve made after injection of standard proteins, the molecular masses of which are known. Before purification, the column is equilibrated with at least two column volumes (2 x 340 = 680 ml) of **buffer E** (20 mM Tris-HCl (pH 8), 150 mM NaCl, 5% glycerol, 1mM DTT, and 1 mM PMSF) at a flow rate of 1 ml/min. The protein eluted from the second step is concentrated to less than 6 ml then loaded onto the column at a flow rate of 1 ml/min. The elution is carried out with **buffer E**. AlbC-IMI is eluted at 160 ml, which corresponds to its molecular mass.

4.2.2.6 Concentration and conservation of the protein

The protein is concentrated with Amicon® Ultra-15 10K centrifugal filter device (Millipore) by centrifugation at 5000 g with a fixed angle rotor. Every 10 minutes, the protein is homogenized to avoid the aggregation. Then the concentration of the protein is determined by absorption of UV at 280 nm. The pure protein is analyzed with SDS-PAGE and western blot then conserved at -80 °C. The protein should beforehand be frozen in liquid nitrogen before conservation at -80 °C.

4.2.3 Substrate order of AlbC-IMI

As described in Gondry et al. (Gondry et al. 2009), the experiment is globally carried out in three steps (**Figure 45**): firstly, the acyl-enzyme-forming reaction takes place in appropriate reaction medium *in vitro* which contains radioactivity labeled amino acid ([^{14}C]Phe or [^{14}C]Leu). Secondly, the reaction medium is deposited on two SDS-PAGE gels and AlbC-IMI is separated from the other components of reaction by electrophoresis. Finally, one gel is stained in silver nitrate in order to reveal proteins whereas the proteins on the other gel are transferred onto a PVDF

membrane by western blot. The latter one is analyzed by a radioactivity analyzer "β-Imager" so as to reveal the ^{14}C labeled acyl-enzyme. In this way, we show which aa-tRNA interacts with AlbC-IMI as the first substrate.

It should be noted that $tRNA^{Phe}$ and $tRNA^{Leu}$ used in our *in vitro* experiments are obtained by *in vitro* transcription (Villet et al. 2007) performed by M. Fonvielle (UMR S 872, INSERM UPMC UPD) on the basis of the sequences of *E. coli* tRNAs. In *E. coli*, there exist six $tRNA^{Leu}$ isoacceptors. A manuscript dealing with the recognition of each of these isoacceptors by AlbC from *S. nousei* is being prepared in the laboratory. This study showed that the $tRNA^{Leu}$ CAG isoacceptor is the best substrate for AlbC from *S. nousei* compared to the other five isoacceptors. We thus used this $tRNA^{Leu}$ isoacceptor for the study of AlbC-IMI after having made sure that it is a good substrate for AlbC-IMI, which means, in the presence of this $tRNA^{Leu}$ isoacceptor and $tRNA^{Phe}$, AlbC-IMI produces cFL (data not shown).

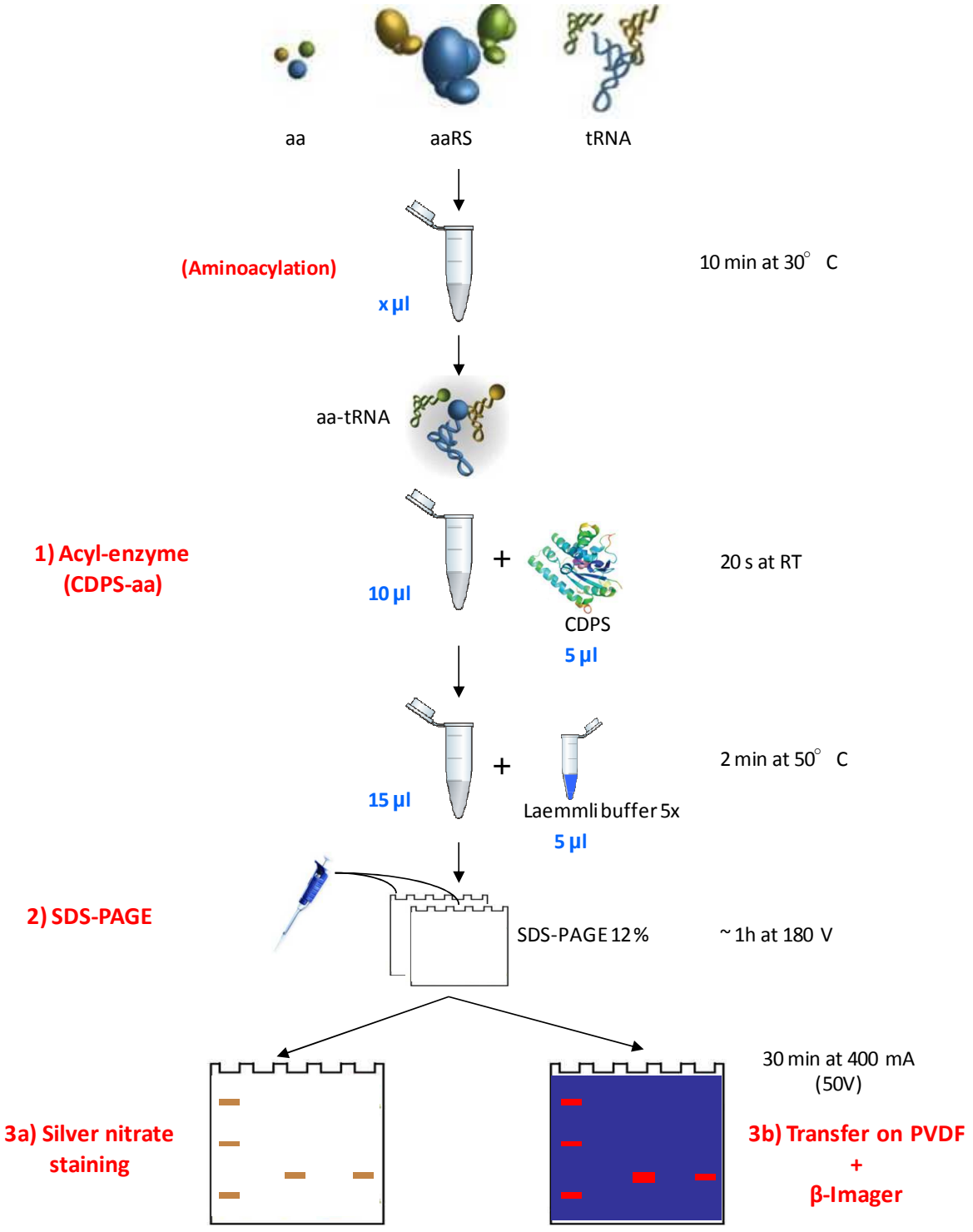


Figure 45: Procedure of acyl-enzyme experiment.

4.2.3.1 Acyl-enzyme-forming reaction

The formation of acyl-enzyme is realized by a coupled reaction. The total volume of one reaction is 10 µl. In the reaction medium, we mix 50mM Hepes-KOH (pH 7.5), 150 mM KCl, 14 mM $MgCl_2$, 2 mM ATP, 10 mM β-mercaptoethanol, 0.1 mM EDTA, 1.5 µM tRNA, 50 µM radioactivity labeled amino acid from PerkinElmer ([^{14}C]Phe or [^{14}C]Leu, the specific radioactivities of which are respectively 90 and 153 nCi/nmol), and RNase free H_2O. The aminoacylation reaction is initiated by adding 1 µM cognate aaRS. The reaction medium is then incubated at 30 °C during 10 min. The formation of acyl-enzyme is initiated by adding 5 µl of 3 µM pure AlbC-IMI. So the final concentration of AlbC-IMI in the reaction medium is 1 µM. The reaction lasts for 20 s at room temperature. Then the reaction is stopped by adding 5 µl of Laemmli buffer 5X at 50°C for 2 min. The Laemmli buffer 5X contains 0.1% (w/v) SDS, 10% (v/v) glycerol, 50 mM Tris-HCl (pH 6.8) and 100 mM DTT.

4.2.3.2 SDS-PAGE

The reaction medium is firstly mixed with a blue solution containing 62.5 mM Tris-HCl (pH 6.8), 0.02% (w/v) bromophenol blue, 2% (w/v) SDS, 2% (v/v) β-mercaptoethanol and 20% (v/v) glycerol. The sample is denatured at 95 °C for 5 min. Then 9 µl of denatured sample is deposited on two SDS-PAGE gels containing 12% acrylamide for two different revelation methods (silver nitrate staining and β-Imager) described later. 3 µl of protein marker from Bio-Rad (Precision Plus Protein™ All Blue Standards) is deposited on each gel. The migration is performed with the system Mini-Protean3 Dodeca™ Cell (Bio-Rad) at 200 volt during 45 min.

4.2.3.3 Silver nitrate staining of SDS-PAGE gel

One of the two SDS-PAGE gels is stained in silver nitrate. Le gel is firstly incubated for 30 min in the fixing solution (50% methanol, 10% acetic acid, and 100 mM ammonium acetate), then washed twice with H_2O for 30 min each time. The gel is incubated with 0.005% (w/v) sodium thiosulfate solution for 30 min then with 0.1% (w/v) silver nitrate solution for another 30 min. After that, the gel is washed with

H_2O for several seconds. Then the gel is soaked in a developer (0.036% formaldehyde, 2% sodium carbonate) for several minutes until the protein bands appear. The staining is stopped by incubating the gel with 50 mM EDTA during 30 min. Finally, the gel is washed twice with H_2O.

4.2.3.4 Transfer to the PVDF membrane and revelation by β-Imager

The protein bands of the other SDS-PAGE gel are transferred to a PVDF membrane. Firstly, the membrane is activated by 100% ethanol for 5 seconds then rinsed in the transfer buffer composed of 25 mM Tris-HCl (pH 7.5), 192 mM glycine and H_2O. The transfer is performed with a semi-dry transfer cell (Trans-Blot® SD) from Bio-Rad at 400 mA during 30 min. After transfer, the membrane is rinsed with H_2O then with a solution composed of 5% TCA and 0.5% casamino acids three times with 10 min each time in order to remove the non-specific fixation of amino acid. Finally, the membrane is dried with a hair dryer.

Before analysis of the membrane by β-Imager, the bands of molecular weight standard are highlighted by toothpicks with a ^{14}C labeled solution in order that the protein marker can be revealed by β-Imager. After that, the membrane is analyzed for 4 hours with Biospace Beta imager 2000 digital system for auto-radiography which records every beta particle and generates images on screen in real time at low levels of radioactivity.

4.3 Results

4.3.1 AlbC-IMI, an active member of the CDPS family

AlbC-IMI (279 residues) is a little longer than most of the ten characterized CDPSs (216-249 residues in general, with only one CDPS having 289 residues). It displays 35% identity and 48% similarity to AlbC from *S. noursei*. In addition, the isolated gene cluster containing *albC-IMI* from *S.* sp. IMI 351155 shows *albC-IMI* has a similar genetic environment to *albC* of *S. noursei* (**Figure 44B**). As mentioned in the

introduction section (**4.1**), AlbC-IMI was supposed to synthesize cFL but not cFF. In order to confirm this hypothesis, recombinant C-terminal histidine-tagged AlbC-IMI was expressed in *E. coli*, and then the culture supernatant was analyzed by LC-MS/MS as previously described (Gondry et al. 2009).

Five cyclodipeptides were detected by LC-MS/MS in the culture supernatant of AlbC-IMI (**Figure 46**). They are: cFL (21.2 ± 3.0 mg l^{-1} of culture supernatant), cFM (1.6 ± 0.4 mg l^{-1}), cYL (0.2 ± 0 mg l^{-1}), cLM (0.2 ± 0.1 mg l^{-1}) and trace amounts of cFF (< 0.1 mg l^{-1}) (**Figure 47A**). Thus, the result demonstrated that AlbC-IMI is an active CDPS that predominantly synthesizes cFL and produces almost no cFF. Analysis of the total expression of AlbC-IMI by SDP-PAGE and western blot showed that AlbC-IMI was much less produced than AlbC in *E. coli* (**Figure 47 B**). It could explain that AlbC-IMI synthesizes less cFL than AlbC does (44.5 ± 5.7 mg l^{-1} of culture supernatant) (**Figure 47A**).

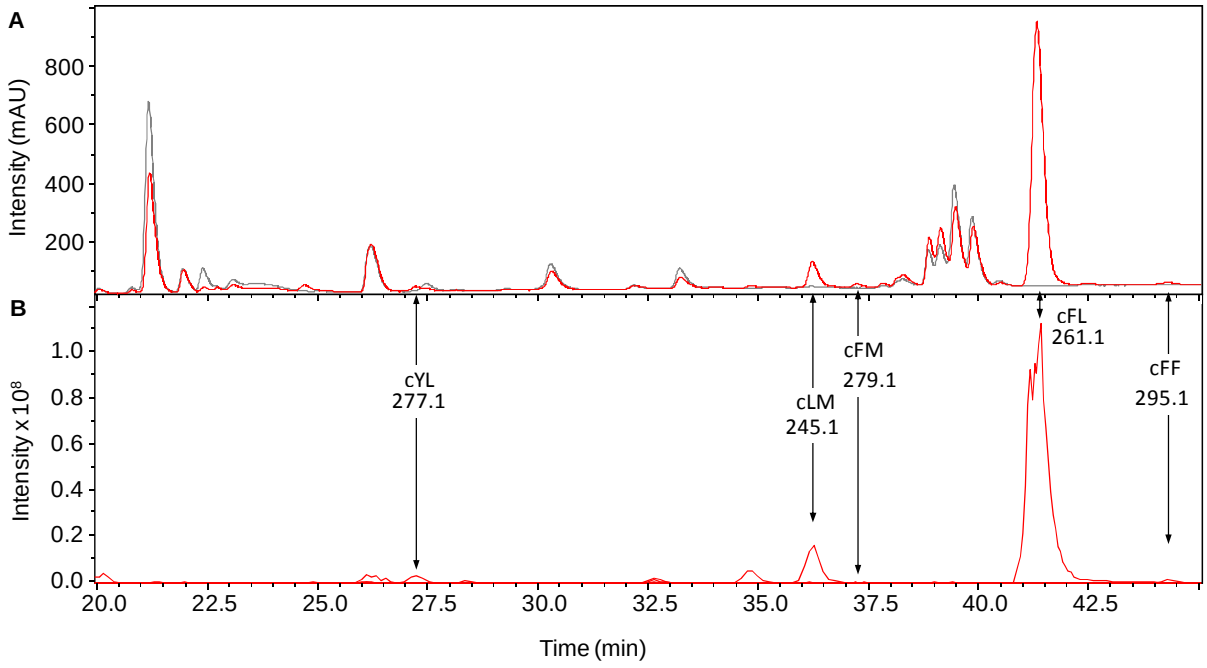


Figure 46: LC-MS/MS analyses of the cyclodipeptides secreted into the culture supernatant of *E. coli* cells producing AlbC-IMI. (A) UV traces (λ = 214 nm) of the culture medium of AlbC-IMI-producing cells (red) and cells containing the empty cloning vector (gray). (B) Corresponding EIC chromatograms with the same color codes.

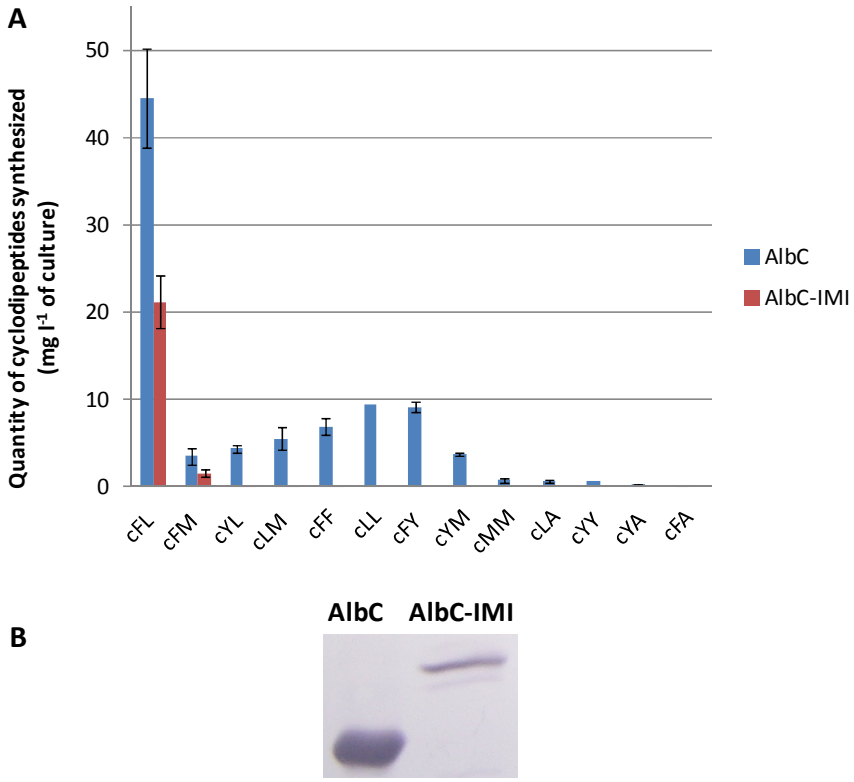


Figure 47: Comparison of cyclodipeptide-synthesizing activities and protein expression profiles for AlbC-IMI and AlbC. (A) Histogram of the amounts of the various cyclodipeptides synthesized by AlbC-IMI (red) and AlbC (blue). Cyclodipeptide amounts are shown with errors bars because the values are obtained from at least two independent experiments, except for cLL for which only one experiment allowed its quantification. (B) Expression levels of AlbC-IMI and AlbC in *E. coli* revealed by a western blot with an immunorevelation.

4.3.2 Purification of AlbC-IMI

AlbC-IMI was expressed in *E. coli* BL21-AI™ cells (Invitrogen) transformed with the plasmid pQE60-AlbC-IMI. We replaced *E. coli* M15 [pREP4], used for expression of AlbC from *S. noursei*, by *E. coli* BL21-AI™ cells in order to increase the expression level of AlbC-IMI (data not shown). The expressed histidine-tagged AlbC-IMI protein was purified with a three-step chromatographic procedure as describe by Gondry et al. (Gondry et al. 2009). However, some precaution should be taken because AlbC-IMI is susceptible to proteolysis. Addition of the protease inhibitors, phenylmethylsulfonyl fluoride (PMSF) and phosphoramidon, is indispensable during the purification procedure (see below).

The purification results of AlbC-IMI protein is shown in **Figure 48**. Analysis with SDS-PAGE and western blot showed that AlbC-IMI was well expressed in *E. coli* BL21-AI™ and well soluble (**Figure 48D**, lanes "Homogenate" and "Soluble fraction"). The first step of purification by HisTrap allowed separating His-tagged AlbC-IMI from other proteins in the cellular soluble fraction but it was contaminated by nucleic acids because the absorbance of the fraction eluted at 260 nm was greater than that at 280 nm (**Figure 48A** and **Figure 48D**, lane 1). The chromatogram of Heparin (**Figure 48B**) showed that nucleic acids were eliminated during the washing process and AlbC-IMI protein was eluted from the column during the NaCl gradient. The absorbance of the protein eluted at 280 nm was about 1.5 folds of that at 260 nm, which indicated no nucleic acid contamination was present. The protein was further purified with a gel filtration column and only one peak was observed. The elution volume (160 ml) corresponded to that expected for AlbC-IMI. SDS-PAGE and western blot analysis confirmed the purity of the protein (**Figure 48D**, lane 4). It should be noted that proteolysis of AlbC-IMI was efficiently inhibited by PMSF and phosphoramidon. The precise molecular mass of the protein was obtained by Maldi-Tof (31535.2 Da). It corresponded to the mass of AlbC-IMI. Finally, we harvested 5.3 mg pure AlbC-IMI from two liters of culture.

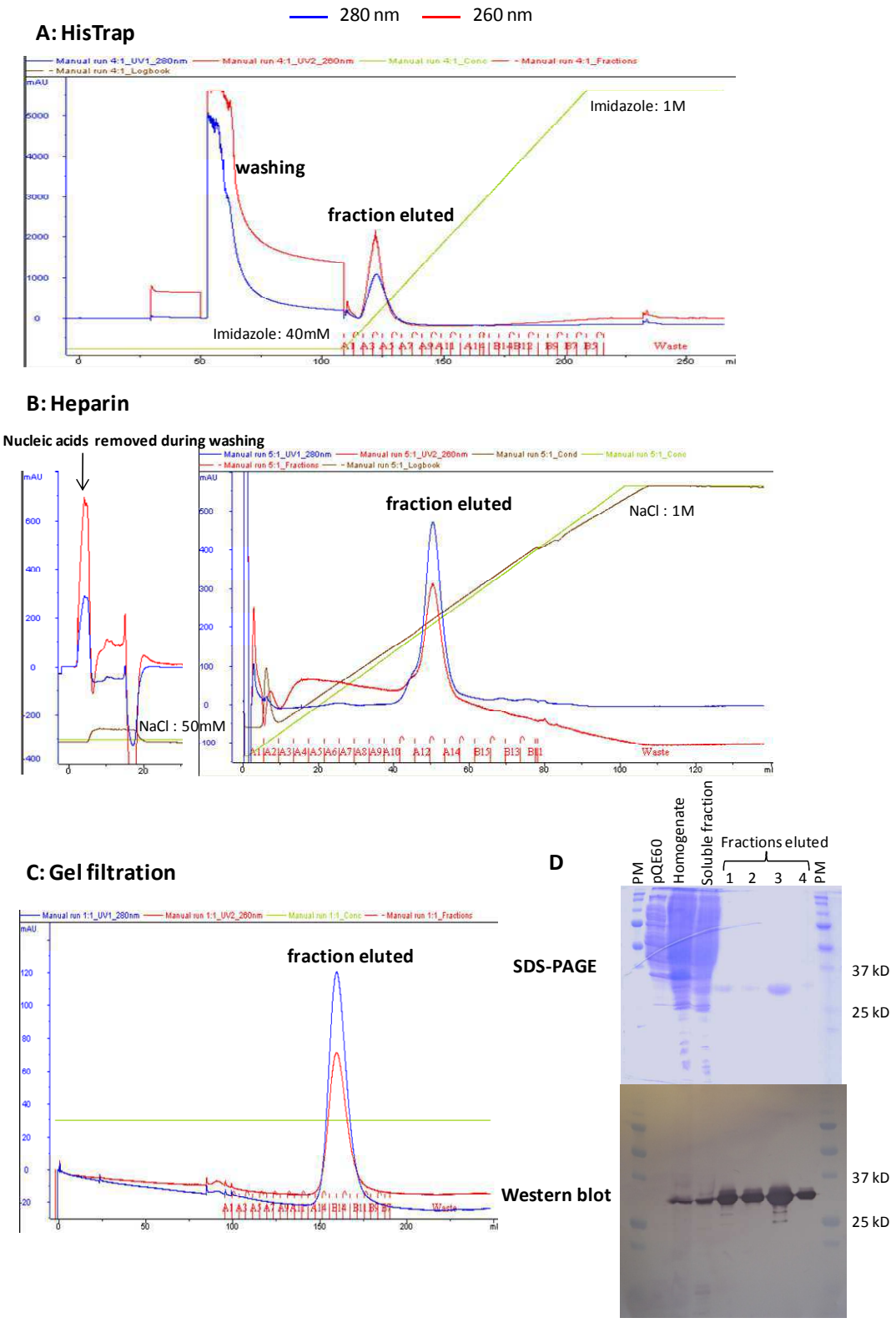


Figure 48: Three-step purification procedure of AlbC-IMI and analysis with SDS-PAGE and Western blot. Chromatograms of purification on columns HisTrap (A), Heparin (B) and Gel filtration (C). Blue and red lines indicate UV absorptions at 280 and 260 nm, respectively. Expression of AlbC-IMI and fractions eluted from the three purification steps are analyzed by SDS-PAGE and Western blot (D).1, 2, 3, and 4 lanes indicate the fractions eluted from HisTrap, Heparin, concentrated fraction from Heparin, and Gel filtration, respectively.

4.3.3 Substrate order of AlbC-IMI

Our *in vivo* experiments in *E. coli* have demonstrated that cFL is the predominant product of AlbC-IMI. AlbC-IMI should use Phe-tRNA^{Phe} and Leu-tRNA^{Leu} as substrates to catalyze the formation of cFL *via* a ping-pong mechanism. In order to know which aa-tRNA is the first substrate to interact with the protein, that means which amino acyl residue is covalently bound to AlbC-IMI to form the aminoacyl-enzyme intermediate, we carried out *in vitro* experiments with each of the radioactivity labeled substrates. As described in the section "materials and methods" (section **4.2.3**), the tRNA^{Leu} isoacceptor used in the experiments is tRNA^{Leu} CAG.

The results of the experiments are shown in the **Figure 49**. The silver nitrate-stained gel confirmed the presence of AlbC-IMI in the two reaction media (**Figure 49A**). The PVDF membrane analyzed by β-Imager showed that AlbC-IMI incubated with [^{14}C]Phe-tRNA^{Phe} retained a high amount of the radiolabel, whereas AlbC-IMI incubated in the presence of [^{14}C]Leu-tRNA^{Leu} did not retain any radiolabel (**Figure 49B**). These results demonstrate that AlbC-IMI forms acyl-enzyme with the phenylalanine but not with the leucine. So it reveals that Phe-tRNA^{Phe} is the first substrate of AlbC-IMI in its catalysis.

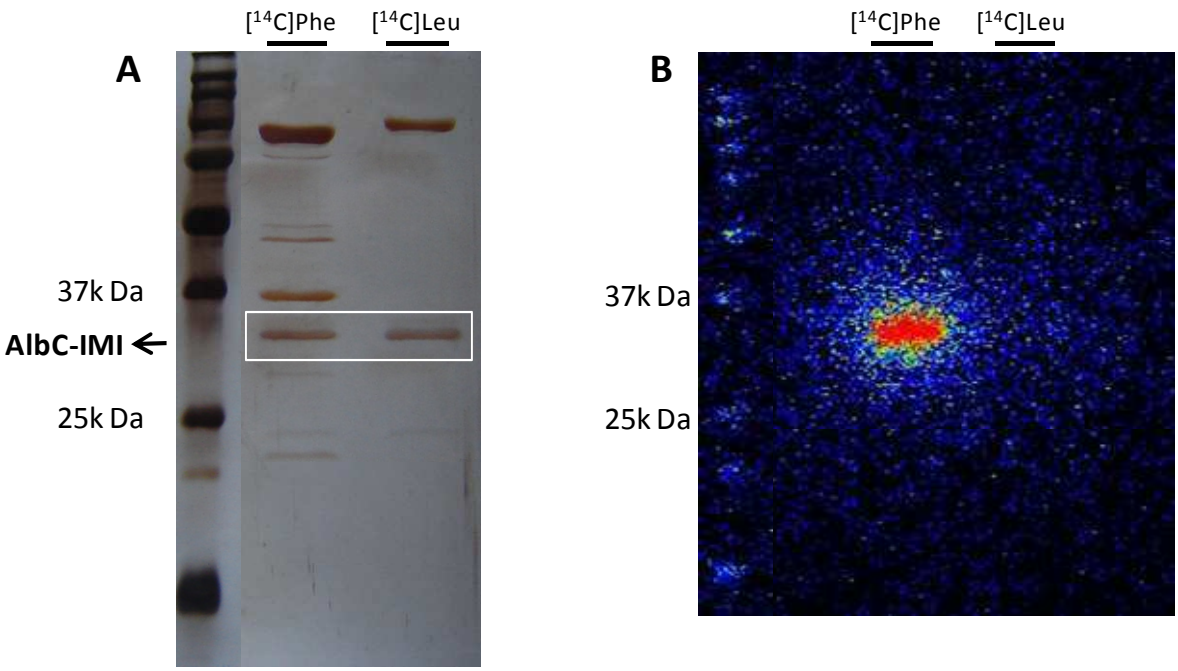


Figure 49: Acyl-enzyme-forming experiments. Enzymes were incubated with [^{14}C]Phe-tRNA^{Phe} or [^{14}C]Leu-tRNA^{Leu}, as described in "Materials and Methods" section, separated on SDS-PAGE. One gel was stained with silver nitrate and AlbC-IMI bands were highlighted with white frame (A). Enzymes on the other gel were transferred onto a PVDF membrane which was then analyzed with β-Imager (B).

4.4 Discussion

We have shown that, as expected, AlbC-IMI essentially synthesizes cFL with cFF produced in trace amounts and no cLL produced at all. This CDPS is therefore appropriate for subsequent studies. AlbC-IMI has proven difficult to obtain in large quantity and without proteolysis. Nevertheless, we have still developed suitable conditions to obtain it. We were also able to start *in vitro* studies especially to determine the order of the two substrates used by AlbC-IMI. With this done, the kinetic characterization studies can be considered and performed.

5 XP_001636126 FROM *NEMATOSTELLA VECTENSIS*, THE FIRST ACTIVE CDPS IDENTIFIED IN ANIMAL

5.1 Introduction

With studies on the CDPS family going on, we are aware that members of the family are persistently growing. In addition to the bacterial members of the CDPS family described above, several other putative members of this family have been detected in various bacterial, and even in eukaryotes (Aravind et al. 2010; Seguin et al. 2011). In this chapter, we will mainly describe, in the form of article, the first characterized eukaryotic CDPS (XP_001636126) from the sea anemone *Nematostella vectensis* in our laboratory. We have named it *Nvec*-CDPS2. A homology-based structural model for *Nvec*-CDPS2 is generated by using the crystal structure of AlbC as a template. By comparing with the structure of AlbC, the structural model gives us preliminary structural knowledge on *Nvec*-CDPS2. The cyclodipeptides produced by recombinant *Nvec*-CDPS2 *in vivo* are also identified. In addition, we will present the results of experiments carried out *in vitro* to characterize *Nvec*-CDPS2.

My participation in this work concerned analysis by LC-MS/MS of the *in vivo* activity of *Nvec*-CDPS2.

5.2 Article

Chemistry & Biology
Brief Communication

Nonribosomal Peptide Synthesis in Animals: The Cyclodipeptide Synthase of *Nematostella*

Jérôme Seguin,[1,3] Mireille Moutiez,[1,3] Yan Li,[1] Pascal Belin,[1] Alain Lecoq,[1] Matthieu Fonvielle,[1,4] Jean-Baptiste Charbonnier,[1] Jean-Luc Pernodet,[2] and Muriel Gondry[1,*]
[1]CEA, Institut de Biologie et Technologies de Saclay, F-91191 Gif-sur-Yvette, France
[2]Université Paris-Sud 11, CNRS, UMR8621, Institut de Génétique et Microbiologie, F-91405 Orsay, France
[3]These authors contributed equally to this work
[4]Present address: UMR S 872, Centre de Recherche des Cordeliers, INSERM/UPMC/UPD, F-75270 Paris, France
*Correspondence: muriel.gondry@cea.fr
DOI 10.1016/j.chembiol.2011.09.010

SUMMARY

Cyclodipeptide synthases (CDPSs) are small enzymes structurally related to class-I aminoacyl-tRNA synthetases (aaRSs). They divert aminoacylated tRNAs from their canonical role in ribosomal protein synthesis, for cyclodipeptide formation. All the CDPSs experimentally characterized to date are bacterial. We show here that a predicted CDPS from the sea anemone *Nematostella vectensis* is an active CDPS catalyzing the formation of various cyclodipeptides, preferentially containing tryptophan. Our findings demonstrate that eukaryotes encode active CDPSs and suggest that all CDPSs have a similar aminoacyl-tRNA synthetase-like architecture and ping-pong mechanism. They also raise questions about the biological roles of the cyclodipeptides produced in bacteria and eukaryotes.

INTRODUCTION

Cyclodipeptide synthases (CDPSs) form a family of small enzymes using two aminoacyl-tRNAs (aa-tRNAs) as substrates for formation of the two peptide bonds of cyclodipeptides (Gondry et al., 2009). These enzymes are promiscuous, but have different substrate specificities. Indeed, AlbC from *Streptomyces noursei* mostly produces cyclo(L-Phe-L-Leu) (cFL), Rv2275 from *Mycobacterium tuberculosis* mostly produces cyclo(L-Tyr-L-Tyr) (cYY), and the other six experimentally characterized CDPSs mostly produce cyclo(L-Leu-L-Leu) (cLL) (Gondry et al., 2009; Lautru et al., 2002). The cyclodipeptides produced are modified by the cyclodipeptide-tailoring enzymes (Belin et al., 2009; Gondry et al., 2001) associated with CDPSs, in diketopiperazine (DKP) biosynthesis pathways. The biological role of the final DKPs in the organisms producing them remains unknown, but albonoursin, which is derived from cFL in *S. noursei,* has antibacterial activity (Fukushima et al., 1973), whereas mycocyclosin derived from cYY may be essential for *M. tuberculosis* viability (Belin et al., 2009; McLean et al., 2008; Vetting et al., 2010), and pulcherriminic acid derived from cLL chelates iron in *Bacillus* species (Cryle et al., 2010; Uffen and Canale-Parola, 1972).

CDPSs share a common ping-pong mechanism, involving a covalent intermediate in which an amino acid is transferred from an aa-tRNA to an active site serine, and a common architecture (Bonnefond et al., 2011; Sauguet et al., 2011; Vetting et al., 2010). Their structure resembles that of the catalytic domain of class-Ic aminoacyl-tRNA synthetases (aaRSs), the enzymes catalyzing the activation of amino acids and their transfer to cognate tRNAs to form aa-tRNAs. They contain a highly conserved deep pocket, which accommodates the aminoacyl moiety of the aa-tRNA substrate, and a patch of basic residues that may interact with the tRNA moiety of the substrate, as suggested for AlbC (Sauguet et al., 2011).

New putative CDPSs have recently been identified by iterative PSI-BLAST searches (Aravind et al., 2010) in various bacteria, but also in eukaryotes, such as the filamentous fungus *Gibberella*, the annelid worm *Platynereis* and the sea anemone *Nematostella*. We show here that a putative CDPS identified in *Nematostella vectensis* is an active CDPS, using a catalytic mechanism similar to that of bacterial CDPSs. We also demonstrate that all the most abundant cyclodipeptides synthesized by this new CDPS contain L-tryptophan, suggesting that new CDPSs with different substrate specificities remain to be discovered.

RESULTS

Putative CDPSs Found in *Nematostella vectensis*

Eight bacterial members of the CDPS family have been experimentally characterized to date (Gondry et al., 2009). However, several other putative members of this family have been detected in various bacterial phyla, and even in eukaryotes (Aravind et al., 2010). In the sea anemone *N. vectensis*, three proteins (XP_001636125, XP_001636126, XP_001621211), which we have named *Nvec*-CDPS1-3, respectively, have been detected. These proteins are encoded by three genes containing large introns (see Figure S1 available online). Comparison of the sequences of *Nvec*-CDPS1-3 showed that their C-terminal parts are almost identical and are aligned with the sequences of known CDPSs (Figure 1). The main difference between the three proteins of *N. vectensis* lies at the N-terminal part of the

Chemistry & Biology
Eukaryotic Cyclodipeptide Synthase

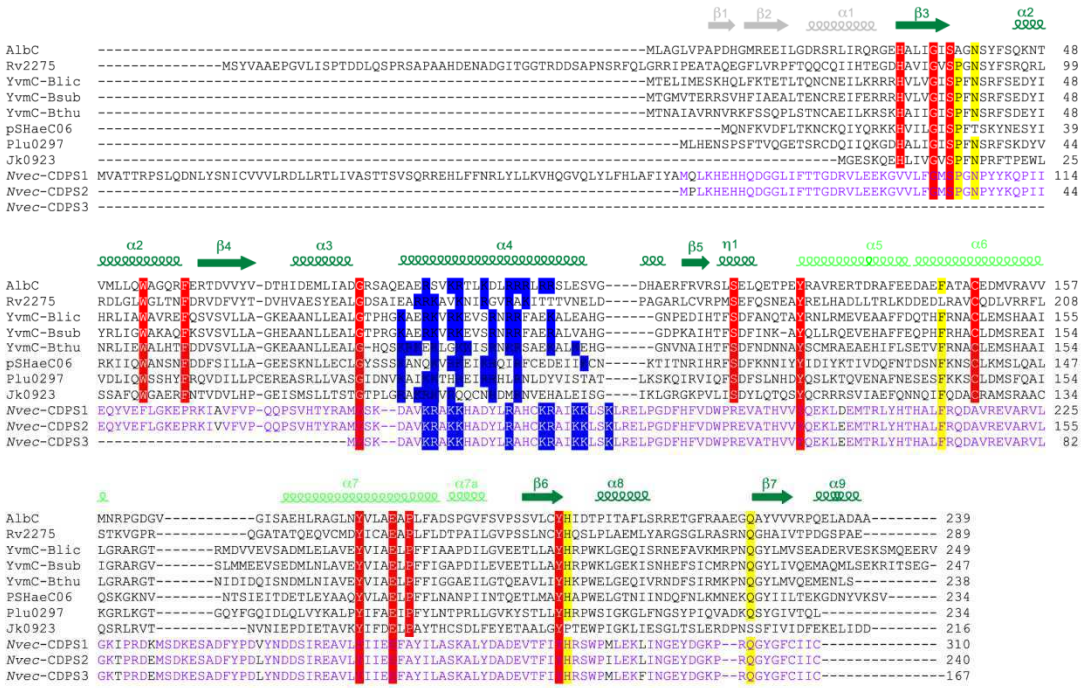


Figure 1. Alignment of the Protein Sequences of the Eight Known CDPSs and the Three Hypothetical Proteins from *N. vectensis*

Positions with identical residues in known CDPSs are indicated by a red background. Positions with identical residues in at least 10 proteins (nine proteins if the position does not exist in the truncated *Nvec*-CDPS3) are indicated by a yellow background. Positions with identical residues in *Nvec*-CDPS1, *Nvec*-CDPS2, and *Nvec*-CDPS3 are shown in purple. The basic residues, which define a patch of positive charge, are indicated by a blue background. The secondary structural elements of AlbC (Sauguet et al., 2011) are indicated above the alignment.
See also Figure S1.

sequences, which differ in length. The *Nvec*-CDPS2 (240 residues) is of a similar size to the eight characterized CDPSs (216–249 residues in general, with only one CDPS having 289 residues), whereas *Nvec*-CDPS1 was longer (310 residues) and *Nvec*-CDPS3 smaller (167 residues). The three proteins displayed only moderate sequence similarity to the known CDPSs. For example, *Nvec*-CDPS2 displayed only 16%–21% identity and 30%–40% similarity to the known CDPSs. However, 7 of the 13 residues conserved among bacterial CDPSs are also found in *Nvec*-CDPS1 and *Nvec*-CDPS2. These residues (G35, S37, G79, Y128, Y178, E182, Y202, AlbC numbering) were previously shown to be involved in CDPS activity. In particular, residue S37 has been identified as the catalytic residue involved in formation of the covalent aminoacyl-enzyme intermediate (Sauguet et al., 2011). The hypothetical protein *Nvec*-CDPS3 appears to be truncated, because its N-terminal sequence starts ~40 residues after the essential G35 and S37 residues (Figure 1).

We generated a homology-based structural model for *Nvec*-CDPS2 with Modeler software (Sali and Blundell, 1993), using the crystal structure of AlbC (Protein Data Bank [PDB] id: 3OQV) as a template (Figure 2; Figure S2). Superimposition of the three-dimensional structures of AlbC and *Nvec*-CDPS2 gave an root-mean-square deviation value of 0.48 Å over 193 Cα. We compared the pocket in *Nvec*-CDPS2 model (Figure 2A) with its equivalent in AlbC, which accommodates the aminoacyl moiety of the aa-tRNA substrate and contains the catalytic residues (Figure 2B). The catalytic residues in *Nvec*-CDPS2, including the catalytic serine residue in particular, are located at positions compatible with their involvement in catalysis. The buried surfaces of the two pockets consist mostly of side chains of hydrophobic residues, but with a higher proportion of aromatic residues for *Nvec*-CDPS2 (Figures 2A and 2B). We also compared the distribution of charged residues on the surfaces of the two proteins (Figures 2C and 2D). The known CDPSs were found to contain a patch of basic residues thought to interact with the tRNA moiety of an aa-tRNA substrate (Sauguet et al., 2011). A similar patch is also present in *Nvec*-CDPS2 (Figure 2C), and probably also in *Nvec*-CDPS1 and *Nvec*-CDPS3 (Figure 1), based on sequence alignment.

The presence of both these basic patches and the conservation of similar residues in the three proteins strongly suggest that these proteins may be members of the CDPS family, with *Nvec*-CDPS3 being a truncated CDPS.

Chemistry & Biology
Eukaryotic Cyclodipeptide Synthase

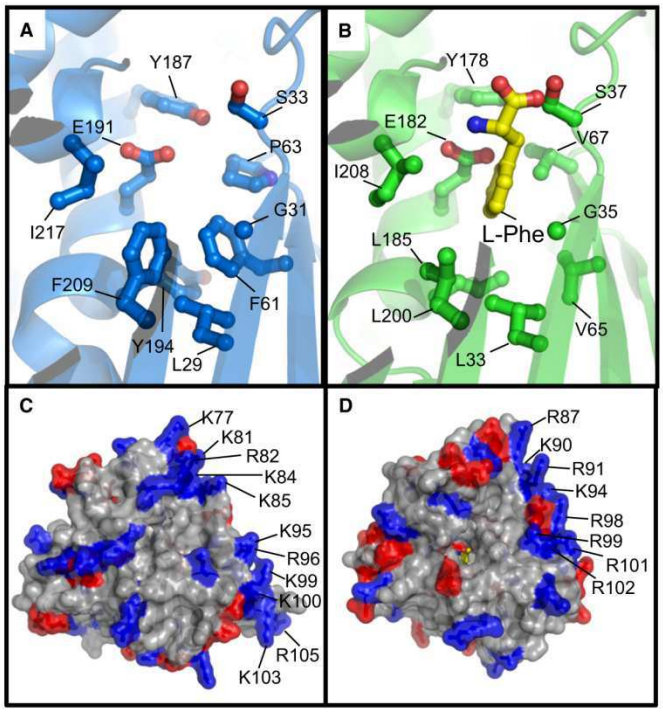


Figure 2. Comparison of the *Nvec*-CDPS2 model with the crystal structure of AlbC

Surface-accessible pockets of (A) *Nvec*-CDPS2 and (B) AlbC with the phenylalanyl moiety of a Phe-tRNA^{Phe}. Distribution of Arg/Lys (blue) and Asp/Glu (red) residues on the surface of (C) *Nvec*-CDPS2 and (D) AlbC. See also Figure S2.

Biosynthesis of Cyclodipeptides by *Nvec*-CDPS2

We investigated the ability of *Nvec*-CDPS1 and *Nvec*-CDPS2 to catalyze the formation of cyclodipeptides. As we had previously shown that all known CDPSs, when produced in *Escherichia coli*, synthesize cyclodipeptides that are then released into the medium (Gondry et al., 2009), we introduced constructs encoding each of the two putative CDPSs separately in this host. *Nvec*-CDPS2 was produced successfully, with a proportion of this protein found in the soluble protein extract, whereas no *Nvec*-CDPS1 production was detected (Figure S3). We then analyzed the cyclodipeptide content of the culture supernatants of *E. coli* cells producing *Nvec*-CDPS2. This analysis was carried out by LC-MS/MS, without prior assumptions about the nature of cyclized amino acids, as previously reported (Gondry et al., 2009). Comparisons of the UV chromatograms for *Nvec*-CDPS2 samples and the control (culture supernatant of *E. coli* containing the empty vector) showed several peaks specific to *Nvec*-CDPS2-producing cells (Figure 3A). An analysis of the corresponding extracted ion current (EIC) chromatograms and MS/MS fragmentation patterns led to the identification of 13 cyclodipeptides (cWF, cWL, cWM, cWA, cWG, cFF, cFL, cFM, cFA, cYL, cYF, cYM, and cLA; Figures 3B and 3C). Most were combinations of an L-Trp, L-Phe or L-Tyr residue with a nonpolar amino acid (L-Leu, or L-Met or L-Ala or Gly). The three cyclodipeptides, cWF, cWL, and cWM, were the major products generated by *Nvec*-CDPS2. They were present at concentrations of 0.8–1 mg l^{-1} of culture supernatant after 44 hr (Figure S4). All other cyclodipeptides were produced at concentrations of 0.1–0.5 mg l^{-1} of culture supernatant, except for cFL and cWG, which were identified from retention times and fragmentation patterns but could not be quantified.

Thus, *Nvec*-CDPS2 was active in *E. coli* and the most abundant cyclodipeptides synthesized by *Nvec*-CDPS2 contained an L-Trp. This is consistent with the presence of large numbers of aromatic residues in the catalytic pocket of our structural model (Figure 2A). This is the first cyclodipeptide-synthesizing enzyme to be shown to display such specificity. Indeed, AlbC from *S. noursei*, Rv2275 from *M. tuberculosis* and the other six CDPSs generate principally L-Phe-, L-Tyr-, and L-Leu-containing cyclodipeptides, respectively (Gondry et al., 2009).

Nvec-CDPS2, a New Member of the CDPS Family

We then investigated the possible use, by *Nvec*-CDPS2, of aa-tRNAs as substrates for the synthesis of cyclodipeptides. We produced *Nvec*-CDPS2 in *E. coli* and purified the protein to homogeneity by a three-step chromatographic procedure, as previously described (Gondry et al., 2009). *Nvec*-CDPS2 appears to be a monomer (Figure S5), as reported for AlbC from *S. noursei* (Sauguet et al., 2011) and YvmC from *Bacillus licheniformis* and *Bacillus subtilis* (Bonnefond et al., 2011). We then assessed the cyclodipeptide-synthesizing activity of the purified enzyme. *Nvec*-CDPS2 preferentially synthesizes Trp-containing cyclodipeptides, but it also produces significant amounts of other cyclodipeptides, including cFF (Figure 3), in vivo. We used an in vitro assay based on cFF detection, which was developed in our laboratory for AlbC, to check whether the purified *Nvec*-CDPS2 generated cFF. The enzyme-dependent formation of cFF was unambiguously detected (Figure 4). Indeed, the amount of cFF increased over time in the presence of 1 μM *Nvec*-CDPS2 (Figures 4A, 4C, and 4D), and increased with the concentration of *Nvec*-CDPS2 after 3 hr of reaction (Figure 4B). As we used *E. coli* tRNA^{Phe} to assess the activity of the purified *Nvec*-CDPS2, accurate kinetic analyses would be meaningless. However, our experiments clearly show that aa-tRNAs are the substrates of *Nvec*-CDPS2, as for previously characterized CDPSs.

We then investigated whether *Nvec*-CDPS2 used a similar catalytic mechanism to bacterial CDPSs. It has been shown that bacterial CDPSs use a ping-pong mechanism and that the first step of their catalytic cycle involves a nucleophilic attack of the catalytic serine on a aminoacyl-tRNA, to form a covalent

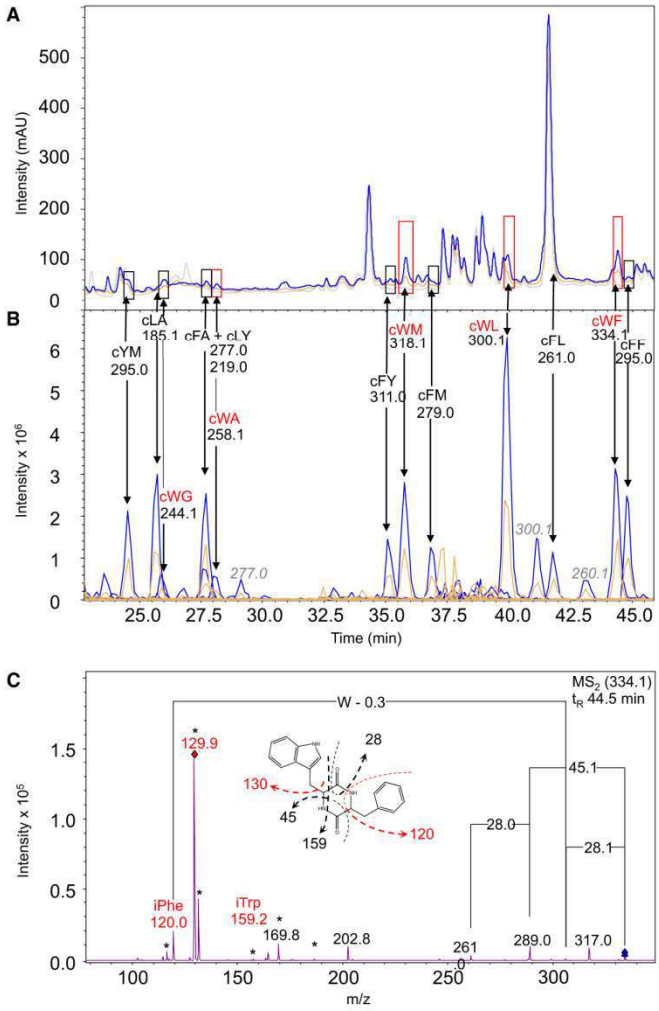


Figure 3. LC-MS/MS Analysis of the Cyclodipeptides Secreted into the Culture Supernatant of *E. coli* Cells Producing *Nvec*-CDPS2

(A) UV traces (λ = 214 nm) of the culture medium of *Nvec*-CDPS2-producing *E. coli* cells after 20 hr (orange) and 44 hr (blue) of expression, and of cells containing the empty vector (gray). Cyclodipeptide peaks are surrounded by a rectangle.

(B) Corresponding EIC peaks, showing the cyclodipeptide peaks and their identity.

(C) MS/MS of one of the major compounds, identified as cWF. The fragmentation pattern is characteristic of a cyclodipeptide and red labels at 120.0 ± 0.1 and 159 ± 0.1 match the immonium ions of Phe (iPhe) and Trp (iTrp), respectively. The m/z peaks marked with an asterisk are specific to Trp fragmentation (Falick et al., 1993; Papayannopoulos, 1995).

See also Figures S3 and S4.

acyl-enzyme intermediate (Vetting et al., 2010; Sauguet et al., 2011; Bonnefond et al., 2011). As *Nvec*-CDPS2 can produce cFF, this suggests that it can recognize Phe-tRNA^{Phe} as a substrate. We therefore incubated *Nvec*-CDPS2 with substoichiometric amounts of [^{3}H]Phe-tRNA^{Phe}, as previously described for AlbC. Positive and negative controls were performed with AlbC and its inactive variant S37A, respectively (Sauguet et al., 2011). The proteins were separated by SDS-PAGE, transferred to a PVDF membrane and analyzed with a radioimager (Figure 4E). Radioactivity associated with *Nvec*-CDPS2 is clearly visible, indicating the formation of a covalently bound phenylalanyl-intermediate. The intensity of the signal is lower than that for AlbC, but the specificity of the two enzymes should be taken into account: AlbC mostly synthesizes Phe-containing cyclodipeptides, whereas cFF is a minor product of *Nvec*-CDPS2 activity. Our results support the hypothesis that *Nvec*-CDPS2, like its bacterial counterparts, uses a ping-pong mechanism involving a covalent aminoacyl-enzyme intermediate.

DISCUSSION

The starlet sea anemone *N. vectensis* belongs to the phylum Cnidaria. *In silico* analysis of the draft genome of this emerging Cnidaria model (Putnam et al., 2007) led to the identification of three proteins presenting sequence similarities to bacterial CDPSs. We show here that *Nvec*-CDPS2 is a member of the CDPS family. *Nvec*-CDPS1 is probably a functional CDPS but we were unable to produce it in our experimental conditions. With the sequence data available, *Nvec*-CDPS3 appears to be a truncated CDPS, lacking essential residues.

Nvec-CDPS2 differs in specificity from other characterized CDPSs as it produces mostly L-Trp-containing cyclodipeptides, whereas the others synthesize L-Phe-, L-Tyr-, and L-Leu-containing cyclodipeptides (Gondry et al., 2009). The study of other putative CDPSs would extend the range of substrates known to be used by these enzymes and increase our understanding of CDPS specificity.

Cyclodipeptides are synthesized by CDPSs or nonribosomal peptide synthetases (NRPSs) (Gondry et al., 2009). NRPSs are widespread in bacteria and fungi but have not been identified so far in plants or animals. CDPSs are present in bacteria and putative CDPSs have been identified in eukaryotes, such as the fungus *Gibberella zeae* and the annelid *Platynereis dumerilii* (Aravind et al., 2010). *Nvec*-CDPS2 from *N. vectensis* is the first enzyme from animal origin experimentally demonstrated to be involved in nonribosomal peptide synthesis.

The cyclodipeptides synthesized by the known bacterial CDPSs are precursors of more complex DKPs generated by tailoring reactions (Belin et al., 2009; Gondry et al., 2001). In bacteria, the genes encoding the CDPSs and the associated

Chemistry & Biology
Eukaryotic Cyclodipeptide Synthase

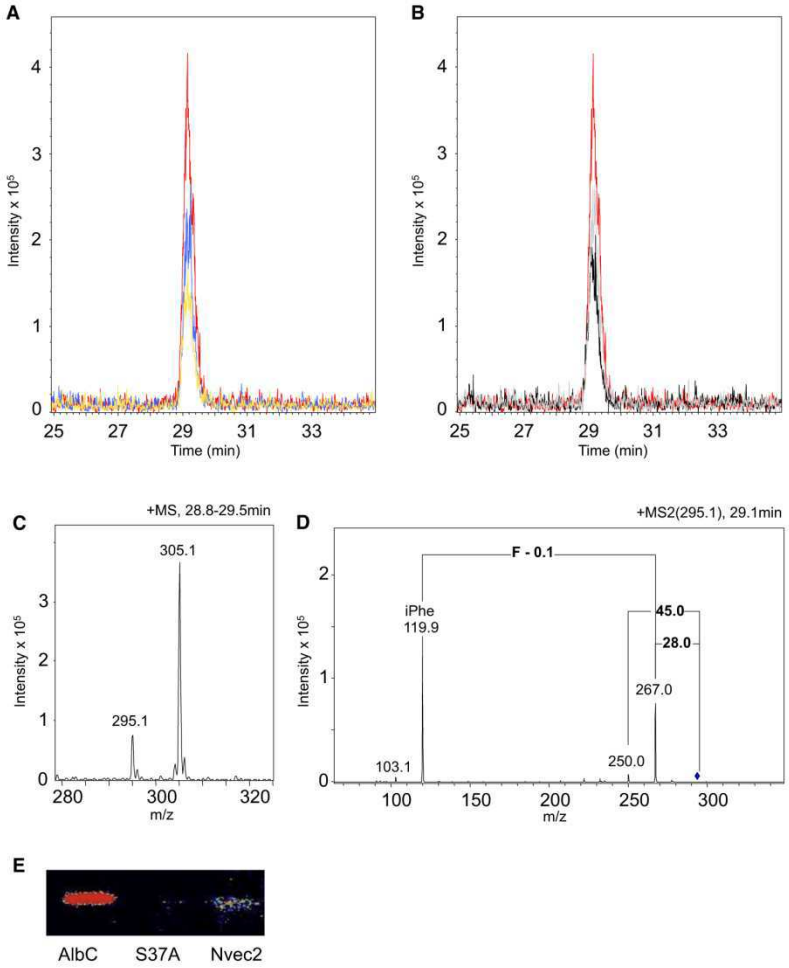


Figure 4. In Vitro Enzymatic Assays with Purified *Nvec*-CDPS2

(A) EIC chromatograms corresponding to the formation of cFF (m/z = 295) catalyzed by 1 μM *Nvec*-CDPS2 at 30 min (yellow), 1 hr (blue), and 3 hr (red).

(B) EIC chromatograms corresponding to the formation of cFF (m/z = 295) at 3 hr catalyzed by 125 nM *Nvec*-CDPS2 (black), 500 nM *Nvec*-CDPS2 (gray), and 1 μM *Nvec*-CDPS2 (red).

(C) Formation of cFF catalyzed by 1 μM *Nvec*-CDPS2, after 3 hr: MS spectrum of products eluted at 29.1 min, identified as cFF (m/z = 295.1) and $^{13}C_9$,^{15}N-labeled cFF (m/z = 305.1).

(D) MS/MS spectrum of the product formed by *Nvec*-CDPS2, identified as cFF (m/z = 295.1).

(E) Covalent labeling of *Nvec*-CDPS2 by tritiated Phe transferred from [^{3}H]Phe-tRNA^{Phe}, as described in Experimental Procedures. Wild-type AlbC and the S37A variant are used as positive and negative controls, respectively.

See also Figure S5.

cyclodipeptide-tailoring enzymes are clustered, facilitating the identification of tailoring enzymes and, thus, of the final DKPs. No such genetic organization is observed in the *N. vectensis* genome, so it remains unknown whether the cyclodipeptides are the final DKPs.

The physiological role of the final DKPs remains unclear. It has been suggested that DKPs act as small diffusible molecules involved in cell-to-cell communication. In bacteria, they may constitute a new class of quorum-sensing signals (Degrassi et al., 2002; Holden et al., 1999; Park et al., 2006), or even interspecies signals (Li et al., 2011). Furthermore, as DKPs have bioactive effects on their plant or animal hosts, a role in transkingdom signaling has also been suggested (Ortiz-Castro et al., 2011; Prasad, 1995). Moreover, expression of the CDPS gene in

Chemistry & Biology
Eukaryotic Cyclodipeptide Synthase

P. dumerilii increases in response to septic injury (Altincicek and Vilcinskas, 2007), suggesting a possible role of the CDPS in *Platynereis* immunity (Aravind et al., 2010). The nature and role of the final DKPs remain unknown for *N. vectensis*.

SIGNIFICANCE

Cyclodipeptide synthases (CDPSs) form a family of small enzymes structurally similar to class-I aminoacyl-tRNA synthetases (aaRSs). They use aminoacyl-tRNAs (aa-tRNAs) as substrates to catalyze the formation of various cyclodipeptides, which are the precursors of many natural products exhibiting noteworthy biological activities. So far, all characterized CDPSs originated from bacteria. We show here that a predicted protein identified in the sea anemone *N. vectensis*, namely *Nvec*-CDPS2, belongs to the CDPS family. *Nvec*-CDPS2 is the first enzyme involved in nonribosomal peptide synthesis to be identified in animals. This finding raises questions about the biological roles of the cyclodipeptides produced in eukaryotes. *Nvec*-CDPS2, like its bacterial counterparts, exhibits an aminoacyl-tRNA synthetase-like architecture and uses a ping-pong mechanism involving a covalent aminoacyl-enzyme intermediate. However, *Nvec*-CDPS2 differs in specificity from other characterized CDPSs as it produces mostly L-Trp-containing cyclodipeptides, whereas the others synthesize L-Phe-, L-Tyr-, and L-Leu-containing cyclodipeptides. This finding suggests that new CDPSs with different substrate specificities remain to be discovered. The characterization of new CDPSs opens the way to CDPS engineering to further increase the natural diversity of cyclodipeptides, a family of compounds with diverse biological properties.

EXPERIMENTAL PROCEDURES

Structural Model of *Nvec*-CDPS2

Alignments of protein sequences were obtained with MUSCLE and HHPred (Söding et al., 2005) and used for model building with Modeler (Sali and Blundell, 1993). The crystal structure of AlbC at 1.9 Å resolution (PDB id: 3OQV) was used as a template for this homology modeling. The quality of the *Nvec* model was checked with the Qmean server (Benkert et al., 2009) (Figure S2).

Cloning, Production, and Purification of *N. vectensis Nvec*-CDPS1 and *Nvec*-CDPS2

Synthetic genes encoding *Nvec*-CDPS1 and *Nvec*-CDPS2 were purchased from GENEART and inserted into the pQE60 vector (QIAGEN), as previously described (Gondry et al., 2009). The His_6-tagged *Nvec*-CDPS2 protein was then produced and purified as described by Braud et al. (2005) and Gondry et al. (2009).

Analysis of Cyclodipeptide Synthesis by *Nvec*-CDPS2

In vivo assays for *Nvec*-CDPS2 activity were performed as described by (Gondry et al., 2009). Cyclodipeptides were detected and identified from both their m/z value (MS) and their daughter ion spectra (MS/MS), as a result of their common fragmentation patterns. The nature of the detected cyclodipeptides was unambiguously confirmed by comparison with authentic standards.

Detection of *Nvec*-CDPS2 Activity

The cFF-forming activity of *Nvec*-CDPS2 was measured by a coupled assay, as described by (Sauguet et al., 2011). Reactions were initiated by adding *Nvec*-CDPS2. Aliquots were withdrawn at various times and mixed with $^{13}C_9$,^{15}N-labeled cFF solution as a stable isotope internal standard, before cFF analysis by LC-MS.

Detection of the Covalent Aminoacyl-Enzyme Intermediate

Purified *Nvec*-CDPS2, wild-type AlbC (positive control) and the AlbC variant S37A (negative control) were incubated with [^{3}H]Phe-tRNA^{Phe}. The tritiated substrate was obtained as previously described (Sauguet et al., 2011). The enzyme was added at a final concentration of 16 μM for wild-type AlbC or 80 μM for the variant S37A and *Nvec*-CDPS2. After 30 s of incubation for AlbC wild-type and 120 s for *Nvec*-CDPS2 and S37A, the reaction was quenched and analyzed (Sauguet et al., 2011).

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and can be found with this article online at doi:10.1016/j.chembiol.2011.09.010.

ACKNOWLEDGMENTS

We thank Robert Thai and Bertrand Czarny for helpful advice on mass spectrometry and use of the Beta-imager, respectively. This work was supported by the CEA and the ANR (ANR 2010 BLAN 1501 01). Y.L. holds a doctoral fellowship from the CEA. The authors have no conflict of interest to declare.

Received: July 22, 2011
Revised: September 2, 2011
Accepted: September 23, 2011
Published: November 22, 2011

REFERENCES

Altincicek, B., and Vilcinskas, A. (2007). Analysis of the immune-related transcriptome of a lophotrochozoan model, the marine annelid *Platynereis dumerilii*. Front. Zool. *4*, 18.

Aravind, L., de Souza, R.F., and Iyer, L.M. (2010). Predicted class-I aminoacyl tRNA synthetase-like proteins in non-ribosomal peptide synthesis. Biol. Direct *5*, 48.

Belin, P., Le Du, M.H., Fielding, A., Lequin, O., Jacquet, M., Charbonnier, J.B., Lecoq, A., Thai, R., Courçon, M., Masson, C., et al. (2009). Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. USA *106*, 7426–7431.

Benkert, P., Künzli, M., and Schwede, T. (2009). QMEAN server for protein model quality estimation. Nucleic Acids Res. *37* (Web Server issue), W510–W514.

Bonnefond, L., Arai, T., Sakaguchi, Y., Suzuki, T., Ishitani, R., and Nureki, O. (2011). Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog. Proc. Natl. Acad. Sci. USA *108*, 3912–3917.

Braud, S., Moutiez, M., Belin, P., Abello, N., Drevet, P., Zinn-Justin, S., Courçon, M., Masson, C., Dassa, J., Charbonnier, J.B., et al. (2005). Dual expression system suitable for high-throughput fluorescence-based screening and production of soluble proteins. J. Proteome Res. *4*, 2137–2147.

Cryle, M.J., Bell, S.G., and Schlichting, I. (2010). Structural and biochemical characterization of the cytochrome P450 CypX (CYP134A1) from *Bacillus subtilis*: a cyclo-L-leucyl-L-leucyl dipeptide oxidase. Biochemistry *49*, 7282–7296.

Degrassi, G., Aguilar, C., Bosco, M., Zahariev, S., Pongor, S., and Venturi, V. (2002). Plant growth-promoting *Pseudomonas putida* WCS358 produces and secretes four cyclic dipeptides: cross-talk with quorum sensing bacterial sensors. Curr. Microbiol. *45*, 250–254.

Falick, A.M., Hines, W.M., Medzihradszly, K.F., Baldwin, M.A., and Gibson, B.W. (1993). Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass spectrometry. J. Am. Soc. Mass Spectrom. *4*, 882–893.

Fukushima, K., Yazawa, K., and Arai, T. (1973). Biological activities of albonoursin. J. Antibiot. *26*, 175–176.

Chemistry & Biology
Eukaryotic Cyclodipeptide Synthase

Gondry, M., Lautru, S., Fusai, G., Meunier, G., Ménez, A., and Genet, R. (2001). Cyclic dipeptide oxidase from *Streptomyces noursei*. Isolation, purification and partial characterization of a novel, amino acyl alpha,beta-dehydrogenase. Eur. J. Biochem. *268*, 1712–1721.

Gondry, M., Sauguet, L., Belin, P., Thai, R., Amouroux, R., Tellier, C., Tuphile, K., Jacquet, M., Braud, S., Courçon, M., et al. (2009). Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. Nat. Chem. Biol. *5*, 414–420.

Holden, M.T., Ram Chhabra, S., de Nys, R., Stead, P., Bainton, N.J., Hill, P.J., Manefield, M., Kumar, N., Labatte, M., England, D., et al. (1999). Quorum-sensing cross talk: isolation and chemical characterization of cyclic dipeptides from *Pseudomonas aeruginosa* and other gram-negative bacteria. Mol. Microbiol. *33*, 1254–1266.

Lautru, S., Gondry, M., Genet, R., and Pernodet, J.L. (2002). The albonoursin gene Cluster of *S noursei* biosynthesis of diketopiperazine metabolites independent of nonribosomal peptide synthetases. Chem. Biol. *9*, 1355–1364.

Li, J., Wang, W., Xu, S.X., Magarvey, N.A., and McCormick, J.K. (2011). *Lactobacillus reuteri*-produced cyclic dipeptides quench *agr*-mediated expression of toxic shock syndrome toxin-1 in staphylococci. Proc. Natl. Acad. Sci. USA *108*, 3360–3365.

McLean, K.J., Carroll, P., Lewis, D.G., Dunford, A.J., Seward, H.E., Neeli, R., Cheesman, M.R., Marsollier, L., Douglas, P., Smith, W.E., et al. (2008). Characterization of active site structure in CYP121. A cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. J. Biol. Chem. *283*, 33406–33416.

Ortiz-Castro, R., Díaz-Pérez, C., Martínez-Trujillo, M., del Río, R.E., Campos-García, J., and López-Bucio, J. (2011). Transkingdom signaling based on bacterial cyclodipeptides with auxin activity in plants. Proc. Natl. Acad. Sci. USA *108*, 7253–7258.

Papayannopoulos, I.A. (1995). The interpretation of collision-induced dissociation tandem mass spectra of peptides. Mass Spectrom. Rev. *14*, 49–73.

Park, D.K., Lee, K.E., Baek, C.H., Kim, I.H., Kwon, J.H., Lee, W.K., Lee, K.H., Kim, B.S., Choi, S.H., and Kim, K.S. (2006). Cyclo(Phe-Pro) modulates the expression of *ompU* in *Vibrio spp*. J. Bacteriol. *188*, 2214–2221.

Prasad, C. (1995). Bioactive cyclic dipeptides. Peptides *16*, 151–164.

Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V.V., et al. (2007). Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science *317*, 86–94.

Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. *234*, 779–815.

Sauguet, L., Moutiez, M., Li, Y., Belin, P., Seguin, J., Le Du, M.H., Thai, R., Masson, C., Fonvielle, M., Pernodet, J.L., et al. (2011). Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis. Nucleic Acids Res. *39*, 4475–4489.

Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. *33* (Web Server issue), W244–W248.

Uffen, R.L., and Canale-Parola, E. (1972). Synthesis of pulcherriminic acid by *Bacillus subtilis*. J. Bacteriol. *111*, 86–93.

Vetting, M.W., Hegde, S.S., and Blanchard, J.S. (2010). The structure and mechanism of the *Mycobacterium tuberculosis* cyclodityrosine synthetase. Nat. Chem. Biol. *6*, 797–799.

SUPPLEMENTAL INFORMATION

Supplemental Data

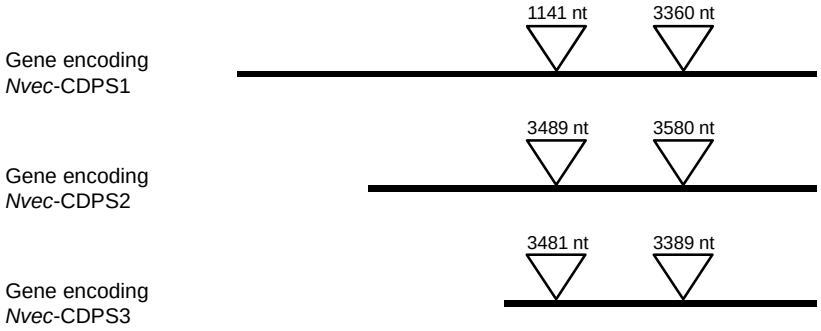


Figure S1, related to Figure 1. Schematic representation of the three genes encoding *Nvec*-CDSP1-3
The coding regions of the three genes are represented by lines. Their relative positions correspond to the alignment of their sequences. The positions of the introns, indicated by inverted triangles, are the same in the three sequences and the lengths of the introns are given. The sequences of the genes, of partial mRNAs and of the deduced proteins are available in GenBank (*Nvec*-CDPS1: accession numbers for the gene NW_001834372 region 1122446..1128760, mRNA XM_001636075, protein XP_001636125; *Nvec*-CDPS2 accession numbers for the gene NW_001834372 region 1133497..1141771, mRNA XM_001636076, protein XP_001636126; *Nvec*-CDPS3: accession numbers for the gene NW_001833027 region 6667..14036, mRNA XM_001621161, protein XP_001621211).

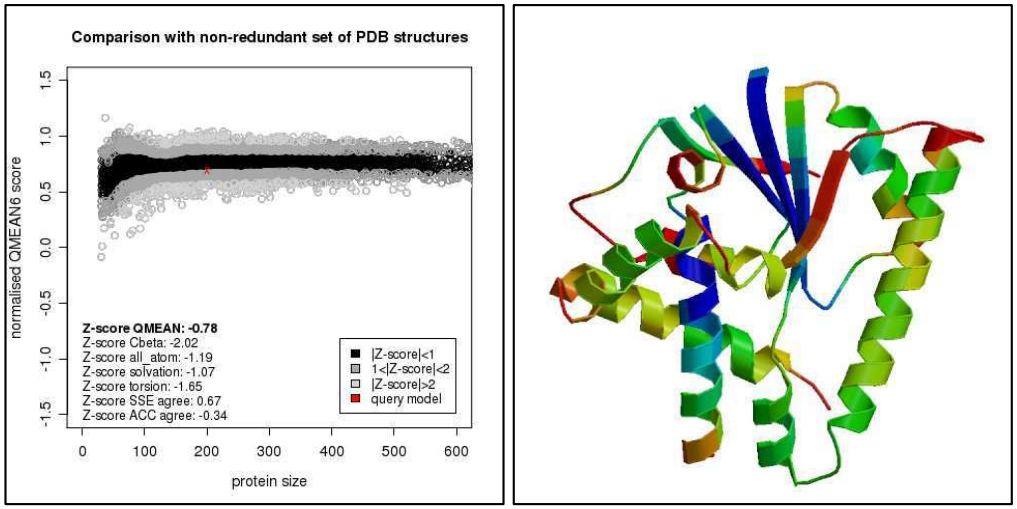


Figure S2, related to Figure 2. *Nvec*-CDPS2 model reliability by Qmean Server
The quality of the model of *Nvec*-CDPS2 was checked with Qmean server (Benkert *et al*, 2009). (A) The QMEAN score of the query model is related to the scores of a non-redundant set of high-resolution X-rays structures of similar size and a Z-score is calculated. The QMEAN score is a global score of the whole model reflecting the predicted model reliability ranging from 0 to 1. (B) The estimated residue error is visualized using a color gradient from blue (more reliable regions) to red (potentially unreliable regions, estimated error above 3.5 Å).

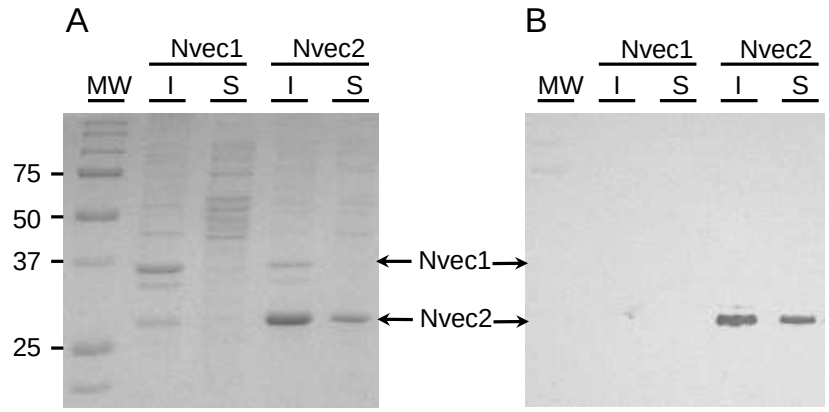


Figure S3, related to Figure 3. Expression of *Nvec*-CDPS1 and *Nvec*-CDPS2 in *E. coli*
Analysis of soluble (S) and insoluble (I) protein fractions after bacterial lysis and centrifugation. Protein separations were performed on 12% acrylamide gels. Then, the proteins were stained by Coomassie blue (A) or transferred onto a polyvinylidene fluoride (PVDF) membrane visualized using alkaline phosphatase-conjugated anti-His antibody and NBT/BCIP solution (Sigma) (B). Arrows indicate the expected position of *Nvec*-CDPS1 at 26 kDa, and of *Nvec*-CDPS2 at 28 kDa. Molecular weight standards (MW) are in kDa.

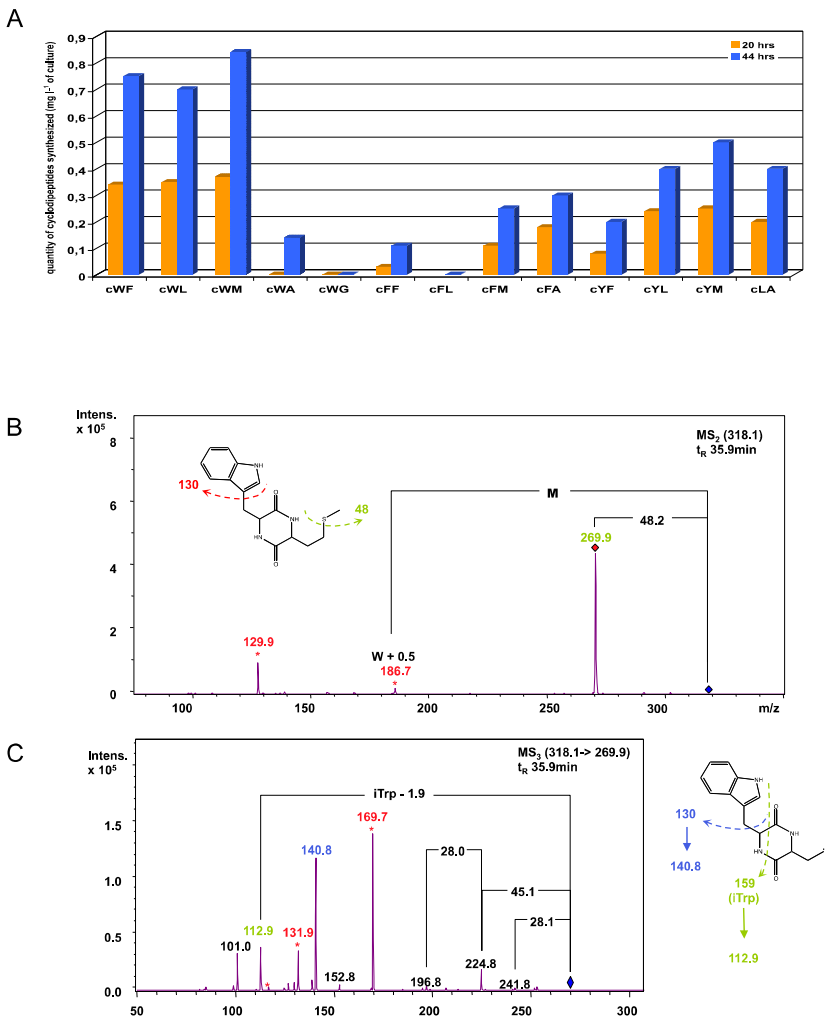


Figure S4, related to Figure 3. Analyses of *in vivo* assays for *Nvec*-CDPS2 activity
(A) Histogram of the amounts of the various cyclodipeptides synthesized by *Nvec*-CDPS2 after 20 hrs (orange) and 44 hrs (blue) of expression in *E. coli*. Cyclodipeptides were quantified on the basis of their peak area at 214 nm using calibration curves obtained with authentic standards, as previously performed in Gondry *et al* (2009). The amounts of cWM and cWA (light blue and light orange) were estimated using cWL calibration.
(B) Characterization of cWM by MS/MS. Signals specific for Trp are in red. The loss of 48.0 is specific for methionine's fragmentation.
(C) MS/MS/MS of the major daughter ion resulting from cWM fragmentation. The fragmentation pattern is characteristic for the cyclodipeptide structure.

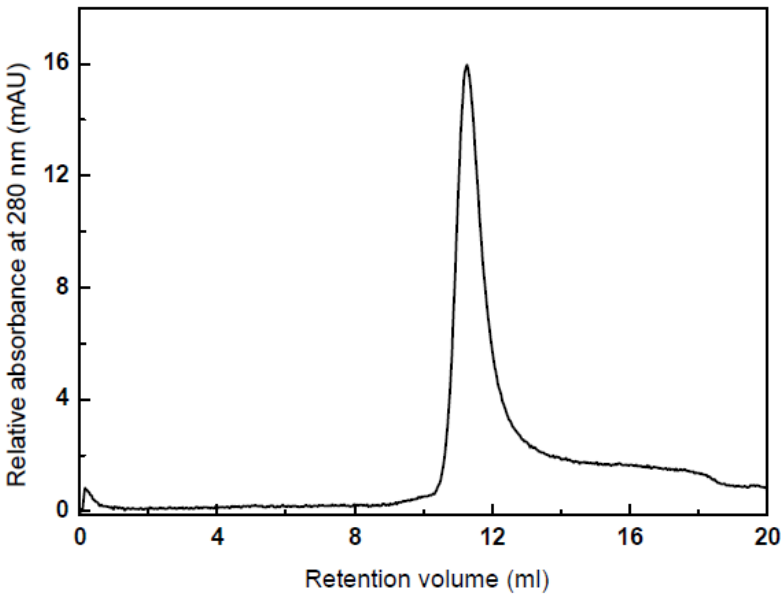


Figure S5, related to Figure 4. Analysis of the oligomeric state of the purified *Nvec*-CDPS2
280nm-chromatogram obtained when *Nvec*-CDPS2-containing fractions were submitted to a size exclusion chromatography. Fractions were loaded onto a Superdex 75 HR 10/30 column (GE-Healthcare BioSciences). *Nvec*-CDPS2 eluted as a single peak, with estimated *Mr* of 21 kDa, a value very close to its theorical *Mr* of 28 kDa, revealing a monomeric state of the enzyme.

Supplemental Experimental Procedures

Purification of *N. vectensisNvec*-CDPS2

The His_6-tagged protein *Nvec*-CDPS2 was expressed and purified as previously described (Braud et al. 2005; Gondry et al. 2009), except that the HiPrep Sephacryl S-200 column was replaced by a Superdex HR 75 column (GE- Healthcare Bio-Sciences). Protein production was followed by SDS-PAGE analyses; the proteins were stained by Coomassie blue or transferred onto a polyvinylidene fluoride (PVDF) membrane visualized using alkaline phosphatase-conjugated anti-His antibody and NBT/BCIP solution (Sigma).

Analysis of cyclodipeptide synthesis by *Nvec*-CDPS2

In vivo assays for *Nvec*-CDPS2 activity were performed as previously described(Gondry et al. 2009), except that culture durations were respectively 20 hrs and 44 hrs after addition of IPTG. The control strain was M15[pREP4] (Qiagen) harboring the empty vector pQE60. Cyclodipeptides were detected and identified by both their m/z value (MS) and their daughter ion spectrums (MS/MS), as a result of their common fragmentation patterns. The nature of the

detected cyclodipeptides was unambiguously confirmed by comparison with authentic standards except for cWM and cWA. The standards cWL and cWF were obtained from Bachem, cLA from Sigma-Aldrich, and others have already been described (Gondry et al. 2009).

Detection of *Nvec*-CDPS2 activity

cFF-forming activity of *Nvec*-CDPS2 was measured by a coupled assay detailed in(Sauguet et al. 2011). Assay mixtures contained 2 mM L-Phe, 1.2 µM *E. coli* $tRNA^{Phe}$ (Sigma), 1 µM of *E. coli* PheRS, in 50 mM Hepes-KOH, pH 7 containing 50 mM KCl, 10 mM ATP, 1 mM DTT and 20 mM $MgCl_2$. After 20 min of incubation at 30°C, reactions were initiated by the addition of *Nvec*-CDPS2 (125 nM-1 µM). 90 µl aliquots were withdrawn at different times and quenched by the addition of 2 µl of TFA. After addition of 10 µl of a 12.5 µM $^{13}C_9$,^{15}N-labeled cFF solution (see below), 90 µl of these aliquots were injected in LC-MS on the same system as *in vivo* assays, with a linear gradient from 20 to 40% (v/v) acetonitrile in 0.1% (v/v) formic acid for 20 min (flow rate, 0.6 ml min^{-1}). The detection of cFF synthesized was achieved by mass spectrometry using $^{13}C_9$,^{15}N-labeled cFF as a stable isotope internal standard. The synthesis of the standard was performed according to (Jeedigunta et al. 2000), with the exception that Boc-Phe-OH $^{13}C_9$,^{15}N was used instead of Boc-Phe-OH. The resulting $^{13}C_9$,^{15}N-labeled cFF (m/z = 305.1) and cFF (m/z = 295.1) have a difference of 10 mass units.

5.3 Additional information

We have described the first active CDPS, *Nvec*-CDPS2, identified in animal. *Nvec*-CDPS2 is probably not the only eukaryotic member of the CDPS family. Up to now, except in the sea anemone *Nematostella vectensis,* iterative PSI-BLAST searches have detected other four putative CDPSs in eukaryotes: the annelid *Platynereisdumerilii*, the fungi *Gibberella zeae* and *Fusarium oxysporum*, and the protozoa *Ichthyophthirius multifiliis* (Aravind et al. 2010; Belin et al. 2012). Not only in eukaryotes, are more and more bacterial putative CDPSs also being discovered in different phyla. A further 25 putative CDPSs have been recently identified, most of which belong to actinobacteria and proteobacteria. Aravind *et a*l. had already noted the large number of parasitic bacteria among CDPS-encoding organisms, which are not specialized in the synthesis of secondary metabolites (Aravind et al. 2010). Taking for example, Rv2275 is a CDPS from a human pathogen *M. tuberculosis* which persists within the macrophage through a complex host-pathogen relationship. Knowledge on these CDPSs from the parasitic organisms along with their cyclodipeptide products could help us in studies of related diseases.

6 CONCLUSIONS AND PERSPECTIVES

At the beginning of this manuscript, we have emphasized the developing interest in the drug discovery research of the diketopiperazines (DKPs) which include the cyclodipeptides and their derivatives. The structural complexity and diversity of this molecule family make it difficult to synthesize some DKPs with complex modifications *via* chemical approach. It is thus of great interest to elucidate the biosynthetic pathways of the DKPs. However, our knowledge on this aspect is still poor. Nonribosomal peptide synthetases (NRPSs) were the first enzymes discovered of being responsible for the synthesis of some DKPs. In the **chapter 1**, we listed several DKP biosynthetic pathways dependent of NRPSs as examples, and detailed structures of NRPSs as well as their involvement in the nonribosomal peptide bond formation. Lately, the cyclodipeptide synthases (CDPS) family was discovered to be dedicated to the formation of cyclodipeptides (Gondry et al. 2009). Studies on CDPSs are thus of great interest to elucidate the biosynthesis of DKPs.

The first part of our results on AlbC (**Chapter 2**), which is published in *Nucleic Acid Research* journal, reveals that AlbC displays high structural similarity to the catalytic domain of class-I aminoacyl-tRNA synthetases (aaRSs), especially class-Ic TyrRSs and TrpRSs. AlbC contains a catalytic pocket which is highly conserved among CDPSs. Our studies indicate that this pocket accommodates the aminoacyl moiety of the aa-tRNA substrate whereas the tRNA moiety interacts with AlbC *via* at least one patch of basic residues. Moreover, we have demonstrated that AlbC catalyzes its two-substrate reaction *via* a sequential ping-pong mechanism with a covalent intermediate in which Phe is shown to be transferred from Phe-tRNA^{Phe} to an active serine situated in the catalytic pocket. A mechanism of covalent phenylalanyl-enzyme formation is also proposed. All these findings provide insight

into the molecular bases of the interactions between CDPSs and their aa-tRNA substrates, and the catalytic mechanism used by CDPSs to achieve the nonribosomal synthesis of cyclodipeptides.

The following part of my thesis work was, by using the newly characteraized CDPSs, to deepen our knowledge on specificity of CDPSs and their catalytic mechanism.

The second part of our results (**Chapter 3**) is on Ndas_1148, a recently characterized CDPS in *Nocardiopsis dassonvillei*. Compared to the other nine identified CDPS (Gondry et al. 2009; Seguin et al. 2011), Ndas_1148 has the closest protein sequence to AlbC and a genomic context similar to those of AlbC. However, in spite of all these similarities, the two CDPSs do not have the same substrate specificity. Ndas_1148 produces cFY as the major cyclodipeptide instead of cFL, which is the major product of AlbC. After a series of studies, we showed that their difference in profiles of cyclodipeptide production is mainly due to the difference in residues constituting the two CDPS pockets because the AlbC chimera obtained by grafting the Ndas_1148 pocket on AlbC has the substrate specificity of Ndas_1148. We also showed that the CDPS pocket contains specificity determinants responsible for the incorporation of a leucyl residue into cyclodipeptides. These findings are significant for understanding of the molecular bases of CDPS specificity, especially those located in the pockets. The chimera that corresponds to AlbC having the Ndas_1148 pocket has acquired the substrate specificity of Ndas_1148, suggesting that reprogramming of CDPS specificity can be achieved by replacing the entire pocket of a given CDPS with that of another one.

The third part of our results (**Chapter 4**) consists of studies on AlbC-IMI, a CDPS identified in *Streptomyces* sp. IMI 351155. AlbC-IMI is a good study model for its sole cyclodipeptide produced, cFL, in the presence of Phe-tRNA^{Phe} and Leu-tRNA^{Leu} substrates. As CDPSs have been shown to use a sequential ping-pong mechanism, the two different substrate-binding sites on AlbC-IMI can thus be distinguished. According to results on acyl-enzyme intermediate-forming

experiments, we demonstrated that Phe-tRNA^{Phe} is the first substrate to interact with AlbC-IMI rather than Leu-tRNA^{Leu} substrate. Our preliminary results demonstrate that the CDPS AlbC-IMI is of great interest and studies on it should pursue.

The last part of our results (**Chapter 5**) concerns published work on the first characterized CDPS from an animal, that is *Nvec*-CDPS2 identified in the sea anemone *Nematostella vectensis*. *Nvec*-CDPS2 mostly produces *in vivo* Trp-containing cyclodipeptides, thus exhibits a different specificity than other characterized bacterial CDPSs which synthesize Phe-, Tyr-, and Leu-containing cyclodipeptides. This difference in specificity could be attributed to residues constituting the catalytic pocket because *Nvec*-CDPS2 contains in the pocket a higher proportion of aromatic residues than the other characterized CDPSs. This finding also suggests that the substrate specificity of CDPSs is related to their catalytic pockets. The characterization of *Nvec*-CDPS2 is of great significance because we have experimentally proved the existence of CDPS in the animal. Otherwise, other four eukaryotic genomes have been found to encode CDPSs: the fungi *Gibberella zeae* and *Fusarium oxysporum*, the annelid *Platynereis dumerilii* and the protozoa *Ichthyophthirius multifiliis* (Aravind et al. 2010; Belin et al. 2012).

Genome mining revealed the presence of putative CDPSs in approximately 60 different genome from bacteria, fungi or animal (**Figure 50**), including numerous parasitic bacteria (*F. oxysporum*, *Candidatus* Glomeribacter gigasporarum and *Candidatus* Odyssella thessalonicensis, for example) (Aravind et al. 2010; Belin et al. 2012). It was suggested that cyclodipeptides might enable the parasite to reside within the host (Aravind et al. 2010). The earliest isolations of CDPSs included that of Rv2275 from *Mycobacterium tuberculosis*, which is a human pathogen that persists within the macrophage through a complex host-pathogen relationship. Rv2275 and its associated tailoring enzyme, CYP121, are highly conserved among the other pathogenic tuberculosis-causing mycobacteria of the tuberculosis complex (*M. africanum*, *M. bovis*, *M. canetii* and *M. microti*). Research on CDPSs originated from pathogens could be useful for studies on related diseases.

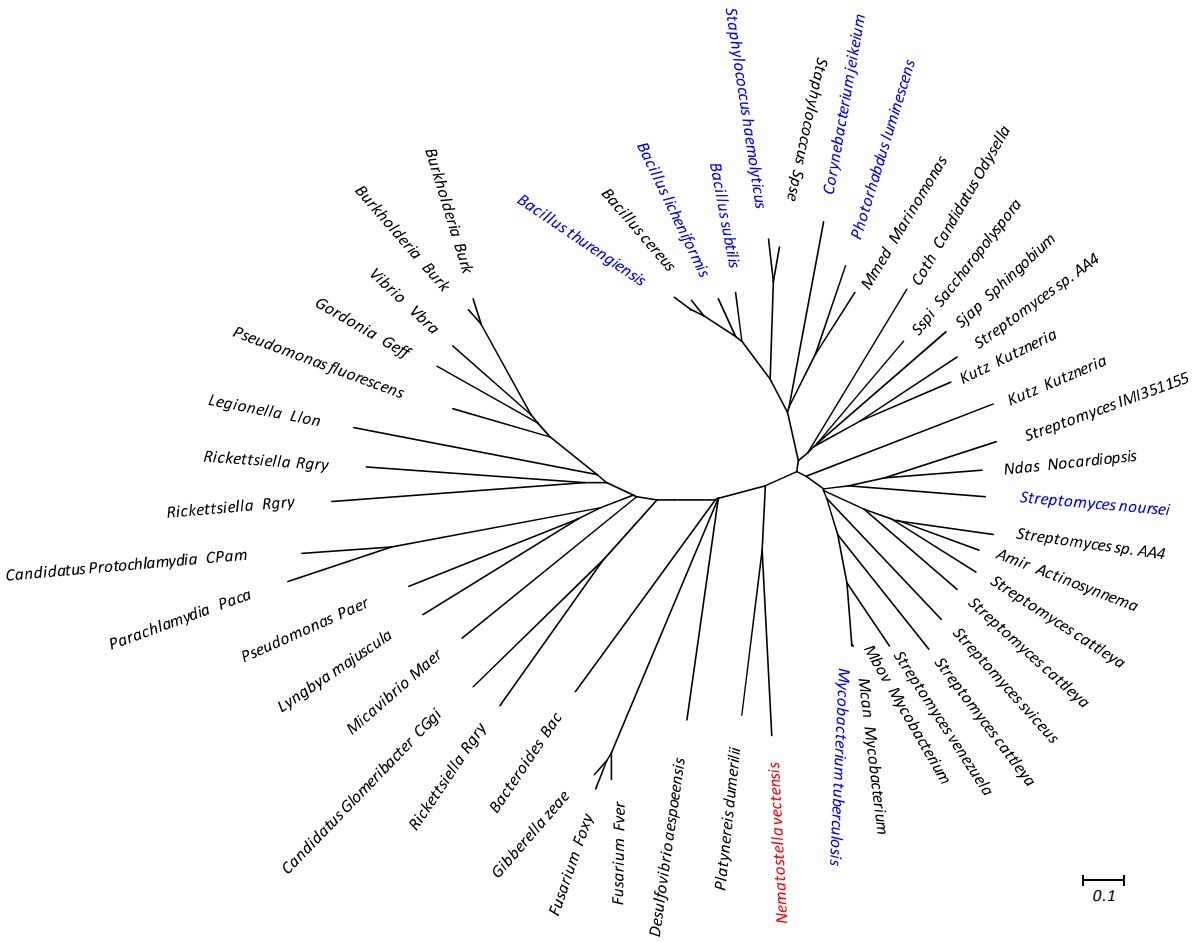


Figure 50: The developing CDPS-encoding organisms shown in the form of phylogenetic tree. The eight bacterial organisms and the eukaryotic organism encoding characterized CDPSs are shown in blue and in red, respectively. The organisms colored in black represent those encoding putative CDPSs.

In bacteria, cyclodipeptides produce by CDPSs are usually precursors of more complex DKPs generated by tailoring reactions. The genes encoding the CDPSs and the associated cyclodipeptide-tailoring enzymes are clustered. Only three CDPS-dependent pathways haven been fully elucidated so far, in which the cyclodipeptide-tailoring enzymes are responsible for α,β–dehydrogenation, C-C aryl coupling, and DKP ring oxidation (Belin et al. 2012), as described in the introduction of the manuscript. New putative cyclodipeptide-tailoring enzymes are being discovered such as methyl-transferases, oxydoreductases, acyl-CoA and hypothetical proteins not yet related to any known function (Belin et al. 2012). For example, genes encoding methyl transferases have been seperetely found in the vicinity of the genes encoding Ndas_1148 and AlbC-IMI (sections 3 and **4**). These proteins may catalyze

alternative modifications of cyclodipeptides, thereby enlarging the diversity of DKPs synthesized through CDPS-dependent pathways.

As previously described, in addition to the CDPS family, NPRS systems are also used to synthesize DKP metabolites in nature (sections **1.1.2.2.1** and **1.2.2.1**). However, these two enzyme families are very different. One of the obvious differences is their size: CDPSs (about 26 kDa) are about ten times smaller than NRPSs involved in DKP biosynthesis. CDPSs are simple systems whereas NRPSs are complex, multimodular enzymes. This size difference probably reflects the different strategies used to activate the amino acid carboxyl group required for peptide bond formation: NRPSs use adenylation and pepetidyl carrier protein (PCP) domains to recognize and activate amino acids in the form of PCP-bound aminoacyl thioesters (section **1.2.2.1**), whereas CDPSs directly use activated amino acids in the form of aa-tRNAs, thereby eliminating the need to activate amino acids (Belin et al. 2012). For that reason, CDPSs can only use the 20 canonical L-amino acids charged on tRNAs as substrates. However, amino acids incorporated by NRPSs are not restricted to the 20 canonical L-amino acids. They can also incorporate some unusual amino acids (section **1.2.2.1**, **Figure 32**) such as 4-nitrotryptophan (Johnson et al. 2009) into thaxtomin A (**Figure 6F**) biosynthesis (Belin et al. 2012). In addition, NRPS substrates can be altered on the enzyme by accessory domains introducing chemical modifications, such as methylation (methylation domains, thaxtomin synthetase) or configuration changes (epimerization domains, erythrochelin synthetase), whereas in CDPS-dependent pathways, chemical modifications can only be introduced after cyclodipeptide formation by cyclodipeptide-tailoring enzymes. Consequently, it coud lead to a wider structural complexity of DKPs synthesized through NRPS-dependent pathways than that produced *via* CDPS-dependent pathways (Belin et al. 2012).

Otherwise, It is interesting to note that all the NRPS-dependent DKP biosynthetic pathways characterized to date are only found in bacteria and fungi, whereas CDPS-dependent pathways have been identified in bacteria, fungi (*G. zeae*, *F. oxysporum*), protozoa (*I. multifiliis*), and animals (*N. vectensis*, *P. dumerilii*) (Belin et al.

2012), as described above. The first characterized animal-originated CDPS from *N. vectensis*is has been experimentally demonstrated to be involved in nonribosomal cyclodipeptide synthesis (Seguin et al. 2011). Furthermore, the functions of the DKPs known to be synthesized by NRPS pathways are usually identified (roles of siderophores or toxins involved in pathogenicity), whereas those from CDPS pathways are poorly known (Belin et al. 2012). Interestingly, it is noted that in the marine annelid *Platynereis dumerilii* (Aravind et al. 2010), the expression of the CDPS gene increases in response to septic injury (Altincicek and Vilcinskas 2007). A role in immunity can thus be suggested for the DKP produced *via* the CDPS pathway (Belin et al. 2012).

Regarding the reprogramming of enzymes, NRPS genes are particularly amenable to genetic engineering to modify the sequence or the nature of the peptide synthesized because of the modular architecture of the NRPS enzymes (Belin et al. 2012), knowing that each module of NRPS, divided into domains, is responsible for the incorporation of one amino acid (section **1.2.2.1**). Several groups have succeeded in deleting, inserting or replacing modules, and site directed mutagenesis has been used to modify the substrate specificity of some adenylation domains (Eppelmann et al. 2002; Mootz et al. 2002; Hahn and Stachelhaus 2004; Lautru and Challis 2004; Miao et al. 2006; Nguyen et al. 2006; Wilkinson and Micklefield 2007).

Studies on CDPSs began much latter than NRPSs. Our knowledge on CDPSs is still limited. However, the small size of CDPSs makes them good candidates for enzyme engineering. Such small proteins are generally easy to produce and to manipulate and could therefore be used to synthesize new cyclodipeptides. With our deepening understanding on molecular bases of the substrate specificity, it is conceivable to reprogram CDPSs by altering their substrate specificity. We already succeeded in modifying the nature of the cyclodipeptides produced by the CDPS AlbC by mutating a residue of its amino acyl-binding pocket (Sauguet et al. 2011). In the **chapter 3**, we also demonstrated that the specificity of AlbC can be switched to that of Ndas_1148 by grafting the pocket of the latter on AlbC. Furthermore, the structural homology of CDPSs with class-Ic aaRSs is a precious source of information

for such a reprogramming since extensive work has been done on aaRSs, especially on the amino acid-binding pocket of TyrRSs to accommodate non natural amino acids (Liu and Schultz 2010). The strategies developed for TyrRSs could thus be adapted to CDPSs (Belin et al. 2012). In addition, in bacteria, the CDPS-dependent biosynthetic pathways are organized into gene clusters, providing useful information for combinatorial approaches. The characterization of new and versatile CDPSs and of various cyclodipeptide-tailoring enzymes, combined with CDPS engineering, opens up new possibilities for pathway engineering and combinatorial approaches to increase the natural diversity of DKPs, thus to generate products with interesting biological or pharmaceutical properties (Belin et al. 2012).

References

Abramochkin G. and Shrader T. E. (1996). "Aminoacyl-tRNA Recognition by the Leucyl/Phenylalanyl-tRNA-Protein Transferase." Journal of Biological Chemistry **271**(37): 22901-22907.

Adamczeski M., Reed A. R. and Crews P. (1995). "New and Known Diketopiperazines from the Caribbean Sponge, Calyx cf. podatypa." Journal of Natural Products **58**(2): 201-208.

Altincicek B. and Vilcinskas A. (2007). "Analysis of the immune-related transcriptome of a lophotrochozoan model, the marine annelid Platynereis dumerilii." Frontiers in Zoology **4**(1): 18-27.

Ambrogelly A., Korencic D. and Ibba M. (2002). "Functional Annotation of Class I Lysyl-tRNA Synthetase Phylogeny Indicates a Limited Role for Gene Transfer." Journal of Bacteriology **184**(16): 4594-4600.

Aravind L., de Souza R. F. and Iyer L. M. (2010). "Predicted class-I aminoacyl tRNA synthetase-like proteins in non-ribosomal peptide synthesis." Biology Direct **5**(48): 48-58.

Arnez J. G. and Moras D. (1997). "Structural and functional considerations of the aminoacylation reaction." Trends in Biochemical Sciences **22**(6): 211-216.

Belin P., Le Du M. H., Fielding A., Lequin O., Jacquet M., Charbonnier J.-B., Lecoq A., Thai R., Courçon M., Masson C., Dugave C., Genet R., Pernodet J.-L. and Gondry M. (2009). "Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in Mycobacterium tuberculosis." Proceedings of the National Academy of Sciences of the United States of America **106**(18): 7426-7431.

Belin P., Moutiez M., Lautru S., Seguin J., Pernodet J.-L. and Gondry M. (2012). "The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways." Natural Product Reports **29**(9): 961-979.

Benson T. E., Prince D. B., Mutchler V. T., Curry K. A., Ho A. M., Sarver R. W., Hagadorn J. C., Choi G. H. and Garlick R. L. (2002). "X-Ray Crystal Structure of Staphylococcus aureus FemA." Structure **10**(8): 1107-1115.

Beringer M. and Rodnina M. V. (2007). "The Ribosomal Peptidyl Transferase." Molecular Cell **26**(3): 311-321.

Biarrotte-Sorin S., Maillard A. P., Delettré J., Sougakoff W., Arthur M. and Mayer C. (2004). "Crystal Structures of Weissella viridescens FemX and Its Complex with UDP-MurNAc-Pentapeptide: Insights into FemABX Family Substrates Recognition." Structure **12**(2): 257-267.

Bieling P., Beringer M., Adio S. and Rodnina M. V. (2006). "Peptide bond formation does not involve acid-base catalysis by ribosomal residues." Nature Structural & Molecular Biology **13**(5): 423-428.

Blow D. M., Birktoft J. J. and Hartley B. S. (1969). "Role of a Buried Acid Group in the Mechanism of Action of Chymotrypsin." Nature **221**(5178): 337-340.

Bonnefond L., Arai T., Sakaguchi Y., Suzuki T., Ishitani R. and Nureki O. (2011). "Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog." Proceedings of the National Academy of Sciences of the United States of America **108**(10): 3912-3917.

Boojamra C. G., Lemoine R. C., Lee J. C., Léger R., Stein K. A., Vernier N. G., Magon A., Lomovskaya O., Martin P. K., Chamberland S., Lee M. D., Hecker S. J. and Lee V. J. (2001). "Stereochemical Elucidation and Total Synthesis of Dihydropacidamycin D, a Semisynthetic Pacidamycin." Journal of the American Chemical Society **123**(5): 870-874.

Bradley E. L., Herbert R. B., Lawrie K. W. M., Khan J. A., Moody C. M. and Young D. W. (1996). "The biosynthesis of the Streptomyces antibiotic bicyclomycin." Tetrahedron Letters **37**(38): 6935-6938.

Braud S., Moutiez M., Belin P., Abello N., Drevet P., Zinn-Justin S., Courçon M., Masson C., Dassa J., Charbonnier J.-B., Boulain J.-C., Ménez A., Genet R. and Gondry M. (2005). "Dual Expression System Suitable for High-Throughput Fluorescence-Based Screening and Production of Soluble Proteins." Journal of Proteome Research **4**(6): 2137-2147.

Bruner S. D., Weber T., Kohli R. M., Schwarzer D., Marahiel M. A., Walsh C. T. and Stubbs M. T. (2002). "Structural Basis for the Cyclization of the Lipopeptide Antibiotic Surfactin by the Thioesterase Domain SrfTE." Structure **10**(3): 301-310.

Caboche S., Pupin M., Leclère V., Fontaine A., Jacques P. and Kucherov G. (2008). "NORINE: a database of nonribosomal peptides." Nucleic Acids Research **36**(suppl 1): D326-D331.

Cai S., Kong X., Wang W., Zhou H., Zhu T., Li D. and Gu Q. (2012). "Aspergilazine A, a diketopiperazine dimer with a rare N-1 to C-6 linkage, from a marine-derived fungus Aspergillus taichungensis." Tetrahedron Letters **53**(21): 2615-2617.

Cain C. C., Lee D., Waldo R. H., Henry A. T., Casida E. J., Wani M. C., Wall M. E., Oberlies N. H. and Falkinham J. O. (2003). "Synergistic Antimicrobial Activity of Metabolites Produced by a Nonobligate Bacterial Predator." Antimicrobial Agents and Chemotherapy **47**(7): 2113-2117.

Chen R., Buko A., Whittern D. and McAlpine J. (1989). "Pacidamycins, a novel series of antibiotics with anti-Pseudomonas aeruginosa activity. II. Isolation and structural elucidation." The Journal of Antibiotics **42**(4): 512-520.

Choi E. J., Park J. S., Kim Y. J., Jung J. H., Lee J. K., Kwon H. C. and Yang H. O. (2011). "Apoptosis-inducing effect of diketopiperazine disulfides produced by Aspergillus sp. KMD 901 isolated from marine sediment on HCT116 colon cancer cell lines." Journal of Applied Microbiology **110**(1): 304-313.

Cole S. T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S. V., Eiglmeier K., Gas S., Barry C. E., Tekaia F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver K., Osborne J., Quail M. A., Rajandream M. A., Rogers J., Rutter S., Seeger K., Skelton J., Squares R., Squares S., Sulston J. E., Taylor K., Whitehead S. and Barrell B. G. (1998). "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence." Nature **393**(6685): 537-544.

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

Crary S. M., Niranjanakumari S. and Fierke C. A. (1998). "The Protein Component of Bacillus subtilis Ribonuclease P Increases Catalytic Efficiency by Enhancing Interactions with the 5‘ Leader Sequence of Pre-tRNAAsp." Biochemistry **37**(26): 9409-9416.

Crick F. H. C. (1968). "The origin of the genetic code." Journal of Molecular Biology **38**(3): 367-379.

Cryle M. J., Bell S. G. and Schlichting I. (2010). "Structural and Biochemical Characterization of the Cytochrome P450 CypX (CYP134A1) from Bacillus subtilis: A Cyclo-l-leucyl-l-leucyl Dipeptide Oxidase." Biochemistry **49**(34): 7282-7296.

Dare K. and Ibba M. (2012). "Roles of tRNA in cell wall biosynthesis." Wiley Interdisciplinary Reviews: RNA **3**(2): 247-264.

De Carvalho M. P. and Abraham W. R. (2012). "Antimicrobial and Biofilm Inhibiting Diketopiperazines." Current Medicinal Chemistry **19**(21): 3564-3577.

de Kievit T. R. and Iglewski B. H. (2000). "Bacterial Quorum Sensing in Pathogenic Relationships." Infection and Immunity **68**(9): 4839-4849.

Degrassi G., Aguilar C., Bosco M., Zahariev S., Pongor S. and Venturi V. (2002). "Plant Growth-Promoting *Pseudomonas putida* WCS358 Produces and Secretes Four Cyclic Dipeptides: Cross-Talk with Quorum Sensing Bacterial Sensors." Current Microbiology **45**(4): 250-254.

Dong X., Kato-Murayama M., Muramatsu T., Mori H., Shirouzu M., Bessho Y. and Yokoyama S. (2007). "The crystal structure of leucyl/phenylalanyl-tRNA-protein transferase from Escherichia coli." Protein Science **16**(3): 528-534.

Dramsi S., Magnet S., Davison S. and Arthur M. (2008). "Covalent attachment of proteins to peptidoglycan." FEMS Microbiology Reviews **32**(2): 307-320.

Ehlert K., Schroder W. and Labischinski H. (1997). "Specificities of FemA and FemB for different glycine residues: FemB cannot substitute for FemA in Staphylococcal peptidoglycan pentaglycine side chain formation." Journal of Bacteriology **179**(23): 7573-7576.

Eppelmann K., Stachelhaus T. and Marahiel M. A. (2002). "Exploitation of the Selectivity-Conferring Code of Nonribosomal Peptide Synthetases for the Rational Design of Novel Peptide Antibiotics." Biochemistry **41**(30): 9718-9726.

Felnagle E. A., Jackson E. E., Chan Y. A., Podevels A. M., Berti A. D., McMahon M. D. and Thomas M. G. (2008). "Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products." Molecular Pharmaceutics **5**(2): 191-211.

Fenical W. (1993). "Chemical studies of marine bacteria: developing a new resource." Chemical Reviews **93**(5): 1673-1683.

Finking R. and Marahiel M. A. (2004). "Biosynthesis of nonribosomal peptides." Annual Review of Microbiology **58**(1): 453-488.

Fischbach M. A. and Walsh C. T. (2006). "Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms." Chemical Reviews **106**(8): 3468-3496.

Forman H. J., Zhang H. and Rinna A. (2009). "Glutathione: Overview of its protective roles, measurement, and biosynthesis." Molecular Aspects of Medicine **30**(1–2): 1-12.

Freist W., Logan D. T. and Gauss D. H. (1996). "Glycyl-tRNA synthetase." Biological chemistry Hoppe-Seyler **377**(6): 343-356.

Fronko R., Lee J., Galazzo J., Chamberland S., Malouin F. and Lee M. (2000). "New pacidamycins produced by Streptomyces coeruleorubidus, NRRL 18370." The Journal of Antibiotics **53**(12): 1405-1410.

Fung A. W., Ebhardt H. A., Abeysundara H., Moore J., Xu Z. and Fahlman R. P. (2011). "An Alternative Mechanism for the Catalysis of Peptide Bond Formation by L/F Transferase: Substrate Binding and Orientation." Journal of Molecular Biology **409**(4): 617-629.

Galperin M. Y. and Koonin E. V. (1997). "A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity." Protein Science **6**(12): 2639-2643.

Gardiner D. M., Cozijnsen A. J., Wilson L. M., Pedras M. S. C. and Howlett B. J. (2004). "The sirodesmin biosynthetic gene cluster of the plant pathogenic fungus Leptosphaeria maculans." Molecular Microbiology **53**(5): 1307-1318.

Gardiner D. M., Waring P. and Howlett B. J. (2005). "The epipolythiodioxopiperazine (ETP) class of fungal toxins: distribution, mode of action, functions and biosynthesis." Microbiology **151**(4): 1021-1032.

Garg R. P., Qian X. L., Alemany L. B., Moran S. and Parry R. J. (2008). "Investigations of valanimycin biosynthesis: Elucidation of the role of seryl-tRNA." Proceedings of the National Academy of Sciences of the United States of America **105**(18): 6543-6547.

Gogos A. and Shapiro L. (2002). "Large Conformational Changes in the Catalytic Cycle of Glutathione Synthase." Structure **10**(12): 1669-1676.

Gondry M., Lautru S., Fusai G., Meunier G., Ménez A. and Genet R. (2001). "Cyclic dipeptide oxidase from Streptomyces noursei." European Journal of Biochemistry **268**(6): 1712-1721.

Gondry M., Sauguet L., Belin P., Thai R., Amouroux R., Tellier C., Tuphile K., Jacquet M., Braud S., Courcon M., Masson C., Dubois S., Lautru S., Lecoq A., Hashimoto S., Genet R. and Pernodet J. L. (2009). "Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes." Nature Chemical Biology **5**(6): 414-420.

González-López M. A., González-Vela M. C., Salas-Venero C. A., Conde R. and Val-Bernal J. F. (2011). "Cutaneous infection caused by Nocardiopsis dassonvillei presenting with sporotrichoid spread." Journal of the American Academy of Dermatology **65**(3): e90-e91.

Grosjean H., Edqvist J., Stråby K. B. and Giegé R. (1996). "Enzymatic Formation of Modified Nucleosides in tRNA: Dependence on tRNA Architecture." Journal of Molecular Biology **255**(1): 67-85.

Grovel O., Kerzaon I., Petit K., Robiou Du Pont T. and Pouchus Y.-F. (2006). "A new and rapid bioassay for the detection of gliotoxin and related epipolythiodioxopiperazines produced by fungi." Journal of Microbiological Methods **66**(2): 286-293.

Grundmann A. and Li S. M. (2005). "Overproduction, purification and characterization of FtmPT1, a brevianamide F prenyltransferase from Aspergillus fumigatus." Microbiology **151**(7): 2199-2207.

Guengerich F. P. (2001). "Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity." Chemical Research in Toxicology **14**(6): 611-650.

Guerrier-Takada C., Gardiner K., Marsh T., Pace N. and Altman S. (1983). "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme." Cell **35**(3): 849-857.

Guo H. J., Sun B. D., Gao H., Chen X. L., Liu S. C., Yao X. S., Liu X. Z. and Che Y. S. (2009). "Diketopiperazines from the Cordyceps-Colonizing Fungus Epicoccum nigrum." Journal of Natural Products **72**(12): 2115-2119.

Gurney K. A. and Mantle P. G. (1993). "Biosynthesis of 1-N-methylalbonoursin by an Endophytic Streptomyces sp. Isolated from Perennial Ryegrass." Journal of Natural Products **56**(7): 1194-1198.

Hahn M. and Stachelhaus T. (2004). "Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains." Proceedings of the National Academy of Sciences of the United States of America **101**(44): 15585-15590.

Hartman G. and Wise R. (1998). "Quorum sensing: potential means of treating gram-negative infections?" The Lancet **351**(9106): 848-849.

Hartmann R. K., Gößringer M., Späth B., Fischer S. and Marchfelder A. (2009). Chapter 8 The Making of tRNAs and More – RNase P and tRNase Z. Progress in Molecular Biology and Translational Science. C. Ciaran, Academic Press. **Volume 85:** 319-368.

Harvey A. L. (2008). "Natural products in drug discovery." Drug Discovery Today **13**(19–20): 894-901.

Healy F. G., Krasnoff S. B., Wach M., Gibson D. M. and Loria R. (2002). "Involvement of a Cytochrome P450 Monooxygenase in Thaxtomin A Biosynthesis by Streptomyces acidiscabies." Journal of Bacteriology **184**(7): 2019-2029.

Healy F. G., Wach M., Krasnoff S. B., Gibson D. M. and Loria R. (2000). "The txtAB genes of the plant pathogen Streptomyces acidiscabies encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity." Molecular Microbiology **38**(4): 794-804.

Hegde S. S. and Blanchard J. S. (2003). "Kinetic and Mechanistic Characterization of Recombinant Lactobacillus viridescens FemX (UDP-N-acetylmuramoyl Pentapeptide-lysine N6-Alanyltransferase)." Journal of Biological Chemistry **278**(25): 22861-22867.

Hegde S. S. and Shrader T. E. (2001). "FemABX family members are novel nonribosomal peptidyltransferases and important pathogen-specific drug targets." Journal of Biological Chemistry **276**(10): 6998-7003.

Henke J. M. and Bassler B. L. (2004). "Bacterial social engagements." Trends in Cell Biology **14**(11): 648-656.

Herrera K., Cahoon R. E., Kumaran S. and Jez J. M. (2007). "Reaction Mechanism of Glutathione Synthetase from Arabidopsis thaliana." Journal of Biological Chemistry **282**(23): 17157-17165.

Hoagland M. B., Stephenson M. L., Scott J. F., Hecht L. I. and Zamecnik P. C. (1958). "A soluble ribonucleic acid intermediate in protein synthesis." Journal of Biological Chemistry **231**(1): 241-257.

Holden M. T. G., Ram Chhabra S., De Nys R., Stead P., Bainton N. J., Hill P. J., Manefield M., Kumar N., Labatte M., England D., Rice S., Givskov M., Salmond G. P. C., Stewart G. S. A. B., Bycroft B. W., Kjelleberg S. and Williams P. (1999). "Quorum-sensing cross talk: isolation and chemical characterization of cyclic dipeptides from Pseudomonas aeruginosa and other Gram-negative bacteria." Molecular Microbiology **33**(6): 1254-1266.

Hong K. W., Ibba M., WeygandDurasevic I., Rogers M. J., Thomann H. U. and Soll D. (1996). "Transfer RNA-dependent cognate amino acid recognition by an aminoacyl-tRNA synthetase." The EMBO Journal **15**(8): 1983-1991.

Horinishi H., Hashizume S., Seguchi M. and Takahashi K. (1975). "Incorporation of methionine by a soluble enzyme system from Escherichia coli." Biochemical and Biophysical Research Communications **67**(3): 1136-1143.

Huang C. S., Chang L. S., Anderson M. E. and Meister A. (1993). "Catalytic and regulatory properties of the heavy subunit of rat kidney gamma-glutamylcysteine synthetase." Journal of Biological Chemistry **268**(26): 19675-19680.

Huang R., Zhou X., Xu T., Yang X. and Liu Y. (2010). "Diketopiperazines from Marine Organisms." Chemistry & Biodiversity **7**(12): 2809-2829.

Ibba M., Hong K. W., Sherman J. M., Sever S. and Soll D. (1996). "Interactions between tRNA identity nucleotides and their recognition sites in glutaminyl-tRNA synthetase determine the cognate amino acid affinity of the enzyme." Proceedings of the National Academy of Sciences of the United States of America **93**(14): 6953-6958.

Ibba M. and Söll D. (2000). "AMINOACYL-tRNA SYNTHESIS." Annual Review of Biochemistry **69**(1): 617-650.

Isin E. M. and Guengerich F. P. (2007). "Complex reactions catalyzed by cytochrome P450 enzymes." Biochimica et Biophysica Acta (BBA) - General Subjects **1770**(3): 314-329.

Jayatilake G. S., Thornton M. P., Leonard A. C., Grimwade J. E. and Baker B. J. (1996). "Metabolites from an Antarctic Sponge-Associated Bacterium, Pseudomonas aeruginosa." Journal of Natural Products **59**(3): 293-296.

Jeedigunta S., Krenisky J. M. and Kerr R. G. (2000). "Diketopiperazines as advanced intermediates in the biosynthesis of ecteinascidins." Tetrahedron **56**(21): 3303-3307.

Jia J., Ma X., Wu C., Wu L. and Hu G. (2005). "Cordycedipeptide A, a new cyclodipeptide from the culture liquid of cordyceps sinensis (BERK.) SACC." Chemical and Pharmaceutical Bulletin **53**(5): 582-583.

Johnson E. G., Krasnoff S. B., Bignell D. R. D., Chung W.-C., Tao T., Parry R. J., Loria R. and Gibson D. M. (2009). "4-Nitrotryptophan is a substrate for the non-ribosomal peptide synthetase TxtB in the thaxtomin A biosynthetic pathway." Molecular Microbiology **73**(3): 409-418.

Jovine L., Djordjevic S. and Rhodes D. (2000). "The crystal structure of yeast phenylalanine tRNA at 2.0 Å resolution: cleavage by Mg^{2+} in 15-year old crystals." Journal of Molecular Biology **301**(2): 401-414.

Kanoh K., Kohno S., Katada J., Hayashi Y., Muramatsu M. and Uno I. (1999). "Antitumor activity of phenylahistin *in vitro* and *in vivo*." Bioscience, Biotechnology, and Biochemistry **63**(6): 1130-1133.

Kanzaki H., Yanagisawa S. and Nitoda T. (2004). "Enzymatic Synthesis of Dehydro Cyclo(His–Phe)s, Analogs of the Potent Cell Cycle Inhibitor, Dehydrophenylahistin, and Their Inhibitory Activities toward Cell Division." Bioscience, Biotechnology, and Biochemistry **68**(11): 2341-2345.

Khokhlov A. S. and Lokshin G. B. (1963). "The structure or albonoursin." Tetrahedron Letters **4**(27): 1881-1885.

Kim S., You S. and Hwang D. (2011). "Aminoacyl-tRNA synthetases and tumorigenesis: more than housekeeping." Nature Reviews Cancer **11**(10): 708-718.

King R. R. and Calhoun L. A. (2009). "The thaxtomin phytotoxins: Sources, synthesis, biosynthesis, biotransformation and biological activity." Phytochemistry **70**(7): 833-841.

King R. R., Lawrence C. H. and Calhoun L. A. (1992). "Chemistry of phytotoxins associated with Streptomyces scabies the causal organism of potato common scab." Journal of Agricultural and Food Chemistry **40**(5): 834-837.

Kirsebom L. A. (2007). "RNase P RNA mediated cleavage: Substrate recognition and catalysis." Biochimie **89**(10): 1183-1194.

Koglin A., Mofid M. R., Löhr F., Schäfer B., Rogov V. V., Blum M.-M., Mittag T., Marahiel M. A., Bernhard F. and Dötsch V. (2006). "Conformational switches modulate protein interactions in peptide antibiotic synthetases." Science **312**(5771): 273-276.

Kohn H. and Widger W. (2005). "The molecular basis for the mode of action of bicyclomycin." Current Drug Targets - Infectious Disorders **5**(3): 273-295.

Konz D. and Marahiel M. A. (1999). "How do peptide synthetases generate structural diversity?" Chemistry & Biology **6**(2): R39-R48.

Kopp F. and Marahiel M. A. (2007). "Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis." Natural Product Reports **24**(4): 735-749.

Lautru S. and Challis G. L. (2004). "Substrate recognition by nonribosomal peptide synthetase multi-enzymes." Microbiology **150**(6): 1629-1636.

Lautru S., Gondry M., Genet R. and Pernodet J. L. (2002). "The albonoursin gene cluster of S-noursei: Biosynthesis of diketopiperazine metabolites independent of nonribosomal peptide synthetases." Chemistry & Biology **9**(12): 1355-1364.

Li J., Wang W., Xu S. X., Magarvey N. A. and McCormick J. K. (2011). "Lactobacillus reuteri-produced cyclic dipeptides quench agr-mediated expression of toxic shock syndrome toxin-1 in staphylococci." Proceedings of the National Academy of Sciences of the United States of America **108**(8): 3360-3365.

Li J. W.-H. and Vederas J. C. (2009). "Drug discovery and natural products: End of an era or an endless frontier?" Science **325**(5937): 161-165.

Linne U. and Marahiel M. A. (2000). "Control of directionality in nonribosomal peptide synthesis: Role of the condensation domain in preventing misinitiation and timing of epimerization." Biochemistry **39**(34): 10439-10447.

Liu C. C. and Schultz P. G. (2010). "Adding New Chemistries to the Genetic Code." Annual Review of Biochemistry **79**(1): 413-444.

Loria R., Bignell D., Moll S., Huguet-Tapia J., Joshi M., Johnson E., Seipke R. and Gibson D. (2008). "Thaxtomin biosynthesis: the path to plant pathogenicity in the genus *Streptomyces*." Antonie van Leeuwenhoek **94**(1): 3-10.

Magyar A., Zhang X., Kohn H. and Widger W. R. (1996). "The antibiotic bicyclomycin affects the secondary RNA binding site of *Escherichia coli* transcription termination factor Rho." Journal of Biological Chemistry **271**(41): 25369-25374.

Mainardi J.-L., Villet R., Bugg T. D., Mayer C. and Arthur M. (2008). "Evolution of peptidoglycan biosynthesis under the selective pressure of antibiotics in Gram-positive bacteria." FEMS Microbiology Reviews **32**(2): 386-408.

Maiya S., Grundmann A., Li S.-M. and Turner G. (2006). "The Fumitremorgin Gene Cluster of *Aspergillus fumigatus*: Identification of a Gene Encoding Brevianamide F Synthetase." ChemBioChem **7**(7): 1062-1069.

Marahiel M. A. (2009). "Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis." Journal of Peptide Science **15**(12): 799-807.

Martinis S. A., Plateau P., Cavarelli J. and Florentz C. (1999). "Aminoacyl-tRNA synthetases: A new image for a classical family." Biochimie **81**(7): 683-700.

Martins M. B. and Carvalho I. (2007). "Diketopiperazines: biological activity and synthesis." Tetrahedron **63**(40): 9923-9932.

Meister A. (1974). "Glutathione; Metabolism and function via the γ-glutamyl cycle." Life Sciences **15**(2): 177-190.

Miao V., Coëffet-Le Gal M.-F., Nguyen K., Brian P., Penn J., Whiting A., Steele J., Kau D., Martin S., Ford R., Gibson T., Bouchard M., Wrigley S. K. and Baltz R. H. (2006). "Genetic Engineering in Streptomyces roseosporus to Produce Hybrid Lipopeptide Antibiotics." Chemistry & Biology **13**(3): 269-276.

Milne P. J., Hunt A. L., Rostoll K., Van Der Walt J. J. and Graz C. J. M. (1998). "Medicinal Chemistry: The Biological Activity of Selected Cyclic Dipeptides." Journal of Pharmacy and Pharmacology **50**(12): 1331-1337.

Minelli A., Conte C., Grottelli S., Bellezza M., Cacciatore I. and Bolaños J. P. (2009). "Cyclo(His-Pro) promotes cytoprotection by activating Nrf2-mediated up-regulation of antioxidant defence." Journal of Cellular and Molecular Medicine **13**(6): 1149-1161.

Minelli A., Grottelli S., Mierla A., Pinnen F., Cacciatore I. and Bellezza I. (2012). "Cyclo(His-Pro) exerts anti-inflammatory effects by modulating NF-κB and Nrf2 signalling." The International Journal of Biochemistry & Cell Biology **44**(3): 525-535.

Miyashita K., Murakami M., Yamada M., Iriuchijima T. and Mori M. (1993). "Histidyl-proline diketopiperazine. Novel formation that does not originate from thyrotropin-releasing hormone." Journal of Biological Chemistry **268**(28): 20863-20865.

Mocibob M., Ivic N., Bilokapic S., Maier T., Luic M., Ban N. and Weygand-Durasevic I. (2010). "Homologs of aminoacyl-tRNA synthetases acylate carrier proteins and provide a link between ribosomal and nonribosomal peptide synthesis." Proceedings of the National Academy of Sciences of the United States of America **107**(33): 14585-14590.

Mogk A., Schmidt R. and Bukau B. (2007). "The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies." Trends in Cell Biology **17**(4): 165-172.

Mootz H. D., Kessler N., Linne U., Eppelmann K., Schwarzer D. and Marahiel M. A. (2002). "Decreasing the Ring Size of a Cyclic Nonribosomal Peptide Antibiotic by In-Frame Module Deletion in the Biosynthetic Genes." Journal of the American Chemical Society **124**(37): 10980-10981.

Mootz H. D., Schwarzer D. and Marahiel M. A. (2002). "Ways of Assembling Complex Natural Products on Modular Nonribosomal Peptide Synthetases." ChemBioChem **3**(6): 490-504.

Müller G., Barclay S. J. and Raymond K. N. (1985). "The mechanism and specificity of iron transport in Rhodotorula pilimanae probed by synthetic analogs of rhodotorulic acid." Journal of Biological Chemistry **260**(26): 13916-13920.

Musetti R., Polizzotto R., Vecchione A., Borselli S., Zulini L., D'Ambrosio M., Toppi L. S. d. and Pertot I. (2007). "Antifungal activity of diketopiperazines extracted from Alternaria alternata against Plasmopara viticola: An ultrastructural study." Micron **38**(6): 643-650.

Nakatsu T., Kato H. and Oda J. (1998). "Crystal structure of asparagine synthetase reveals a close evolutionary relationship to class II aminoacyl-tRNA synthetase." Nature Structural Biology **5**(1): 15-19.

Neuwald A. F. and Landsman D. (1997). "GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein." Trends in Biochemical Sciences **22**(5): 154-155.

Nguyen K. T., Ritz D., Gu J.-Q., Alexander D., Chu M., Miao V., Brian P. and Baltz R. H. (2006). "Combinatorial biosynthesis of novel antibiotics related to daptomycin." Proceedings of the National Academy of Sciences of the United States of America **103**(46): 17462-17467.

O'Donoghue P. and Luthey-Schulten Z. (2003). "On the Evolution of Structure in Aminoacyl-tRNA Synthetases." Microbiology and Molecular Biology Reviews **67**(4): 550-573.

Ortiz-Castro R., Díaz-Pérez C., Martínez-Trujillo M., del Río R. E., Campos-García J. and López-Bucio J. (2011). "Transkingdom signaling based on bacterial cyclodipeptides with auxin activity in plants." Proceedings of the National Academy of Sciences of the United States of America **108**(17): 7253-7258.

Park D.-K., Lee K.-E., Baek C.-H., Kim I. H., Kwon J.-H., Lee W. K., Lee K.-H., Kim B.-S., Choi S.-H. and Kim K.-S. (2006). "Cyclo(Phe-Pro) Modulates the Expression of ompU in Vibrio spp." Journal of Bacteriology **188**(6): 2214-2221.

Park H. B., Kwon H. C., Lee C.-H. and Yang H. O. (2009). "Glionitrin A, an Antibiotic–Antitumor Metabolite Derived from Competitive Interaction between Abandoned Mine Microbes." Journal of Natural Products **72**(2): 248-252.

Phizicky E. M. and Hopper A. K. (2010). "tRNA biology charges to the front." Genes & Development **24**(17): 1832-1860.

Polekhina G., Board P. G., Gali R. R., Rossjohn J. and Parker M. W. (1999). "Molecular basis of glutathione synthetase deficiency and a rare gene permutation event." The EMBO Journal **18**(12): 3204-3213.

Prasad C. (1995). "Bioactive cyclic dipeptides." Peptides **16**(1): 151-164.

RajBhandary U. L. and Söll D. (2008). "Aminoacyl-tRNAs, the bacterial cell envelope, and antibiotics." Proceedings of the National Academy of Sciences of the United States of America **105**(14): 5285-5286.

Rausch C., Weber T., Kohlbacher O., Wohlleben W. and Huson D. H. "Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs)." Nucleic Acids Research **33**(18): 5799-5808.

Reuter K., Mofid M. R., Marahiel M. A. and Ficner R. (1999). "Crystal structure of the surfactin synthetase-activating enzyme Sfp: a prototype of the 4[prime]-phosphopantetheinyl transferase superfamily." The EMBO Journal **18**(23): 6823-6831.

Ribas de Pouplana L. and Schimmel P. (2001). "Two Classes of tRNA Synthetases Suggested by Sterically Compatible Dockings on tRNA Acceptor Stem." Cell **104**(2): 191-193.

Ribas de Pouplana L. s. and Schimmel P. (2001). "Aminoacyl-tRNA synthetases: potential markers of genetic code development." Trends in Biochemical Sciences **26**(10): 591-596.

Roback P., Beard J., Baumann D., Gille C., Henry K., Krohn S., Wiste H., Voskuil M. I., Rainville C. and Rutherford R. (2007). "A predicted operon map for Mycobacterium tuberculosis." Nucleic Acids Research **35**(15): 5085-5095.

Rodriguez P. L. and Carrasco L. (1992). "Gliotoxin: inhibitor of poliovirus RNA synthesis that blocks the viral RNA polymerase 3Dpol." Journal of Virology **66**(4): 1971-1976.

Rohrer S., Ehlert K., Tschierske M., Labischinski H. and Berger-Bachi B. (1999). "The essential Staphylococcus aureus gene fmhB is involved in the first step of peptidoglycan pentaglycine interpeptide formation." Proceedings of the National Academy of Sciences of the United States of America **96**(16): 9351-9356.

Ruff M., Krishnaswamy S., Boeglin M., Poterszman A., Mitschler A., Podjarny A., Rees B., Thierry J. C. and Moras D. (1991). "Class II aminoacyl transfert RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp)." Science **252**(5013): 1682-1689.

Saks M. E., Sampson J. R. and Abelson J. (1998). "Evolution of a Transfer RNA Gene Through a Point Mutation in the Anticodon." Science **279**(5357): 1665-1670.

Salmond G. P. C., Bycroft B. W., Stewart G. S. A. B. and Williams P. (1995). "The bacterial 'enigma': cracking the code of cell–cell communication." Molecular Microbiology **16**(4): 615-624.

Samel S. A., Wagner B., Marahiel M. A. and Essen L.-O. (2006). "The Thioesterase Domain of the Fengycin Biosynthesis Cluster: A Structural Base for the Macrocyclization of a Non-ribosomal Lipopeptide." Journal of Molecular Biology **359**(4): 876-889.

Sauguet L., Moutiez M., Li Y., Belin P., Seguin J., Le Du M.-H., Thai R., Masson C., Fonvielle M., Pernodet J.-L., Charbonnier J.-B. and Gondry M. (2011). "Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis." Nucleic Acids Research **39**(10): 4475-4489.

Schmeing T. M. and Ramakrishnan V. (2009). "What recent ribosome structures have revealed about the mechanism of translation." Nature **461**(7268): 1234-1242.

Schultz A. W., Oh D.-C., Carney J. R., Williamson R. T., Udwary D. W., Jensen P. R., Gould S. J., Fenical W. and Moore B. S. (2008). "Biosynthesis and Structures of Cyclomarins and Cyclomarazines, Prenylated Cyclic Peptides of Marine Actinobacterial Origin." Journal of the American Chemical Society **130**(13): 4507-4516.

Schuwirth B. S., Borovinskaya M. A., Hau C. W., Zhang W., Vila-Sanjurjo A., Holton J. M. and Cate J. H. D. (2005). "Structures of the bacterial ribosome at 3.5 angstrom resolution." Science **310**(5749): 827-834.

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

Schwarzer D., Mootz H. D. and Marahiel M. A. (2001). "Exploring the impact of different thioesterase domains for the design of hybrid peptide synthetases." Chemistry & Biology **8**(10): 997-1010.

Seguin J., Moutiez M., Li Y., Belin P., Lecoq A., Fonvielle M., Charbonnier J.-B., Pernodet J.-L. and Gondry M. (2011). "Nonribosomal Peptide Synthesis in Animals: The Cyclodipeptide Synthase of Nematostella." Chemistry & Biology **18**(11): 1362-1368.

Selmer M., Dunham C. M., Murphy F. V., Weixlbaumer A., Petry S., Kelley A. C., Weir J. R. and Ramakrishnan V. (2006). "Structure of the 70S ribosome complexed with mRNA and tRNA." Science **313**(5795): 1935-1942.

Sieber S. A. and Marahiel M. A. (2005). "Molecular Mechanisms Underlying Nonribosomal Peptide Synthesis: Approaches to New Antibiotics." Chemical Reviews **105**(2): 715-738.

Sievers A., Beringer M., Rodnina M. V. and Wolfenden R. (2004). "The ribosome as an entropy trap." Proceedings of the National Academy of Sciences of the United States of America **101**(21): 7897-7901.

Sipiczki M. (2006). "Metschnikowia Strains Isolated from Botrytized Grapes Antagonize Fungal and Bacterial Growth by Iron Depletion." Applied and Environmental Microbiology **72**(10): 6716-6724.

Sissler M., Delorme C., Bond J., Ehrlich S. D., Renault P. and Francklyn C. (1999). "An aminoacyl-tRNA synthetase paralog with a catalytic role in histidine biosynthesis." Proceedings of the National Academy of Sciences of the United States of America **96**(16): 8985-8990.

Spikes S., Xu R., Nguyen C. K., Chamilos G., Kontoyiannis D. P., Jacobson R. H., Ejzykowicz D. E., Chiang L. Y., Filler S. G. and May G. S. (2008). "Gliotoxin Production in Aspergillus fumigatus Contributes to Host-Specific Differences in Virulence." Journal of Infectious Diseases **197**(3): 479-486.

Stachelhaus T., Mootz H. D., Bergendahl V. and Marahiel M. A. (1998). "Peptide Bond Formation in Nonribosomal Peptide Biosynthesis." Journal of Biological Chemistry **273**(35): 22773-22781.

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

Stachelhaus T. and Walsh C. T. (2000). "Mutational Analysis of the Epimerization Domain in the Initiation Module PheATE of Gramicidin S Synthetase." Biochemistry **39**(19): 5775-5787.

Steitz T. A. and Shulman R. G. (1982). "CRYSTALLOGRAPHIC AND NMR-STUDIES OF THE SERINE PROTEASES." Annual Review of Biophysics and Bioengineering **11**: 419-444.

Strieker M., Tanović A. and Marahiel M. A. (2010). "Nonribosomal peptide synthetases: structures and dynamics." Current Opinion in Structural Biology **20**(2): 234-240.

Ström K., Sjögren J., Broberg A. and Schnürer J. (2002). "Lactobacillus plantarum MiLAB 393 Produces the Antifungal Cyclic Dipeptides Cyclo(l-Phe-l-Pro) and Cyclo(l-Phe-trans-4-OH-l-Pro) and 3-Phenyllactic Acid." Applied and Environmental Microbiology **68**(9): 4322-4327.

Strom M. S. and Paranjpye R. N. (2000). "Epidemiology and pathogenesis of Vibrio vulnificus." Microbes and Infection **2**(2): 177-188.

Sun L., Campbell F. E., Zahler N. H. and Harris M. E. (2006). "Evidence that substrate-specific effects of C5 protein lead to uniformity in binding and catalysis by RNase P." The EMBO Journal **25**(17): 3998-4007.

Suto K., Shimizu Y., Watanabe K., Ueda T., Fukai S., Nureki O. and Tomita K. (2006). "Crystal structures of leucyl/phenylalanyl-tRNA-protein transferase and its complex with an aminoacyl-tRNA analog." The EMBO Journal **25**(24): 5942-5950.

Tang M. R., Sternberg D., Behr R. K., Sloma A. and Berka R. M. (2006). "Use of transcriptional profiling & bioinformatics to solve production problems: Eliminating red pigment production in a Bacillus subtilis strain producing hyaluronic acid " Industrial Biotechnology **2**(1): 66-74.

Tanovic A., Samel S. A., Essen L.-O. and Marahiel M. A. (2008). "Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthetase." Science **321**(5889): 659-663.

Teixidó M., Zurita E., Malakoutikhah M., Tarragó T. and Giralt E. (2007). "Diketopiperazines as a Tool for the Study of Transport across the Blood−Brain Barrier (BBB) and Their Potential Use as BBB-Shuttles." Journal of the American Chemical Society **129**(38): 11802-11813.

Tobias J., Shrader T., Rocap G. and Varshavsky A. (1991). "The N-end rule in bacteria." Science **254**(5036): 1374-1377.

Ton-That H., Labischinski H., Berger-Bachi B. and Schneewind O. (1998). "Anchor structure of staphylococcal surface proteins III. Role of the femA, femB, and femX factors in anchoring surface proteins to the bacterial cell wall." Journal of Biological Chemistry **273**(44): 29143-29149.

Torres-Larios A., Swinger K. K., Krasilnikov A. S., Pan T. and Mondragon A. (2005). "Crystal structure of the RNA component of bacterial ribonuclease P." Nature **437**(7058): 584-587.

Tran L., van Baarsel J. A., Washburn R. S., Gottesman M. E. and Miller J. H. (2011). "Single-Gene Deletion Mutants of Escherichia coli with Altered Sensitivity to Bicyclomycin, an Inhibitor of Transcription Termination Factor Rho." Journal of Bacteriology **193**(9): 2229-2235.

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

Vetting M. W., Hegde S. S. and Blanchard J. S. (2010). "The structure and mechanism of the Mycobacterium tuberculosis cyclodityrosine synthetase." Nature Chemical Biology **6**(11): 797-799.

Villet R., Fonvielle M., Busca P., Chemama M., Maillard A. P., Hugonnet J.-E., Dubost L., Marie A., Josseaume N., Mesnage S., Mayer C., Valéry J.-M., Ethève-Quelquejeu M. and Arthur M. (2007). "Idiosyncratic features in tRNAs participating in bacterial cell wall synthesis." Nucleic Acids Research **35**(20): 6870-6883.

Vogel A., Schilling O., Späth B. and Marchfelder A. (2005). "The tRNase Z family of proteins: physiological functions, substrate specificity and structural properties." Biological Chemistry **386**(12): 1253-1264.

Vollmer W., Blanot D. and De Pedro M. A. (2008). "Peptidoglycan structure and architecture." FEMS Microbiology Reviews **32**(2): 149-167.

von Dohren H. (2009). "Charged tRNAs charge into secondary metabolism." Nature Chemical Biology **5**(6): 374-375.

Walsh C. T. (2007). "The Chemical Versatility of Natural-Product Assembly Lines." Accounts of Chemical Research **41**(1): 4-10.

Waring P. and Beaver J. (1996). "Gliotoxin and related epipolythiodioxopiperazines." General Pharmacology: The Vascular System **27**(8): 1311-1316.

Watanabe K., Toh Y., Suto K., Shimizu Y., Oka N., Wada T. and Tomita K. (2007). "Protein-based peptide-bond formation by aminoacyl-tRNA protein transferase." Nature **449**(7164): 867-871.

Weber T., Baumgartner R., Renner C., Marahiel M. A. and Holak T. A. (2000). "Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases." Structure **8**(4): 407-418.

Weber T., Rausch C., Lopez P., Hoof I., Gaykova V., Huson D. H. and Wohlleben W. (2009). "CLUSEAN: A computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters." Journal of Biotechnology **140**(1–2): 13-17.

Weinger J. S., Parnell K. M., Dorner S., Green R. and Strobel S. A. (2004). "Substrate-assisted catalysis of peptide bond formation by the ribosome." Nature Structural & Molecular Biology **11**(11): 1101-1106.

Wenzel S. C. and Müller R. (2005). "Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic." Current Opinion in Chemical Biology **9**(5): 447-458.

Wilkinson B. and Micklefield J. (2007). "Mining and engineering natural-product biosynthetic pathways." Nature Chemical Biology **3**(7): 379-386.

Williams D. E., Bombuwala K., Lobkovsky E., de Silva E. D., Karunaratne V., Allen T. M., Clardy J. and Andersen R. J. (1999). "ChemInform Abstract: Ambewelamides A and B, Antineoplastic Epidithiapiperazinediones Isolated from the Lichen Usnea sp." ChemInform **30**(12): no-no.

Winn M., Goss R. J. M., Kimura K.-i. and Bugg T. D. H. (2010). "Antimicrobial nucleoside antibiotics targeting cell wall assembly: Recent advances in structure-function studies and nucleoside biosynthesis." Natural Product Reports **27**(2): 279-304.

Woese C. R., Olsen G. J., Ibba M. and Söll D. (2000). "Aminoacyl-tRNA Synthetases, the Genetic Code, and the Evolutionary Process." Microbiology and Molecular Biology Reviews **64**(1): 202-236.

Yamaguchi H., Kato H., Hata Y., Nishioka T., Kimura A., Oda J. i. and Katsube Y. (1993). "Three-dimensional Structure of the Glutathione Synthetase from Escherichia coli B at 2·0 Å Resolution." Journal of Molecular Biology **229**(4): 1083-1100.

Yang E. J. and Chang H. C. (2010). "Purification of a new antifungal compound produced by Lactobacillus plantarum AF1 isolated from kimchi." International Journal of Food Microbiology **139**(1–2): 56-63.

Yonus H., Neumann P., Zimmermann S., May J. J., Marahiel M. A. and Stubbs M. T. (2008). "Crystal Structure of DltA." Journal of Biological Chemistry **283**(47): 32484-32491.

Yusupov M. M., Yusupova G. Z., Baucom A., Lieberman K., Earnest T. N., Cate J. H. D. and Noller H. F. (2001). "Crystal structure of the ribosome at 5.5 angstrom resolution." Science **292**(5518): 883-896.

Zhang W., Ntai I., Bolla M. L., Malcolmson S. J., Kahne D., Kelleher N. L. and Walsh C. T. (2011). "Nine Enzymes Are Required for Assembly of the Pacidamycin Group of Peptidyl Nucleoside Antibiotics." Journal of the American Chemical Society **133**(14): 5240-5243.

Zhang W., Ntai I., Kelleher N. L. and Walsh C. T. (2011). "tRNA-dependent peptide bond formation by the transferase PacB in biosynthesis of the pacidamycin group of pentapeptidyl nucleoside antibiotics." Proceedings of the National Academy of Sciences of the United States of America **108**(30): 12249-12253.

Zhao S., Smith K. S., Deveau A. M., Dieckhaus C. M., Johnson M. A., Macdonald T. L. and Cook J. M. (2002). "Biological activity of the tryprostatins and their diastereomers on human carcinoma cell lines." Journal of Medicinal Chemistry **45**(8): 1559-1562.

Zwiefka A., Kohn H. and Widger W. R. (1993). "Transcription termination factor rho: The site of bicyclomycin inhibition in Escherichia coli." Biochemistry **32**(14): 3564-3570.