



HAL
open science

New insights into the recognition of the substrates of cyclodipeptide synthases

Nicolas Canu

► **To cite this version:**

Nicolas Canu. New insights into the recognition of the substrates of cyclodipeptide synthases. Biochemistry, Molecular Biology. Université Paris Saclay (COMUE), 2019. English. NNT: 2019SACLS580 . tel-03506235

HAL Id: tel-03506235

<https://theses.hal.science/tel-03506235>

Submitted on 2 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

New insights into the recognition of the substrates of cyclodipeptide synthases

Thèse de doctorat de l'Université Paris-Saclay préparée à
l'Université Paris-Sud

École doctorale n°569 Innovation thérapeutique :
du fondamental à l'appliqué (ITFA)
Spécialité de doctorat: biochimie et biologie structurale

Thèse présentée et soutenue à Gif-sur-Yvette le 12 décembre 2019, par

Nicolas CANU

Composition du Jury :

Pr. Stéphanie Bury-Moné Professeure, Université Paris-Sud	Présidente
Dr. Magali Frugier Directrice de Recherche, Université de Strasbourg	Rapporteuse
Dr. Yan-Yan Li Chargée de Recherche, Muséum National d'Histoire Naturelle	Rapporteuse
Dr. Gilles Truan Directeur de Recherche, INSA de Toulouse	Examineur
Dr. Axel Innis Directeur de Recherche, Institut Européen de Chimie et de Biologie	Examineur
Dr. Muriel Gondry Directrice de Recherche, CEA	Directrice de thèse

*A mon papa, pour m'avoir appris par l'exemple
l'abnégation et l'amour du travail bien fait,*

A mon cousin Clément,

REMERCIEMENTS

Avant tout, merci aux membres de mon jury d'avoir accepté de prendre le temps de lire et d'évaluer ce travail de thèse.

Merci au CEA (et donc aux contribuables français) pour avoir financé ma thèse et croire (encore un peu) que la recherche scientifique vaut la peine qu'on y consacre de l'argent et de l'énergie.

Merci à toi Muriel pour m'avoir accueilli dans ton équipe et m'avoir donné l'opportunité de jouer avec les CDPS pendant ces trois belles années. La bienveillance et l'optimisme qui émanent de toi ont beaucoup joué dans ma décision de faire une thèse et *a fortiori* de la faire avec vous. Ta rigueur et ton exigence m'ont amené à me dépasser et je suis très fier d'avoir contribué à écrire une petite page de la belle aventure scientifique que tu mènes.

Merci Carine pour ton aide et ta gentillesse. Tu sais tout le respect et l'admiration que j'ai pour toi. Merci Jérôme pour les bons moments qu'on a passé ensemble comme voisins à tous les sens du terme. Merci Pascal pour ta bonne humeur permanente, ta rigueur et ton aide dans les moments un peu durs. Merci à Morgan et à Inès pour votre aide sur mon projet et à Fabien pour ta gentillesse et tes encouragements.

Merci Bob pour ta bonne humeur, ta gentillesse, ton professionnalisme et ta disponibilité. C'était vraiment un plaisir de bosser avec toi.

Un grand merci à tous les gens que j'ai croisé de près ou de loin pendant ces trois années au CEA. En particulier à tous les thésards/post-docs/permanents du 144 pour les bons moments passés ensemble, les footings du midi, les raclettes de Noël, aux collègues du SIMOPRO pour tous les déjeuners avec eux. Un immense merci à Coline et Raphael, c'était top de partager cette aventure avec vous deux.

Merci à toi Jean-Christophe pour m'avoir ouvert les portes de ton labo et m'avoir donné les bases du ~~parfait~~ petit chimiste. Je n'étais sans doute pas le plus doué des organiciens mais j'ai quand même pris un grand plaisir à travailler avec vous. Merci à Robin, Mathilde, Florian, Karine, Émilie, David pour m'avoir guidé dans ce monde un peu nouveau pour moi, et à Olivier Cinquin pour son aide dans la synthèse des cyclodipeptides.

Merci à Olivier Lequin et Isabelle Correia pour leur réactivité et leur pédagogie quand il a fallu analyser des composés dans une certaine précipitation.

Merci à Magali Frugier, Joëlle Rudinger-Thirion et Matthieu Fonvielle pour m'avoir fait profiter de leur expérience du monde parfois un peu déroutant des ARN.

Thanks to Hiroaki Suga for his amazing work on the development of flexizymes.

Many thanks to Taek-Soon Lee and Aram Kang for accepting me in Berkeley, introducing me to experimental research and to the unique feeling of discovering something, even when it's a very small something.

Enfin sur une note plus personnelle, merci à mes amis de prépa, quel plaisir de constater que malgré les années on se retrouve toujours sur l'essentiel. Devenir un grand à vos côtés a été une belle aventure, pourvu que ça dure ! Merci à mes amis de 187 qui sont devenus au fil des années une deuxième famille et aux Biotech pour les bons moments passés ensemble.

Merci à Jean-Marc Jancovici, Gaël Giraud, Pablo Servigne, Philippe Bihouix, Arthur Keller pour m'avoir accompagné durant mes longs trajets en navette, avoir aiguisé ma curiosité, stimulé mon esprit critique et amélioré un peu ma compréhension du monde.

Merci à mes frères et sœurs pour avoir été des exemples de courage et d'amour pour le petit dernier que je suis. Merci à mes grands-parents pour les bons moments passés chez eux. Merci à mes parents pour nous avoir tant apporté et nous avoir permis de nous réaliser. Merci Maman de ta présence à nos côtés toutes ces années et pour nous avoir transmis ton sens de l'accueil et du carpe diem. Merci Papa de t'être tant donné pour que l'on ne manque de rien, nous avoir transmis ton amour de la musique et de la montagne.

Enfin merci à toi ma Justine d'être la personne extraordinaire, intelligente, altruiste et généreuse que tu es, de me supporter dans tous les sens du terme et d'avoir accepté de me suivre à Paris. J'ai hâte de partager la suite des aventures avec toi et de découvrir ce que la vie nous réserve.

ABSTRACT

Cyclodipeptides and their complex derivatives, the diketopiperazines (DKPs), constitute a large class of natural products mostly synthesized by microorganisms. Owing to the diverse and noteworthy pharmacological activities observed for many naturally occurring DKPs (essentially antibiotic, anticancer or antiviral activities), the cyclodipeptide-ring has long been recognized as a privileged scaffold for the generation of therapeutic compounds. Since various chemical approaches have led to DKPs with improved pharmacological activities, a promising and complementary approach is the use and the manipulation of diketopiperazine biosynthetic pathways. Two types of pathways have been described, which respectively use gigantic nonribosomal peptide synthetases or cyclodipeptide synthetases (CDPSs). CDPSs are of particular interest: these small enzymes catalyse the formation of cyclodipeptides by using aminoacyl-tRNAs (AA-tRNAs) as substrates, which they hijack from their canonical function in ribosomal protein synthesis, and they are often associated with discrete cyclodipeptide-tailoring enzymes. CDPSs have been shown to use almost all proteinogenic amino acids to generate an important diversity of cyclodipeptides and they constitute an interesting tool for the biological synthesis of cyclodipeptides. However, in order to unlock the biosynthetic potential of these enzymes, better understanding their specificity, in particular towards non-natural substrates, is required.

In my doctoral work, we first introduce a new method for the functional annotation of CDPSs, using Sequence Similarity Networks. We showed that the activity of about 80% of putative CDPSs can be predicted using this high-throughput automated method. Then, we significantly expanded the diversity of cyclodipeptides accessible with CDPSs by showing that CDPSs could incorporate non-canonical amino acids (ncAAs). We took advantage of the promiscuity of *E. coli* aminoacyl-tRNA synthetases (the enzymes responsible for amino acid loading on tRNAs) to study the promiscuity of CDPSs towards ncAAs *in vivo*, using strains of *E. coli* auxotrophic for proteinogenic amino acids. 26 ncAAs were incorporated and about 200 ncAA-containing cyclodipeptides are produced. Finally, we gave new insights into the recognition by CDPSs of the tRNA moieties of their substrates. We introduced an *in vitro* enzymatic assay that allows to study separately the recognition of the two different substrates of a CDPS and to estimate kinetic parameters for these complex, bi-substrates enzymes. By using an innovative RNA acylation strategy based on a class of ribozymes called flexizymes, we generated analogues of AA-tRNAs with truncated RNA moieties. Among these "AA-minitRNAs", we showed that those mimicking the entire 7 bp stems of tRNAs are as good substrates as AA-tRNAs, which suggests that CDPSs interact mainly with the acceptor arms of tRNAs and paves the way for promising biophysical and structural studies.

TABLE OF CONTENTS

REMERCIEMENTS	5
TABLE OF CONTENTS.....	9
INTRODUCTION	11
I) 2,5-Diketopiperazines and their biosynthetic pathways	15
A) 2,5-diketopiperazines: a large class of noteworthy natural products	15
B) 2,5-diketopiperazines biosynthesis.....	19
i. 2,5-diketopiperazines formation by Nonribosomal Peptide Synthetases.....	19
ii. Cyclodipeptide synthases: an original biosynthetic route dedicated to the production of 2,5-diketopiperazines.....	24
Article published in Natural Product Research	26
II) Cyclodipeptide synthase substrates: aminoacyl-tRNAs.....	36
A) Aminoacyl-tRNAs and their cognate aminoacyl-tRNA synthetases.....	37
i. Aminoacyl-tRNA molecules.....	37
ii. Aminoacyl-tRNA synthetases and their specificity	39
B) Different methods to produce non-canonical AA-tRNAs.....	42
i. Exploiting the natural promiscuity of aminoacyl-tRNA synthetases	42
ii. Engineering of aminoacyl-tRNA synthetases with improved promiscuity	44
iii. Chemical ligation of amino acids to tRNAs	47
iv. Versatile RNA acylation using aminoacyl-tRNA synthetase-like ribozymes, the flexizymes.....	48
III) Objectives of this thesis	53
CHAPTER 1: HIJACKING THE HIJACKERS: INCORPORATION OF NON-CANONICAL AMINO ACIDS INTO 2,5-DIKETOPIPERAZINES BY CYCLODIPEPTIDE SYNTHASES.....	54
I) Introduction.....	54
II) Article	56
Article published in Angewandte Chemie International Edition	57
III) Supplementary information	62
IV) Discussion and perspectives.....	123

CHAPTER 2: TOWARDS A BETTER UNDERSTANDING OF RNA RECOGNITION BY CYCLODIPEPTIDE SYNTHASES	127
I) Introduction.....	127
II) Article	129
Manuscript to be submitted to Nucleic Acid Research	130
III) Supplementary Information	153
IV) Discussion and perspectives.....	190
GENERAL DISCUSSION AND PERSPECTIVES.....	192
Bibliography	194
RESUME EN FRANCAIS	207

Note that each article inserted in the global manuscript has its own reference section.

INTRODUCTION

Exploiting the activities of chemicals produced by living beings has been the basis of all traditional medicines, with traces of natural products in the pharmacopoeias of every ancient civilization. Exploiting these natural products in modern medicine has been one of the greatest achievements of humanity: drugs based on natural products have contributed to save millions of lives in the 20th century. Because of their chemical richness and their ease of manipulation, micro-organisms have been widely exploited and, in particular, they have yielded most of the antibiotics that constitute our weapons towards infectious diseases¹. However, after the tremendous successes obtained during the “Golden Age” of natural product research in the 1940s to 1960s, pharmaceutical companies have largely decreased their research efforts in natural products since the 1980s. This is partly due to the rise of alternative strategies such as combinatorial chemistry or macromolecules-based “biologics” but also to a decrease in the discovery of new chemical scaffold from natural product programs. However, the screening of synthetic chemical libraries produced by combinatorial chemistry brought only limited results, which suggests that the intrinsic chemical complexity of natural products should continue to be exploited in our search for new drugs².

The slowing pace at which new scaffolds emerge from natural product research programs might be explained by the fact that conventional methods, based on the culture of isolated microorganisms, extraction and activity-based assays, only give access to a very small proportion of the actual chemical diversity produced by micro-organisms. One explication is that only a fraction of micro-organisms is cultivable using standard laboratory procedures. The laborious optimization of cultivation conditions for every strain and the difficulty to mimic complex environments (for example for symbiotic organisms³) is not compatible with the tremendous diversity of micro-organism species and of their ecological niches.

After the limits of traditional methods for natural product discovery became obvious, technological advances in the last 20 years have opened up new opportunities and have contributed to a revival of natural product research⁴.

The development of innovative strategies of micro-organism cultivation, which aims at better mimicking natural ecosystem, continues to enable the fermentation of previously uncultivated

species and to lead to the discovery of new natural products. An exciting example was recently given by the discovery of the promising antibacterial teixobactin⁵, which was possible thanks to a microfluidic cultivation device where bacteria are cultivated in their original soil sample. However, these methods remain empirical and often suffer from limited throughputs.

The major revolution in the field has come from the exponential increase in sequencing capacities. Genomes of micro-organisms are now acquired on a daily basis and databases currently contain tens of thousands of microbial genomes or metagenomes. In parallel of this surge in the amount of genetic data available, our understanding of the biosynthetic logic of the main classes of natural products and of their biosynthetic enzymes has improved^{6–10}. This has enabled the development of genome mining, *i.e.* the automated recognition of the genes involved in the biosynthesis of natural products¹¹. Interestingly, in microbial genomes, the different genes coding for a natural product pathway are in most cases clustered, which facilitates their study. The number of biosynthetic gene clusters found in the genomes of natural product producers is often way above the number of natural products isolated from these organisms using conventional methods. This suggests that most of the biosynthetic pathways (about 90%) encoded within microbial genomes are cryptic, *i.e.* not expressed under standard cultivation conditions^{1,12}.

Advances in molecular biology and continuous progress in synthetic biology now enables the community to exploit the booming number of putative biosynthetic gene clusters. By “awakening” cryptic pathways, biotechnological approaches can enable the discovery of new chemical scaffolds. Additionally, if the yields obtained through biotechnological production are high enough, such approaches can be an interesting eco-friendly alternative to chemical synthesis for the industrial production of natural products.

Several biotechnological approaches can be pursued. A first approach is to modify the gene expression of a biosynthetic gene cluster in the original host³, either by triggering a difference at the genome-wide level¹³ or by activating one particular cluster by genetic engineering¹⁴. Another more radical approach is the transfer by heterologous expression of a biosynthetic gene cluster into a more tractable host¹⁵. This method has its limitations and in particular the host of expression has to be chosen very carefully depending on the nature of the pathways. However, it has proven to be a useful tool to assay large numbers of biosynthetic gene clusters in a high-throughput fashion¹⁶ and to produce natural products at high yield¹⁷. Typical synthetic biology

strategies can be applied to increase the yield of products: engineering of the host, of the pathways and/or of the enzymes themselves^{18,19}.

Our understanding of natural product biosynthesis can also be used to manipulate the pathways, in order to produce “unnatural” natural products, analogues whose properties are improved. In combinatorial biosynthesis experiments, enzymatic cascades are modified by withdrawing or adding enzymes from other pathways, leading to modified products²⁰. Unnatural precursors or substrates can also be introduced, in so-called “mutasynthesis” studies¹. In some cases, the engineering of enzymes can be required, in order to modify or extend their specificity.

Biosynthesis-based approaches require a good knowledge of the enzymes involved in biosynthetic pathways, in order to efficiently mine the genomes for putative interesting clusters, to choose and engineer the host of expression and to guide the manipulation of the pathways. Therefore, thorough characterization of biosynthetic enzymes appears to be a key prerequisite before any attempt of manipulation of these pathways.

This thesis falls within the framework of better understanding biosynthetic enzymes for the production of new natural products. My host group is interested in the biosynthesis of 2,5-diketopiperazines (2,5-DKPs), a large family of natural products that share a common cyclodipeptide scaffold²¹. One of the biosynthetic routes to these molecules was discovered about 15 years ago by this group²². It is composed of small intriguing enzymes dedicated to the synthesis of cyclodipeptides, called cyclodipeptide synthases (CDPSs)²³. CDPSs are often clustered with 2,5-DKP-tailoring enzymes which modify the cyclodipeptides and lead to products with interesting bioactivities^{24–26}. The originality of CDPSs is the unusual nature of their substrates, aminoacyl-transfer RNAs (AA-tRNAs), which makes their study and their manipulation challenging but fascinating. My work has been focused on better understanding how CDPSs recognize their substrates, in order to enable their manipulation and the biotechnological production of new 2,5-DKPs. Given the complexity of generating AA-tRNAs, methods inspired from the field of synthetic biology have been used and transposed to the study of CDPSs.

The introduction of this manuscript is composed of a brief presentation of 2,5-DKP structures and bioactivities and a description of their biosynthesis, with a focus on CDPSs and CDPS-dependent pathways. Then, CDPS substrates, AA-tRNAs, are presented, and the different methods to generate them are described.

I) 2,5-Diketopiperazines and their biosynthetic pathways

A) 2,5-diketopiperazines: a large class of noteworthy natural products

2,5-DKPs are cyclic organic compounds that share a common scaffold, composed of a piperazine core bearing two ketone groups²¹ (Figure 1). Since this scaffold can be formed by the condensation of two α -amino acids, simple 2,5-DKPs are also referred as cyclodipeptides. By convention, in this manuscript, cyclodipeptides composed of two L- amino acids are referred as cyclo(Xxx-Xxx), Xxx being the three-letters short form of the amino acids.

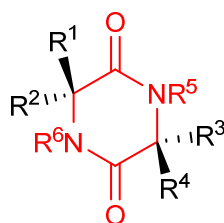


Figure 1: Standard formula of a 2,5-DKP. The 2,5-DKP scaffold is coloured in red.

The 2,5-DKP scaffold confers several interesting features from a medicinal chemistry point of view. The cyclic dipeptide backbone can mimic a peptide conformation without some of the undesirable features of linear peptides (low solubility, low membrane permeability and poor *in vivo* stability due to protease degradation). The cycle can be substituted at six different locations (Figure 1), of which four can be stereospecifically controlled, which explains the tremendous chemical diversity arising from this simple scaffold. These features have sparked the interest of medicinal chemists, which have proposed many synthetic ways to produce and further modify these molecules and of researchers involved in natural product discovery, who reported a tremendous diversity of 2,5-DKP-based molecules and investigated the biosynthesis of these compounds.

Note that several recent comprehensive reviews dedicated to 2,5-DKPs are published on both their chemical properties (conformation, reactivity, syntheses, medicinal chemistry)^{21,27} and their attractivity as diverse and bioactive natural products^{28–32}.

2,5-DKPs were primarily thought to be only by-products of peptide degradation, found in animal fluids³³ or in thermally treated or fermented food and beverages^{30,34}, with small interest from a biological point of view. This changed radically in the 1970s, when researchers involved in natural products research reported the discovery of many complex 2,5-DKPs found to be produced mainly by micro-organisms, bacteria and filamentous fungi and, in a few cases, in marine archaeon³⁵, marine animals³¹ and mammals³³ (Figure 2). Whatever their origin, a lot of natural 2,5-DKPs bear modifications when compared to the simple cyclodipeptide to which they are related. The range of modifications is broad and spans from α,β -dehydrogenations (as in albonoursin, halimide, okaramines, purincyclamide), *N*- or *C*-methylations (cyclomarazines, thaxtomins, nocardioazines, purincyclamide), hydroxylations (thaxtomins, sirodesmin PL), introduction of disulfide bridges (sirodesmin PL), addition of prenylated moieties (cyclomarazines, brevianamides, halimide, drimentines) or nucleobase (purincyclamide) or intramolecular cyclization (okaramines, sirodesmin PL, mycocyclosin, bicyclomycin, drimentines, nocardioazines) (Figure 2).

2,5-DKPs are specialized (secondary) metabolites whose physiological roles for the producing hosts have been investigated but remain often elusive. Several studies have suggested that 2,5-DKPs act as small diffusible molecules involved in cell-to-cell communication, essentially for coordination of community behaviour or inhibition of competing species. In bacteria, they may constitute a specific class of quorum-sensing (QS) signals^{36–38}, or even interspecies signals^{39,40}. Furthermore, since 2,5-DKPs have bioactive effects on their plant or animal hosts, a role in transkingdom signalling has also been suggested^{41,42}.

2,5-DKPs act through other processes. For example, pulcherriminic acid produced by *Bacillus subtilis* is secreted into the extracellular environment where it chelates Fe³⁺ (see pulcherrimin in Figure 2B) from the growth medium through a nonenzymatic reaction, which locally depletes the iron concentration and protects its producer from colonization by neighbouring bacteria⁴³. The diverse thaxtomins (Figure 2A), produced by several *Streptomyces* species, function as cellulose biosynthesis inhibitors, although the specific cellular target of these phytotoxins remains unknown⁴⁴. They are key pathogenicity determinants which cause the common scab disease of potato, with adverse economic consequences.

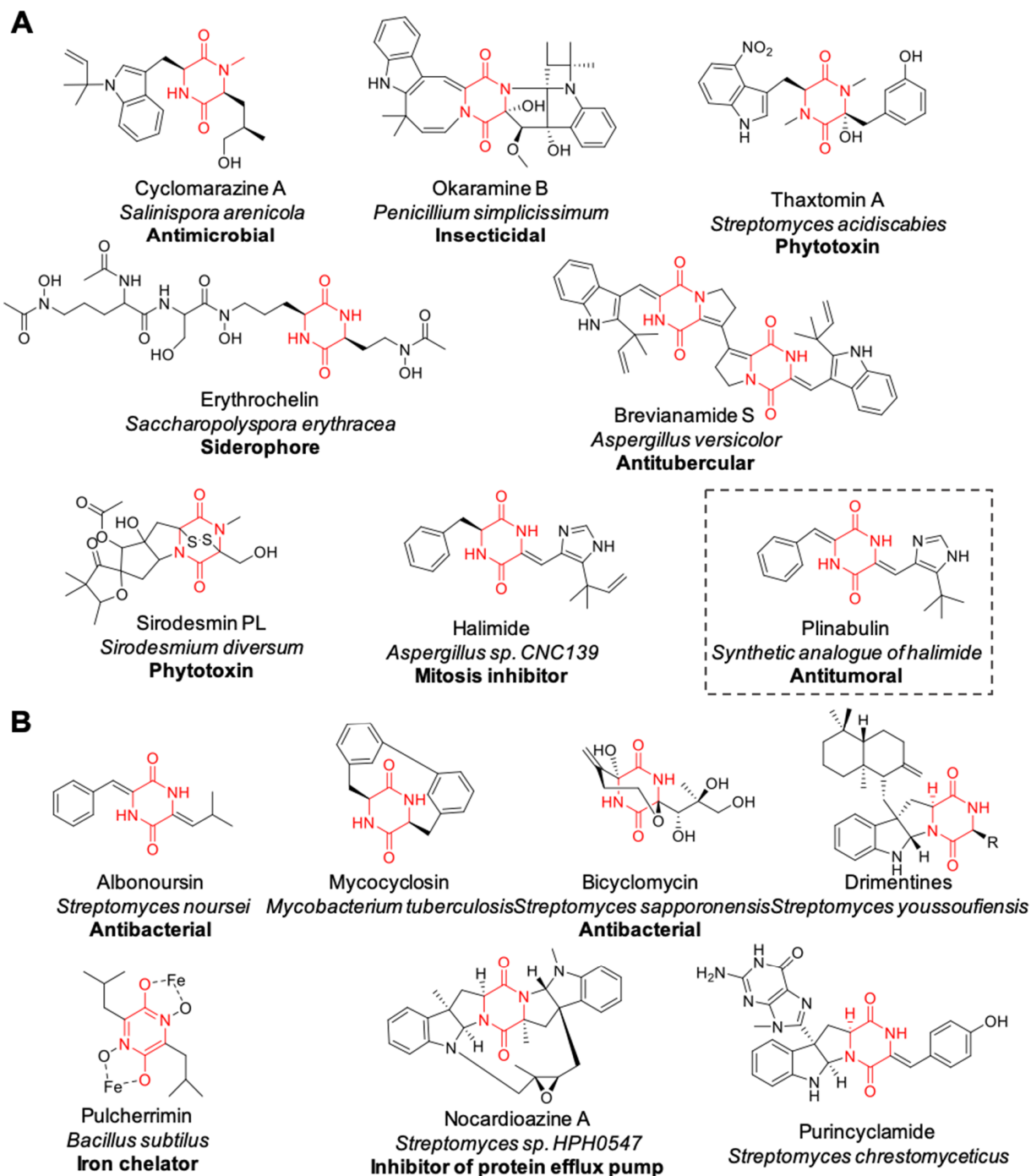


Figure 2: Diversity of 2,5-DKPs. The 2,5-DKP scaffold is coloured in red. The species from which the compound was isolated are in italic type (except for plinabulin which is artificial and is placed in a dotted frame). Demonstrated pharmacological effects are in bold type. Compounds in A) are biosynthesised by NRPS-dependent pathways, whereas compounds in B) are biosynthesized by CDPS-dependent pathways (see section I.B. of the introduction).

Despite the poor understanding of their relevance *in vivo*, many 2,5-DKPs have been shown to exhibit interesting biological activities, such as phytotoxic⁴⁵ or insecticidal activities⁴⁶ but mainly pharmacological activities, antibacterial³², antifungal, antitumor and antiviral activities^{21,28} (Figure 2). In the last few years, strategies using 2,5-DKPs as perturbators of QS pathways in bacterial pathogens to suppress their virulence have also emerged. As an example, some molecules were shown to inhibit the opportunistic respiratory pathogen *Burkholderia cenocepacia* QS, rendering the pathogen unable to produce virulence factors and thus less able to colonize its host⁴⁷.

In some cases, optimization of natural bioactive 2,5-DKPs by medicinal chemistry has led to promising molecules. Halimide is a derivative of cyclo(Phe-His) which is produced by an *Aspergillus* fungi and which inhibits mitosis in human cells (Figure 2A)⁴⁸. Extensive structure-activity relationship studies and optimization of halimide led to plinabulin (Figure 2A)^{49,50}. Plinabulin is currently in phase III clinical development in combination with docetaxel for the treatment of lung cancer and the only candidate drug from a marine fungi to reach clinical trial to date⁵¹. Another example is bicyclomycin (Figure 2B) also known as bicozamycin, which was first discovered to be produced by *Streptomyces sapporonensis*⁵² and which has the ability to disrupt the activity of the bacterial termination factor Rho⁵³. Its mechanism of action together with its activity against clinically relevant Gram-negative bacterial pathogens, like *Escherichia coli*, *Acinetobacter baumannii* and *Klebsiella pneumoniae*, makes it a clinically relevant antibiotic^{54,55}. This is confirmed when it is used in combination with bacteriostatic concentrations of antibiotics inhibiting protein synthesis, which leads to a rapid bactericidal synergy⁵⁵. Furthermore, structure-activity relationship studies showed that the bicyclomycin derivative modified on its exomethylene group has enhanced activities⁵⁶. Such examples highlight the complementarity between natural product discovery and medicinal chemistry in the frame of 2,5-DKPs.

B) 2,5-diketopiperazines biosynthesis

Biosynthesis of 2,5-DKPs can be divided into two steps: the assembly of the cyclodipeptide scaffold by condensation of two amino acids and, in some cases, the modification of the cyclodipeptide by tailoring enzymes. Two biosynthetic routes for cyclodipeptide assembly have been described. One involves gigantic multi-enzymatic complexes, nonribosomal peptide synthetases (NRPSs), which catalyse the formation of an important diversity of peptide natural products. The other involves a family of small enzymes, dedicated to cyclodipeptide formation, named cyclodipeptide synthases (CDPSs). 2,5-DKP synthesis by NRPSs is first briefly presented then a more comprehensive description of CDPSs and CDPS-dependent pathways is presented.

i. 2,5-diketopiperazines formation by Nonribosomal Peptide Synthetases

NRPSs belong to the megasynthase superfamily, together with polyketide synthases (PKSs) and fatty acid synthases (FASs). They are multi-enzymatic complexes responsible for the formation of small peptides (from 2 to 25 amino acids)⁸ and are mainly found in bacteria and fungi. The core architecture of NRPSs is composed of modules, long peptide chains which are themselves composed of several domains, each domain bearing one enzymatic activity (Figure 3). Canonical NRPS machineries are composed of one initiation module, several elongation modules, each responsible for the addition of one amino acid to the nascent peptide chain, and one termination module. A typical NRPS elongation module contains at least three essential domains. An adenylation domain, or A domain, selects a particular amino acid and catalyses its ATP-dependent activation, leading to a high-energy adenylated amino acid intermediate. The activated amino acid is then transferred to the peptidyl carrier protein domain, or PCP domain (also referred as thiolation domain or T domain), through a thioester bond with a 4'-phosphopantetheine (4'PP) cofactor. The tethered amino acid is then shuttled to a condensation domain, or C domain, which catalyses its peptide bond coupling to the nascent peptide chain. Once the full peptide chain is assembled, a thioesterase domain, or TE domain, present in the termination module disconnects the peptide from the NRPS machinery. An intermediate ester bond is formed between the C-terminus of the peptide and a conserved serine of the TE domain. Hydrolysis or intramolecular attack of a nucleophilic moiety yields a linear or macrocyclic product, respectively. The canonical architecture described above is found in most NRPSs but many

exceptions have been reported: replacement of TE domains by other termination domains⁵⁷, stand-alone NRPS domain⁸, reuse of one module several times in a biosynthetic cycle (iterative NRPS)^{58,59}, fusion with PKS (PKS/NRPS hybrid),...

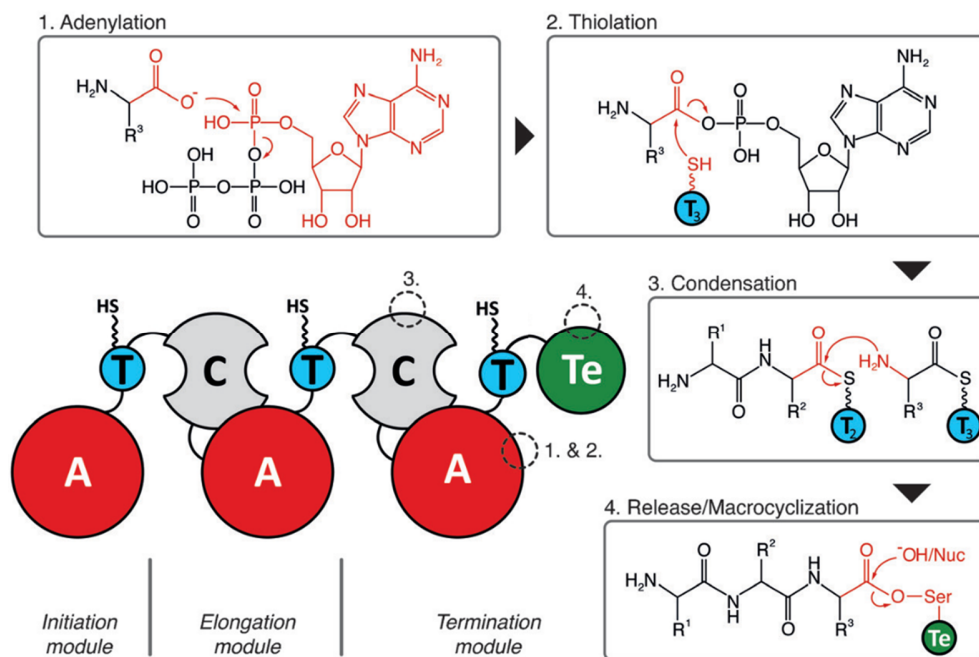


Figure 3: Domain arrangement of canonical bacterial NRPSs and their catalysed reactions. 1) Selection and adenylation of the amino acid by the A domain 2) Subsequent thiolation of the activated amino acid yielding an aminoacyl thioester attached to the 4'PP of the T domain 3) Formation of a peptide bond by the C domain 4) Release of the oligopeptide by the TE domain. Adapted from Süssmuth et Mainz, 2017⁸.

The modular organization of NRPSs leads to gigantic enzymatic complexes. Since the size of one canonical elongation module is about 120kDa (approximately 60kDa for A domain, 10kDa for PCP domain and 50kDa for C domain), the size of a complete NRPS machinery is usually in the range of several hundreds of kDa and can reach up to two millions kDa⁶⁰.

Specificity of NRPSs is notably ensured by the A domains of the different modules, which are responsible for the recognition and the activation of the amino acids. Intensive investigation of the structure/function relationship of adenylation domains have led to the identification of the residues involved in substrate side-chain recognition and to the establishment of a specificity-conferring code for A domains⁶¹. This code can be used for the prediction of the nonribosomal peptides produced and give precious information for the genome mining of NRPSs⁶².

The tremendous chemical diversity produced by NRPS-dependent pathways is due to the possibility to introduce modifications at every step of the biosynthetic pathways. First, unlike

ribosomal synthesis whose precursors are restricted to the 20 proteinogenic amino acids, NRPSs can use more than 500 different monomers, involving non proteinogenic L-amino acids, D-amino acids, β -amino acids⁶³... These non-proteinogenic amino acids are either common metabolic intermediates or they are specifically biosynthesized by dedicated enzymes. Modifications can also be introduced by modifying enzymes. These enzymes can be NRPS modification domains, inserted within the NRPS machinery, which modify the peptides during their elongation, or discrete tailoring enzymes which act after the release of the peptides. The wide range of modifications introduced by modification domains and tailoring enzymes includes epimerization, cyclization, methylation, halogenation, several types of oxidation...⁸ As a result of these different diversification strategies, more than 1100 NRPS-produced nonribosomal peptide scaffolds have been reported in databases and this number goes up to several tens of thousands when all analogues are considered⁶⁴.

Due to their modular organization, NRPSs are attractive candidates for engineering approaches aiming at biosynthesizing new nonribosomal peptides⁶⁵. Many attempts of modules and domains switching have been reported but most of them were unsuccessful. An improved understanding of the complex rules that govern interdomains and intermodules interaction seems required before these approaches can unleash the full biosynthetic potential of NRPSs⁶⁶.

Among the many diverse peptide scaffolds produced by NRPSs, 2,5-DKPs only account for a fraction, since the NRPS-dependent pathways of less than 20 2,5-DKPs have been reported to date^{24,67} (see a few examples in Figure 2).

In a few cases, the formation of 2,5-DKPs by NRPSs results from the early spontaneous intramolecular cyclisation of a linear peptide during its elongation, due to conformational constraints. These truncated side products include cyclo(D-Phe-L-Pro)⁶⁸ and cyclomarazines (Figure 2)⁶⁹. However, most of 2,5-DKPs are synthesized by dedicated NRPSs.

NRPSs dedicated to the formation of 2,5-DKPs are found in bacteria and filamentous fungi. They all lack the TE domain which is necessary in canonical NRPSs to the release of the peptide. Instead, an additional condensation domain is present at the end of the second module. This domain is supposed to be required for the cyclization of the linear dipeptidyl thioester and the release of the cyclodipeptide. Such an architecture is for example found in the bacterial NRPSs of the thaxtomins pathways (Figure 4A). In many 2,5-DKP-producing NRPSs of fungal origin, this extraneous condensation domain is completed by an additional T domain, which seems necessary

to the cyclization catalysis. This particular mechanism was recently elucidated in the context of the biosynthesis of gliotoxin, a toxic epipolythiodiketopiperazine produced by filamentous fungi and derived from a cyclo(Phe-Ser) scaffold⁷⁰ (Figure 4B). After formation of a Phe-Ser dipeptide by conventional NRPS-mediated elongation, the additional T₃ domain serves as a tether for dipeptide cyclization by the adjacent C domain. Finally, 2,5-DKP-producing NRPSs with atypical free-standing domains have been reported in the pathways of maremycins, a family of 2,5-DKPs containing the rare S-methyl-L-cysteine nonproteinogenic residue⁷¹. The diketopiperazine core of maremycins is formed by three different peptides, one containing an adenylation domain with a methyl transferase inserted, one containing only a free T domain and a third one containing a C domain followed by an A/T couplet and a domain with unknown function (Figure 4C)⁷². The exact mechanism of this NRPS remains to be established.

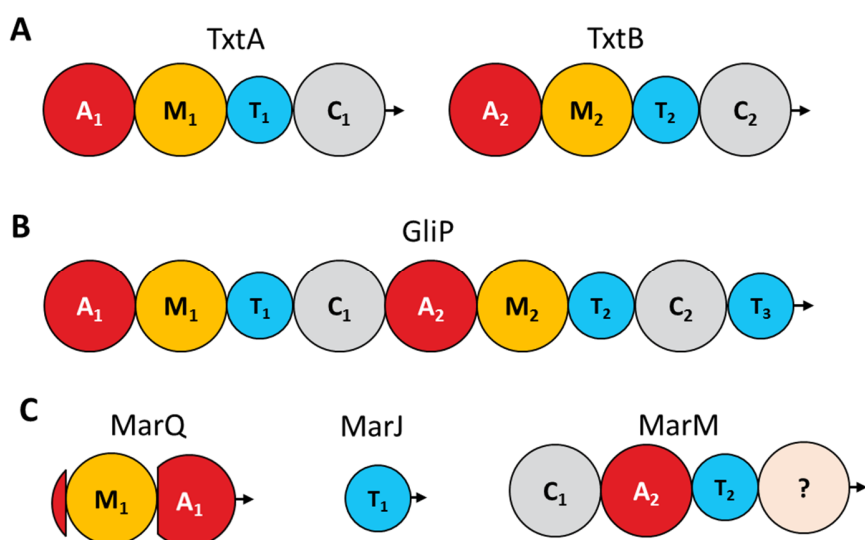


Figure 4: Diversity of architectures of 2,5-DKP-producing NRPSs. A) TxA and TxB from the thaxtomin A pathway. B) GliP from the gliotoxin pathway, containing an additional T₃ domain. C) MarQ, MarJ and MarM, free-standing NRPSs from the maremycin pathway. M corresponds to methylation domains, ? corresponds to a domain with unknown function.

As for other NRPS-dependent pathways, biosynthetic gene clusters containing 2,5-DKP-producing NRPSs often encode other enzymes. These enzymes can be responsible for the formation of non-proteinogenic amino acids used by NRPSs⁸. They can also be 2,5-DKP tailoring enzymes, which act *a posteriori* on the cyclodipeptides to further modify it⁶⁵. The range

of modifications introduced by these tailoring enzymes include indole prenylation, acetylation, hydroxylation and methylation.

An example of how modifications are introduced throughout a 2,5-DKP biosynthetic pathway is given by the pathway of thaxtomin A, a phytotoxin based on a cyclo(L-Phe-L-4-NitroTrp) scaffold and produced by several *Streptomyces* species⁴⁵ (Figure 5). The formation of the non-proteinogenic nitrated tryptophan is catalysed by two proteins whose genes are clustered with the ones of the NRPSs. A nitric oxide synthase TxtD generates nitric oxide NO from L-arginine. This NO is used by the cytochrome P450 TxtE to catalyse the regiospecific nitration of L-Trp at the 4-position⁷³. NRPSs TxtA and TxtB catalyse the condensation and cyclisation of L-4-nitrotryptophan and L-phenylalanine. These two modules comprise one methylation domain (M₁ and M₂ on Figure 5B) which transfer two methyl groups from S-adenosyl-L-methionine (SAM) to the amine of the diketopiperazine core, leading to the release of the dimethylated 2,5-DKP. Finally, a tailoring P450 further modifies the 2,5-DKP by introducing two hydroxylations on the phenylalanine moiety (on the α -carbon and at meta position)⁷⁴.

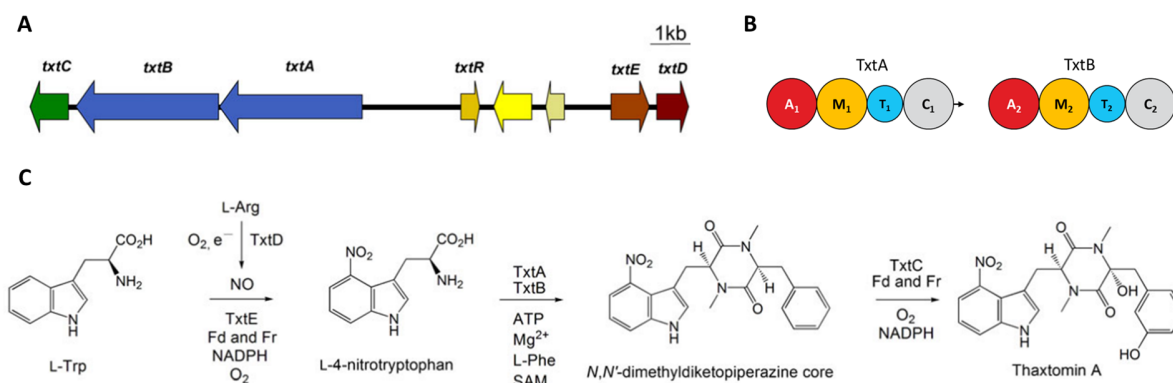


Figure 5: Thaxtomin A biosynthesis. A) Organization of the thaxtomin A biosynthetic gene cluster in *Streptomyces turgidiscabies*. The *txtR* gene encodes a cellobiose-responsive pathway specific activator. B) Organization of the bimodular thaxtomin NRPSs, TxtA and TxtB. C) Proposed pathway for thaxtomin A formation. Adapted from Gu et al., 2013⁶⁷.

To summarize, NRPSs are huge enzymatic machineries (several hundreds of kDa or more and about 250kDa for NRPSs dedicated to 2,5-DKP formation) composed of several interacting subunits. They catalyse the formation of an important number of peptides, with varying sizes and typology (linear, cyclic, branched...), and only a small number is dedicated to the synthesis of 2,5-DKPs. Recognition and activation of the incorporated amino acids rely on adenylation domains which tightly control the amino acids that will be incorporated in the peptide chain. Diversity is

generated through incorporation of non-canonical amino acids in the peptides and through enzymatic modification either during the elongation by modification domains or *a posteriori* by discrete tailoring enzymes.

- ii. Cyclodipeptide synthases: an original biosynthetic route dedicated to the production of 2,5-diketopiperazines

The discovery of the biosynthetic pathway of albonoursin in 2002 led to the identification of AlbC, a small protein of 27kDa, which was shown to be responsible for the formation of the albonoursin cyclodipeptide precursor, cyclo(Phe-Leu). Intriguingly, AlbC did not have significant similarity with neither NRPS domains nor any proteins characterized at that time²². Intensive studies of AlbC and of the similar enzymes that emerged from genomic databases led to the discovery of a new peptide-bond-forming enzyme family: the CDPSs²³. Unlike NRPSs, which activate their amino acid substrates using dedicated adenylation domains, CDPSs use one form of activated amino acids ubiquitous in cells: AA-tRNAs. By doing so, they hijack the activation machinery which canonically delivers its substrates to the ribosomal machinery (see Part II of this introduction). CDPSs have been shown to use a remarkable diversity of AA-tRNAs to produce several dozens of different cyclodipeptides. CDPS genes are usually clustered with genes encoding 2,5-DKP tailoring enzymes which introduce a great variety of modifications into cyclodipeptides and an increasing number of CDPS-dependent biosynthetic pathways is being described²⁶.

A detailed and up-to-date presentation of CDPSs is included in the review, presented hereafter, entitled “Cyclodipeptide synthases: a promising biotechnological tool for the synthesis of diverse 2,5-diketopiperazines” recently published in *Natural Product Reports*⁷⁵. Besides a global overview on CDPSs, the main novelty brought by this review is the use of Sequence Similarity Networks (SSNs) to visualize and classify putative and characterized CDPS sequences.

Author contributions:

Mireille Moutiez carried out the PSI-BLAST analysis that delivered an updated data set of CDPS sequences. I used this data set to generate and analyse the SSNs presented in this review. The manuscript was written by Muriel Gondry and me and was reviewed and approved by all authors,

with major input from Pascal Belin for the section dealing with tailoring enzymes. At the time of the writing of this review, the study on the incorporation of non-canonical amino acid by CDPSs, presented in detail in the first chapter of this manuscript, was already published and it is therefore summarized in this review.



Cite this: DOI: 10.1039/c9np00036d

Cyclodipeptide synthases: a promising biotechnological tool for the synthesis of diverse 2,5-diketopiperazines†

Nicolas Canu,  Mireille Moutiez,  Pascal Belin  and Muriel Gondry *

Covering: Up to mid-2019

Cyclodipeptide synthases (CDPSs) catalyse the formation of cyclodipeptides using aminoacylated-tRNA as substrates. The recent characterization of large sets of CDPSs has revealed that they can produce highly diverse products, and therefore have great potential for use in the production of different 2,5-diketopiperazines (2,5-DKPs). Sequence similarity networks (SSNs) are presented as a new, efficient way of classifying CDPSs by specificity and identifying new CDPS likely to display novel specificities. Several strategies for further increasing the diversity accessible with these enzymes are discussed here, including the incorporation of non-canonical amino acids by CDPSs and use of the remarkable diversity of 2,5-DKP-tailoring enzymes discovered in recent years.

Received 17th June 2019

DOI: 10.1039/c9np00036d

rsc.li/npr

1 Introduction

Cyclodipeptide synthases (CDPSs) constitute a family of enzymes involved in the biosynthesis of 2,5-diketopiperazines (2,5-DKPs), a class of natural products containing many therapeutically promising compounds.^{1–4} Even if 2,5-DKPs have been widely developed and synthesized by medicinal chemists,¹ the understanding and manipulation of their biosynthetic pathways continues to deliver new chemical structures that could lead to new activities.

CDPSs use aminoacyl-tRNAs (AA-tRNAs) as substrates to produce cyclodipeptides, the simplest representatives of 2,5-DKPs, which are frequently modified by the cyclodipeptide-tailoring enzymes of the 2,5-DKP biosynthetic pathways.^{5–8} The other enzymes responsible for the formation of the 2,5-DKP scaffold belong to the multimodular nonribosomal peptide synthetase (NRPS) family.⁹ The atypical mode of function of CDPSs (*i.e.* hijacking AA-tRNAs from the ribosomal machinery), makes these enzymes a fascinating subject of research at the interface of ribosomal and non-ribosomal peptide synthesis.¹⁰

Several recent studies have improved our understanding of CDPSs and opened up new opportunities for using these enzymes and the associated 2,5-DKP-tailoring enzymes for the biotechnological synthesis of many diverse 2,5-DKPs. This highlight covers recent developments in this area and presents the challenges ahead for the growing community of researchers studying CDPSs and CDPS-containing biosynthesis pathways.

2 Improvements to our understanding of the CDPS family

Since the discovery of the first CDPS in 2002 and that of the CDPS family in 2009,^{5,11} the cyclodipeptide-synthesizing activities of more than 120 members of this family have been characterized,¹⁰ including more than 40 within the last two years.^{12–19} These enzymes catalyse the production of a remarkable diversity of cyclodipeptides, containing most of the proteinogenic amino acids (see Section 3.1).

2.1 A growing and diverse family

CDPSs essentially originate from three bacterial phyla — Actinobacteria, Firmicutes and Proteobacteria — with very few found in other domains of life.¹³ All but one of the active CDPSs identified to date originate from bacteria, the exception being an enzyme from a sea anemone.²⁰

CDPSs are small enzymes, most having 200 to 300 residues, and they display variable degrees of sequence identity, less than 10% in some cases. Nevertheless, iterative homology searching can be used for the efficient mining of genomic sequence databases for putative CDPSs. In recent years, the number of

Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS Univ. Paris-Sud, Université Paris-Saclay, 91198 Gif-sur-Yvette cedex, France. E-mail: muriel.gondry@cea.fr

† Electronic supplementary information (ESI) available: Note on SSN construction: CDPS sequences were obtained by PSI-BLAST, with a non-redundancy threshold set at 98% identity. SSNs were generated with EFI-EST, using the specified similarity thresholds, and were visualised with Cytoscape coupled to the BridgeDB app. A data set is available. See DOI: 10.1039/c9np00036d

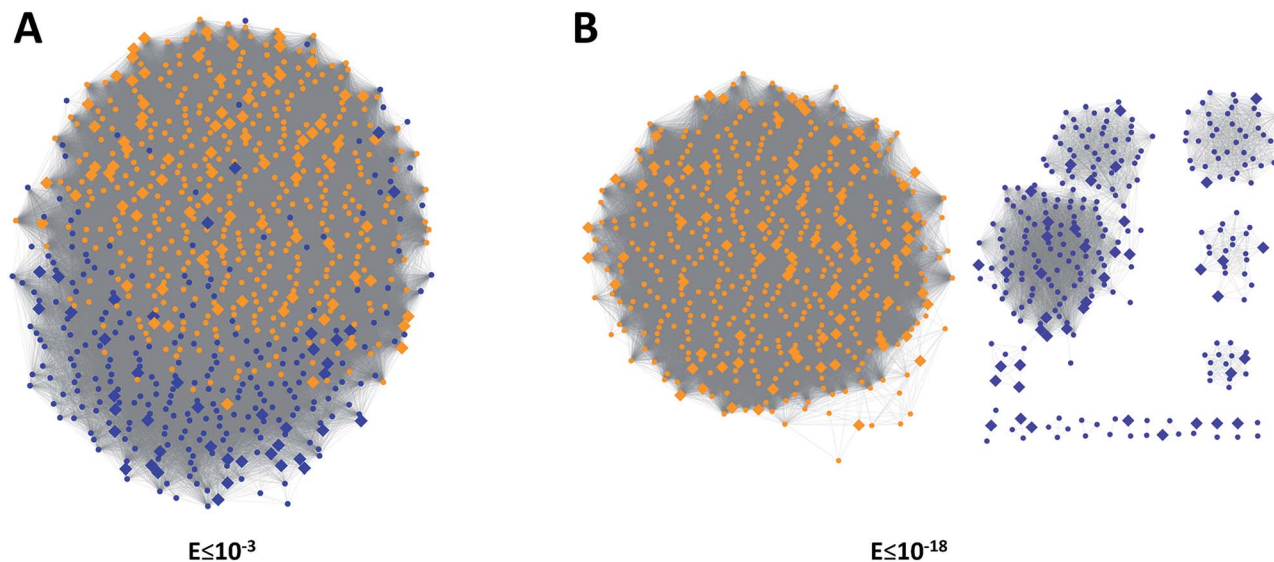


Fig. 1 SSNs of 775 CDPS sequences obtained with very low (A) and low (B) similarity thresholds. Characterised CDPSs are represented as diamonds and putative CDPSs as dots. The colour of the node indicates the subfamily: NYH (orange) or XYP (blue).

putative CDPSs identified in genomic databases has increased exponentially, and almost 800 different sequences had been identified by February 2019 (with a 98% minimal identity threshold to discriminate between very closely related CDPSs).

Fig. 1 shows the sequence similarity networks (SSNs) for a set of 775 characterized or putative CDPSs. SSN is a visualization tool in which the interrelationships between protein sequences are described using independent pairwise alignments between sequences.^{21,22} This tool is particularly adapted to the study of large set of proteins such as entire enzyme families or super-families.²³ Each individual sequence is represented by a symbol at a node in the network. Two nodes with sequence similarities exceeding the user-specified threshold are linked by a line (referred to as an edge). The threshold used is set as an E -value of the independent pairwise BLAST analysis of the two sequences.

In the first analyses, this E -value threshold was set to 10^{-3} . With this high E -value, even pairs of CDPSs with very low levels of sequence similarity were connected, resulting in all CDPS sequences being grouped into a single big cluster (Fig. 1A). Characterised CDPSs (nodes represented as large diamonds) account for only about 15% of sequences, but are well distributed throughout the cluster. The phylogenetic tree of CDPSs has been shown to split into two branches corresponding to two subfamilies. These subfamilies were named “NYH” and “XYP”, according to the identity of a pair of specific catalytic residues (see Section 2.2).²⁴ All 775 sequences of the set were assigned to subfamilies, with each subfamily indicated by a different colour on the SSN. Members of the same subfamily tend to partition into the same area of the cluster, indicating greater sequence identity between members of the same subfamily than with members of the other subfamily (Fig. 1A). The E -value threshold was then set at 10^{-18} , corresponding to a mean threshold sequence identity of 28%. With this threshold, the CDPS sequences fan out into one large cluster and several smaller groups (Fig. 1B). The large cluster contains all members of the

NYH subfamily, whereas the members of the XYP subfamily form several groups or are completely isolated. This distribution reflects the higher sequence diversity in the XYP subfamily.^{13,24}

2.2 Similar functioning throughout the CDPS family

The partitioning of CDPSs between these two subfamilies raises questions about potential functional and structural differences, but recent studies have strongly suggested that CDPSs from the two subfamilies are similar enough to be likely to work in the same way.

The underlying mechanism for the most widely studied CDPS, AlbC from *Streptomyces noursei*, was elucidated mostly through extensive experimental studies.^{25,26} A recent computational study defined the role of four essential catalytic residues more clearly²⁷ (Fig. 2). AlbC uses phenylalanyl-tRNA and leucyl-tRNA as the first and the second substrates, respectively, for the synthesis of cyclo(L-Phe-L-Leu) (cFL).²⁸ The binding of the first AA-tRNA by AlbC leads to the covalent attachment of the amino acid to residue S37, generating an aminoacyl-enzyme intermediate.²⁵ The second AA-tRNA binds to the intermediate, and its aminoacyl moiety is transferred to the aminoacyl-enzyme, resulting in the formation of a dipeptidyl-enzyme intermediate.²⁶ Finally, the dipeptidyl moiety undergoes intramolecular cyclisation, in which the Y202 residue serves as a proton relay, generating the cyclo-dipeptide.²⁷ Another two residues, Y178 and E182, are essential for the anchoring of the aminoacyl moiety of the first substrate throughout the catalytic cycle and the maintenance of the dipeptidyl moiety in a suitable conformation for the cyclisation reaction²⁷ (Fig. 2). The four catalytic residues are strictly (S37, Y202) or very highly (Y178, E182) conserved in all CDPSs. Two additional AlbC residues, N40 and H203, are conserved in the NYH subfamily (which gets its name from the N40, Y202 and

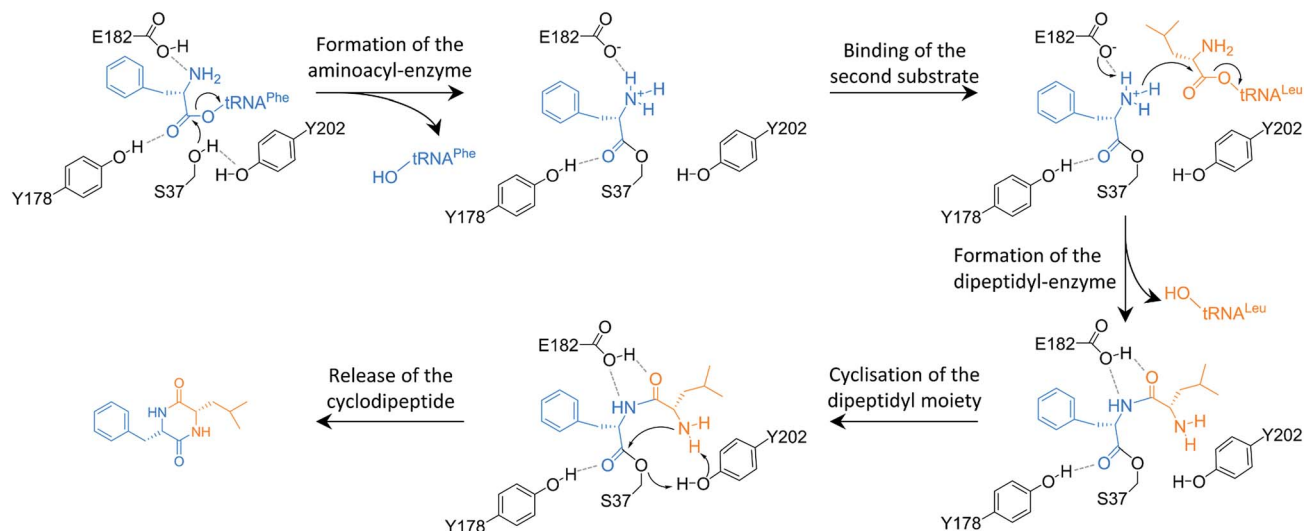


Fig. 2 Proposed mechanism for cFL synthesis by AlbC. The mechanism is adapted from the one previously proposed²⁶ thanks to recent crystallographic data³¹ and computational studies.²⁷ The first and second substrates are highlighted in blue and orange, respectively. The four catalytic residues conserved in CDPSS are in black. Possible interactions are indicated by dashed lines.

H203 residues) and are replaced by a non-conserved residue and a proline residue, respectively, in the XYP subfamily (named after the X40, Y202, P203 residues).²⁴

The crystal structures of three CDPSS from the NYH subfamily were reported in three independent papers in the early 2010s.^{25,29,30} The crystal structures of another four CDPSS, three of which belong to the XYP subfamily, were published very recently.³¹ All CDPSS, regardless of the subfamily to which they belong, adopt a common architecture consisting of a monomer built around a Rossmann-fold domain (Fig. 3). The NYH and XYP subfamilies differ mostly in terms of the structural organisation of the first half of this domain (Fig. 3A and B). The pair of catalytic residues differing between the two subfamilies (N40, H203 *versus* X40, P203) induces two structural solutions for stabilizing the conserved catalytic residues S37 and Y202. In NYH CDPSS structures, N40 is involved in contacts with S37 and

H203 is observed in packing interactions with Y202. In XYP CDPSS structures, X40 also interacts with S37 and P203 is stacked onto Y202.³¹ As the conserved residues adopt similar positions in all available CDPSS structures (Fig. 3C), CDPSS are likely to share the same catalytic mechanism.³¹

One of the most remarkable structural features of CDPSS is their high degree of structural similarity to the catalytic domains of class Ic aminoacyl-tRNA synthetases (AARSS), TyrRS and TrpRS, suggesting that CDPSS are likely to derive from a class-I AARS-like precursor.^{30,32} CDPSS have two solvent-accessible pockets located close to each other: the P1 pocket corresponding to the aminoacyl binding pocket in TyrRS and TrpRS, and the P2 pocket, which has no equivalent in these AARSS. The deep and narrow P1 pocket accommodates the aminoacyl moiety of the first AA-tRNA, whereas the larger, flatter P2 pocket interacts with the aminoacyl moiety of the

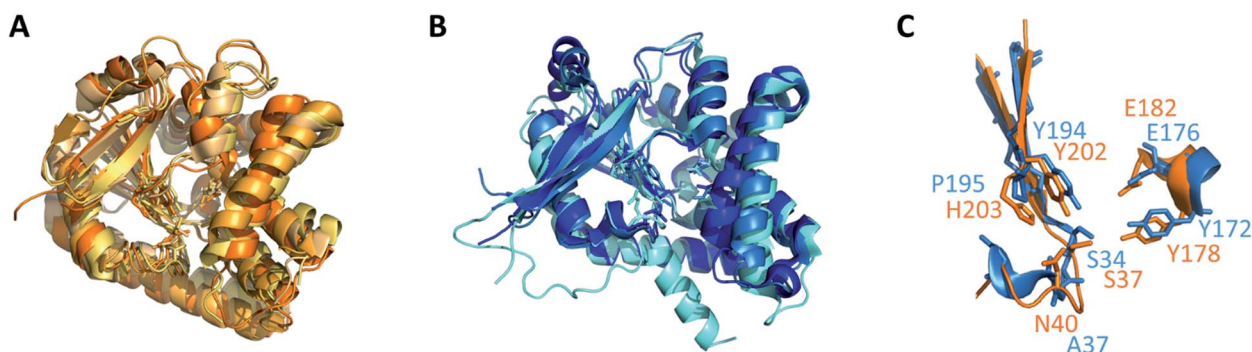


Fig. 3 Crystal structures of CDPSS. (A) Superimposition of the structures of the four NYH CDPSS: AlbC (PDB 4Q24), Rv2275 (PDB 2X9Q), *Shae*-CDPS (PDB 6EZ3) and YvmC (PDB 3OQH), shown as a colour gradient extending from dark orange to light yellow. (B) Superimposition of the structures of the three XYP CDPSS: *Fdm*-CDPS (PDB 5OCD), *Nbra*-CDPS (PDB 5MLQ) and *Rgry*-CDPS (PDB 5MLP), shown as a colour gradient extending from dark to light blue. (C) Structural alignment of the active sites of the NYH AlbC (initially obtained with the S37C variant) and the XYP *Nbra*-CDPS, shown with the same colour code. The regions displayed comprise residues 30–44, 178–182 and 197–203 (according to AlbC numbering).

second AA-tRNA.^{25,26} On the other hand, much remains unknown about the tRNA-binding regions of the two substrates on CDPs, and therefore further structural studies are required to elucidate these key features of substrate recognition.

3 Using CDPs as a tool for the production of diverse cyclodipeptides

3.1 Exploring and predicting CDPs activities

The heterologous expression of CDPs in *Escherichia coli* has proved efficient as a means of characterising the cyclodipeptide-synthesising activities of these enzymes. Biosynthesised cyclodipeptides are recovered from culture supernatants, identified and quantified by LC-MS.⁵ The characterisation of diverse CDPs in the last few years has revealed the production of a large number of cyclodipeptides. Seventy-six of the 210 possible natural cyclodipeptides have been shown to be produced by at least one CDPs (Fig. 4). Most combinations of aromatic and hydrophobic amino acids are produced. Cyclodipeptides containing polar and charged amino acids are less frequently found. The first arginine-containing cyclodipeptide produced by a CDPs (cPR) has recently been described, making lysine and aspartate the last two amino acids not known to be accepted by any characterised CDPs.¹³

Promiscuity in the CDPs family is variable: some CDPs produce only one cyclodipeptide, whereas others produce several different cyclodipeptides. These promiscuous CDPs generally produce one major and several minor cyclodipeptides, all with one amino acid in common. These CDPs are named according to their common amino acid. For example, “cWX-producing enzymes” for CDPs producing several different tryptophan-containing cyclodipeptides. The characterized CDPs include a large proportion of promiscuous enzymes, and recent characterisations have led to the identification of many cFX-, cPX-, and cWX-producing CDPs.^{13–15}

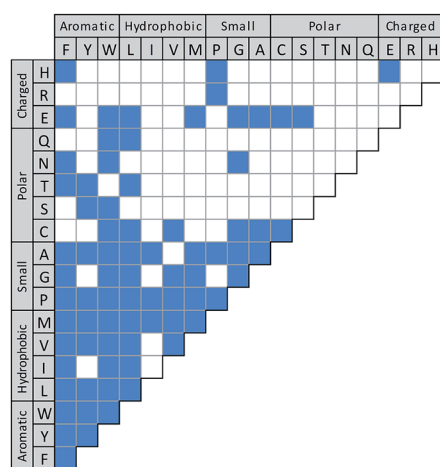


Fig. 4 Diversity of cyclodipeptides accessible through CDPs expression in *E. coli*. Each square represents one cyclodipeptide. Cyclodipeptides produced by CDPs are shown in colour.

The characterised CDPs constitute an interesting toolbox for the generation of diverse cyclodipeptides. However, one key challenge remains: in the context of a steady rise in the number of putative CDPs in databases, is there a rapid efficient way to predict CDPs activity and to identify uncharacterised CDPs likely to display previously unknown activities?

Prediction strategies based on the sequence specificity motifs of the P1 and P2 pockets have been presented in two previous studies.^{13,24} They are based on multiple sequence alignments of CDPs manually adjusted for predicted secondary structures, to identify the residues lining the P1 and P2 pockets in uncharacterised CDPs. These approaches effectively predict the main product for about 80% of putative CDPs, leaving about 20% of sequences with no prediction. However, in many cases, only the amino acid recognised by the first pocket can be predicted. An algorithm has recently been developed for predicting the main cyclodipeptides produced by CDPs. It is based on a similar multiple sequence alignment strategy, but without adjustment, due to the automated workflow using a huge set of sequences.³³ This would explain the low prediction rates obtained for recently characterised CDPs.¹³ This finding highlights the need to investigate other tools likely to be more suitable for the growing number of CDPs.

SSNs are a powerful tool for enzyme classification that may be useful for predicting the cyclodipeptides produced by CDPs. Indeed, in many cases, the choice of appropriate thresholds can make it possible to segregate an enzyme family into isofunctional clusters.^{22,34,35} The combination of SSN visualisations of large data sets with the biochemical characterisation of a few well-chosen enzymes is an efficient, high-throughput, proven strategy for the functional annotation of enzyme families.²³

Fig. 5 presents SSNs generated with the sequence data set used in Fig. 1 and lower *E*-value thresholds. Node colour indicates the nature of the main cyclodipeptide characterised (large diamonds) or predicted¹³ (circles). The light grey circles in this representation correspond to 211 new putative CDPs sequences that appeared in databases between June 2017 and February 2019 and have not yet been manually annotated. The *E*-value threshold was initially set to 10^{-55} , corresponding to a mean sequence identity threshold of 45% (Fig. 5A). More than 80% of the sequences cluster into groups of at least five different sequences, and almost all these groups contain at least one characterised CDPs. Almost all the clusters contain characterised and predicted CDPs with similar activities, suggesting that SSNs efficiently classify CDPs into isofunctional groups. Only two groups are not isofunctional: one contains a mixture of cYY- and cWX-producing CDPs, and the other contains cAE-, cLE- and cGE-producing CDPs. In both cases, setting lower *E*-value thresholds leads to the segregation of these two groups into perfectly isofunctional clusters (Fig. 5B and C). Such data demonstrate that SSNs can be used to predict the activity of putative CDPs on the basis of sequence similarity to characterised CDPs. For example, about 80% of the new CDPs corresponding to unannotated sequences can be predicted from the SSNs presented. This fraction is identical to the overall predictability of CDPs sequences from manual, sequence motif-based annotations.¹³

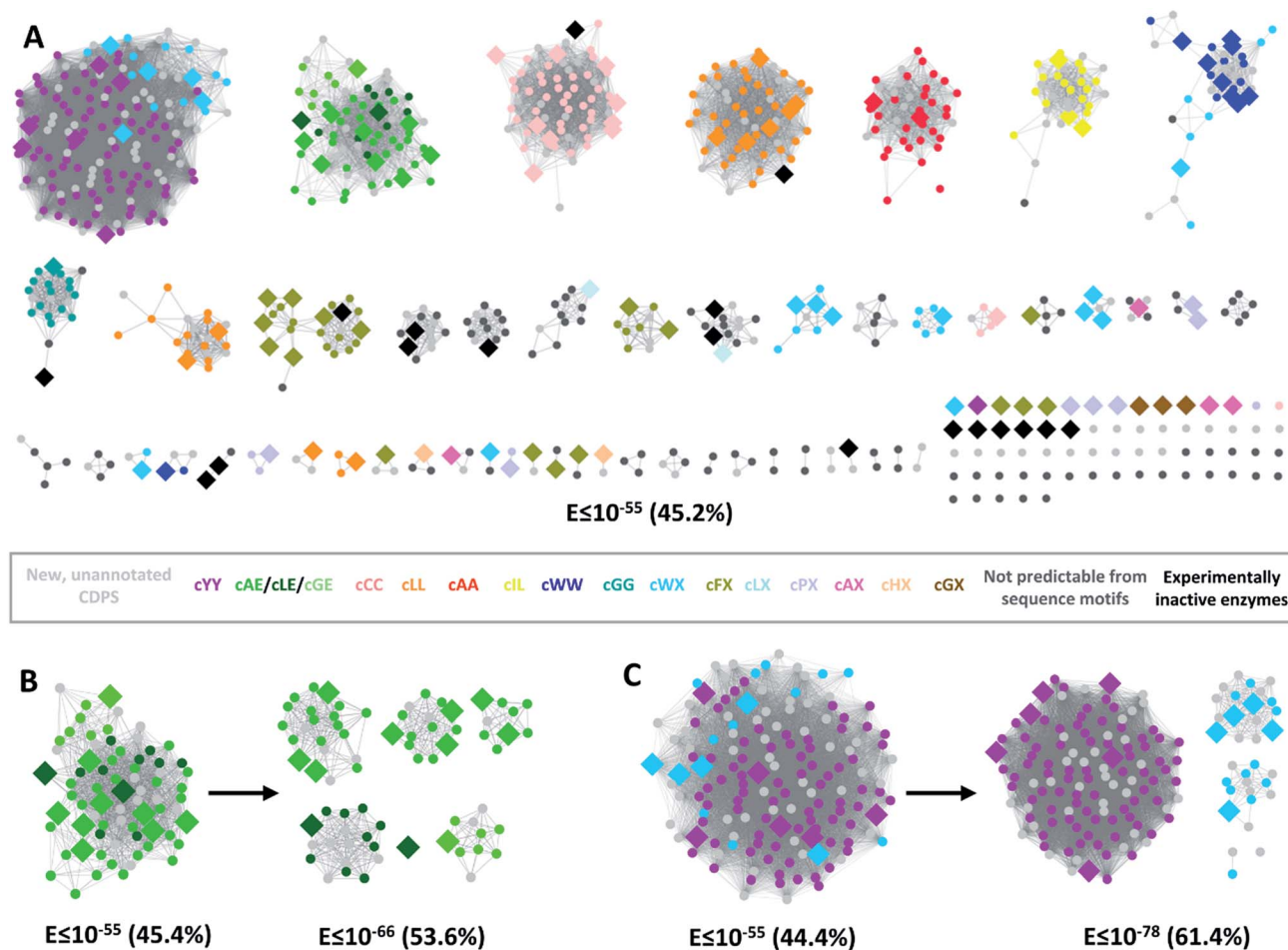


Fig. 5 SSNs illustrating the distribution of CDPs into isofunctional clusters. (A) Full set of 775 CDPs. (B) Subset of 72 cXE-synthesising enzymes. (C) Subset of 152 cYY- and cWX-synthesising enzymes. Characterised CDPs are represented as diamonds and putative CDPs as dots. Colours correspond to the main cyclodeptide produced (for experimentally characterized CDPs) or predicted (from sequence motifs), see the caption in the box. Percentages in parentheses indicate the mean sequence identity at the E -value threshold.

Taking these new predictions into account, more than 60% of CDPs are predicted to synthesise one of the 10 cyclodeptides for which the two amino acids can be accurately predicted (cYY, cLL, cCC, cIL, cWW, cGG, cAA, cAE, cLE and cGE) as a major product. The corresponding CDP sequences form fairly homogeneous groups, resulting in large clusters in SSNs. The activity of about 20% of CDPs can be predicted for only one of the two amino acids incorporated (*i.e.* as cWX- or cFX-synthesising enzymes). These enzymes tend to fan out in smaller clusters, reflecting the higher diversity of these groups.

A few characterised CDPs have rare activities, such as cLX-, cPX-, cAX-, cHX- or cGX-synthesising activities, for which prediction from sequence motifs is not possible.^{13,24} These CDPs are either totally isolated or form very small clusters in SSNs. These clusters will probably expand in the future, as increasing amounts of sequencing data become available. Further characterisation efforts targeting these future enriched clusters should improve our understanding and the prediction of these activities, as for the other groups.

About 13% CDPs have never been characterised and their activities cannot be predicted on the basis of clustering in SSNs.

Future characterisation efforts should target these enzymes as a matter of priority. SSN visualisation allows the straightforward identification of this “unexplored” enzymatic space.

Finally, about 2% of CDPs were found to have no detectable activity in *E. coli*. A minority of these inactive enzymes (4/18) have mutations in the typical quatuor of catalytic residues, which might explain their lack of activity in our assay. For the others, genome misannotations or inability of these enzymes to use tRNAs from *E. coli* may account for the observed lack of activity. Some of these inactive enzymes belong to the same clusters as characterised active enzymes, suggesting that they may have similar activities. Other inactive CDPs are isolated or fall into unexplored clusters, and their activity is totally unknown and difficult to predict.

In conclusion, SSNs appear to be the fastest and most efficient method for dealing with large numbers of CDPs and these networks predict the major product in a straightforward manner for about 80% of sequences. As for other prediction tools, the prediction of the second amino acid recognised by promiscuous CDPs is not possible, and rare activities, for

which only a few enzymes have been described, cannot yet be predicted.

3.2 Hijacking the hijackers: incorporation of non-canonical amino acids by CDPSSs

Thorough characterisation efforts aiming to describe CDPS diversity have shown that these enzymes can incorporate most proteinogenic amino acids into cyclodipeptides. However, the diversity of products arising from CDPS catalysis appears to be intrinsically limited by the nature of the substrates of these enzymes, AA-tRNAs, which limits the potential precursors to the 20 proteinogenic amino acids. One attractive strategy for unlocking the biosynthetic potential of CDPSs is the application of numerous methods for tRNA aminoacylation with non-canonical amino acids (ncAAs). These methods, originally developed for ncAA incorporation into proteins translated by ribosomes, include use of the natural promiscuity of AARS,³⁶ the heterologous expression of AARS/tRNA variants engineered for greater promiscuity,³⁷ the chemical ligation of ncAAs to tRNAs³⁸ and

the use of AARS-like ribozymes.³⁹ Methods based on AARS promiscuity are of particular interest from a biotechnological point of view, due to the compatibility of this system with *in vivo* production.

We have used such a method in combination with the heterologous expression of CDPSs.⁴⁰ Using strains of *E. coli* auxotrophic for one or several amino acids, with ncAAs added to the culture medium, we artificially “replaced” proteinogenic amino acids with non-canonical surrogates and tested these ncAAs as substrates for the expressed CDPS (Fig. 6).

In total, 60 ncAAs previously shown to be substrates of *E. coli* AARS^{36,41–43} were tested. Incorporation was demonstrated for 26 of the 60 ncAAs tested, resulting in the production of about 200 ncAA-containing cyclodipeptides, most of which had never been synthesized before. The ncAAs incorporated display a wide range of chemical modifications with respect to their proteinogenic counterparts, from simple substitutions, such as fluorination, hydroxylation and methylation, to more complex modifications, such as the presence of “clickable” azido groups.⁴⁴ Fig. 6 illustrates the diversity achieved with this method.

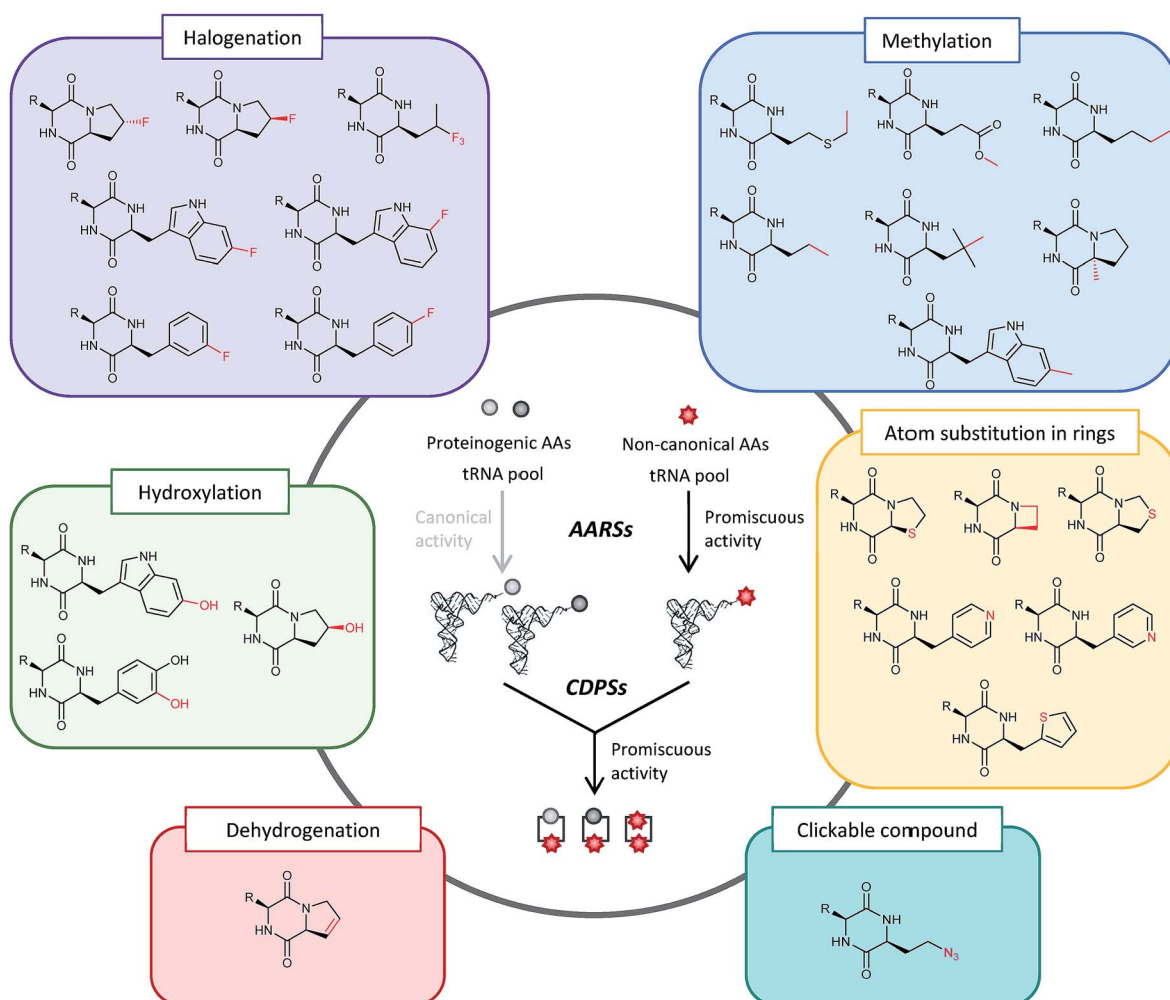


Fig. 6 Principles of ncAA incorporation into cyclodipeptides and diversity of ncAAs incorporated. Modifications relative to proteinogenic amino acids are shown in red.

This work clearly demonstrates that the promiscuity of CDPSs for non-canonical substrates can be used to generate a great diversity of compounds. The use of more efficient and promiscuous strategies for nCAA-loading onto tRNAs, such as the expression of engineered AARSs,^{37,45} appears to be an attractive strategy for increasing diversity still further. In some cases in which CDPS tolerance to nCAAs is the main limitation to their incorporation, CDPS engineering may prove necessary.

3.3 CDPSs: attractive but tricky candidates for enzyme engineering

Identification of the sequence motifs of the P1 and P2 aminoacyl-binding pockets makes it possible to predict the main cyclodipeptides produced by CDPSs (see Section 3.1). It is, thus, tempting to compare these sequence specificity motifs, with the aim of identifying and modifying the key residues apparently responsible for a given specificity. Several studies based on this rational engineering approach have shown that the substitution of one or two P1 or P2 residues can influence cyclodipeptide yields, but these studies were unable to generate novel specificities.^{12,15,25} Attempts to swap all the residues of the P1 or P2 pockets of two chosen CDPSs did not modify substrate specificity either (Y. Li, unpublished results). These results indicate that further studies are required to obtain a full understanding of the molecular basis of CDPS specificity and that more radical engineering strategies will be required to generate novel specificities.

Many studies have focused on AARSs, which have been successfully engineered to load a wide range of nCAAs onto tRNAs. The approaches used were often based on library construction, with randomization of the residues in the aminoacyl-binding sites.³⁷ The similarity between CDPSs and class Ic AARSs has raised hopes that such strategies could be successfully implemented with CDPSs in the near future. However, two main challenges lie ahead. First, the role of the tRNA moieties in substrate recognition remains unclear. Better insight into this recognition and the CDPS residues involved is required for the targeting of appropriate residues by semi-rational approaches. Second, the throughputs of the LC-MS methods currently used to determine cyclodipeptide-synthesising activities are too low for the screening of large libraries of CDPS variants. Recent advances in MALDI-TOF-MS-based analyses of small molecules^{46,47} may facilitate the application of this higher-throughput method to the study of CDPSs, making it possible to overcome this analytical obstacle.

4 Using CDPS-dependent pathways to expand 2,5-DKP diversity

The use of CDPS sequences as “biosynthetic hooks” is an efficient strategy for identifying genes encoding 2,5-DKP-tailoring enzymes in genome sequence databases.^{6,48,49} Nine clusters have been functionally characterised, including four over the last year (for a recent comprehensive review, see Lane, 2019 (ref. 8)). The characterised tailoring enzymes belong to several classes: oxidoreductases, including cytochromes

P450,^{16,18,19,48,50,51} flavin-dependent oxidases^{51–54} or 2-oxoglutarate and Fe(II)-dependent oxygenases,^{16,18} transferases including methyltransferases^{51,53,55} and prenyltransferases,¹⁵ and probable isomerases with an atypical membrane-associated terpene cyclase¹⁵ (Fig. 7). The activity of these enzymes with their natural substrates in biosynthetic pathways has been determined. The chemical modifications introduced include dehydrogenations between C atoms,^{16,18,52–54} O- and N-methylations,^{53,55} N- and C-hydroxylations,^{16,18,50} intramolecular cyclisations through ether formation^{16,18} or C–C coupling,⁴⁸ dimerisation through C–C coupling,¹⁹ the addition of a guaninyl moiety through C–N bond or through C–C bond followed by N-methylation of the nucleobase^{51,56} and the addition of a farnesyl moiety followed by its cyclisation¹⁵ (Fig. 7).

A few CDPS-associated tailoring enzymes have been evaluated for specificity, to assess their potential for the generation of chemical diversity. Some, such as CYP121 from the mycocyclusin biosynthetic pathway of *Mycobacterium tuberculosis*, which introduces an intramolecular C–C bond specifically in cYY, were found to be highly specific.^{48,57} In contrast, others, such as the cyclodipeptide oxidases (CDOs) of *Streptomyces noursei* and *Nocardioopsis dassonvillei*, which catalyse C α –C β dehydrogenations on diverse cyclodipeptides preferentially carrying hydrophobic side chains, have a much broader range of substrates.^{52,55,58} The cytochrome P450 NascB of the nase-seazine biosynthetic pathway of a marine *Streptomyces* strain provides a striking example of a less-specific tailoring enzyme.¹⁹ Its native activity is the catalysis of C2–N10 intramolecular coupling in the tryptophanyl moiety of cWP, followed by intermolecular C3–C6' coupling with the tryptophanyl moiety of another cWP molecule. Interestingly, the authors were able to reconstitute the activity of this cytochrome P450 both *in vitro* and in the whole-cell extract of an engineered *E. coli* strain. This cell extract approach made it possible to probe the promiscuity of the P450 towards several cWX analogues and to generate about 30 nase-seazine analogues containing halogenated tryptophan surrogates and various second amino acids. The possibility of expressing functional 2,5-DKP-tailoring P450s in *E. coli* cells opens up interesting new opportunities, given the wide range of activities performed by these P450s.

Tailoring enzymes from non-CDPS-containing pathways could also be used to increase diversity. For example, a number of NRPS-associated fungal prenyltransferases have been shown to catalyse the transfer of the C5-allyl moiety from the ubiquitous prenyl donor dimethylallyl diphosphate to the indole ring of the tryptophan of various cyclodipeptides.^{59,60} Several of these fungal prenyltransferases were recently expressed with bacterial CDPSs *in vivo* in *E. coli*.⁶¹ Tryptophan-containing cyclodipeptides prenylated at various sites, and in a regular or reverse fashion, were produced. The yields of prenylated products increased considerably following the introduction, into the expression host, of a hybrid mevalonate pathway for increasing intracellular levels of the prenyl donor. This first demonstration of the successful use of a combination of CDPSs with NRPS-associated 2,5-DKP-tailoring enzymes paves the way for exploitation of a huge reservoir of biosynthetic diversity.

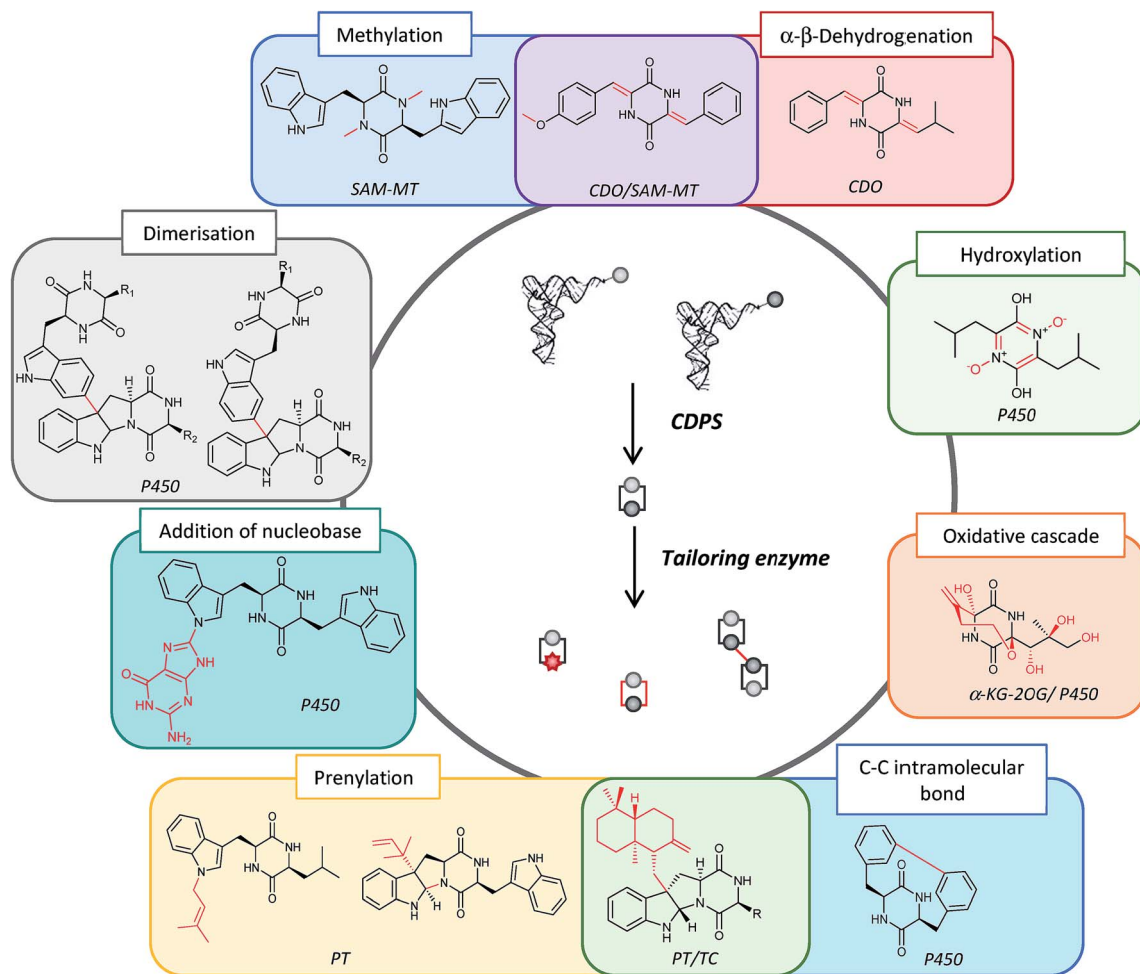


Fig. 7 Diversity of the modifications accessible with 2,5-DKP tailoring enzymes. Modifications relative to cyclodipeptides are shown in red. Italic abbreviations indicate the enzymes involved in the modifications. SAM-M: *S*-adenosylmethionine-dependent methyltransferase; CDO: cyclic dipeptide oxidase; P450: cytochrome P450; PT: prenyl transferase; TC: terpene cyclase; α -KG-2OG: 2-oxoglutarate and Fe(II)-dependent oxygenase.

Overall, these studies highlight the utility of 2,5-DKP-tailoring enzymes for generating chemical diversity, *in vitro* or *in vivo*. Bioinformatic analyses of putative CDPS-containing clusters have suggested that a large proportion of this biosynthetic diversity remains unexplored and that enzymes from many families other than those already characterised are closely associated with CDPSs. These enzymes may include *N*-acetyltransferases, sulfotransferases and glycosyltransferases.³³ Given the diversity of the cyclodipeptides synthesised by CDPSs and their capacity to incorporate nAAs, the association of CDPSs with promiscuous tailoring enzymes in unnatural pathways appears to be a promising way of expanding the chemical range of 2,5-DKPs accessible to medicinal chemists.

5 Conclusion and perspectives

The broad specificity spectrum of CDPSs, which include proteinogenic and non-canonical amino acids, make them attractive catalysts for 2,5-DKP biosynthesis. They constitute an interesting toolbox that is steadily expanding with the ever-

growing number of putative CDPS sequences available. SSN visualisation is an efficient strategy for tackling this increasing diversity and functionally annotating CDPSs.

Unlike NRPSs which are massive, multimodular complexes that are not particularly easy to manipulate, CDPSs are a tractable solution for the biosynthesis of cyclodipeptides, thanks to their small size and the relatively high yield of cyclodipeptides obtained when they are expressed in *E. coli*.

The bioactivities of 2,5-DKPs have for long sparked the interest of medicinal chemists, which have delivered inventive synthetic routes to these molecules.¹ Although the synthesis of simple cyclodipeptides is straightforward in most cases and can be done at high yield from standard protected amino acids, replicating the range of available 2,5-DKP-tailoring activities is a challenge for the chemists interested in total synthesis of complex 2,5-DKPs.⁶² In particular, the regio- and stereo-specificity of the enzymes involved and the constraints of polycyclic 2,5-DKPs are tedious to replicate in synthetic chemistry.^{1,63–66} A prime example is given by the dimerization and pyrroloindoline formation which occur during naseaezines

biosynthesis. The total synthesis of these molecules requires a complex, multi-step reaction scheme to control the regio- and stereospecificity of the modifications.⁶⁶ On the other hand, these molecules and their derivatives can be obtained in a single step using a P450, which has been shown to tolerate a wide range of cyclodipeptides as precursors.¹⁹ Therefore, 2,5-DKP-tailoring enzymes appear to be a valuable tool either for purely biosynthetic approaches but also for conversion of synthetic cyclodipeptides as exemplified with many recent studies.^{19,67}

The untapped potential of biosynthetic pathways as sources of new catalysts combined with dramatic technological advances in the field of synthetic biology should significantly broaden the chemical diversity of 2,5 DKPs and give rise to novel and interesting bioactive molecules.

6 Conflicts of interest

The authors have no conflicts of interest to declare.

7 Acknowledgements

Work in the laboratory of Dr Muriel Gondry is funded by grants from the French National Research Agency (ANR-14-CE09-0021; ANR-16-CE29-0026). N. C. holds a PhD scholarship from the CEA. We thank Prof. John A. Gerlt and Dr Rémi Zallot for helpful advice concerning Cytoscape.

8 Notes and references

- 1 A. D. Borthwick, *Chem. Rev.*, 2012, **112**, 3641–3716.
- 2 R. M. Huang, X. X. Yi, Y. Zhou, X. Su, Y. Peng and C. H. Gao, *Mar. Drugs*, 2014, **12**, 6213–6235.
- 3 X. Wang, Y. Li, X. Zhang, D. Lai and L. Zhou, *Molecules*, 2017, **22**, 2026.
- 4 A. Ortiz and E. Sansinenea, *Curr. Med. Chem.*, 2017, **24**, 2773–2780.
- 5 M. Gondry, L. Sauguet, P. Belin, R. Thai, R. Amouroux, C. Tellier, K. Tuphile, M. Jacquet, S. Braud, M. Courçon, C. Masson, S. Dubois, S. Lautru, A. Lecoq, S. Hashimoto, R. Genet and J.-L. Pernodet, *Nat. Chem. Biol.*, 2009, **5**, 414–420.
- 6 P. Belin, M. Moutiez, S. Lautru, J. Seguin, J.-L. Pernodet and M. Gondry, *Nat. Prod. Rep.*, 2012, **29**, 961–979.
- 7 T. Giessen and M. Marahiel, *Int. J. Mol. Sci.*, 2014, **15**, 14610–14631.
- 8 P. Borgman, R. D. Lopez and A. L. Lane, *Org. Biomol. Chem.*, 2019, **17**, 2305–2314.
- 9 B. Gu, S. He, X. Yan and L. Zhang, *Appl. Microbiol. Biotechnol.*, 2013, **97**, 8439–8453.
- 10 M. Moutiez, P. Belin and M. Gondry, *Chem. Rev.*, 2017, **117**, 5578–5618.
- 11 S. Lautru, M. Gondry, R. Genet and J.-L. Pernodet, *Chem. Biol.*, 2002, **9**, 1355–1364.
- 12 K. Brockmeyer and S.-M. Li, *J. Nat. Prod.*, 2017, **80**, 2917–2922.
- 13 M. Gondry, I. Jacques, R. Thai, M. Babin, N. Canu, J. Seguin, P. Belin, J.-L. Pernodet and M. Moutiez, *Front. Microbiol.*, 2018, **9**, 1–14.
- 14 J. Liu, H. Yu and S. M. Li, *Appl. Microbiol. Biotechnol.*, 2018, 1–10.
- 15 T. Yao, J. Liu, Z. Liu, T. Li, H. Li, Q. Che, T. Zhu, D. Li, Q. Gu and W. Li, *Nat. Commun.*, 2018, **9**, 4091.
- 16 J. B. Patteson, W. Cai, R. A. Johnson, K. C. Santa Maria and B. Li, *Biochemistry*, 2017, **57**, 61–65.
- 17 N. M. Vior, R. Lacroix, G. Chandra, D.-R. Siobhan, M. Trick and A. W. Truman, *Appl. Environ. Microbiol.*, 2018, **84**, 1–17.
- 18 S. Meng, W. Han, J. Zhao, X. H. Jian, H. X. Pan and G. L. Tang, *Angew. Chem., Int. Ed.*, 2018, **57**, 719–723.
- 19 W. Tian, C. Sun, M. Zheng, J. R. Harmer, M. Yu, Y. Zhang, H. Peng, D. Zhu, Z. Deng, S. L. Chen, M. Mobli, X. Jia and X. Qu, *Nat. Commun.*, 2018, **9**, 1–9.
- 20 J. Seguin, M. Moutiez, Y. Li, P. Belin, A. Lecoq, M. Fonvielle, J. B. Charbonnier, J. L. Pernodet and M. Gondry, *Chem. Biol.*, 2011, **18**, 1362–1368.
- 21 H. J. Atkinson, J. H. Morris, T. E. Ferrin and P. C. Babbitt, *PLoS One*, 2009, **4**, e4345.
- 22 J. A. Gerlt, J. T. Bouvier, D. B. Davidson, H. J. Imker, B. Sadkhin, D. R. Slater and K. L. Whalen, *Biochim. Biophys. Acta, Proteins Proteomics*, 2015, **1854**, 1019–1037.
- 23 R. Zallot, N. O. Oberg and J. A. Gerlt, *Curr. Opin. Chem. Biol.*, 2018, **47**, 77–85.
- 24 I. B. Jacques, M. Moutiez, J. Witwinowski, E. Darbon, C. Martel, J. Seguin, E. Favry, R. Thai, A. Lecoq, S. Dubois, J.-L. Pernodet, M. Gondry and P. Belin, *Nat. Chem. Biol.*, 2015, **11**, 721–731.
- 25 L. Sauguet, M. Moutiez, Y. Li, P. Belin, J. Seguin, M. H. Le Du, R. Thai, C. Masson, M. Fonvielle, J. L. Pernodet, J. B. Charbonnier and M. Gondry, *Nucleic Acids Res.*, 2011, **39**, 4475–4489.
- 26 M. Moutiez, E. Schmitt, J. Seguin, R. Thai, E. Favry, P. Belin, Y. Mechulam and M. Gondry, *Nat. Commun.*, 2014, **5**, 1–7.
- 27 E. Schmitt, G. Bourgeois, M. Gondry and A. Aleksandrov, *Sci. Rep.*, 2018, **8**, 1–11.
- 28 M. Moutiez, M. Fonvielle, P. Belin, E. Favry, M. Arthur and M. Gondry, *Nucleic Acids Res.*, 2014, **42**, 7247–7258.
- 29 M. W. Vetting, S. S. Hegde and J. S. Blanchard, *Nat. Chem. Biol.*, 2010, **6**, 797–799.
- 30 L. Bonnefond, T. Arai, Y. Sakaguchi, T. Suzuki, R. Ishitani and O. Nureki, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 3912–3917.
- 31 G. Bourgeois, J. Seguin, M. Babin, P. Belin, M. Moutiez, Y. Mechulam, M. Gondry and E. Schmitt, *J. Struct. Biol.*, 2018, **203**, 17–26.
- 32 L. Aravind, R. F. de Souza and L. M. Iyer, *Biol. Direct*, 2010, **5**, 1–11.
- 33 M. A. Skinnider, C. W. Johnston, N. J. Merwin, C. A. Dejong and N. A. Magarvey, *BMC Genomics*, 2018, **19**, 1–11.
- 34 S.-H. Dong, N. D. Frane, Q. H. Christensen, E. P. Greenberg, R. Nagarajan and S. K. Nair, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 9092–9097.
- 35 K. L. Dunbar, J. R. Chekan, C. L. Cox, B. J. Burkhart, S. K. Nair and D. A. Mitchell, *Nat. Chem. Biol.*, 2014, **10**, 823–829.

- 36 M. C. T. Hartman, K. Josephson and J. W. Szostak, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 4356–4361.
- 37 L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498–500.
- 38 T. G. Heckler, L. H. Chang, Y. Zama, T. Naka, M. S. Chorghade and S. M. Hecht, *Biochemistry*, 1984, **23**, 1468–1473.
- 39 T. Passioura and H. Suga, *Top. Curr. Chem.*, 2014, **11**, 331–346.
- 40 N. Canu, P. Belin, R. Thai, I. Correia, O. Lequin, J. Seguin, M. Moutiez and M. Gondry, *Angew. Chem., Int. Ed.*, 2018, **57**, 3118–3122.
- 41 C. Fan, J. M. L. Ho, N. Chirathivat, D. Söll and Y. S. Wang, *ChemBioChem*, 2014, **15**, 1805–1809.
- 42 K. Kirshenbaum, I. S. Carrico and D. A. Tirrell, *ChemBioChem*, 2002, **3**, 235–237.
- 43 F. Oldach, R. Altoma, A. Kuthning, T. Caetano, S. Mendo, N. Budisa and R. D. Süßmuth, *Angew. Chem., Int. Ed.*, 2012, **51**, 415–418.
- 44 C. Wang, D. Ikhlef, S. Kahlal, J. Y. Saillard and D. Astruc, *Coord. Chem. Rev.*, 2016, **316**, 1–20.
- 45 P. Kast and H. Hennecke, *J. Mol. Biol.*, 1991, **222**, 99–124.
- 46 H. Wang, Z. Zhao and Y. Guo, *Top. Curr. Chem.*, 2013, 165–192.
- 47 M. Lu, X. Yang, Y. Yang, P. Qin, X. Wu and Z. Cai, *Nanomaterials*, 2017, **7**, 87.
- 48 P. Belin, M. H. Le Du, A. Fielding, O. Lequin, M. Jacquet, J.-B. Charbonnier, A. Lecoq, R. Thai, M. Courçon, C. Masson, C. Dugave, R. Genet, J.-L. Pernodet and M. Gondry, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 7426–7431.
- 49 T. W. Giessen and M. A. Marahiel, *Front. Microbiol.*, 2015, **6**, 1–11.
- 50 M. J. Cryle, S. G. Bell and I. Schlichting, *Biochemistry*, 2010, **49**, 7282–7296.
- 51 J. Liu, X. Xie and S.-M. Li, *Angew. Chem., Int. Ed.*, 2019, **58**, 2–9.
- 52 M. Gondry, S. Lautru, G. Fusai, G. Meunier, A. Ménez and R. Genet, *Eur. J. Biochem.*, 2001, **268**, 1712–1721.
- 53 T. W. Giessen, A. M. Von Tesmar and M. A. Marahiel, *Biochemistry*, 2013, **52**, 4274–4283.
- 54 Y. Li, Y. M. Lai, Y. Lu, Y. L. Yang and S. Chen, *Arch. Microbiol.*, 2014, **196**, 765–774.
- 55 T. W. Giessen, A. M. Von Tesmar and M. A. Marahiel, *Chem. Biol.*, 2013, **20**, 828–838.
- 56 H. Yu, X. Xie and S.-M. Li, *Org. Lett.*, 2018, **20**, 4921–4925.
- 57 M. Fonvielle, M. H. Le Du, O. Lequin, A. Lecoq, M. Jacquet, R. Thai, S. Dubois, G. Grach, M. Gondry and P. Belin, *J. Biol. Chem.*, 2013, **288**, 17347–17359.
- 58 H. Kanzaki, D. Imura, T. Nitoda and K. Kawazu, *J. Biosci. Bioeng.*, 2000, **90**, 86–89.
- 59 J. Winkelblech, A. Fan and S. M. Li, *Appl. Microbiol. Biotechnol.*, 2015, **99**, 7379–7397.
- 60 A. Fan, J. Winkelblech and S. M. Li, *Appl. Microbiol. Biotechnol.*, 2015, **99**, 7399–7415.
- 61 P. Dubois, I. Correia, F. Le Chevalier, S. Dubois, I. Jacques, N. Canu, M. Moutiez, R. Thai, M. Gondry, O. Lequin and P. Belin, *Sci. Rep.*, 2019, **9**, 1–13.
- 62 J. Kim and M. Movassaghi, *Acc. Chem. Res.*, 2015, **48**, 1159–1171.
- 63 T. Amatov, R. Pohl, I. Císařová and U. Jahn, *Angew. Chem., Int. Ed.*, 2015, **54**, 12153–12157.
- 64 X. Zhu, C. C. McAtee and C. S. Schindler, *Org. Lett.*, 2018, **20**, 2862–2866.
- 65 X. Shen, J. Zhao, Y. Xi, W. Chen, Y. Zhou, X. Yang and H. Zhang, *J. Org. Chem.*, 2018, **83**, 14507–14517.
- 66 M. Wada, H. Suzuki, M. Kato, H. Oikawa, A. Tsubouchi and H. Oguri, *ChemBioChem*, 2019, **20**, 1273–1281.
- 67 V. Wohlgemuth, F. Kindinger and S. M. Li, *Appl. Microbiol. Biotechnol.*, 2018, **102**, 2671–2681.

II) Cyclodipeptide synthase substrates: aminoacyl-tRNAs

As substrates of the ribosomes, AA-tRNAs have been present in cells since the very early steps of evolution. Given this ubiquitous presence throughout evolution, it is not surprising that several enzyme families, which do not belong to the translational apparatus, have evolved the ability to use AA-tRNAs as substrates for the catalysis of diverse chemical reactions. This kind of mechanism is found in a growing number of biological processes (Figure 6): in natural product biosynthesis (2,5-DKPs biosynthesis by CDPs, site-specific dehydration of ribosomally synthesized and posttranslationally modified peptides (RiPPs), amino acid transfer during various antibiotics biosynthetic pathways) but also in primary metabolism (amino acid and tetrapyrrole biosynthesis) and in house-keeping processes⁷⁶.

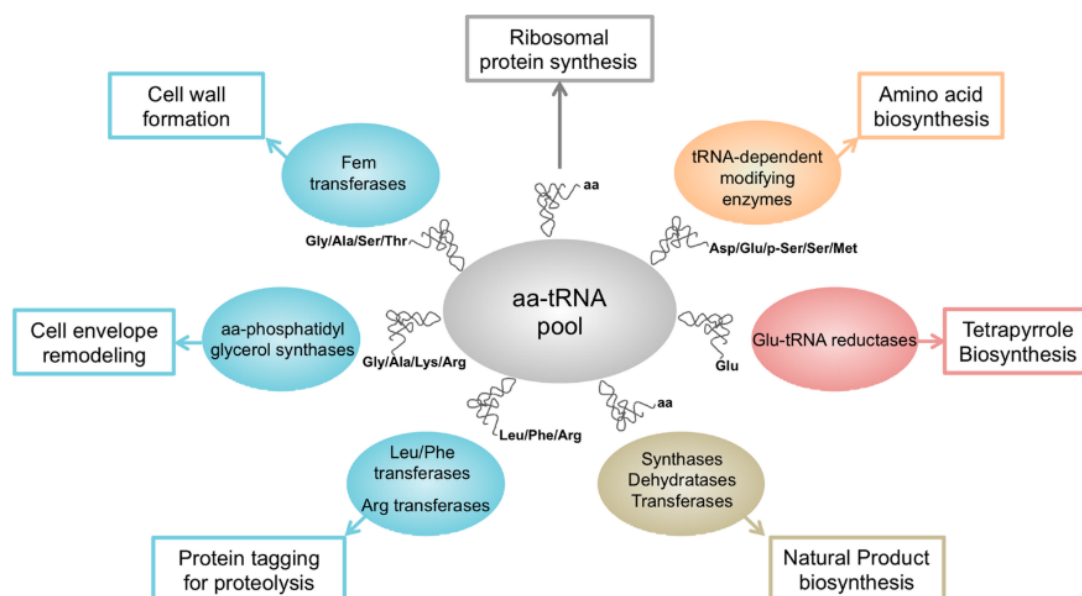


Figure 6: Diversity of AA-tRNA-utilizing enzymes. Some AA-tRNAs are hijacked for use as substrates by diverse enzymes (schematically represented by external circles) involved in various cellular processes (indicated in boxes). Reproduced from Moutiez, Belin et al., 2017⁷⁶.

The study and the manipulation of AA-tRNA-utilizing enzymes, and in particular of CDPs, require generating AA-tRNAs and AA-tRNA analogues. tRNA aminoacylation by the dedicated enzymes, aminoacyl-tRNA synthetases (AARSs) has been studied for more than 60 years and we globally have a good understanding of how this reaction is performed. Additionally, motivated by the perspective of introducing new chemistries into proteins and peptides, several groups have

developed strategies to perform the RNA aminoacylation reaction with more flexibility than what is allowed using conventional AARSs. This part of the introduction aims at giving an overview of AA-tRNA characteristics and of the different methods that can be used to generate these molecules and non-canonical analogues.

A) Aminoacyl-tRNAs and their cognate aminoacyl-tRNA synthetases

tRNA research has been very fruitful in the last six decades and the amount of knowledge accumulated in the field is huge. The following part summarizes some of the information that seem relevant for the study of CDPSs.

i. Aminoacyl-tRNA molecules

tRNAs are small single-stranded RNAs with a typical size of 76 nucleotides. Almost all tRNAs exhibit the typical cloverleaf secondary structures, composed of the acceptor stem, the dihydrouridine (D) stem-loop, the anticodon stem-loop which bears the anticodon trinucleotide, a variable loop and the T ψ C stem-loop where ψ represents pseudouridine (Figure 7A). Several residues are conserved in almost all tRNAs and in particular the 3' terminus C₇₄C₇₅A₇₆ (numeration corresponds to a typical tRNA such as yeast tRNA^{Phe})⁷⁷. Thorough biochemical characterizations have revealed the presence of many modified residues⁷⁸, some of them being widely conserved between species and cognate amino acids, such as the dihydrouridine of D stem-loop and the pseudouridine of the T ψ C stem-loop. Another universal feature of tRNAs is the 5' monophosphate terminus which is due to the digestion steps which occur during tRNA biosynthesis⁷⁷.

All tRNAs exhibit a similar tertiary structure and fold into a L-shaped structure⁷⁹, in which the T ψ C stem-loop stacks onto the acceptor stem to create the 12 bp acceptor–T ψ C minihelix, and the anticodon stem–loop stacks onto the D-stem to give the 9–10 bp stem of the anticodon–D loop dumbbell (Figure 7B and 7C). This architecture is flexible, as revealed by the crystal structures of free tRNAs compared to these of tRNAs in macromolecular complexes⁸⁰. This intrinsic plasticity allows tRNA-utilizing enzymes to bend their substrates to adopt the proper folding necessary for catalysis.

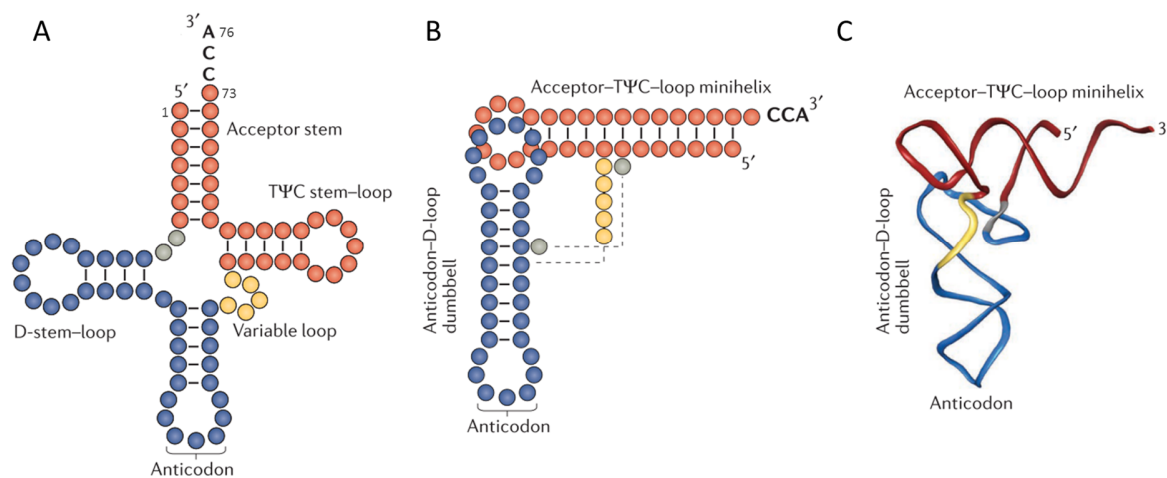


Figure 7: Overview of tRNA structure. Secondary structure (A), schematic (B) and ribbon 3D-structures (C) of a typical tRNA. Reproduced from Schimmel, 2018⁸¹.

tRNA aminoacylation corresponds to the formation of an ester link between the carboxylic acid function of an amino acid backbone and one of the two free hydroxyls of the ribose of the terminal adenosine. The forms esterified on the 2'- or the 3'- hydroxyl are rapidly trans-esterified from one to the other through a base-assisted mechanism, reaching an equilibrium where the form bearing the amino acid on the 3'-hydroxyl of the sugar is slightly more abundant than its isomer⁸² (Figure 8).

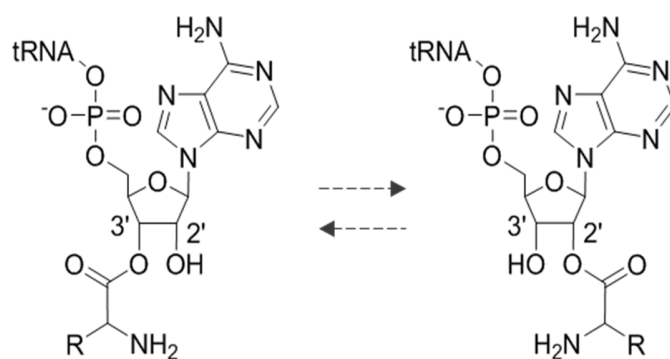


Figure 8: The two co-existing forms of ester link between a tRNA and an amino acid.

When AA-tRNAs are free in solution, this ester link is easily hydrolysed and therefore unstable. The rate of hydrolysis is highly dependent on pH (acidic pH tend to limit spontaneous deacylation)⁸² and temperature (half-lives of Phe-tRNA^{Phe} can vary from 30h at 20°C to 15min at 60°C)⁸³. Finally, the nature of esterified amino acids is also an important determinant of AA-tRNAs stability: prolyl-tRNA is particularly instable whereas bulky hydrophobic amino acids such as

valine or isoleucine tend to have a protective effect against diacylation (Figure 9). On the other hand, tRNA sequences do not appear to influence significantly the stability of AA-tRNAs (Figure 9)⁸⁴.

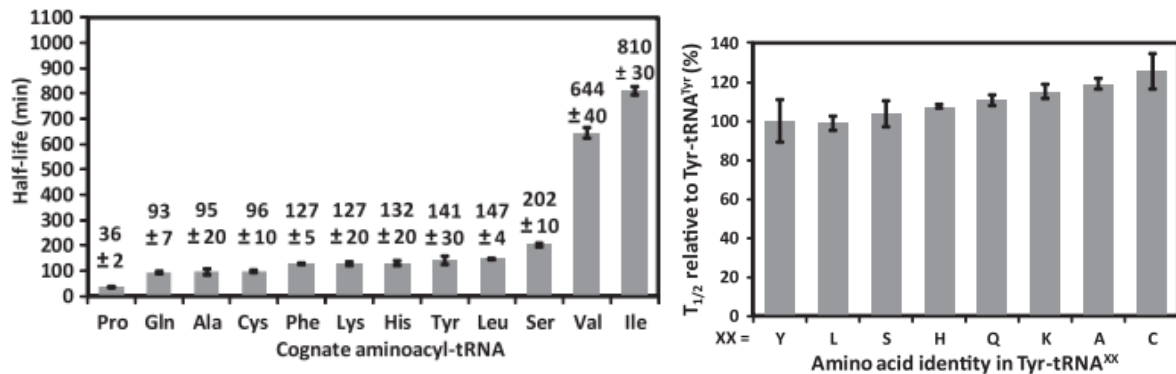


Figure 9: Influence of amino acid side chain (left) and tRNA sequences (right) on AA-tRNA stability. Reproduced from Peacock et al., 2014⁸⁴.

In prokaryotic cells, the aminoacylation of tRNAs triggers the interaction of AA-tRNAs with the elongation factor Tu (EF-Tu) (also named EF1A), a GTPase which protects AA-tRNAs from hydrolysis and delivers AA-tRNAs to the ribosome machinery⁸⁵. The affinity of EF-Tu for AA-tRNAs is really high (K_Ds in the nanomolar range⁸⁶) and it is estimated that almost all AA-tRNAs in cells are in complex with EF-Tu. How nonribosomal AA-tRNA-utilizing enzymes can hijack their substrates from this powerful protection remains an unanswered question^{76,87}.

ii. Aminoacyl-tRNA synthetases and their specificity

An accurate matching between tRNA sequences and loaded amino acids is required to maintain the accuracy of protein translation. Each proteinogenic amino acid is recognized and loaded by one cognate AARS on the corresponding tRNA isoacceptors. tRNA aminoacylation is performed by AARSs through a two-step mechanism: first, the cognate amino acid is accommodated in a binding pocket and activated with adenosine triphosphate (ATP), which leads to an aminoacyl-adenylate intermediate. Then, this intermediate is transferred on a hydroxyl of the ribose of the terminal adenosine of the cognate tRNA. This ester link is formed on the 2'-hydroxyl for class I AARSs or the 3'-hydroxyl for class II AARSs⁸⁸. AARS specificity is highly sophisticated and responds to a complex problem: each AARS must specifically recognize its

substrates (amino acid and tRNA) from a pool of potential substrates and reject the close analogues of its substrates.

Recognition of the cognate amino acid by AARSs is based on the architecture of the amino acid binding pockets. Complex interaction networks (involving hydrogen bonds, hydrophobic and π -stacking interactions, salt bridges and metal complexes) are established between the binding pocket residues and the cognate amino acid substrate (Figure 10)⁸⁹. In some cases such as in GlyRS, specificity is not based on specific interaction with the cognate amino acid, but on the rejection of amino acids bulkier than the one of interest⁹⁰.

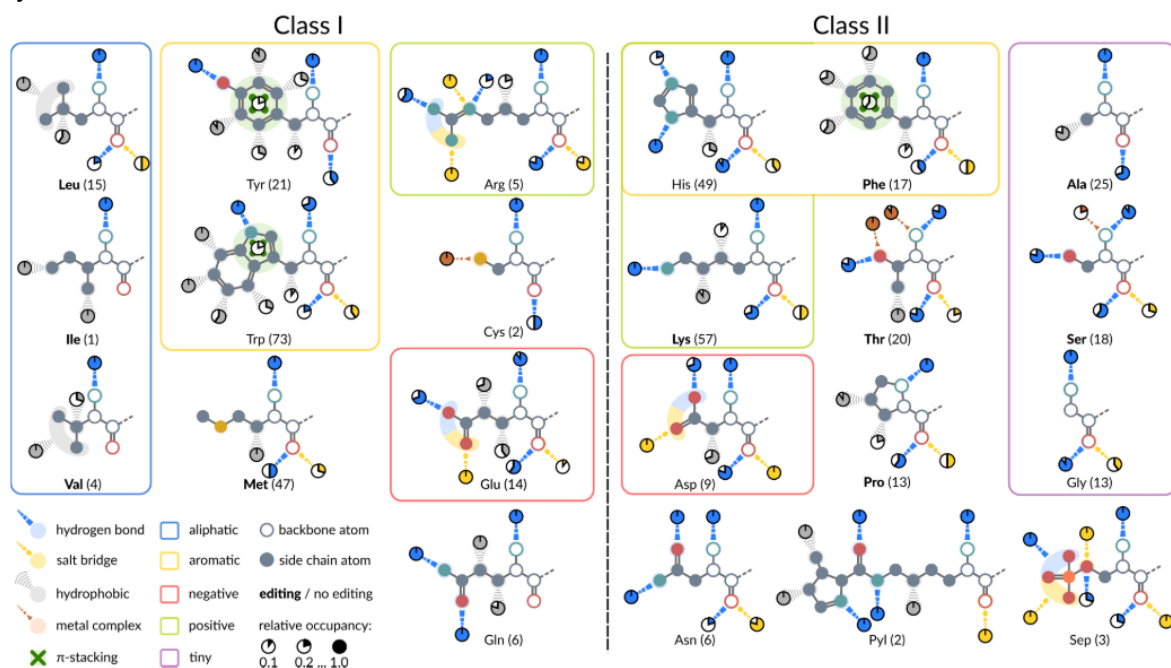


Figure 10: Compilation of the interactions involved in the recognition of proteinogenic amino acids by cognate AARSs. Different types of non-covalent protein-ligand interactions were determined through bioinformatics analyses using all AARSs:amino acids crystal structures available in the PDB (424 structures). The relative frequency of each interaction in respect to the total number of investigated structures (number in parentheses for each AARS) is given by pie charts. Interactions with a frequency below 0.1 are neglected. Interactions for which a unique mapping to an individual atom is not possible due to ambiguous isomorphism, *e.g.* for the side chain of valine, were assigned to multiple atoms. The AARSs conducting error correction *via* editing mechanisms are typeset in bold. Backbone atoms of the amino acids are depicted as circles without filled interior. Reproduced from Kaiser et al., 2019.⁸⁹

Despite this first “sieve” based on interactions between amino acids and the binding pockets, some AARSs do not perfectly discriminate between their cognate amino acids and other structurally related proteinogenic amino acids. For example, *E. coli* IleRS has been shown to bind

valine instead of isoleucine⁹¹ and ProRS is known to produce Ala-tRNA^{Pro92}. In order to ensure the fidelity of the translation, these misacylations are corrected. About half of AARSs have editing domains, distinct of the acylation sites, which recognize misacylated AA-tRNAs and specifically hydrolyse them^{91,93}. In other cases, some proteins distinct from AARSs also achieve this quality control *a posteriori*⁹⁴. An example of such *trans* editing factor is the ProXp-ala family, which recognizes and deacylates Ala-tRNA^{Pro}, without acting on Pro-tRNA^{Pro} and Ala-tRNA^{Ala}⁹⁶.

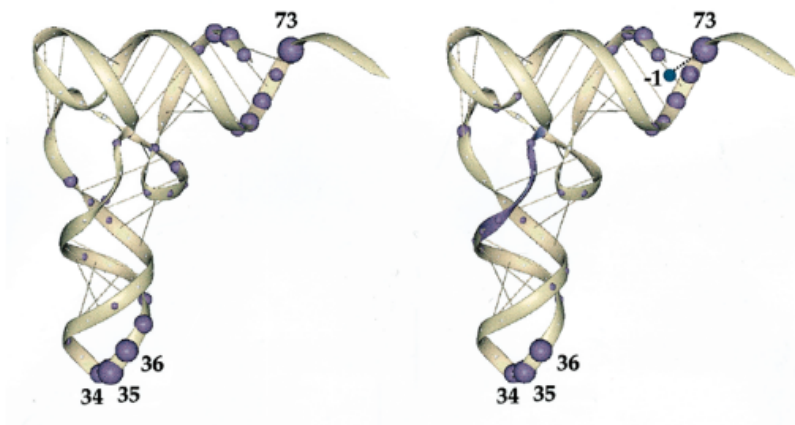


Figure 11: Distribution of identity elements for tRNA aminoacylation in *E. coli*. Spheres are positioned on the nucleotides which act as identity determinants, for tRNAs charged by the 10 class I AARSs (right) and 10 class II AARSs (left). The size of purple spheres is proportional to the frequency of identity nucleotides at a given position (five decreasing sizes of purples coloured spheres corresponding to 9–10-fold, 7–8-fold, 4–6-fold, 2–3-fold and 1-fold presence of an identity element). Spheres are represented in the 3D ribbon model of tRNA^{Phe}. Reproduced from Giegé et al. 1998⁹⁵.

The recognition of the cognate tRNA isoacceptors is also of great importance because the matching between the tRNA anticodon will be used to decipher the mRNA by the ribosome. However, discriminating the cognate tRNA isoacceptors from a pool of dozens of tRNAs which all share similar structural features is highly complex. Therefore, AARSs have evolved sophisticated strategies to ensure that they recognize the cognate tRNAs. Summarizing the huge amount of studies made in this area is not possible here (see for example the review of reference by Giegé et al. 1998⁹⁵). To simplify, it has been shown that some nucleotides of tRNAs have major roles in the recognition: they are identity elements and taken together constitute the “identity set” of each tRNA. These identity sets are sometimes referred as the “second genetic code”. Even if tRNA recognition by AARSs is highly idiosyncratic and strategies differ largely between AARS/tRNA couples, several general rules tend to apply⁸⁰. First, identity elements are mainly located on the acceptor arm and in the anticodon loop (Figure 11). Second, the residue N73 (tRNA^{Phe} numbering),

called the discriminator base, is an identity determinant in almost every tRNA. Finally, when a tRNA exhibits an atypical specific structural feature, this structural element is often an identity element. Besides these positive identity elements, which are required by an AARS to allow aminoacylation, antideterminants were also discovered, which prevent interactions between the AARSs and noncognate tRNAs.

Most AARSs require full tRNAs to perform aminoacylation. However, in a few cases, no identity elements are present in large part of tRNAs and truncated tRNAs can be accepted as RNA acceptors by AARSs. The most documented example is AlaRS, which tolerates with a reduced efficiency tRNA minihelix, composed of the amino-acid acceptor-T Ψ C helix only and even microhelix, corresponding to the seven base pairs of the acceptor arm, as substrates⁹⁷.

B) Different methods to produce non-canonical AA-tRNAs

The amino acid activation machinery has evolved to be very accurate, which contributes to the low error rate of protein translation (estimated at 10^{-3} to 10^{-4} error per codon⁹⁸). However, for some applications, it can be interesting to generate non-canonical AA-tRNAs (canonical or mutated tRNAs loaded with non-cognate or non-canonical amino acids). In this part of the manuscript, the different methods that have been developed to produce non-canonical AA-tRNAs (ncAA-tRNAs) are briefly described, with a focus on the two methods that have been applied to the study of CDPs in this manuscript, *i.e.* residue-specific incorporation using the natural promiscuity of AARSs and flexizymes.

i. Exploiting the natural promiscuity of aminoacyl-tRNA synthetases

In the 1950s⁹⁹, Cowie and Cohen demonstrated that selenomethionine could serve as an effective surrogate for methionine in the whole translational process. This landmark experiment demonstrated that AARSs can exhibit promiscuity towards ncAAs and produce ncAA-tRNAs that are subsequently used by the translational apparatus. AARSs have been challenged through evolution to discriminate between proteinogenic amino acids, but no evolutionary pressure was exerted on the rejection of ncAA which were not present. This observation paved the way for the residue-specific incorporation of ncAAs into proteins, *i.e.* the replacement in a whole protein of one proteinogenic amino acid by a non-canonical surrogate.

In order to extend the chemical diversity accessible through this approach, extensive exploration of the promiscuity of AARSs towards ncAAs was required. Early experiments used indirect experiments to demonstrate that ncAAs could act as substrates of the whole translational machinery, and thus of AARSs. The toxicity of some close analogues of proteinogenic amino acids on bacterial cultures was soon interpreted as a proof that they could penetrate the cell and be incorporated into proteins which were therefore inactive. In 1962, more than 30 ncAAs had been shown to be toxic to bacterial cultures because of their incorporation into proteins, including some natural analogues such as ethionine or canavanine, and some artificial ones, such as fluorinated analogues¹⁰⁰. In 2006, Hartman and colleagues published a landmark study describing a high-throughput *in vitro* screening approach, based on the MALDI-TOF detection of digested, derivatized ncAA-tRNAs generated by AARSs¹⁰¹. They identified 59 new ncAAs as substrates of various *E. coli* AARSs. A thorough review of the literature up to 2017 yielded about 100 ncAAs shown to be substrates of at least one AARS, either through indirect, *in vivo* based experiments,^{102–104} or through *in vitro* assays on purified AARSs¹⁰¹.

The main practical strategy employed for residue-specific incorporation involves the use of auxotrophic strains, unable to produce the proteinogenic amino acid that is replaced by the ncAA. The cells are first grown in a media complemented with the proteinogenic amino acid, in order to restore a normal phenotype and to allow growth and production of the diverse components of the translation apparatus. Then, the media is rinsed, and cells are incubated in a media supplemented with the ncAA to be incorporated. Upon uptake of the ncAA from the media to the cell, the ncAA is recognized by its cognate AARS, loaded on tRNA and the ncAA-tRNA can be used by the translational apparatus. Interestingly, since the biosynthetic pathways for all the proteinogenic amino acids are well established in *E. coli*, auxotrophy for almost all the proteinogenic amino acids can be easily obtained, and strains with such phenotypes sharing common genotypes are easily accessible^{105,106}.

This approach has been used in particular for the incorporation of ncAAs into RiPPs^{104,107} and even for the replacement of tryptophan by a non-canonical surrogate in the whole proteome of *E. coli*¹⁰⁸. The transposition of this strategy to CDPSs is presented in the first chapter of this manuscript. A list of all the ncAAs that have been reported to be loaded by wild-type *E. coli* AARSs can be found in this chapter (Table 1 and Table S6 of the Supplementary Information of Chapter 1).

ii. Engineering of aminoacyl-tRNA synthetases with improved promiscuity

The ncAAs that are substrates of wild type *E. coli* AARSs display an interesting but limited diversity. Thanks to the insights into AARS specificity brought by structural studies, biochemists have attempted to engineer AARSs in order to increase their substrate spectrum.

Engineering of E. coli AARSs

A first strategy is to enlarge the binding pocket of AARSs, thus allowing the accommodation of bulkier substrates. Some interesting results have been obtained with mutations enlarging the amino acid binding pocket of *E. coli* PheRS. The mutation A294G, discovered by Kast and Hennecke¹⁰⁹, allows the loading of several *para*-substituted Phe/Tyr analogues, including some halogenated and click-chemistry-compatible surrogates¹¹⁰ and even one bicyclic aromatic amino acid (Figure 12A)¹¹¹. Another strategy for reaching more promiscuous AARSs consists in deleting the proof-reading activity. In their search for more promiscuous ValRS variants, Döring et al. found a mutation, T222P, which dramatically increases the activity of the mutant ValRS towards non-cognate and non-canonical amino acids^{112,113} (Figure 12B). They demonstrated that this mutation does not increase the promiscuity of the acylation activity of ValRS, but that it impairs the proof-reading, editing activity.

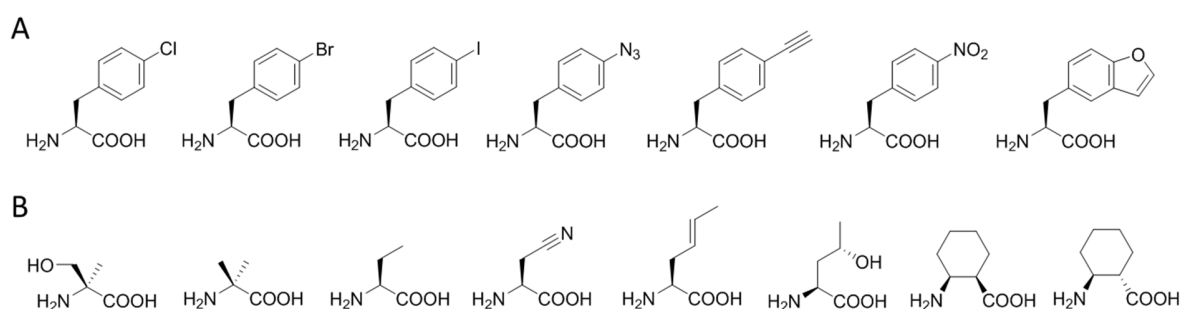


Figure 12: Substrates of *E. coli* AARS variants with improved promiscuity. A) Aromatic ncAAs that are substrates of A294G PheRS. B) Small ncAAs that are substrates of T222P ValRS.

Engineering of heterologous AARS/tRNA pairs with improved promiscuity

Instead of using the host AARS to load ncAA on native tRNAs, another approach is the heterologous expression of an AARS/tRNA pair. A major advantage of this strategy is that it allows the use of mutant tRNAs, designed to target a rarely used codon (most commonly the amber stop codon). AARSs are engineered to have the ability to load an ncAA of interest on a tRNA bearing an amber stop codon. If amber stop codons are introduced in the mRNA sequence, the ncAA-tRNA will suppress the stop codon and will lead to the site-specific incorporation of the ncAA. The main criteria for the engineered AARS/tRNA pair is its orthogonality to the native translational apparatus, which means that no proteinogenic AAs are loaded on the suppressor tRNA by the introduced AARS and that the ncAAs is not substrate of any AARSs of the host.

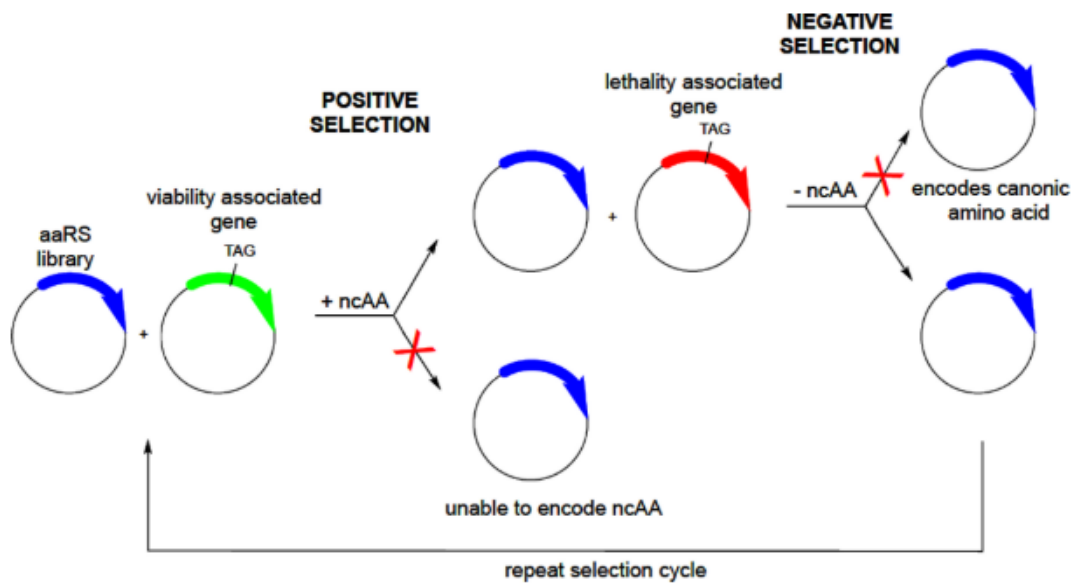


Figure 13: Principle of AARS evolution for the discovery of promiscuous and orthogonal AARS variants. Reproduced from Young et Schultz 2018¹¹⁴.

This method was pioneered by the group of Prof. Schultz, which developed and improved over the years an evolution platform where AARS variant banks are subjected to positive and negative rounds of selection to ensure activity and orthogonality of the evolved system (Figure 13)^{115,116}. Since the AARS must not recognize any of the endogenous tRNAs, the AARS/tRNA pair is

commonly chosen from a different life domain than the host organism. Concerning systems compatible with the *E. coli* machinery (*i.e.* orthogonal to *E. coli* translational system), most successes have been obtained with two AARS/tRNA pairs from archaeal origin: the TyrRS/tRNA^{Tyr} pair from *Methanococcus jannaschii*¹¹⁷ and the pyrrolysine-specific PylRS/tRNA^{Pyl} pair from *Methanosarcina barkeri* and *Methanosarcina mazei*¹¹⁸.

Many excellent reviews comprehensively assess the different ncAAs and corresponding AARS/tRNA pairs that have been developed by this really productive community (see for example Liu et Schultz, 2010¹¹⁹ ; Dumas et al, 2015¹²⁰ ; Young et Schultz, 2018¹¹⁴). More than 150 ncAAs, with side chains bearing a remarkable diversity of chemical reactivity, have been successfully loaded on tRNAs by evolved AARSs. However, the evolution process remains tedious and the diversity of ncAAs that can be loaded by evolved AARSs remains limited. Most of the ncAAs that can be loaded by evolved AARSs are analogues of aromatic amino acids, bearing substitutions on the aromatic lateral chains, or analogues of acetylated-lysine¹²⁰. Comparatively, fewer or no successes have been obtained at generating AARS variants able to load ncAAs with modifications of the amino acid backbone, such as β -amino acids¹²¹ α -hydroxy acids¹²², and N-alkyl amino acids or D-amino acids.

The ability to expand the genetic code and to introduce site-specifically ncAA with novel reactivity opens up fascinating opportunities¹¹⁴. First, it enables new approaches to probe protein structure and function, innovative cross-linking strategies to map protein interactions or new fluorescent probes to label proteins in living cells for example. It also allows the direct production of proteins bearing post-translational modifications in order to better understand and characterize the role of these modifications in proteins. Finally, it allows to introduce ncAAs with bioorthogonal chemical reactivities, which enables the modification of proteins site-specifically with synthetic moieties and paves the way for promising therapeutical applications.

The perspective of combining CDPSs with engineered AARSs with improved promiscuity is discussed in the 'Discussion and Perspectives' section of chapter 1 of the manuscript.

iii. Chemical ligation of amino acids to tRNAs

As seen above, RNA acylation methods based on AARSs allow a limited degree of flexibility in regards of both the amino acids and the RNA acceptor. In some cases, more flexibility is required, for example when heavily mutated/truncated tRNAs or exotic ncAAs are involved or when a modified bond between the amino acid and the RNA is required.

One of the approaches offering more flexibility is to synthesize AA-tRNAs by chemical approaches. However, from a chemist point of view, the regioselectivity of the aminoacylation reaction is challenging: every free hydroxyl groups of a tRNA can potentially be prone to esterification by the amino acid. To ensure that only the 3' terminal adenosine will be esterified, a chemoenzymatic route was developed by the group of Hetch¹²⁴ and improved by the group of Schultz in the 1990s¹²⁵. First, a 5'-phospho-2'-deoxycytidineadenine (pdCpA) is esterified on the terminal adenine. The presence of a 2'-deoxyriboctosine (instead of ribocytosine) prevents the esterification to take place on the free hydroxyl group of the cytidine. Then, this aminoacylated dinucleotide is ligated by T4 RNA ligase to an *in vitro*-produced tRNA missing the terminal cytidine and adenosine moieties (Figure 14). This chemoenzymatic approach was originally proposed for ncAA incorporation into proteins¹²⁵, but it has also been successfully used to study AA-tRNA recognition by transferases from the Fem family^{126,127}.

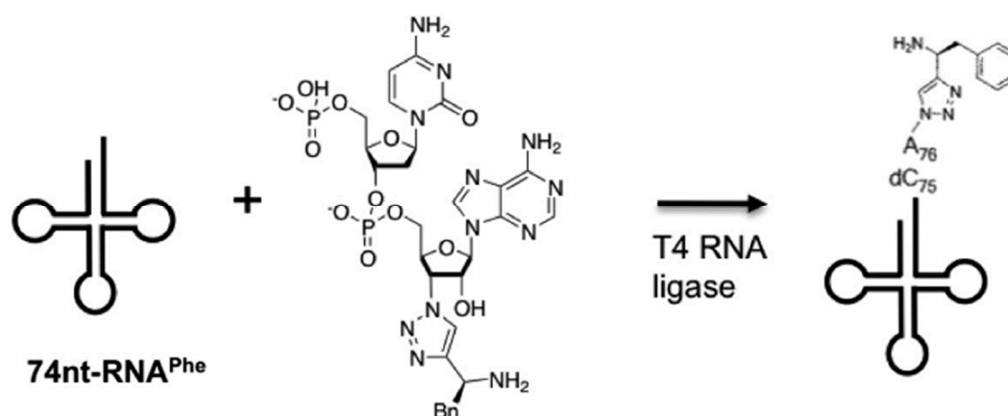


Figure 14: Chemoenzymatic synthesis of a Phe-tRNA^{Phe} mimic bearing a triazole ring instead of the canonical ester bond. Reproduced from Santarem et al., 2014¹²³.

The major advantage of this approach is its flexibility: it allows to load any ncAA on RNA acceptors of any sequence and length. Interestingly, it also allows to modify the nature of the bond between the amino acid and the terminal ribose. The ester bond can for example be replaced by a triazole ring, which mimics the geometry of the ester bond and leads to stable, non-isomerizable analogues useful in biophysical studies¹²⁸ (Figure 14). Such molecules were proposed as substrate analogues for the study of CDPSs but technical difficulties in obtaining the substrates in sufficient amount limited this application. Additionally, preliminary experiments suggested that the absence of the 2'hydroxyl in the C75 limited the interaction between the substrate analogue and the CDPS.

- iv. Versatile RNA acylation using aminoacyl-tRNA synthetase-like ribozymes, the flexizymes

In the 1980s, Nobel Prize winners Prof. Altman and Prof. Cech demonstrated that some small RNA oligonucleotides could exhibit a catalytic activity. These “RNA enzymes” were called ribozymes and were seen as indicators of a pre-protein “RNA-world”, where RNA was the basis of both the catalysis and the heredity. Several classes of ribozymes have been found in living beings¹²⁹. Apart from the ribosome that can be seen as a ribozyme¹³⁰, most natural ribozymes are RNA-processing catalysts. Some ribozymes are self-modifying catalysts, such as the self-cleaving hammerhead¹³¹ or hepatitis delta virus (HDV)¹³² ribozymes, whereas others act in *trans*, like true catalysts, such the RNase P¹³³ involved in tRNA maturation.

Soon after the discovery of natural RNA-processing ribozymes, researchers investigated the possibility to design artificial ribozymes that could perform other reactions than RNA-cleavage¹³⁴. They were inspired by the field of aptamers, oligonucleotides having high-affinity towards specific ligands that can be selected from large combinatorial RNA libraries¹³⁵. The method of reference for aptamers/ribozymes selection is based on iterative rounds of selection and amplification and is called SELEX, for Systematic Evolution of Ligands by EXponential enrichment^{136,137}. By adapting the selection steps of SELEX to screen for a particular catalytic activity, several classes of artificial ribozymes were evolved.

Discovery, optimisation and description of flexizymes

Attempts to apply the SELEX methodology to the discovery of aminoacylating-ribozymes started in the 1990s. Several groups reported the discovery of ribozymes with RNA aminoacylating activities^{138–140} but the first report of an actual tRNA-aminoacylating ribozyme acting in *trans* was made by the group of Prof. Suga in the early 2000s¹⁴¹. Starting from a randomized sequence fused at the 5' terminus of a tRNA, several rounds of SELEX allowed the identification of a ribozyme which specifically aminoacylates the terminal adenosine of a tRNA. This ribozyme used as substrate a phenylalanine activated with a cyano methyl ester (CME) on its carboxylic acid. This promising ribozyme was then thoroughly studied and sequence-optimized through rational engineering and new rounds of randomisation/evolution¹⁴². This led to the flexizyme system, a set of three small RNAs (45 to 47 nucleotides) that recognize tRNAs through interaction with their conserved 3' terminus and that catalyze the transfer of activated amino acids to the terminal adenosine of tRNA or tRNA-like molecules (Figure 15A).

By testing different tRNA acceptors, Prof. Suga and his group demonstrated that the interaction between flexizymes and acceptor RNAs relies on base pairing of only three nucleotides, involving the conserved 5'-XCCA-3' terminus of tRNA and the 5'-GGY-3' terminus of flexizyme (for optimal recognition, Y must be chosen as the complementary of X)¹⁴³. This was confirmed by the crystal structure of a flexizyme:tRNA-minihelix fusion, reported in 2008 (Figure 15B and 15C).¹⁴⁴ As a consequence, flexizymes have the ability to accept any tRNA sequences. Interestingly, tRNA-like truncated RNAs, such as mini-/micro-tRNA helix, or even some 4-mer RNA oligos mimicking the 5'-XCCA-3' terminus of tRNAs are accepted by flexizymes as efficiently as full tRNAs¹⁴⁵.

Flexizymes use as substrates activated amino acids, bearing activating groups esterified to the backbone carboxylic acids. The recognition of the amino acids by flexizymes relies on the presence of an aromatic cycle, which is found either in the lateral chain or in the activating group. Different flexizymes have been selected to work best with different activating groups. Their sequences are very similar, except for mutations of a few nucleotides in the amino acid binding pocket (Figure 15D). For non-aromatic amino acid, the most common choice is the use of a dinitrobenzyl ester (DBE) as activating group (Figure 15E) which has to be used with the dFx flexizyme. For amino acids bearing an aromatic ring in their lateral chain, a non-aromatic activating group, a cyanomethyl ester (CME) (Figure 15E) can be used, in combination with the

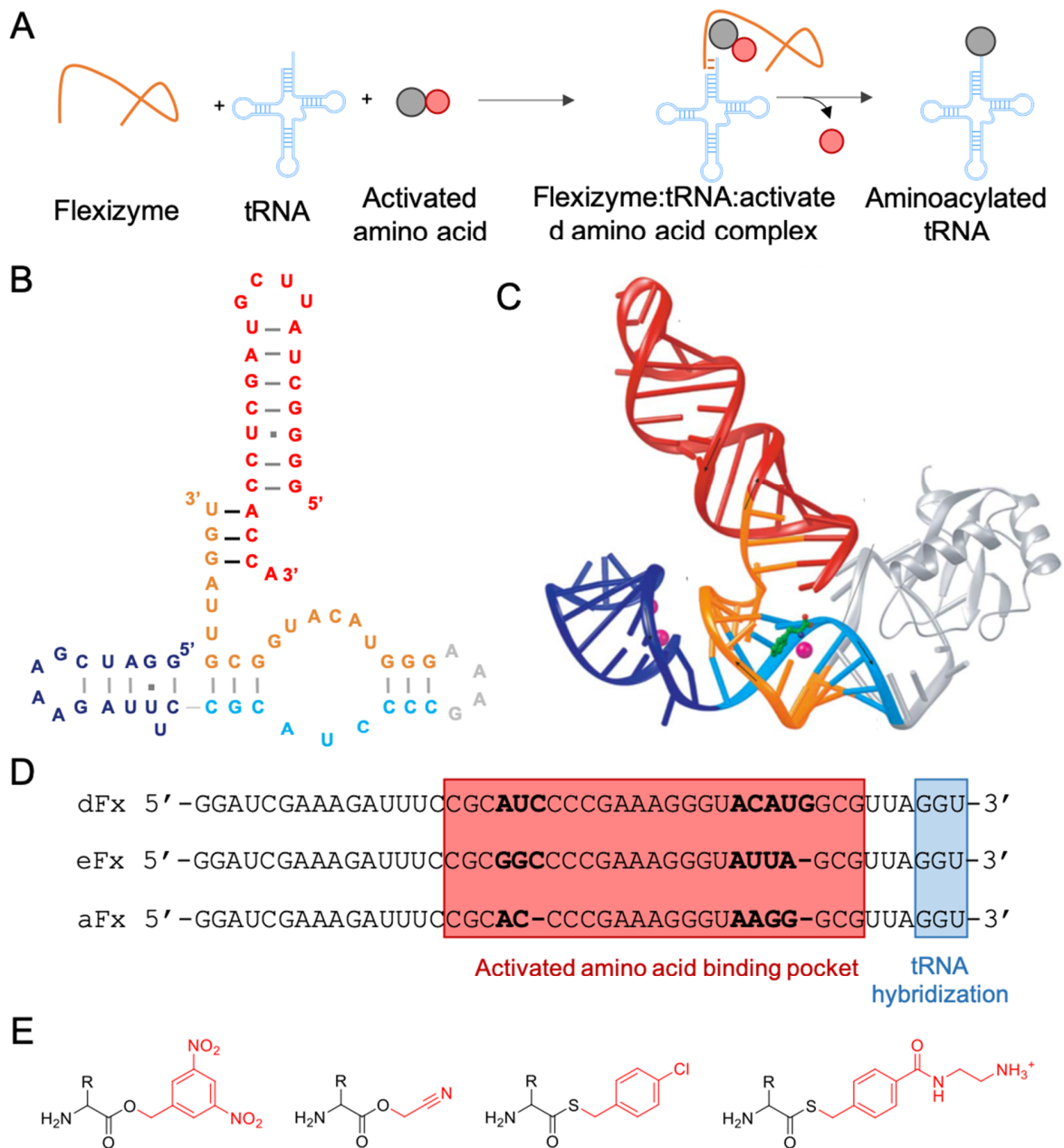


Figure 15: The flexizyme system. A) Principle of flexizyme-dependent aminoacylation. B) Secondary structure of the complex between dFx and a tRNA microhelix mimicking the acceptor arm of *E. coli* tRNA^{Ala} (see Chapter 2) using a similar colour code as in C. C) 3D structure of a flexizyme:miniRNA complex obtained by crystallography (PDB: 3CUN, figure from Xiao et al., 2008¹⁴⁴). Magnesium ions are in pink and a mimick of Phe-CME is shown in green. Note that the sequence used for 3D structure determination does not correspond to a functional flexizyme:miniRNA complex (fusion between the flexizyme and the acceptor RNA and addition of a linker to a protein domain to facilitate crystallisation). D) Sequences of the three different flexizymes. Nucleotides in bold correspond to mutations in the activated amino acid binding pockets between the different flexizymes. E) The four different activating groups used by flexizymes, from left to right, DBE, CME, CBT and ABT.

flexizyme eFx. This flexizyme is also able to use non-aromatic amino acids activated with a chlorobenzyl thioester (CBT) group (Figure 15E), it can be tested if the combination DBE/dFx leads to poor aminoacylation yields, for example for amino acids with bulky side chains. Finally, a last activating group have been developed to address the issue of the low solubility of some particularly hydrophobic amino acids. This aromatic group, (2-aminoethyl)-amidocarboxybenzyl thioester (ABT) (Figure 15E), bears a positive charge at neutral pH, which improves solubility of activated amino acids¹⁴⁶ and can be used with another flexizyme sequence, aFx. Using one of these four activating groups with the cognate flexizyme sequence, an almost infinite diversity of ncAAs or amino acid-like molecules can be loaded on acceptor RNAs. This include amino acids with bulky side chains¹⁴⁷, D-amino acids^{148,149}, β -amino acids^{150,151}, N-alkyl amino acids^{152,153}, α -hydroxy amino acids¹⁵⁴ and even peptides^{155,156}.

Applications of flexizymes

Thanks to its versatility towards acceptor RNAs and acyl donors, flexizymes have proven to be a useful tool and has been applied in a number of studies. The most successful application of flexizymes has been the ribosomal production of peptides and proteins containing ncAAs. Prof. Suga and his team developed a flexible *in vitro* translation system (or FIT), in which the withdrawal of amino acid/AARS pairs leads to codon vacancies that can be filled using ncAA-tRNAs produced by flexizymes¹⁴². Using this method, the translational apparatus was shown to be highly tolerant and to be able to incorporate several hundred ncAAs. Unlike other methods, flexizymes allow the incorporation of ncAA bearing modifications of the backbone itself, such as D-amino acids^{148,149}, β -amino acids^{150,151}, N-alkyl amino acids^{152,153} or α -hydroxy amino acids¹⁵⁴. In the case of some particularly challenging peptides (such as those with several consecutive D-amino acids), an exciting strategy is the engineering of tRNA sequence which is allowed by flexizyme versatility towards RNA sequences and can dramatically increase yield in some cases¹⁵⁷⁻¹⁵⁹. Such backbone-modified ncAAs enlarges significantly the chemical space accessible through ribosomal peptide production. The combination of this FIT system with mRNA libraries can yield peptide libraries with tremendous diversity ($> 10^{12}$ molecules), which are highly attractive for drug discovery¹⁶⁰. This was the basis for Peptidream, a highly successful drug-discovery company founded by Prof. Suga.

Apart from their use at generating substrates of the translational apparatus, flexizymes can be used to generate diverse AA-tRNAs for other applications. They have been used to study the stability of AA-tRNAs towards spontaneous desacylation⁸⁴ or the enzymatic deacylation of misloaded AA-tRNAs^{96,161}. An exciting application of the flexizyme technology would be to generate AA-tRNAs and AA-tRNA analogues for the study of AA-tRNA utilizing enzymes. The second chapter of this manuscript presents an application of flexizymes to the study of RNA recognition by CDPSs.

III) Objectives of this thesis

The diversity of cyclodipeptides normally accessible with CDPSs is limited. Given the nature of their substrates, CDPSs can only use the 20 proteinogenic amino acids in a normal physiological context. In order to unlock the synthetic potential of CDPSs, an attractive strategy is therefore to expand the range of precursors that they can use. However, at the beginning of my thesis, the tolerance of CDPSs towards non-natural substrates remained to be investigated.

In the first part of my PhD, I tested the natural promiscuity of CDPSs towards a wide range of ncAAs. NcAAs were attached on tRNAs *in vivo*, using the promiscuity of *E. coli* wild-type AARSs, and the corresponding ncAA-tRNAs were tested as substrates for CDPSs. The encouraging results obtained (see Chapter 1) could have been pursued by using modified AARSs with increased promiscuity (see Section II.B.ii. of the introduction) or by coexpressing 2,5-DKP tailoring enzymes (see Section I.B.ii. of the introduction).

However, we chose to take advantage of the second part of my PhD to pursue a more challenging approach, aiming at refining our understanding of how CDPSs recognize their substrates. Indeed, our comprehension of CDPS substrate recognition is still not complete: structural studies has allowed the identification of the binding sites for the aminoacyl moieties of AA-tRNA substrates but the recognition of the tRNA moieties is still poorly understood. Better understanding the residues of CDPSs and the tRNA nucleotides involved in tRNA/CDPS interaction is a prerequisite to the engineering of CDPSs with more relaxed specificity or to the combination of CDPSs with orthologous AARS/tRNA pairs. To get more insight into this interaction, I implemented the flexizyme technology in my host lab and took advantage of its versatility to produce truncated AA-tRNA analogues of diverse sizes. By testing the activity of a CDPS on these reduced substrates, I improved our understanding of tRNA/CDPS interaction and paved the way for interesting further biochemical and biophysical studies. This should facilitate the manipulation of CDPSs for the production of new non-canonical cyclodipeptides.

CHAPTER 1: HIJACKING THE HIJACKERS: INCORPORATION OF NON-CANONICAL AMINO ACIDS INTO 2,5-DIKETOPIPERAZINES BY CYCLODIPEPTIDE SYNTHASES

I) Introduction

Diversifying peptide natural products by introducing ncAAs is a common strategy to deliver molecules with improved properties^{162–164}. NcAAs are present in some natural 2,5-DKPs and can be necessary for their activities, as for example the nitrated tryptophan moiety in thaxtomins. Additionally, medicinal chemists have used ncAAs to generate 2,5-DKPs with improved properties. The idea that CDPSs could be used to incorporate ncAAs into cyclodipeptides was suggested soon after the discovery of CDPSs^{24,165}, but no attempt was reported prior to my arrival in the group of Muriel Gondry.

The most straightforward way to generate ncAA-tRNAs and to test them as substrates for CDPSs is to perform the whole experiment *in vivo*. As described in the introduction, two strategies were possible: either exploiting the promiscuity of *E. coli* AARSs by using auxotrophic strains or expressing orthologous engineered AARS/tRNA pairs. The first option was a good start, because it required no additional AARS expression and it allowed us to test different classes of ncAAs and therefore to take advantage of the diversity of CDPS specificity patterns.

This chapter is presented as an article published in *Angewandte Chemie International Edition*.

Author contributions:

Prior to my arrival in the group, Pascal Belin had realized the proof-of-concept that a few CDPSs producing proline-containing cyclodipeptides could incorporate some proline analogues, using an *in vivo* strategy close to the one presented herein after. This study consisted in applying this concept to a larger scale. To do so, I gathered a bank of ncAAs and a bank of auxotrophic

strains sharing the same genetic background. Combinations of ncAAs and CDPSs to be tested were chosen with the advice of Mireille Moutiez. I optimized the analytical conditions in collaboration with Robert Thai and organized the LC-MS analysis pipeline so that it was compatible with the throughput this project required. I realized all the bacterial cultures and LC-MS analyses, including scale-up and purification for three cyclodipeptides. Isabelle Correia and Olivier Lequin realized the NMR characterization of these three cyclodipeptides. Muriel Gondry and I wrote the manuscript which was reviewed and approved by all authors.

II) Article

Incorporation of Non-canonical Amino Acids into 2,5-Diketopiperazines by Cyclodipeptide Synthases

Nicolas Canu, Pascal Belin, Robert Thai, Isabelle Correia, Olivier Lequin, Jérôme Seguin, Mireille Moutiez,* and Muriel Gondry*

Abstract: The manipulation of natural product biosynthetic pathways is a powerful means of expanding the chemical diversity of bioactive molecules. 2,5-diketopiperazines (2,5-DKPs) have been widely developed by medicinal chemists, but their biological production is yet to be exploited. We introduce an *in vivo* method for incorporating non-canonical amino acids (ncAAs) into 2,5-DKPs using cyclodipeptide synthases (CDPSs), the enzymes responsible for scaffold assembly in many 2,5-DKP biosynthetic pathways. CDPSs use aminoacyl-tRNAs as substrates. We exploited the natural ability of aminoacyl-tRNA synthetases to load ncAAs onto tRNAs. We found 26 ncAAs to be usable as substrates by CDPSs, leading to the enzymatic production of approximately 200 non-canonical cyclodipeptides. CDPSs constitute an efficient enzymatic tool for the synthesis of highly diverse 2,5-DKPs. Such diversity could be further expanded, for example, by using various cyclodipeptide-tailoring enzymes found in 2,5-DKP biosynthetic pathways.

The smallest of all cyclic peptides, 2,5-diketopiperazines (2,5-DKPs), are formed by the condensation of two α -amino acids.^[1] They constitute a large class of microbial natural products, and many possess noteworthy biological and pharmacological activities.^[1,2] The highly stable and constrained 2,5-DKP-ring has long been recognized as a privileged structural scaffold. Generating diversity around this scaffold has proven to be an efficient strategy to generate 2,5-DKPs with novel or improved bioactivities.^[1,3]

The biosynthesis of 2,5-DKPs was formerly thought to be catalyzed by the gigantic nonribosomal peptide synthetases.^[4] We identified another biosynthetic route containing the small cyclodipeptide synthases (CDPSs) responsible for 2,5-DKP-ring assembly.^[5] As substrates, CDPSs use activated L-amino acids in the form of aminoacyl-tRNAs (aa-tRNAs; Figure 1,

left),^[6,7] which are formed by aminoacyl-tRNA synthetases (AARSs) to feed the translational machinery, and they act through a ping-pong catalytic mechanism for the stereospecific formation of various cyclodipeptides [cyclo(L-AA1-L-AA2), with AA1 and AA2 corresponding to the two

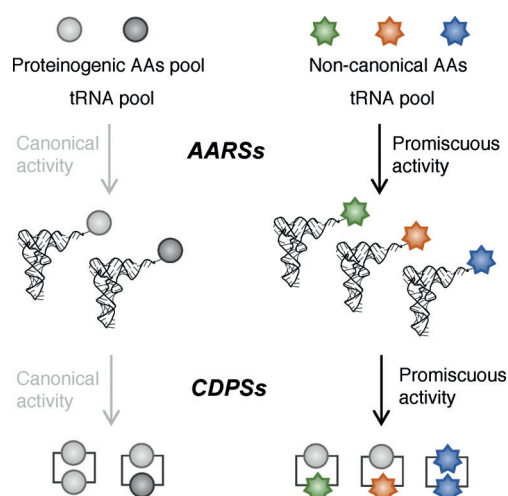


Figure 1. General principle of incorporation of non-canonical amino acids by CDPSs. Left: proteinogenic AAs are loaded onto tRNAs by AARSs. CDPSs divert aa-tRNAs from the translational machinery to produce cyclodipeptides. Right: ncAAs are added to the culture medium, taken up by the cells, and acylated on the cognate tRNAs by AARSs. ncAA-tRNAs serve as substrates for CDPSs to produce non-canonical 2,5-DKPs.

incorporated aminoacyl moieties; these are hereafter referred to as cyclo(AA1-AA2)).^[6–10] CDPSs are good candidates for the biological production of 2,5-DKPs because their heterologous expression in *Escherichia coli* is easy to implement and leads up to high amounts of cyclodipeptides recovered in culture supernatants. CDPSs are often found within biosynthetic gene clusters containing diverse tailoring enzymes responsible for further chemical modifications of the produced cyclodipeptides.^[4,11] However, the diversity of 2,5-DKPs arising from CDPS-based pathways appears to be intrinsically limited, due to the restriction of potential precursors to the 20 proteinogenic amino acids.

We investigated whether CDPSs can incorporate non-canonical amino acids (ncAAs) into their products. CDPSs exhibit diverse specificity patterns, thus making them promising tools to generate diversity.^[12] CDPSs use aa-tRNAs as substrates, and thus the chosen strategy included the loading of ncAAs onto tRNAs. Several methods have been developed

[*] N. Canu, Dr. P. Belin, Dr. J. Seguin, Dr. M. Moutiez, Dr. M. Gondry
Institute for Integrated Biology of the Cell (I2BC), CEA, CNRS
Univ. Paris-Sud, Université Paris-Saclay
91198 Gif-sur-Yvette cedex (France)
E-mail: mireille.moutiez@cea.fr
muriel.gondry@cea.fr

Dr. R. Thai
SIMOPRO, CEA-Saclay, 91198 Gif-sur-Yvette cedex (France)

Dr. I. Correia, Prof. O. Lequin
Sorbonne Université, Ecole Normale Supérieure, PSL University
CNRS, Laboratoire des Biomolécules (LBM), 75005 Paris (France)

Supporting information (including experimental methods) and the ORCID identification number(s) for the author(s) of this article can be found under:
<https://doi.org/10.1002/anie.201712536>.

to aminoacylate tRNA with ncAAs.^[9,13–15] Here, we describe an assay exploiting the promiscuity of AARSs for the medium-throughput screening of CDPS-dependent ncAA incorporation into cyclodipeptides (Figure 1, right).

We developed an *in vivo* method to maximize the throughput of the assay. It relies on the *E. coli* biosynthetic machinery for CDPSs synthesis and ncAAs loading onto tRNAs. We used auxotrophic strains to deplete endogenous production of proteinogenic amino acids to be replaced by ncAAs, in a manner analogous to that used for residue-specific incorporation of ncAAs into proteins and ribosomal peptides.^[16] Strains of *E. coli* auxotrophic for one or several proteinogenic amino acids, were obtained from the Keio Collection (Table S1 in supporting information),^[17] and transformed with plasmids encoding CDPSs (Table S2).^[12] We screened a set of 60 commercial ncAAs among the more than

100 ncAAs reported to be substrates of *E. coli* AARSs.^[13,16,18–20] (Table S3).

A ncAA incorporation assay was performed using a two-step culture method for each selected CDPS (Part I of the Supporting Information, Figure S1). First, cells were grown in minimum medium supplemented with the proteinogenic amino acid required to complement the auxotrophy. After CDPS expression, the culture was rinsed, resuspended in non-supplemented minimum medium and divided into 24-well plates. One different ncAA was added to each well. Second, the cells were incubated overnight to allow the production of non-canonical cyclodipeptides and their release into the culture media. The cyclodipeptides released were then detected, identified, and quantified by LC-MS/MS (Part II of Supporting Information).

In one set of experiments (Figure 2), 10 CDPSs were selected for their ability to synthesize proline-containing

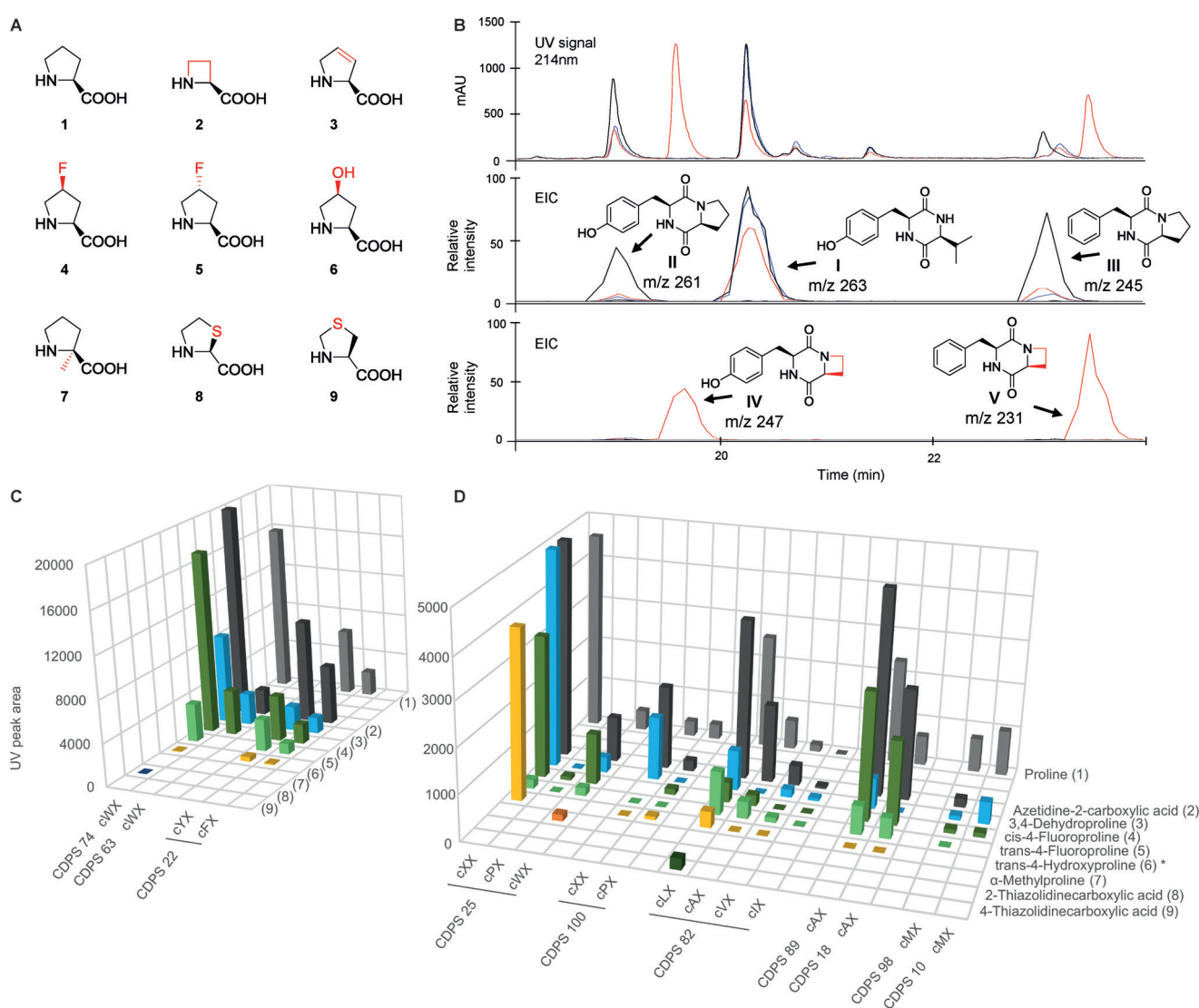


Figure 2. *In vivo* enzymatic synthesis of 2,5-DKPs containing proline analogues by CDPSs. A) The proline analogues that were incorporated. B) LC-MS analysis of cultures of *E. coli* Pro- strain expressing CDPS 22. Cultures were complemented with either proline (black), azetidine carboxylic acid (red), or no AA (blue). EIC profiles corresponding to the m/z of the canonical cyclodipeptides (middle) and ncAA-containing cyclodipeptides (bottom) are displayed. C, D) Histograms showing UV peak areas of non-canonical cyclodipeptides produced by the set of 10 CDPSs. * Low cellular uptake of hydroxyproline was improved using osmotic-stress conditions.

cyclodipeptides: they produce one or several cyclo(Xaa-Pro) (Xaa = Phe, Tyr, Trp, Leu, Ile, Val, Ala, Met, or Pro), with a few also producing other cyclodipeptides. We tested their capacity to incorporate eight different proline analogues in place of proline (Figure 2A). An example of the analytical approach and the results obtained is given in Figure 2B, which shows the replacement of proline (Pro, **1**) with azetidino-2-carboxylic acid (Aze, **2**) in the cyclodipeptides produced by a CDPS from *Diplorickettsia massiliensis* (CDPS 22). UV chromatograms and extracted ion chromatograms (EIC) obtained for supernatants of cultures complemented with Pro, Aze, or no amino acid are shown. The expected cyclodipeptides were produced upon proline complementation, namely cyclo(Tyr-Val) (compound **I**), cyclo(Tyr-Pro) (**II**), and cyclo(Phe-Pro) (**III**). Without proline complementation, the production of cyclo(Tyr-Pro) and cyclo(Phe-Pro) was markedly reduced, whereas the production of cyclo(Tyr-Val) remained unchanged: this confirms the efficiency of the auxotroph-based depletion strategy. The addition of Aze to the medium led to the appearance of two new peaks, with *m/z* corresponding to putative cyclo(Tyr-Aze) (**IV**) and cyclo(Phe-Aze) (**V**). The mass fragmentation analysis of these compounds allowed their unambiguous identification as the expected ncAA-containing cyclodipeptides (Figure S2).^[12,21,22] The production of cyclo(Tyr-Val) was slightly decreased by the addition of Aze, which may reflect competition of Aze over Val.

Figure 2C shows the incorporation of the eight proline analogues by the 10 selected CDPSs. Five ncAAs (**2–6**) were incorporated by nearly all tested CDPSs, with varying efficiency depending on the CDPS and the ncAA. Every expected non-canonical cyclodipeptide was formed for three (**2–4**). One or two cyclodipeptides were not detected for the remaining two (**5** and **6**), but the corresponding canonical cyclodipeptides were only poorly produced. Overall, we identified 47 of the 50 non-canonical cyclodipeptides that could potentially result following incorporation of compounds **2–6**. Compounds **7–9** were not efficiently incorporated, since each was incorporated by only one CDPS, at very low levels.

We extended this method to other sets of ncAAs: analogues of phenylalanine and tyrosine (17 ncAAs tested on 9 CDPSs), tryptophan (10 ncAAs tested on 6 CDPSs), and aliphatic amino acids (25 ncAAs tested on 16 CDPSs). We sometimes used different combinations of ncAAs and auxotrophic strains to specifically replace one of the amino acids incorporated by the CDPSs.

Eighteen ncAAs (in addition to proline analogues) were successfully incorporated (Figure 3), whereas we detected no incorporation for the other 34 (Table S6). We obtained 73 of the 84 potentially expected ncAA-containing cyclodipeptides for the set of tyrosine and phenylalanine analogues, 23 of 36 for the set of tryptophan analogues, and 52 of 84 for the set of aliphatic amino acid analogues (Figure 4). A total of 198 ncAA-containing 2,5-DKPs were produced and identified by LC-MS/MS (Tables S4, S5, Figures S3, S4).

Of the 198 cyclodipeptides, 80 were fluorinated, 21 hydroxylated, and 37 differed from their canonical counterparts by the presence of a methyl group. The remaining 60

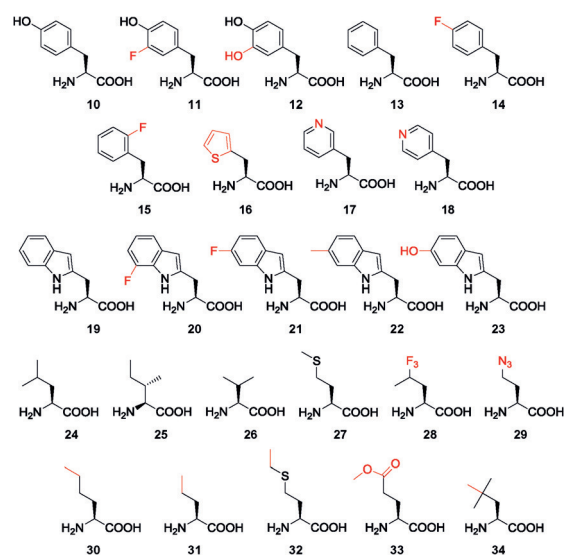


Figure 3. ncAAs incorporated (in addition to proline analogues).

exhibited a wide range of modifications. This is the first reported enzymatic synthesis for almost all of these cyclodipeptides. Of note, azidohomoalanine **29** (Aha), an amino acid compatible with click chemistry reactions, was incorporated with good efficiency by CDPSs, leading to nine different cyclodipeptides.

Our results demonstrate that CDPSs can accept a significant number of non-natural substrates, thus making them an efficient tool for generating a valuable diversity of 2,5-DKPs. However, a large number of ncAAs (36 of 60) were not incorporated (Table S6), regardless of the CDPSs used. This raises questions concerning the potential hindrance to their incorporation.

The substrate specificity of the CDPSs is presumably involved in the rejection of some ncAAs, in particular those incorporated at high levels into proteins and ribosomal peptides using in vivo auxotrophy-based strategies similar to ours. The ongoing investigation of the molecular bases of CDPS substrate specificity^[8,23] will be crucial to the engineering of more promiscuous CDPSs to further increase the chemical diversity accessible with these enzymes.

Insufficient levels of ncAA-tRNAs may also explain the absence of incorporation observed for some ncAAs. ncAA uptake may be deficient. *trans*-4-Hydroxyproline (**6**) is reported to be poorly taken up in *E. coli*.^[24] Using osmotic-stress conditions (600 mM NaCl) that have been shown to increase the uptake of **6**,^[25,26] we were able to efficiently improve the incorporation of this compound into cyclodipeptides (Figure 2C; Figure S5). We implemented this strategy for other poorly incorporated ncAAs likely to use the same import system, such as 2-thioprolino (**8**) and 3-thioprolino (**7**) or small aliphatic amino acids,^[24] but this did not lead to any observable improvements. Metabolic or chemical instability of the ncAAs under intracellular conditions could also hinder the loading of ncAAs by AARSs. All of the ncAAs tested have been reported to be loaded by *E. coli* AARSs but loading efficiencies have been quantified mostly by in vitro assays.^[13,16,18–20] In fact, there does not appear to be any global

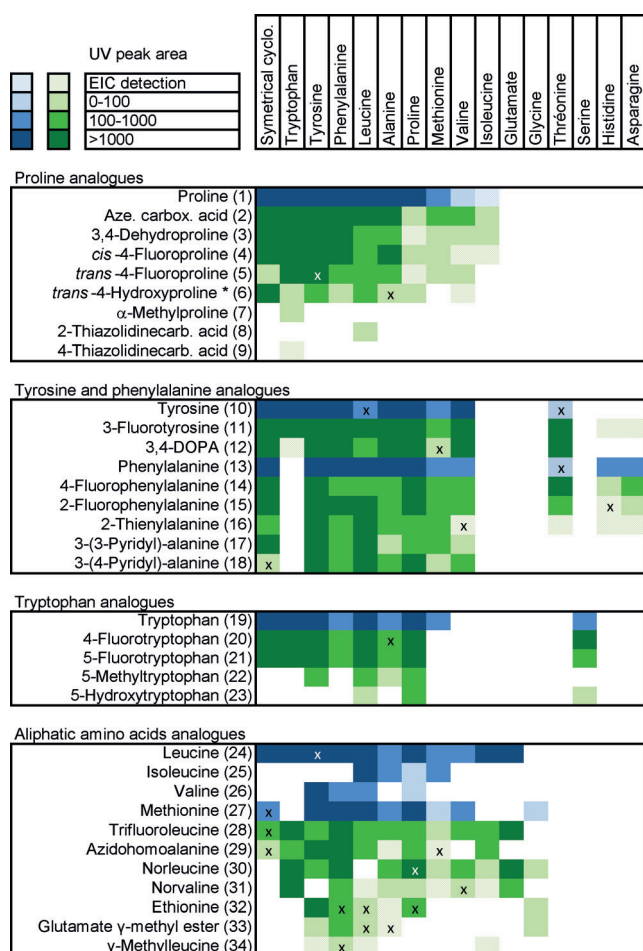


Figure 4. Overview of the diversity created by CDPs-dependent nCAA incorporation. Each colored cell represents a compound produced. Cells are colored according to the UV peak area measured (highest area obtained among the different CDPs and auxotrophic strains couples tested). Gradients of blue and green refer to canonical and non-canonical cyclodipeptides, respectively. Compounds marked with an x could not be completely separated under our chromatographic conditions: an estimation of the UV peak area is given. * Low cellular uptake of hydroxyproline was improved using osmotic stress conditions.

correlation between low loading efficiencies of nCAAs by AARSs and absence of incorporation in our assay (Table S6). Overexpression of AARS variants engineered to exhibit higher efficiency towards nCAAs^[27] may further increase the quantity of substrate available for CDPs.

In summary, we have developed a method for screening for nCAA incorporation into cyclodipeptides by CDPs, which allowed the production of significant amounts of many non-canonical cyclodipeptides. This method can be easily adapted to the production of cyclodipeptides in sufficient quantities to conduct further characterizations, as exemplified with the full molecular structure characterization by NMR of cyclo(5-FluoroTrp-Pro), cyclo(Aha-Phe) and cyclo(Aha-Leu) (Part IV of Supporting Information). This work increases the diversity of cyclodipeptides accessible through enzymatic synthesis and provides a new method to produce a large variety of 2,5-DKPs. The diversity obtained

could be further increased through the action of tailoring enzymes present in 2,5-DKPs biosynthetic pathways. In addition, some of the compounds produced are interesting intermediates for chemical hemisynthesis, in particular click chemistry reactions.^[16]

Acknowledgements

This work was supported by the CEA, the CNRS, the Paris-Sud University, and two grants from the French National Research Agency (ANR-14-CE09-0021; ANR-16-CE29-0026). N.C. was supported by a doctoral fellowship from the CEA. We thank Dr. Jean-Christophe Cintrat for helpful advice concerning cyclodipeptide purification.

Conflict of interest

The authors declare no conflict of interest.

Keywords: biocatalysis · cyclodipeptides · diketopiperazines · natural products · non-canonical amino acids

How to cite: *Angew. Chem. Int. Ed.* **2018**, *57*, 3118–3122
Angew. Chem. **2018**, *130*, 3172–3176

- [1] A. D. Borthwick, *Chem. Rev.* **2012**, *112*, 3641–3716.
- [2] C. Prasad, *Peptides* **1995**, *16*, 151–164.
- [3] Y. Hayashi, Y. Yamazaki-Nakamura, F. Yakushiji, *Chem. Pharm. Bull.* **2013**, *61*, 889–901.
- [4] P. Belin, M. Moutiez, S. Lautru, J. Seguin, J.-L. Pernodet, M. Gondry, *Nat. Prod. Rep.* **2012**, *29*, 961–979.
- [5] S. Lautru, M. Gondry, R. Genet, J.-L. Pernodet, *Chem. Biol.* **2002**, *9*, 1355–1364.
- [6] M. Gondry, L. Sauguet, P. Belin, R. Thai, R. Amouroux, C. Tellier, K. Tiphile, M. Jacquet, S. Braud, M. Courçon, et al., *Nat. Chem. Biol.* **2009**, *5*, 414–420.
- [7] M. Moutiez, P. Belin, M. Gondry, *Chem. Rev.* **2017**, *117*, 5578–5618.
- [8] M. Moutiez, E. Schmitt, J. Seguin, R. Thai, E. Favry, P. Belin, Y. Mechulam, M. Gondry, *Nat. Commun.* **2014**, *5*, 5141.
- [9] T. W. Giessen, F. Altegoer, A. J. Nebel, R. M. Steinbach, G. Bange, M. A. Marahiel, *Angew. Chem. Int. Ed.* **2015**, *54*, 2492–2496; *Angew. Chem.* **2015**, *127*, 2522–2526.
- [10] J. B. Patteson, W. Cai, R. A. Johnson, K. C. Santa Maria, B. Li, *Biochemistry* **2017**, <https://doi.org/acs.biochem.7b00943>.
- [11] T. W. Giessen, M. A. Marahiel, *Front. Microbiol.* **2015**, *6*, 1–11.
- [12] I. B. Jacques, M. Moutiez, J. Witwinowski, E. Darbon, C. Martel, J. Seguin, E. Favry, R. Thai, A. Lecoq, S. Dubois, et al., *Nat. Chem. Biol.* **2015**, *11*, 721–731.
- [13] N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann, R. Huber, *Eur. J. Biochem.* **1995**, *230*, 788–796.
- [14] L. Wang, J. Xie, P. G. Schultz, *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249.
- [15] H. Murakami, A. Ohta, H. Ashigai, H. Suga, *Nat. Methods* **2006**, *3*, 357–359.
- [16] F. Oldach, R. Altoma, A. Kuthning, T. Caetano, S. Mendo, N. Budisa, R. D. Süßmuth, *Angew. Chem. Int. Ed.* **2012**, *51*, 415–418; *Angew. Chem.* **2012**, *124*, 429–432.
- [17] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, et al., *Mol. Syst. Biol.* **2006**, *2*, 1–11.
- [18] K. Kirshenbaum, I. S. Carrico, D. A. Tirrell, *ChemBioChem* **2002**, *3*, 235–237.

- [19] M. C. T. Hartman, K. Josephson, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4356–4361.
- [20] C. Fan, J. M. L. Ho, N. Chirathivat, D. Söll, Y. S. Wang, *ChemBioChem* **2014**, *15*, 1805–1809.
- [21] A. J. C. Furtado, R. Vessecchi, J. C. Tomaz, S. Galembeck, J. Bastos, N. Lopes, A. Crotti, *J. Mass Spectrom.* **2007**, *42*, 1279–1286.
- [22] J. Xing, Z. Yang, B. Lv, L. Xiang, *Rapid Commun. Mass Spectrom.* **2008**, *22*, 1415–1422.
- [23] M.M. Moutiez, J. Seguin, M. Fonvielle, P. Belin, I. B. Jacques, E. Favry, M. Arthur, M. Gondry, *Nucleic Acids Res.* **2014**, *42*, 7247–7258.
- [24] I. Rowland, H. Tristram, *J. Bacteriol.* **1975**, *123*, 871–877.
- [25] D. D. Buechter, D. N. Paoletta, B. S. Leslie, M. S. Brown, K. A. Mehos, E. A. Gruskin, *J. Biol. Chem.* **2003**, *278*, 645–650.
- [26] W. Kim, A. George, M. Evans, V. P. Conticello, *ChemBioChem* **2004**, *5*, 928–936.
- [27] A. Hadd, J. J. Perona, *ACS Chem. Biol.* **2014**, *9*, 2761–2766.

Manuscript received: December 11, 2017

Accepted manuscript online: January 29, 2018

Version of record online: February 21, 2018

III) Supplementary information

Supporting Information

Incorporation of Non-canonical Amino Acids into 2,5-Diketopiperazines by Cyclodipeptide Synthases

Nicolas Canu, Pascal Belin, Robert Thai, Isabelle Correia, Olivier Lequin, Jérôme Seguin, Mireille Moutiez, and Muriel Gondry**

anie_201712536_sm_miscellaneous_information.pdf

Supporting Information

I)	General methods and material	3
a.	General methods	3
	Figure S1 Workflow of the two-step culture strategy, with the example of proline replacement	3
b.	Strains used in the study	5
	Table S1 Strains used in this study	5
c.	Plasmids used in the study	5
	Table S2 CDPSs used in this study	6
d.	Non-canonical amino acids used in the study	8
	Table S3 Non-canonical amino acids used in this study	8
II)	Identification of non-canonical cyclodipeptides	10
a.	General method of identification	10
	Figure S2 Common fragmentation patterns of cyclodipeptides	11
b.	Example of mass spectra analysis	12
	Figure S3 Mass spectra analysis of canonical cyclodipeptides and their non-canonical counterparts	12
	Figure S4 Analysis of mass spectra for cyclodipeptides with non-standard mass fragmentation behavior	16
c.	Summary of the cyclodipeptides identified	19
	Table S4 Non-canonical cyclodipeptides produced in the study	19
	Table S5 Summary of frequently observed ions or mass losses depending on the residues	28
III)	Incorporation of non-canonical amino acids by CDPSs	29
a.	Production of canonical and non-canonical cyclodipeptides by CDPSs	29
	Figure S5 Histograms showing the production of canonical and non-canonical cyclodipeptides by CDPSs	29
	Figure S6 Calibration curves for canonical cyclodipeptides	34
b.	Influence of an hyperosmotic stress on the incorporation of trans-4-hydroxyproline	36
	Figure S7 Comparison of hydroxyproline incorporation with and without osmotic stress	36
c.	Comparative overview of loading efficiencies of ncAAs by AARs and their incorporation by CDPSs	37
	Table S6 Comparative overview of loading efficiencies of ncAAs by AARs and their incorporation by CDPSs	38
IV)	Isolation of selected cyclodipeptides	42
a.	<i>In vivo</i> cyclodipeptide production	42
b.	NMR analysis	43
1)	NMR spectroscopy	43
2)	NMR assignments of cyclo(5-Fluoro-L-Tryptophan-L-Proline)	44
3)	NMR assignments of cyclo (L-Azidohomoalanine-L-Phenylalanine)	45
4)	NMR assignments of cyclo (L-Azidohomoalanine-L-Leucine)	46
5)	Conformational and configurational analysis	47

6)	NMR spectra of cyclo(5-Fluoro-L-Tryptophane-L-Proline) in DMSO (298.6 K).....	48
	Figure S8 1D ¹ H spectrum of cyclo(5-Fluoro-L-Tryptophane-L-Proline) in DMSO	48
	Figure S9 1D ¹³ C DEPTQ spectrum of cyclo(5-Fluoro-L-Tryptophane-L-Proline) in DMSO ...	49
	Figure S10 1D ¹⁹ F spectrum of cyclo(5-Fluoro-L-Tryptophane-L-Proline) in DMSO.....	49
	Figure S11 2D ¹⁵ N- ¹ H HMBC spectrum of cyclo(5-Fluoro-L-Tryptophane-L-Proline) in DMSO	50
	Figure S12 Stereochemical analysis of cyclo(5-Fluoro-L-Tryptophan-L-Proline) in DMSO	51
7)	NMR spectra of cyclo(L-Azidohomoalanine-L-Phenylalanine) in DMSO (298.6 K)	52
	Figure S13 1D ¹ H spectrum of cyclo(L-Azidohomoalanine-L-Phenylalanine) in DMSO.....	52
	Figure S14 1D ¹³ C DEPTQ spectrum of cyclo(L-Azidohomoalanine-L-Phenylalanine) in DMSO	53
	Figure S15 2D ¹⁵ N- ¹ H HMBC spectra of cyclo(L-Azidohomoalanine-L-Phenylalanine) in DMSO	53
	Figure S16 Stereochemical analysis of cyclo(L-Azidohomoalanine-L-Phenylalanine) in DMSO	54
8)	NMR spectra of cyclo(L-Azidohomoalanine-L-Leucine) in DMSO (298.6 K)	55
	Figure S17 1D ¹ H spectrum of cyclo(L-Azidohomoalanine-L-Leucine) in DMSO.	55
	Figure S18 1D ¹³ C DEPTQ spectrum of cyclo(L-Azidohomoalanine-L-Leucine) in DMSO.....	56
	Figure S19 2D ¹⁵ N- ¹ H HMBC spectra of cyclo(L-Azidohomoalanine-L-Leucine) in DMSO	56
	Figure S20 Stereochemical analysis of cyclo(L-Azidohomoalanine-L-Leucine) in DMSO.....	57
	References	58

I) General methods and material

a. General methods

Non-canonical amino acid (ncAA) incorporation was performed using a two-step culture method (Figure S1).

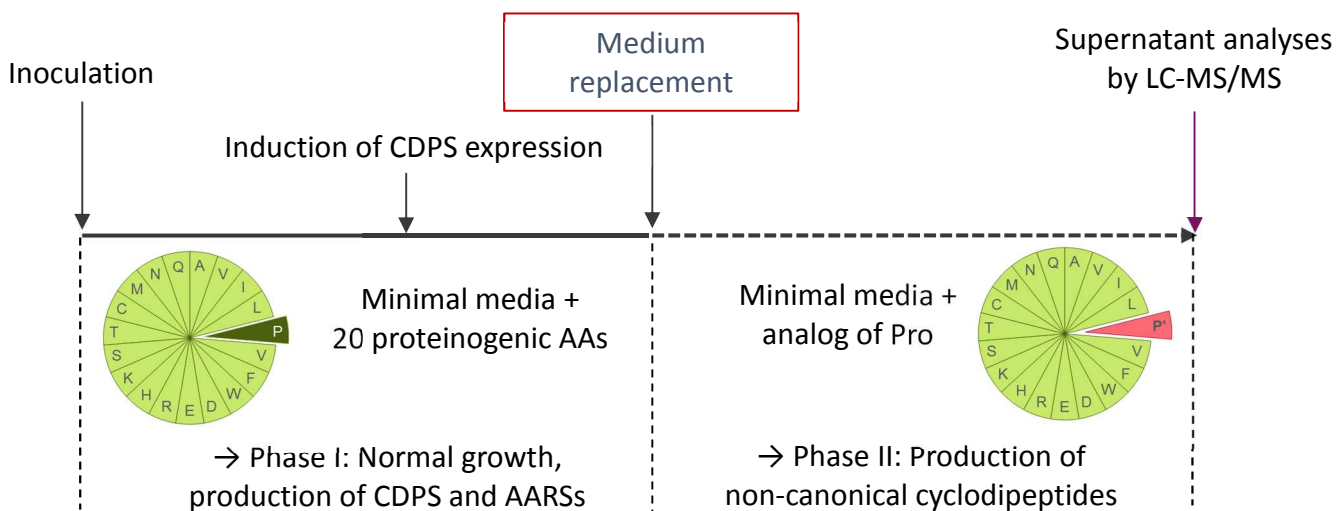


Figure S1 Workflow of the two-step culture strategy, with the example of proline replacement

Phase I:

Auxotrophic strains (Table S1) were transformed by conventional methods with plasmids encoding cyclodipeptide synthases (CDPSs) (Table S2) and were selected on LB plates with 200 µg/ml ampicillin, 25 µg/ml kanamycin. Starter cultures (5-10mL) were prepared with M9 minimum medium supplemented with trace elements and vitamins,^[1] proteinogenic amino acids (50mg/L of each 20), ampicillin (200µg/ml), kanamycin (25µg/ml) and 0.5% glucose. They were inoculated with several colonies from competent bacteria freshly transformed with plasmids encoding CDPSs. After an overnight incubation at 37°C, starter cultures were used to inoculate Phase I cultures (same medium than preculture except that glucose was replaced by 0.5% glycerol; volume was adapted depending on the number of ncAAs to be tested and was usually between 80 and 200mL). When the OD600 reached 0.4-0.6, expression of the CDPS was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG, 1mM final concentration). Cultures were transferred to 20°C and incubation was continued for four hours.

Medium replacement:

After four hours of CDPS expression, cultures were pelleted at 4000rpm for 15min and rinsed with an M9-derived medium, identical to the one used in Phase I except that it contained no vitamins, no amino acids and no antibiotics (whithdrawing of vitamins and antibiotics from the medium in this second step of culture leads to an important improvement in the cleanliness of HPLC analysis profiles, with no differences in cyclodipeptides production). This rinsing operation was repeated twice. In order to artificially increase the biomass concentration, the

pellets were eventually resuspended in a volume equal to half the initial culture volume. These Phase II cultures (40 to 100mL) were used for non-canonical incorporation assays.

Phase II:

Phase II cultures were divided in 24-well plates (Fisher) (2mL per well), covered with a hydrophobic porous film (VWR). Each well was complemented with one ncAA. NcAAs were obtained from various commercial suppliers and used without further purification (Table S3). NcAAs solutions were prepared at 2 g/L in water and stored at -20°C for several months. For three compounds (DL-Phenylglycine, 3-Nitro-L-tyrosine and S-Benzyl-L-cysteine) insoluble at 2 g/L in water, the powder was first dissolved in DMSO then diluted in water in order to achieve a final stock concentration of amino acid of 2 g/L and 10% (v:v) of DMSO. Analogs were added in Phase II cultures at a final concentration of 50 mg/L for proline analogs, 100 mg/L for leucine, isoleucine, valine and methionine analogs, 150 mg/L for phenylalanine and tyrosine analogs and 200mg/L for tryptophan analogs. Plates were incubated at 20°C for 20 hours to allow cyclodipeptide production. After incubation, plates were centrifuged at 1400 rpm for 15 min. Supernatants were collected, acidified with trifluoroacetic acid at 2% v:v final, and stored at -20°C. Supernatants were stable for months at this temperature.

LC-MS/MS analysis:

Cyclodipeptides were detected by LC-MS/MS analyses on an Agilent 1100 HPLC coupled *via* a split system to an Esquire HCT ion trap mass spectrometer (Bruker Daltonik GmbH) set in positive mode. Samples were loaded onto a C18-PFP column (4.6 mm × 150 mm, 3 μm, 100 Å, ACE). The gradient below was used for separation (Solvent A: 0.1% formic acid; Solvent B: 90% acetonitrile, 0.1% formic acid).

100% A	5 min
100% A, 0% B → 50% A, 50 % B	20 min
50% A, 50% B → 100% B	2 min
100% B	3 min
100% B → 100% A	2 min
100% A	18 min

b. Strains used in the study

Auxotrophic strains were obtained from the Coli Genetic Stock Center (Yale University) and are from the Keio Collection^[2]. They derive from BW25113 *E. coli* strain and share a common genetic background, except from one single-gene deletion, where the open reading frame has been replaced with a kanamycin marker. The auxotrophies conferred by the deletions have been previously shown in the literature^[3] and were confirmed in house by growth test in liquid minimal medium with or without complementation by the amino acid concerned (data not shown).

Table S1 Strains used in this study

Number in Keio collection	CGSC accession number	Auxotrophy	Genotype
JW0233-2	8468	Pro	$\Delta proA$
JW5605-1	11483	Leu, Ile, Val	$\Delta ilvD$
JW5807-2	11943	Leu	$\Delta leuB$
JW3745-2	10733	Ile	$\Delta ilvA$
JW3973-1	10856	Met	$\Delta metA$
JW2326-1	9865	Tyr, Phe, Trp	$\Delta aroC$
JW2581-1	10049	Tyr	$\Delta tyrA$
JW2580-1	10048	Phe	$\Delta pheA$
JW1254-2	9131	Trp	$\Delta trpC$

c. Plasmids used in the study

CDPSs used in this study have been previously cloned into the pIJ196, a plasmid previously designed for CDPS expression.^[4] Their activities have been characterized by heterologous expression in *E. coli*. These enzymes and their activities are listed in Table S2.

Table S2: CDPs used in the study

CDPs 10 to 55 were described in reference 4. AlbC was described in reference 5. CDPs 62 to 98 are described in reference 6.

Cyclodipeptides production refers to data obtained in these studies by heterologous expression in *E. coli*.

Proportion of the different compounds, based on UV peak area, are given in parentheses.

CDPS number	Organism	Cyclodipeptides produced	Protein identity	Amino acid sequence	Comments
10	<i>Sphingobium japonicum</i> UT26S	cPM (95%)	WP_013041345	MDVTVLSHRALTVTISIRPGDRMIGDYGHALIAVLSFNYSYTDTRITLVRWAKHTFERSHIPVFDLPHAYTLSARKGHSPGSRVRRARKEGRKLNHNKARIHAEEGIDSDAHLRLTWDDLVNRDDYLQLRVDDVDRAFCTDACFRQAACLTMAADTLLDPEITDEDDRRRAACFLAARYLLDELPLIDLPSIIKCDSSVFLYHRAPSLLKQLFEGRFTLKPSPDRQGYFLSDASEIEAADS	Probably partial because of start codon misannotation.
17	<i>Parabacteroides</i> sp. 20_3	cHF (100%)	WP_005863245	MASLLYERRLESCYRIERYPYSERNAFLGVSIKSRFSEKTLIALFDWCKVNVKSVVYVIADEIQMYTFMASKGLERKEACAKALQIGDIKRYR FIERVIKGGDYDNVRLSSWKAVALPERFKTLQLRLLLYGTEILFRREVIKQILERNRRLPEGFRFARIDSSDYDLASLYILNELAVILYFHLVYFDPVCCYQISPLPMTPLLEILYDGTFLKDLLVPPKIDIGYIEIMEEKDLTGRMIKVNKTGAPL	A GCA codon (Ala) was added after the start codon for cloning facilities.
18	<i>Moorea producens</i>	cAP (100%)	WP_008190827	MASSHRDEETIENFDWNQLSLEQLKLLSQIHKADLKIRDLLKLTQSTANLSKVHRYKASLAKVFPRRRLRSSLDYQNCFFIISLGRSNFVD SERLEASIKWIEHFKAACLVLVCDISYRLTIELRQGLKDEAWLEAIRTGETFINQNRFLFQQYSGSCQFQMASQIENQSEFEIYYKDFQGLYQKDESFQRMVNSFAQTYLNRGEQSEEEVEQLLRQKHLAITYLLEESAVFTCLAKEGWVPVYVPGSIKTFEEIAEGLHPEVPLPLKQMIWVSLRLRKRKATAGESI	A GCA codon (Ala) was added after the start codon for cloning facilities.
19	<i>Pseudomonas aeruginosa</i> PA21_ST175	cIL (93%)	WP_003158562	MANLIAEDFSTRSLLAGERYKAKIAFVSPHTRRSNFEDEPKCFGLVSLERNFTPNRFHSMVEW AARRFEKCSILGDIHHRITLQSTRDMPEEDARRAVRLGDAFMESRLVDAVYRHATDFQYITCSEIQASTECKDFYSSLRNYFLEHSGFRDSEVKEFGRNYHRHNWEKLSSEQQRHRLDFSSRYLLEEFVFAFLVRKGFVMAYPGAFSTLAEIAAGDFPGMLDELQKLTVVSLQLKRR	A GCA codon (Ala) was added after the start codon for cloning facilities.
20	<i>Diplorickettsia massiliensis</i> 20B	cFN (90%)	WP_010598945	MELHAKINGLEKHLKQHPTKRDVSNVIVISMDQPSQRGEKFEAFTRRINLLHKKRGVRKLVIIITDLSQEHYGLNPNLVPTNIKRLAHKKG TDWIRQNSKSLSFCLDPKVNIEIWRWKLSCDDFKKSYREVEQLYEQDKSFQEIENRSLSGDYAQKLSDRRKNWKNPPSPNACFAAAK NYLLEESAIWGPLLNLGDFITYPGCKNEAVNYTYDKLSQSNQFLPWLRYSEFKRSLPGLVSKKLIDPTKSIENLYPQFFSKIPSDNLPEDFELLEESYQGRRSL	Probably partial because of start codon misannotation.
22	<i>Diplorickettsia massiliensis</i> 20B	cYV (30%), cYP (23%), cYT (11%), cYA (10%), cFP (10%), cFV (7%), cFA (3%), cFT (3%)	WP_010598044	MAFPTIKLKHLSADGKTASVLIISIDQIQYQKQEDFEAFCLKMNEEYKGEHCHKVTIETGYLKRHYLALDSKFSSESDLAIAQLGIEW REEOKESINKLMPVKIISWKEILETQLNEKDKPFCYLSDIKTYETDKCFKGVHEHLSKKYAKKLLKEDYDPEKRNKLDKDCLEAARNYLL EESSIIFKLVHYDFTQLYPGAGNAALRYVHRKYFGENNPLPWVEYKIENPNLDSKLKSSNSCFEVEGIDNQNTNIKNILKLEKLKQEE QIKLINELEYIESMKYSSFSVSHK	A GCA codon (Ala) was added after the start codon for cloning facilities.
24	<i>Rickettsiella grylli</i>	cFF (53%), cFL (47%)	WP_006034660	MARYKMKKAKFQGNVDIKNQAQLMIGSMNQTHQSGEELYAFINEIKKYTIKVIIVITDYLHRYVQLETGLPLEEAGKEAEKMGES WLQLNEASLNSLSPVELQLVQWKSLEVEGNSQIEDTYSDDLKSTENCYRDDPFFQMQMVDYTSNEFGQKHCNRLKNRIETLEACQQAANK NYFLEESTIILKFLISLNFVITYPGCKNQGINIYKYGKPLNFISYRSEHVKNSLFSFKKTEESINDAHQFRRNI	A GCA codon (Ala) was added after the start codon for cloning facilities.
25	<i>Parachlamydia acanthamoebae</i> str. Hall's coccus	cPP (75%)	WP_006341088	MANSYIWRQNFATFTDQEKTVRGLAFKTKDITWLETSAVILPISIHNFHEGFQDFKMNAFISTIKKHVKGKITVLLTEKAHVKTLKSLK YSNNFERAFEECLKSAHMLAQRYESIFENCNVVYWSYICQDPYFKKIVELLEEDLYSSDLNFQKLLHEDAISYENLHHRESYMNKSLYIE KCIEDLIEQCACLLVLNKGKGFQYFPGKPLSTEYVNRIFVPKDNRISWIDVLSIEKKTITPHNLHM	A GCA codon (Ala) was added after the start codon for cloning facilities.
30	<i>Rickettsiella grylli</i>	cMA (23%), cMG (27%)	WP_006034819	MAHSMKHHKISGLNITIKKDFKTPFVQEAALIIGVSVGSKKHGDEALAAALVEAINRLHKSVRTNCTIACVCDLQRHNYRIDGKTDEEKALSM SKKAGAEWIEANIEITKHLNVDYITVRWDTWLKDEKRYREAFLEISNLLAQDYAFQKTMDSIQVFSRRFNKRYQELGISAPINDVDLQKS CRDYLLEECALIMKMWPLYKQEHSQYLYPEKMEALEYVYKQVVSQKNLFKWVNRFRKIMQLKGDVLPDPSEPKYFSSSQPPEEG KSSYGFTKADSMNSPVPLEDSVSDOTIFKHT	A GCA codon (Ala) was added after the start codon for cloning facilities.
36	<i>Vibrio brasiliensis</i> LMG 20546	cLE (58%), cLA (20%), cLP (14%)	WP_006880660	MATKARFKHDGKQELVGKKCILAISVGQKYHEGDKLSSITIELKNGFSSTIMLGDTLQRHNLTDPCPEVSAREKANSLGDSWLDNRN ELLDSINPKLEVIRWDDIIDSEYNYERKRIEHEYNINDEYKKNINSTVFTFLQRVRNNGKPFDIEDSFYRGLLEYIEECPAIMPLWAKQGYD FIIYPQPMQAMKATYNIYVKNPFDKANVLSLRFKSSNSNDIPYLKTA	A GCA codon (Ala) was added after the start codon for cloning facilities.
44	<i>Pseudomonas protegens</i> Pf-5	cLE (99%)	WP_011059731	MASKVQEIINKLAYFNKGEEHEFDGKRVVLAVSVGQYEHEDQKLRSTHILNQSGFHVKVVVADTLQRHNKHGKSPGEALSAAIRDGD AWLARNQSLDGLRVPYHITRWQELASDRYAELRQQLNQVYQREELRDAIDSTIGVFTERLRLRDEHADIERAAAQCREYLIEEIPILP LWADGEGYHYVLYPQMTTAMATSRELLIEFSPDRVRWLPLKFKRGIPIPTLPGAHPNWTSAGAI	A GCA codon (Ala) was added after the start codon for cloning facilities.
54	<i>Bacillus</i> sp. 171095_106	cLL (56%), cLF (23%), cLM (20%)	WP_028408934.1	MASQILLEKIAKPLFAIEPLSQNCQRFERRNHILIGISPFNSRFSEYIYRLIEWGIEKFKDVSLLAGNEAANLLEALGTSRVKAERKVRKE VNRNRRFAERALKAHGGNPEAIYTFSDFTQPVYCDMRTKIEESFLHQPNFQACLEMASHAAVGRASGTNMDISVVTSEMLHNAVQYVLAELPFFIGAPNILGIPETVLAYHHPWRLGEQIKDEQFSIRMOQQNGYITVDEIQ	A GCA codon (Ala) was added after the start codon for cloning facilities.
55	<i>Bacillus</i> sp. NSP9.1	cLL (52%), cLF (28%), cLM (20%)	WP_026588781.1	MATELMIESKHSFKTETLSQNCHEILTRRRHVLIGISPFNSRFSDDIYRILGWALDEFQDVSLLAGKEAANLLEALGTPKGAERKVRK EVSRNRRFAERALVAHGGNPEAIHFTSDFAEQATYISLRQEVHAFSEQPHFRQAACLDMSHAALGRARGTQLAIGDVNDAMLNLAVEYV IAEPLFFIGAPDILGVEETLLAYHRPWKLGERISRNEFAVRMRPNQGYLMVSEPDGSMESKNMQEERV	A GCA codon (Ala) was added after the start codon for cloning facilities.
AlbC	<i>Streptomyces noursei</i>	cFL (45%), cFY (17%), cFF (16%), cFM (8%), cYM (5%), cYL (4%), cLM (3%), cLL (1%), cMM (1%)	WP_079142603	MLAGLVPAPDHGMREEILGDRSRLIRQRGEHALIGISAGNSYFSQKNTVMLLQWAGQRFERTDVVYVDTHIDEMLIADGRSAQEASVVKRTLKDLRRRLRRSLESVGDHAEFRFRVRSLSLQETPEYRAVRETRDRAFEEDAFAATACEDMVRAVVMNRPGDVGISAEHLRAGLNYVLAEAPLFDSPGVFVSPSSVLCYHIDTPTIAFLSRRETGFRAAEGQAVVVRPQELADAA	AlbC was cloned in originally used vector pQE60, not pJ196

Table S2 (continued)

CDPS number	Organism	Cyclodeptides produced	Protein identity	Amino acid sequence	Comments
62	<i>Nocardioopsis potens</i>	cFM (30%), cFF (29%), cFA (19%), cFY (8.4%), cFL (7%), cYA (1.7%), cFW (1%), cMM (1%)	WP_017593947.1	MASRNCTSRREAGAAEAHHGSDTTRTSADRPLGGAAFEVDAATPSCRRIREFGRDHLIGVSAGNSYFSAERVGGLLSWAVPFRERIDVVYVDTHIDTMEALGYPPASVSRKRVRKTKLQDLRRRLRRGVAAEPGAAGSVAVRPLSEYVPLPAAYAALESVDRLLEGEDEVLRHACEEQVTDFYKDRADGTALVSSSESEMFAGMAYVRAEFPPFLNTPEMAAGVSSVTCYHKMMPAKALFREESPIPRHPGQGFVIVRPTDPAPEQGR	A GCA codon (Ala) was added after the start codon for cloning facilities.
63	<i>Streptomyces lavendulae</i>	cWW (62%), cWP (15.8%), cWA (9.8%), cWL (6.2%), cWX with X=(S, Q, E, F, C, G, I, N)	WP_051880239.1	MATITADASFEVLPFTRTCRHIWEDGDHVLIGVSPGNSYFSAERIALGLRWATTRFAQVDVYADLHVDRMFAAFGHTRHEAEKRAAKEIKAVRRRILKGVTEGPPQIEIRVRALESEFQSNPVYQLLHRRVLFHLETDDEFKRGCEEMALHFVGSKLPEGESVTDQLRVCFDYMAAELPFVVDTPSILDVPSVAAHYHVRMPLTDVLFARGGGLRATRNRQAYAVVRPEATAPGSATDHAPERTVDERRAA	A GCA codon (Ala) was added after the start codon for cloning facilities.
64	<i>Streptomyces purpureus</i>	cWW (100%)	WP_028798003.1	MAIEDFTVRPFTPLCNIIWEGDHLVIGVSPGNSYFNAARIAELTRWATRRFTAVDFVYADLHVAEMFAALGHDEEHAARRAAKELKAVRRRVVAGVEAAGPPGPIRVRALESEFAANPVYLLHRRVRHFLATDAEFRKGCMDMVDTLFLAPKAVAGDAPVTERQRTACLIDYIAAELPFLLDTPGILDVPSVSCYHQIPLTELLYAKGGGLRAARNQGYAVVCPAEGSTGHDERRAA	Probably partial because of start codon misannotation. The protein tested corresponds to a start at position 3 of the new NCBI sequence WP_019889609.1, with L mutated in M. A GCA codon (Ala) was added after the start codon for cloning facilities.
68	<i>Streptomyces</i> sp. NRRL F-5053	cLW (97%), cLL (1.7%), cFW (1%), cFL (0.7%)	WP_063764781.1	MACEGADVPGPTSGAHFSVSPITARCKDIWERGEHILVGVSTSNFYFNRLRTRLEWLQNHFSAVDVVYADMHIDTSLIALGCEESEARRRAKRRVQVRRRVRQALSAVEDSPTEFRAHALSDFADSPYKALGARVDRALAEEDGLSTACHDMVRSFLSQPGGEGELTAEQIEAGLSVYRAEMFVLVDTPALLGVPSVSCYHMEAQVWRALRRPAGLRAAPQAFAVVAPVDES	A GCA codon (Ala) was added after the start codon for cloning facilities.
69	<i>Streptomyces rimosus</i>	cWY (98%), cWW (1.9%)	WP_030659939.1	MASSGEIRRFMGASRSVAVTFRTPFTDRCDRIQRGDHALIGVSTNNSYFSAERLTALVHWAGGHFRAVDIICADLHIDVTLTAEGAAPENVGRRARRRVTDRRRIRKAVDAASPDGPRPGSHLLSDFQGNPAYQSLRSDIDRALREDPEFANACHDMIRLHLLGRPGQGDIDIFGTSGNGPGDAAEHRRFQAGLQYLGGELPFVDTSPILGVPSVSCYHLFTPLGLPLCMRDNGLRAADNQAFLAVTPDEA	Probably partial because of start codon misannotation. The protein tested corresponds to the first 255 amino acids of the new NCBI entry WP_078897101.1. A GCA codon (Ala) was added after the start codon for cloning facilities.
70	<i>Streptomyces</i> sp. NRRL B-24484	cWW (73.2%), cWL (16.5%), cWS (8.9%), cWX with X=(Q, F, P, E, G)	WP_030265608.1	MASELLSTTIDEFOAFPYSARCHRVFARGDHLIGVSPGNSYFSARRITQLVHWGKEFFATVDIVHADLHVDQAQGAQGYPEAAARRRAKEVKATRRRVERGAADAGRPGVTRTHALSDFGTGETYRRLHAEVLAALADDRPREAERMAALGFLRLDGAAPTADQLAAGLRYIAAELPFLDTPALLGVPSVACVHVELLTPVLFGRPDGLRAAPQAYAVVRPADAAADRRAA	A GCA codon (Ala) was added after the start codon for cloning facilities.
71	<i>Streptomyces roseochromogenes</i> subsp. <i>oscitans</i> DS 12_976	cWY (98%), cWA (2%), cWS	EST20838.1	MAHSEPOVLEHHRVRLPFGASSERILQRGDHALFGVSTGNSYFSRRNLANAMAWAVERFAAVDVVHADIALEAMLEALGYERVAARKSVAKQLRGVRRRIDGALEDIGAAAERVRARPLSAFLDHPYSYREVRARTRALRTDVELRSVRDAMAYQFLAKRLKPDLLTPAQVEAALAYVDAELPFVVDTPRILGVASSVHCYHTVLALGRLLFGERRMALRPADNQGAYAVACRDSGKQIAMASTSTEAA	The protein tested corresponds to the first 250 amino acids. The CDPS domain is fused to a P450 domain.
74	<i>Streptomyces</i> sp. NRRL S-1868	cWP (99.5%), cWA (0.5%)	WP_078873129.1	MANTSLAAVAGDHVPEYSPNCERLLEAADHLIGVSPGNSYFGIPTLTRLTWAHATFRRIDVIVPDRSLAANYRAQGYSEQVAHKKARTIEISSVRRRIRRAWERSGIPEHEQRTHLLSELALLPAYEKLHRRVREALSDEQVDFVQCRATSRAPVVAEQTEGEGTGAAVPAPGPHLALDYLDIDEMPFIDTPTSLILGVPSLVNHYMPVFAAGLLYTGDGPLRAPSHQGLILLDRAAER	A GCA codon (Ala) was added after the start codon for cloning facilities.
77	<i>Streptomyces scabrisporus</i>	cLV (56.2%), cLT (5.9%), cLL (18%), cLI (18.3%), cLA	WP_020551384.1	MGFVWVPTGLSRSVLDHAEHVLIGLSPVNSYKPRVTEALVRWACANFVAVDVVPGDEAAHTLTAAGVQPSAEVRRARYAVGRLRSRARRGLRNAGMADPDRHLHTWTQLSERPAYQHSKLRVHRAYLTDVAARACRRTRARGAVRRVAGQEPSEEQIDQAVGYALAEPLVLDGPAIFGTSSLVFYHREMDLLAPFINGESAHLRPAPGGQYAVVRPPDQATPAGLSQPWPEGRATPPASGRTEPTDPLPRF	
82	<i>Streptomyces varsoviensis</i>	cPL (65.8%), cPA (20%), cPV (7.6%), cPF (1.8%), cPM (1.5%), cLL (1.9%), cLA (0.9%), cPI (0.4%), cAA, cPW	WP_030889157.1	MATTTTPKVEQATFDVAPYTDNCARIYEAHAELIGLSPFNYSYKGLIRSLTEWAARRFATVDVIFPGDEAMLTAAAGWDPGHAERRLHQARKKLRGPARQGRLAAGIDLPHYLHTWTELLDRPAYRHLRAQVGGGCRTPDGLRRPARELSRKAARGLTGDPEDDRQIDHALDYVVAEMPFMIDSPSIFHVKSSVFIYHOPFELAHCLDGREISSVEINPAQGLFVLRPSSGE	A GCA codon (Ala) was added after the start codon for cloning facilities.
85	<i>Actinokineospora</i> sp. EG49	cLI (65.7%), cMI (9.1%), cLL (5.4%), cML (3.3%), cLV (4.9%), cLP (4.6%), cLA (3.3%), cLC (≤3%), cMV (?), cMP (0.5%)	EWC64439.1	MVVDPALSEVRATRYKAKVESVSPKANRTAFEDADECFLGVSLNDSFDVPKLEGVWISRRFRFRCTVLVGDSDIHLRLTSLATAALSPEPALARALELGRFIDRQPVDFLFSRDTFTVTCAEVRWVADHDGFRHRLRSQFAEDPVFRASVESFGRYRHSRGATDLGQELAKRV AISSEYFLEEFVAVFCLRRAGLSVMVYPGSFSSLAEVAEGHPDAPRELRLDVLVSLQLKGR	
89	<i>Algicola sagamiensis</i>	cAP (100%)	WP_018691765.1	MASDIHDLYDFLKKLDYDFDIGNENKKNITLQHVSPFFHEKTSIQEPAFLAVSLSESLAFLNKKFAAIMKYLKGFDPDIALYVDSPIRYTIQIKYDVSEEEATKLSQVVAQTVAHQREVCRMIGCSEPRFVHSTLETRADYVHLQQLRLEFFVEPAFQSVVQHCFEYFLTRIRDEMKSVDKAVNLSRQYLLAELTAVGILNLDYPAMLYPGKIDISGFELELELPEDIATYRDFKFAASLRNRRKKK	A GCA codon (Ala) was added after the start codon for cloning facilities.
91	<i>Methylovulum miyakonense</i>	cFF (51.5%), cFY (36.1%), cFL (8.2%), cFM (3.5%), cYM (0.7%)	WP_019864150.1	MAKVYLAEQRGDCGDSLATSRYRPFMIGSISNRALPLADSLTLKWLARTPSPILPIADEIAVINYRAFKRHGGGGYQAQVQRDAQLHIGQWQEAASQLPAGQGERVFRVREILTPTYRQIEVRAEFAQGGLLQQTIFTLVENYIRSTGKVTTSQRCFLDAEYIIEQLPSLLFGIEVAGIRYQTLIPYTRASDMQGLVLAIRQAGFAGLRSALQNSYVPEDGKPLENSKIQIFILSGQRVAPPAPQSSAQDQENHERLAIS	A GCA codon (Ala) was added after the start codon for cloning facilities.
94	<i>Streptomyces katrae</i>	cYY (71.3%), cYF (14.9%), cYA (10.7%), cYM (2%), cFF (0.6%), cFA, cYL, cYS	WP_030300379.1	MATTAIQGIFAVRPYTPHCQVIHDEGDHAVIGSSGNSYFTHDRVLDLARWGLEHFRQVDLIWDMHVAEMFVALGYPEVAQRKAVKNLRGVRAKVTAAVLAALDPGERLRGRPMSALLELPAYQRISRDLTMTDDPEFRVCDQLAARFLADKLNQGPPTQLQREVCIKYVCAEVPFLFDTPAILGVSSSLNLCYHQALPALLYARGLRGLRASNQGHAVITPAEEPEEQDRSA	WP_030300379.1 has been removed. The new NCBI entry WP_078613504.1 is 29 residues longer in N-ter. A GCA codon (Ala) was added after the start codon for cloning facilities.
98	<i>Vibrio sagamiensis</i>	cPM (95.4%), cMA (4.6%), cPA (0.7%)	WP_039983964.1	MEVTKFKLVNDIVSFKYKHSIQSLQSRALLGLSPGNSFFTEENITNLLTLCQTRHIYILIPDKPHVYNFLGLGYTLKEAKKRADRECRILNRMNSAIDKVKQSQFSNFTVITWQNHIEENEKYTQTKYKLLDEYSNNVSFMDMNNNSYQCLINKAKSRGIKEECVDIEEGVYKYLQELALFNMDNIFGQWTSFAYRSEFRVGLDFLSSKCPDFRDNFLIMYECA	

d. Non-canonical amino acids used in the study

An extensive literature survey was conducted prior to this work in order to get an up-to-date list of ncAAs that had been shown to be substrates of AARSs.^[7-16] 60 ncAAs were eventually selected, obtained from various suppliers and tested towards the corresponding set of CDPSs. ncAA formulas and information about loading by AARSs are given in Table S6.

Table S3: Non-canonical amino acids used in this study

Name	Short form	# in this study	CAS number	MW	Supplier
Proline analogs					
L-Azetidine-2-carboxylic acid	Aze	2	2133-34-8	101.1	Sigma
3,4-Dehydro-L-proline	34DePro	3	3395-35-5	113.1	Sigma
<i>cis</i> -4-Fluoro-L-proline	<i>c</i> 4FPro	4	2438-57-5	133.12	Bachem
<i>trans</i> -4-Fluoro-L-proline	<i>t</i> 4FPro	5	2507-61-1	133.12	Bachem
<i>trans</i> -4-Hydroxy-L-proline	<i>t</i> 4HOPro	6	51-35-4	131.1	Bachem
α -Methyl-L-proline	α MePro	7	42856-71-3	129.2	Sigma
L-2-Thiazolidinecarboxylic acid	2ThPro	8	65126-70-7	133.2	Bachem
L-4-Thiazolidinecarboxylic acid	4ThPro	9	34592-47-7	133.2	Bachem
Phenylalanine/tyrosine analogs					
3-Fluoro-DL-tyrosine	3FTyr	11	403-90-7	199.2	Sigma
3,4-Dihydroxy-L-phenylalanine	DOPA	12	59-92-7	197.2	Sigma
4-Fluoro-DL-phenylalanine	4FPhe	14	51-65-0	183.2	Sigma
2-Fluoro-DL-phenylalanine	2FPhe	15	2629-55-2	183.2	Sigma
3-(2-Thienyl)-DL-alanine	3-ThienA	16	2021-58-1	171.2	Sigma
3-(3-Pyridyl)-L-alanine	3PyrA	17	64090-98-8	166.2	Sigma
3-(4-Pyridyl)-L-alanine	4PyrA	18	37535-49-2	166.2	Sigma
β -(2-Thienyl)-DL-serine			32595-59-8	187.2	Sigma
3,4,5-Trifluoro-L-phenylalanine			646066-73-1	219.2	Chem Impex
β -Methyl-L-phenylalanine hydrochloride			80997-87-1	179.2	Acros Organics
O-Acetyl-L-tyrosine			6636-22-2	223.2	ChemImpex
3-Nitro-L-tyrosine			621-44-3	226.2	Sigma
4-Nitro-L-phenylalanine monohydrate			949-99-5	210.2	Acros Organics
3,5-Dibromo-L-tyrosine monohydrate			1312313-75-9	339	Sigma
3,4-Dichloro-L-phenylalanine			52794-99-7	233.1	Alfa Aesar
DL- β -Homophenylalanine			15099-85-1	179.2	Sigma
L- β -Homotyrosine hydrochloride			336182-13-9	195.2	Sigma
Tryptophan analogs					
4-Fluoro-DL-tryptophan	4FTrp	20	25631-05-4	222.2	Sigma
5-Fluoro-DL-tryptophan	5FTrp	21	154-08-5	222.2	Sigma
5-Methyl-DL-tryptophan	5MeTrp	22	951-55-3	218.25	Sigma
5-Hydroxy-L-tryptophan	5HOTrp	23	4350-09-8	220.2	Sigma
5-Methoxy-L-tryptophan			28052-84-8	234.3	Sigma
S-benzyl-L-Cysteine			3054-01-1	211.3	Sigma
Nin-Boc-L-tryptophan			146645-63-8	304.4	Bachem
DL-Indole-3-lactic acid			832-97-3	205.2	Sigma
N-Methyl-L-tryptophan			526-31-8	218.3	Sigma
5-Bromo-DL-tryptophan			6548-09-0	282.2	Acros Organics

Table S3 (continued)

Name	Short form	# in this study	CAS number	MW	Supplier
Leucine/isoleucine/valine/methionine analogs					
5,5,5-Trifluoro-DL-leucine	TriFLeu	28	2792-72-5	185.2	Alfa Aesar
4-Azido-L-homoalanine	Aha	29	942518-29-8	144.2	Santa Cruz Biotech
L-Norleucine	Norleu	30	327-57-1	131.2	Sigma
DL-Norvaline	Norval	31	760-78-1	117.2	Sigma
DL-Ethionine	Eth	32	67-21-0	163.2	Sigma
L-Glutamic acid methyl ester	MeEGlu	33	1499-55-4	161.2	Sigma
γ -Methyl-L-leucine	γ MeLeu	34	57224-50-7	145.2	Sigma
3-Cyclopentylalanine			99295-82-6	157.2	Sigma
N-Methyl-L-leucine			3060-46-6	145.2	Sigma
DL- β -Homoleucine			3653-34-7	145.2	Sigma
L-Cyclohexylglycine			14328-51-9	157.2	Bachem
DL-Phenylglycine			2835-06-5	151.2	Sigma
DL-3-aminobutyric acid			541-48-0	103.1	Sigma
DL- β -Homovaline			5699-54-7	131.2	Sigma
N-Methyl-DL-valine			2566-32-7	131.2	Sigma
3-fluoro-DL-valine			43163-94-6	135.1	Sigma
O-Methyl-L-threonine			4144-02-9	133.2	ChemImpex
Cycloleucine			52-52-8	129.2	Sigma
Homocycloleucine			2756-85-6	143.2	Sigma
DL-Propargylglycine			64165-64-6	113.1	Acros Organics
DL-2-Allylglycine			7685-44-1	115.1	Sigma
L-Aspartic acid methyl ester hydrochloride			16856-13-6	147.2	Sigma
2-Aminoisobutyric acid			62-57-7	103.1	Sigma
1-Amino-1-cyclobutanecarboxylic acid			22264-50-2	115.1	Sigma
1-Aminocyclopropane-carboxylic acid			22059-21-8	101.1	Sigma

II) Identification of non-canonical cyclodipeptides

a. General method of identification

LC-MS/MS analyses were performed on each supernatant culture and allowed the unambiguous identification of non-canonical cyclodipeptides (see below). LC-MS/MS data do not give information about the stereochemistry of these molecules. However, a L-L configuration can be assigned with confidence to the cyclodipeptides produced due to the stereospecificity of both the *E. coli* tRNA loading apparatus^[17–20] and the CDPSs towards L-configured amino acids.^[21] In all cases where the stereochemistry of cyclodipeptides produced by CDPSs has been assigned experimentally, it has been shown that these cyclodipeptides only contain L-configured amino acids.^[22–25] Since all CDPSs share the same catalytic residues and the same three-dimensional architectures, there is no doubt that these enzymes catalyze the formation of (L-L) cyclodipeptides.^[26–30] In keeping with this, the compounds characterized in Part IV of these supporting information (see page 43) exhibited clear homochirality of the two constitutive aminoacids.

In order to systematically assess ncAA incorporation by CDPS and to validate the identity of the non-canonical cyclodipeptides produced, the following analysis pipeline was applied to each LC-MS/MS dataset:

- The charge ratios (m/z) of every putative non-canonical cyclodipeptide produced in the sample were calculated, on the basis of the canonical cyclodipeptide(s) produced by the CDPS and of the molecular weight of the ncAA added in the medium ;
- EIC profiles corresponding to these m/z ratios were compared to those of an uncomplemented negative standard (*i.e.* same Phase II culture (Figure S1) but no amino acid added in the medium) ;
- When a signal at the expected EIC was observed in the complemented sample and not in the negative standard, the mass spectra of the putative non-canonical cyclodipeptide was analyzed to confirm the identity of the compound.

Mass fragmentation of 2,5-diketopiperazines in positive mode has been widely studied.^[4,31–34] The mass spectra of all the canonical cyclodipeptides produced by the CDPSs used in the study have been previously reported.^[4]

For most cyclodipeptides, three ions corresponding to a loss of neutral fragments of mass 28, 45 and (28+45) are observed. These ions correspond respectively to species $[M - \text{CHO}]^+$ (loss of 28), $[M - \text{CH}_2\text{ONH}_2]^+$ (loss of 45), and $[M - \text{H}-\text{CHOCH}_2\text{ONH}_2]^+$ (loss of 28 + 45), where M stands for the whole molecule. These three ions are indicative of the presence of the diketopiperazine skeleton.^[32,35,36] The loss of fragments containing the lateral chains gives information about the molecular mass of the lateral chains. In most cases, the presence of a daughter ion of formula $[\text{RHNHCH}]^+$ (where R stands for the alkyl form of the lateral chain), called immonium ion, allows the identification of the residues that constitute the cyclodipeptide.

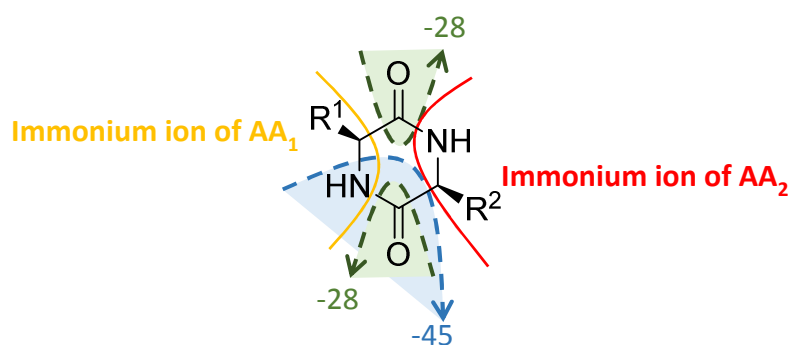


Figure S2 Common fragmentation patterns of cyclodipeptides

Most if not all of the non-canonical cyclodipeptides produced in this study have never been characterized before (with the notable exception of cyclo(Leu/*trans*-4-hydroxy-Pro) and cyclo(Phe/*trans*-4-hydroxy-Pro)^[35]). We used the common fragmentation patterns of cyclodipeptides to check the coherence between the mass spectra and the identity of the putative cyclodipeptide. As an example of the mass spectra analysis method, Figure S3 illustrates the comparative analysis of mass spectra of three non-canonical cyclodipeptides and their corresponding canonical cyclodipeptides.

The ncAAs that were used in this study were commercial compounds, fully characterized and with a claimed purity above 99%. The *de novo* formation of molecules with the *m/z* of expected non-canonical cyclodipeptides in the cultures expressing the CDPSs upon addition of these ncAAs is a strong argument in favor of the assignment of these signals to the expected non-canonical cyclodipeptides. Importantly, a significant proportion of the non-canonical cyclodipeptides (around 70%) characterized in this study were identified independently in several cultures expressing different CDPSs. We could observe similar retention times and mass fragmentation spectra, thus strengthening our assignment process.

b. Example of mass spectra analysis

A

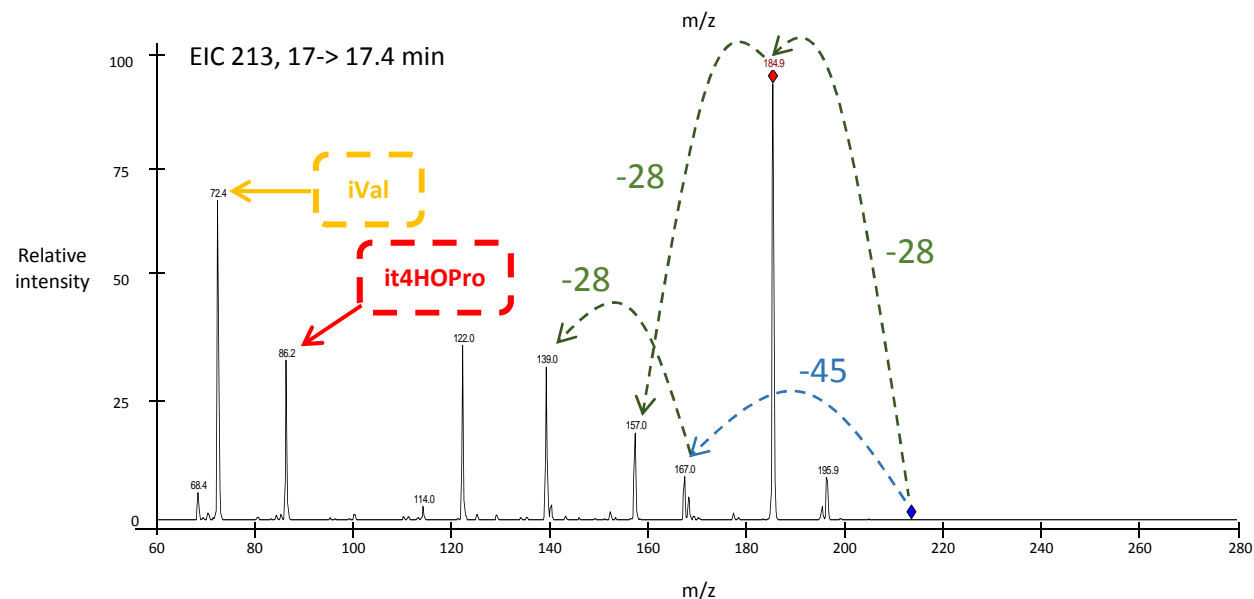
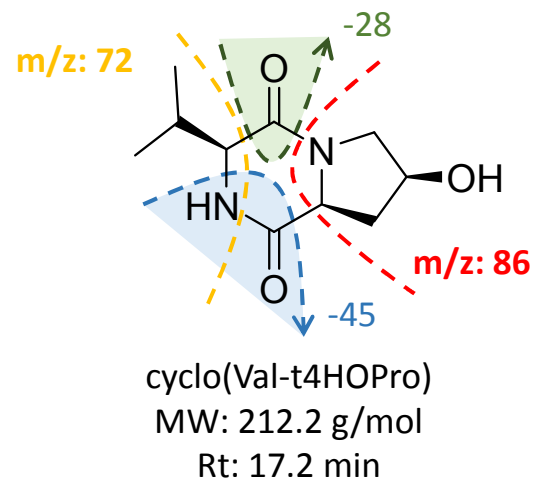
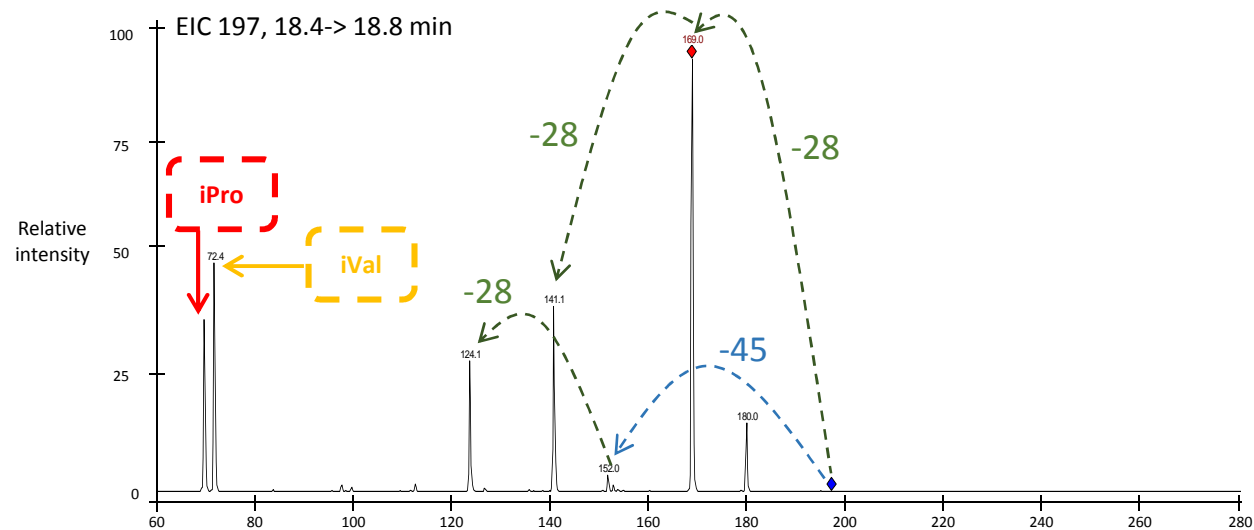
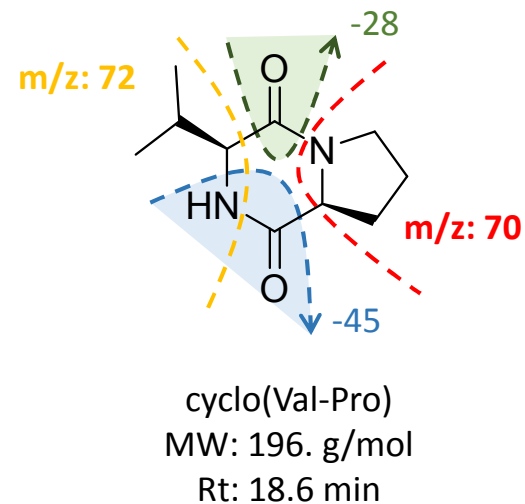


Figure S3 Mass spectra analysis of canonical cyclodipeptides and their non-canonical counterparts

A) Mass spectra of cyclo(Val-Pro) and cyclo(Val-t4HOPro) and the corresponding fragmentation pathways

The 2,5-DKP signature (mass losses of -28, -45, -56 and -73) is visible in both cases, as well as the immonium ion of valine (iVal). The immonium ion of proline (iPro) is present in the mass spectra of cyclo(Val-Pro), so is the immonium ion of *trans*-4-hydroxyproline (it4HOPro).

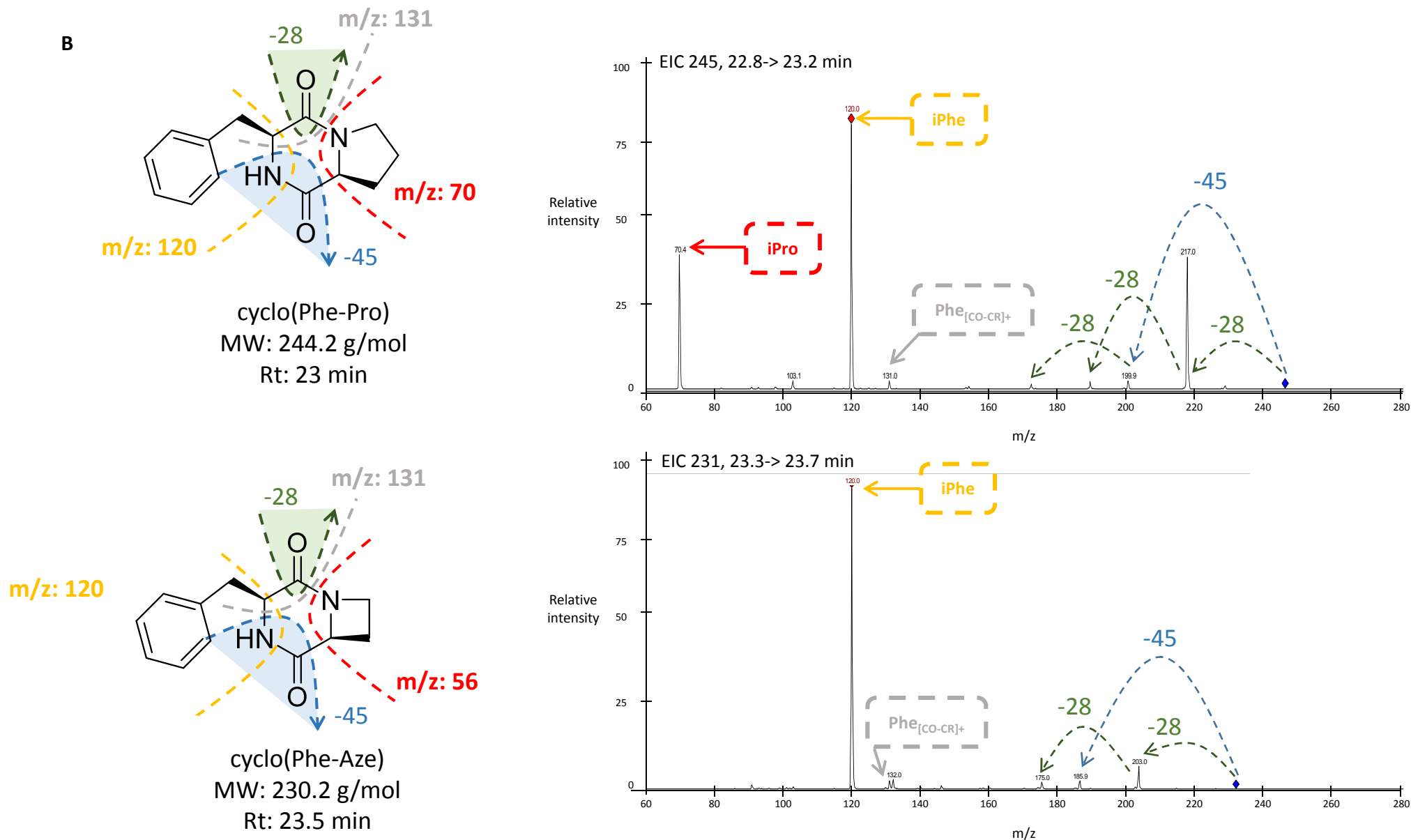


Figure S3 (continued)

B) Mass spectra of cyclo(Phe-Pro) and cyclo(Phe-Aze) and the corresponding fragmentation pathways

The 2,5-DKP signature (mass losses of -28, -45, -56 and -73) is visible in both cases, as well as the immonium ion of phenylalanine (iPhe). The immonium ion of proline (iPro) is present in the mass spectra of cyclo(Phe-Pro). The immonium ion of azetidino-2-carboxylic acid is not visible, which may be due to its low molecular weight (ions below 60 m/z are only rarely detected with our MS settings).

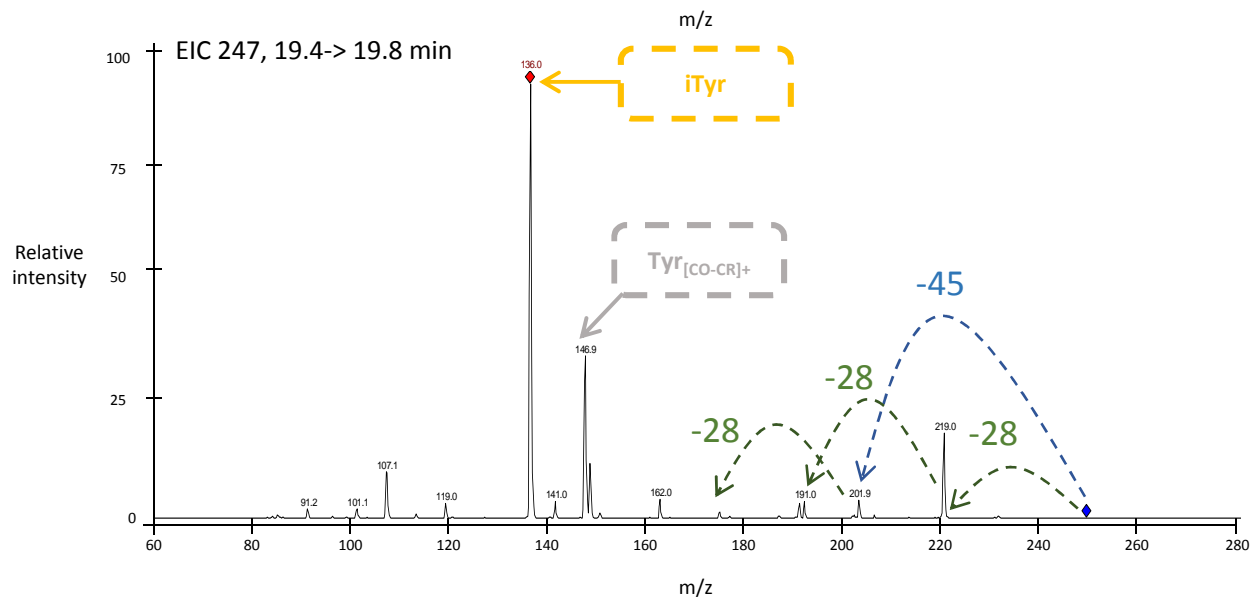
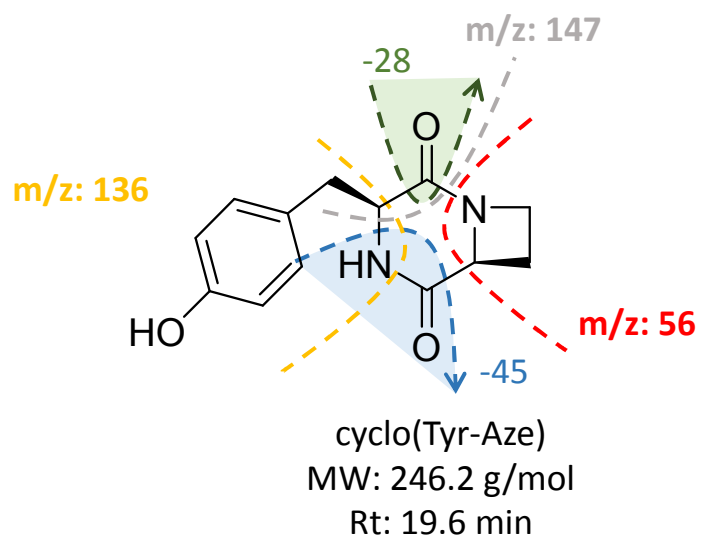
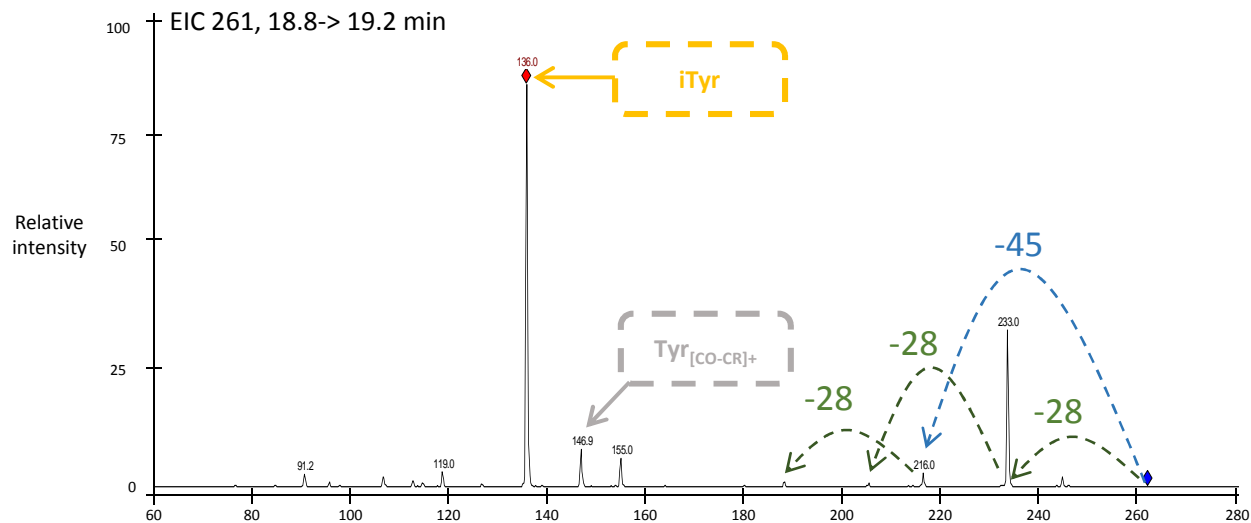
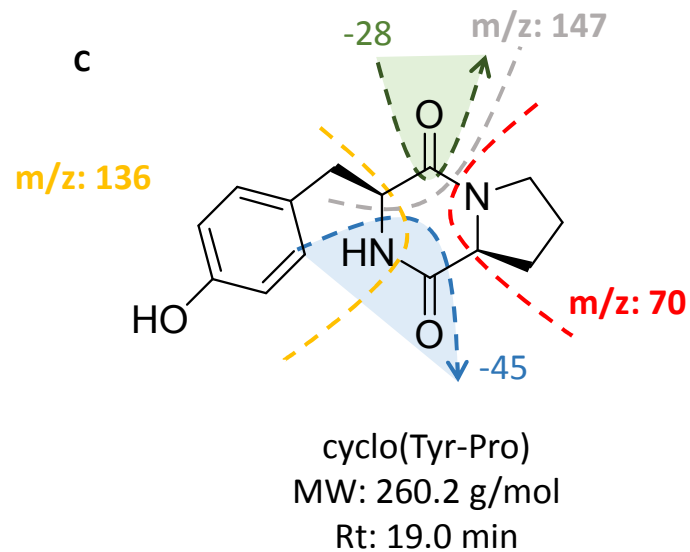


Figure S3 (continued)

C) Mass spectra of cyclo(Tyr-Pro) and cyclo(Tyr-Aze) and the corresponding fragmentation pathways

The 2,5-DKP signature (mass losses of -28, -45, -56 and -73) is visible in both cases as well as the immonium ion of tyrosine (iTyr). The immonium ion of proline is not visible in the mass spectra of cyclo(Tyr-Pro), neither is the immonium ion of azetidine-2-carboxylic acid.

Most ncAAs exhibited mass fragmentation behaviors analogous to their corresponding proteinogenic amino acids when they were incorporated into cyclodipeptides (Figure S3). However, some ncAAs conferred non-standard mass fragmentation behaviors to the cyclodipeptides they were incorporated into. Such non-standard behavior have been previously observed for cyclodipeptides containing amino acids with charged functions or good leaving groups, such as methionine or glutamate. In these cases, the loss of good leaving groups can be predominant over the fragmentation of the 2,5-DKP skeleton or the loss of immonium ion.

When such non-standard mass fragmentation pattern was observed, we systematically compared the mass spectra of every putative cyclodipeptide containing this ncAA and identified fragmentation patterns characteristic of the presence of this ncAA. These patterns allowed us to assign with confidence these signals to the expected non-canonical cyclodipeptides and, in some cases, to propose fragmentation pathways.

Figure S4 gives the example of such analysis process, for cyclodipeptides containing three ncAAs conferring non-standard behavior: ethionine (Eth), glutamic acid methyl ester (MeEGlu) and azidohomoalanine (Aha).

A

m/z XH ⁺			MS2 (XH ⁺)	MS3 (XH ⁺ → XH ⁺ -62)									
cyclo(Gly-Eth)	203	EIC	140.9	113	85.3	84.4							
		Mass loss	-62.1	-27.9	-55.6	-56.5							
		Relative intensity	100	100	34.6	3.1							
cyclo(Ala-Eth)	217	EIC	154.9	127	110	109	99.1	84.3					
		Mass loss	-62.1	-27.9	-44.9	-45.9	-55.8	-70.6					
		Relative intensity	100	100	9.7	1.9	26.6	2.2					
cyclo(Pro-Eth)	243	EIC	180.9	152.9	125	97.2	84.3	82.3	70.4				
		Mass loss	-62.1	-28	-55.9	-83.7	-96.6	-98.6	-110.5				
		Relative intensity	100	46.5	100	4	5.9	1.9	10.1				
cyclo(Leu-Eth)	259	EIC	196.9	169	152	141	140	124.1	113.1	101.1	86.3	85.2	84.3
		Mass loss	-62.1	-27.9	-44.9	-55.9	-56.9	-72.8	-83.8	-95.8	-110.6	-111.7	-112.6
		Relative intensity	100	100	13.9	19.8	2.4	2.8	18.4	2.9	2.1	10.4	4.2
cyclo(Phe-Eth)	293	EIC	230.9	202.9	185.9	185	174.9	157.9	157	156	147	130	91.2
		Mass loss	-62.1	-28	-45	-45.9	-56	-73	-73.9	-74.9	-83.9	-100.9	-139.7
		Relative intensity	100	100	52.3	4.4	41.7	15.6	10.2	4.7	6.7	44.5	17.9
cyclo(Tyr-Eth)	309	EIC	247	218.9	201.9	190.9	173.9	163	145.9	141	113	107.1	85.2
		Mass loss	-62	-28.1	-45.1	-56.1	-73.1	-84	-101.1	-106	-134	-139.9	-161.8
		Relative intensity	100	100	82.4	25.7	20.6	13	63.7	26.5	30	15.6	29

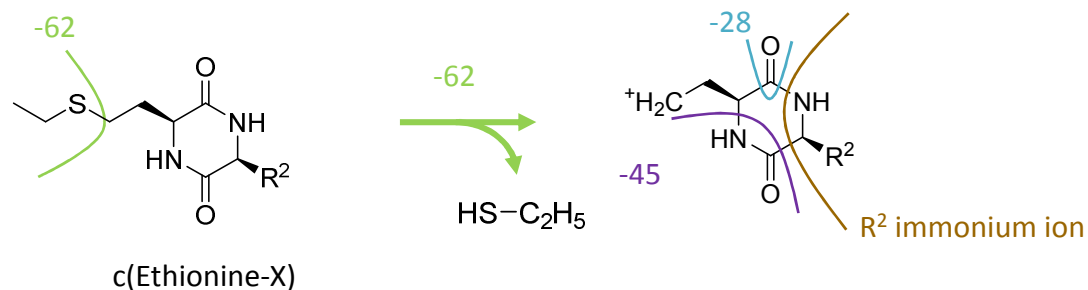


Figure S4 Analysis of mass spectra for cyclodipeptides with non-standard mass fragmentation behavior

A) Mass fragmentation of compounds corresponding to putative ethionine(Eth)-containing cyclodipeptides

A loss of 62 is systematically observed, which can be interpreted as the loss of an ethanethiol fragment. This fragmentation pathway is analogous with the one of methionine, in which the loss of a methanethiol fragment is predominant.

The MS3 fragmentation of the -62 fragment is in accordance with the 2,5-DKP structure of the molecule: mass losses of -28, -45, -56 (28+28) and -73 (28+45) are observed. Fragments corresponding to immonium ions can be observed in the cases of cyclo(Pro-Eth) and cyclo(Leu-Eth).

B

Cyclodipeptide	m/z (XH ⁺)		MS2 (XH ⁺)		MS3 (XH ⁺ -> XH ⁺ -32)								
cyclo(Gly-MeEGlu)	200.9	EIC Mass loss Relative intensity	168.9 -32 100		151 -17.9 4.1	140.9 -28 55.8	124 -44.9 3.5	113 -55.9 70.3	111.7 -57.2 0.5	96.1 -72.8 12	84.2 -84.7 100		
cyclo(Pro-MeEGlu)	241	EIC Mass loss Relative intensity	208.9 -32.1 100	181 -60 3.4	152.9 -56 46.5	125 -83.9 100	97.2 -111.7 4	84.3 -124.6 5.9	82.3 -126.6 1.9	70.4 -138.5 10.1			
cyclo(Leu-MeEGlu)	257	EIC Mass loss Relative intensity	224.9 -32.1 100	197 -60 10.2	206.9 -18 0.8	196.9 -28 100	152 -72.9 0.7	140.9 -84 28.6	129 -95.9 29.2	113.1 -111.8 7	86 -138.9 18.2	84.3 -140.6 9.1	
cyclo(Phe-MeEGlu)	291	EIC Mass loss Relative intensity	259 -32 100	231 -60 6.2	241 -18 0.7	231 -28 100	202.9 -56.1 2	185.9 -73.1 11	120 -139 41.2	84.3 -174.7 1.1			
cyclo(Tyr-MeEGlu)	307	EIC Mass loss Relative intensity	275 -32 100		246.9 -28.1 100	229.9 -45.1 0.4	218.9 -56.1 1.8	201.9 -73.1 6.6	152.9 -122.1 6	136 -139 20.8	125 -150 1.7	107.1 -167.9 9.7	

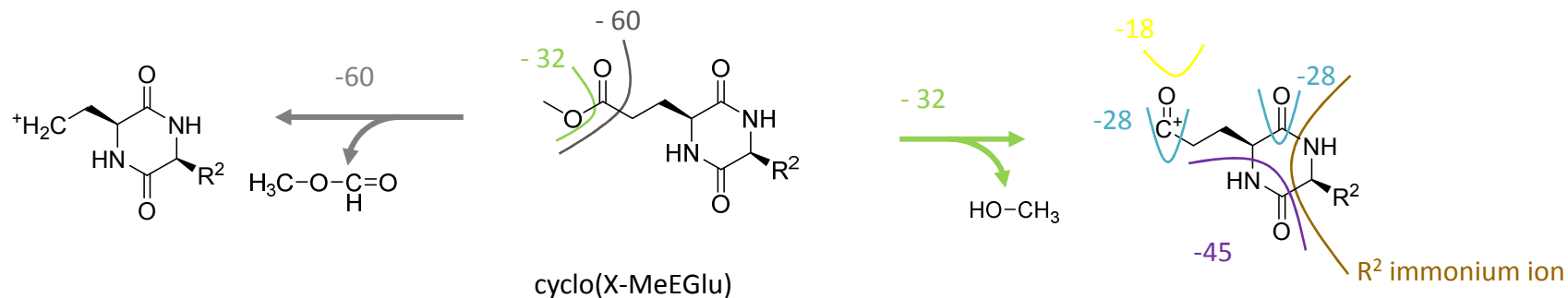


Figure S4 (continued)

B) Mass fragmentation of compounds corresponding to putative methyl ester glutamate (MeEGlu)-containing cyclodipeptides

A loss of 32 is systematically observed and a weak loss of 60 is observed in 3 out of 5 cases. These fragmentation patterns can be interpreted as the predominant loss of a methanol fragment and an additional loss of the complete methyl methanoate fragment. This fragmentation pathway is coherent with the fragmentation of glutamate containing cyclodipeptides, which exhibit a -18 loss corresponding to the loss of a H₂O group.

The MS3 fragmentation of the -32 son ion is in accordance with the 2,5-DKP structure of the molecule: mass losses of -28, -45, -56 (28+28) and -73 (28+45) are observed. A mass loss of 18, observed in some cases can be interpreted as a loss of a water molecule. Fragments corresponding to immonium ions can be observed in the cases of cyclo(Pro-MeGlu), cyclo(Leu-MeGlu), cyclo(Phe-MeGlu) and cyclo(Tyr-MeGlu).

C

Cyclodipeptide	m/z (XH ⁺)	MS2 (XH ⁺)										MS3 (XH ⁺ → YH ⁺)					
		EIC	Mass loss	Relative intensity	YH ⁺	Mass spectra											
cyclo(Phe-Aha)	274	246	217	203	201	189	184	156	144	120	184	156	100	-28	-90		
cyclo(Tyr-Aha)	290	262	234	233	219	217	205	200	172	136	200	172	182	100	46		
cyclo(Ile-Aha)	240	212	183	169	167	155	150	141	139	127	98	86	150	122	100		
cyclo(Leu-Aha)	240	212	183	169	155	150	141	139	127	96	86	150	122	100	-28		
cyclo(Pro-Aha)	224	196	167	165	153	151	139	125	115	93	70	139	111	97	84	70	
cyclo(Trp-Aha)	313	285	256	242	223	202	195	169	156	144	130	223	206	195	194	67	25

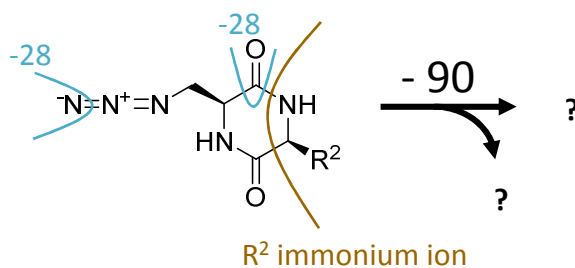


Figure S4 (continued)

C) Mass fragmentation of compounds corresponding to putative azidohomoalanine(Aha)-containing cyclodipeptides

Daughter ions corresponding to mass losses of 28, 57, 71, 73, 85 and 90 are systematically observed, whatever the amino acid with which Aha is condensed. Mass loss of 28 and 73 (28+45) can be interpreted as part of the 2,5-DKP signature. Mass losses of 57, 71, 85 and 90 are likely to be due to the presence of the azide function. However, we could not find references confirming this fragmentation pattern and we are unable to propose precise fragmentation pathways for these molecules.

c. Summary of the cyclodipeptides identified

The identification strategy presented above allowed us to identify 198 non canonical cyclodipeptides which are listed in Table S4. Corresponding canonical cyclodipeptides were introduced in the table as a reference. Residue-dependant fragmentation patterns that emerged from the comparison of the mass spectra of the cyclodipeptides containing each nCAA are summarized in Table S5.

Table S4: Cyclodipeptides identified in this study

For each molecule, the best conditions of production (strain of expression and CDPS) is reported. The retention time and the mass spectrum displayed were obtained in these conditions. Amino acids are referred by their short forms, as found in Table S3 and Table S6. UV peak area noted with a star correspond to compounds that were not totally separated from other compounds in our chromatographic conditions. In these cases, an estimation of the UV peak area is provided.

Cyclodipeptides containing proline analogs								Best conditions of production		
AA1	AA2	Cyclodipeptide	Retention time (min)	Theoretical MW	m/z (+ mode)	MS fragmentation	Auxotrophy used	CDPS	UV peak area	
cyclo(X-X)										
Pro	Pro	cyclo(Pro/Pro)	16.9	194.1	195	MS2 (195): 68.4 (0.4); 70.4 (100); 98.2 (5.9); 134 (0.3); 151 (1.5)	JW0233	25	3657	
Aze	Aze	cyclo(Aze/Aze)	14.4	166.1	166.9	MS2(167): 56.4 (14.1); 79.3 (3.5); 83.3 (6); 84.3 (11.8); 96.2 (8.9); 110.1 (11.9); 111.1 (100); 112.1 (4.3); 124 (3.4); 139 (33.7)	JW0233	25	4915	
34DePro	34DePro	cyclo(34DePro/34DePro)	17.4	190.1	190.9	MS2(191): 68.4 (91.4); 80.3 (0.5); 91.1 (0.4); 96.1 (30.8); 118 (8.8); 124.1 (0.4); 135 (6.7); 144.9 (100); 146.9 (0.5); 163 (13.4)	JW0233	25	3991	
c 4FPro	c 4FPro	cyclo(c4FPro/c4FPro)	17.9	230.1	231	MS2(231): 68.4 (4); 88.2 (53.5); 96.1 (2.6); 155 (0.8); 181.7 (1.3); 182.9 (100); 210.9 (15); 481.2 (0.9)	JW0233	25	158	
t 4FPro	t 4FPro	cyclo(t4FPro/t4FPro)	18.8	230.1	231	MS2 (231): 68.4 (7.7); 80.3 (0.2); 88.2 (100); 116 (11.9); 186.9 (1.6)	JW0233	25	2595	
t4HOPro	t4HOPro	cyclo(t4HOPro/t4HOPro)	14.2	226.1	227	MS2 (227): 68.4 (9.8); 86.2 (100); 96.1 (0.4); 114 (7); 153 (0.5); 164.9 (0.6); 180.9 (5.4); 182.9 (1.2); 198.9 (0.7); 208.9 (0.8)	JW0233	25	3103	
cyclo(Trp-X)										
Trp	Pro	cyclo(Trp/Pro)	24.2	283.3	284.1	MS2 (284): 115.1 (6.2); 117 (2.1); 118.1 (2); 125.1 (2.2); 130 (100); 132 (30.4); 152.9 (4.2); 169.9 (16.6); 239 (10.6); 267 (7.8)	JW0233	74	15837	
Trp	Aze	cyclo(Trp/Aze)	25.2	269.3	270	MS2 (270): 101.1 (3.7); 117 (0.4); 118 (1.5); 130 (100); 132 (6.9); 169.9 (24.6); 198.9 (0.4); 224.9 (0.7); 241 (2.2); 253 (0.6)	JW0233	74	20891	
Trp	34DePro	cyclo(Trp/34DePro)	26.3	281.3	282	MS2 (282): 113.1 (3.4); 118.1 (4.3); 130 (100); 132 (25.2); 135 (2.8); 150.9 (10.8); 164.9 (3.8); 169.9 (22.1); 237 (11.2); 265 (2.8)	JW0233	74	6014	
Trp	c 4FPro	cyclo(Trp/c4FPro)	25.7	301.3	302.1	MS2 (302): 88.2 (1.8); 117.1 (1.2); 130 (100); 132 (24.1); 132.9 (5.1); 145.9 (1.2); 169.9 (14.3); 170.9 (1); 256.9 (2.2); 285 (4)	JW0233	74	3585	
Trp	t 4FPro	cyclo(Trp/t4FPro)	25.7	301.3	302.1	MS2 (302): 88.2 (1.6); 117 (1); 118.1 (1.2); 130 (100); 132 (18.5); 133 (5.6); 169.9 (18); 170.9 (1.7); 257 (1.5); 285.1 (3.9)	JW0233	74	17236	
Trp	t4HOPro	cyclo(Trp/t4HOPro)	22.7	299.3	300	MS2 (300): 118 (8.3); 129.9 (100); 130.8 (4); 132 (36.6); 153.2 (5); 168.9 (9.9); 169.9 (17.2); 250.1 (4.1); 255 (5.7); 282.1 (8.2); 283 (2.6)	JW0233	74	56	
Trp	2ThPro	cyclo(Trp/2ThPro)	25.5	301.3	302	MS2 (302): 103 (0.8); 129.9 (100); 131.9 (10.1); 132.9 (1.3); 169.9 (9); 186.9 (1); 257 (0.8); 268 (0.9); 283.9 (0.8); 285 (1.8)	JW0233	74	199	
cyclo(Tyr-X)										
Tyr	Pro	cyclo(Tyr/Pro)	19	260.3	261.2	MS2 (261): 91.2 (3.1); 119 (3.6); 136 (100); 146.9 (9.5); 155 (7.4); 216 (3.2); 233 (40); 244 (2.7)	JW0233	22	6237	
Tyr	Aze	cyclo(Tyr/Aze)	19.5	246.3	247	MS2 (247): 107.1 (10.6); 136 (100); 146.9 (37.4); 148 (12.6); 162 (4.4); 191 (3.8); 201.9 (4.2); 219 (19.7)	JW0233	22	9708	
Tyr	34DePro	cyclo(Tyr/34DePro)	20.8	258.3	259	MS2 (259): 91.2 (4.1); 107 (1); 119 (4.3); 135 (1.5); 136 (100); 146.9 (1.7); 186 (6.3); 213 (1.1); 214 (20.3); 231 (53.6)	JW0233	22	1814	
Tyr	c 4FPro	cyclo(Tyr/c4FPro)	20.2	278.3	279	MS2 (279): 88.2 (1.5); 91.2 (7.5); 107.1 (11.9); 119 (6.1); 130.9 (1.7); 136 (100); 146.9 (13.1); 172.9 (12); 231 (1.5); 251 (12.1)	JW0233	22	3000*	
Tyr	t 4FPro	cyclo(Tyr/t4FPro)	20	278.3	279.1	MS2 (279): 88.3 (3.1); 91.2 (9.1); 107.1 (17.9); 119 (8.4); 131.1 (1.7); 136 (100); 146.9 (23.6); 172.9 (9.1); 251 (23.1); 262 (2.5)	JW0233	22	4285	
Tyr	t4HOPro	cyclo(Tyr/t4HOPro)	18	276.3	277	MS2 (277): 86.2 (29.2); 91.2 (9.2); 107 (4.8); 119 (7.5); 135.9 (100); 146.9 (18.6); 170.9 (11.5); 231.9 (4.1); 249 (36); 260 (3.5)	JW0234	22	448	
cyclo(Phe-X)										
Phe	Pro	cyclo(Phe/Pro)	23	244.3	245.1	MS2 (245): 70.4 (50.7); 103.1 (2.9); 120 (100); 131 (3); 172 (1.8); 189 (2.7); 199.9 (3); 217 (49.8)	JW0233	22	2325	
Phe	Aze	cyclo(Phe/Aze)	23.5	230.3	231.1	MS2 (231): 91.1 (1.3); 120 (100); 130.9 (2.7); 132 (3.2); 146 (1); 175 (2.2); 185.9 (2.6); 203 (7.4)	JW0233	22	5622	
Phe	34DePro	cyclo(Phe/34DePro)	24.7	242.3	243.1	MS2 (243): 68.5 (0.4); 93.2 (0.4); 102.2 (0.4); 103.1 (1.1); 120 (100); 148 (0.4); 169.9 (0.5); 197 (1.3); 198 (1.4); 215 (27)	JW0233	22	1017	
Phe	c 4FPro	cyclo(Phe/c4FPro)	24.7	262.3	263.1	MS2 (263): 88.2 (0.9); 93.2 (0.6); 103.2 (2.7); 120 (100); 130.9 (1.2); 170 (0.9); 215.1 (1); 217 (2); 218 (0.8); 235 (9.8)	JW0233	22	1000	
Phe	t 4FPro	cyclo(Phe/t4FPro)	24	262.3	263	MS2 (263): 88.3 (2.2); 103.1 (3.1); 120 (100); 131 (5.9); 170 (2.2); 217.9 (2.2); 235 (15.2); 246 (1.2)	JW0233	22	1851	
Phe	t4HOPro	cyclo(Phe/t4HOPro)	21.3	260.3	261	MS2 (261): 86.2 (53); 103.1 (2.4); 120 (100); 130.9 (4); 156 (0.7); 169.9 (3.4); 188 (2.7); 215.9 (2.7); 233 (29.3); 243.9 (1.1)	JW0233	22	117	
cyclo(Leu-X)										
Leu	Pro	cyclo(Leu/Pro)	21.5	210.2	211	MS2 (211): 70.5 (100); 86.3 (40.3); 114.1 (4.4); 127 (9.5); 138.1 (36.6); 155.1 (14.4); 166 (5.6); 167.1 (4.2); 183 (65.1); 194 (25.5)	JW0233	82	2069	
Leu	Aze	cyclo(Leu/Aze)	21.1	196.2	197	MS2 (197): 85.3 (28.7); 86.3 (70.5); 113.1 (100); 125.1 (26.4); 140 (40.1); 141 (33.1); 152 (31.8); 155 (20.1); 169 (78.1); 180 (41.8)	JW0233	82	2965	
Leu	34DePro	cyclo(Leu/34DePro)	22.4	208.2	209	MS2 (209): 68.4 (78.8); 86.3 (100); 114.1 (27.3); 136 (16.4); 163 (15.3); 166 (17.3); 181 (5.8)	JW0233	82	727	
Leu	c 4FPro	cyclo(Leu/c4FPro)	22.5	228.2	229	MS2 (229): 68.5 (20.6); 70.4 (13.6); 86.3 (95.1); 88.2 (100); 114.1 (8.8); 136.1 (18.5); 153 (50.1); 181 (54.2); 201 (25.2); 211.9 (19.8)	JW0233	82	771	
Leu	t 4FPro	cyclo(Leu/t4FPro)	22.9	228.2	229	MS2 (229): 68.5 (6.9); 86.3 (100); 88.2 (48); 114.1 (6.6); 136 (32.8); 145 (6.9); 153 (14.4); 181 (5.1); 201 (41.5); 212 (12.7)	JW0233	82	360	
Leu	t4HOPro	cyclo(Leu/t4HOPro)	19.6	226.2	226.8	MS2 (227): 68.4 (7.4); 86.2 (100); 136 (24.7); 143 (5.4); 153 (9.8); 154 (6.5); 170.9 (4.7); 180.9 (4.8); 199 (41.1); 209.9 (11.6)	JW0233	82	290	
Leu	4ThPro	cyclo(Leu/4ThPro)	23.3	228.3	229.1	MS2 (229): 86.3 (100); 88.2 (10.1); 113 (0.7); 154.9 (1.4); 169 (1.9); 172.9 (2.1); 200.9 (18.3); 210.9 (1.4)	JW0233	82	133	
cyclo(Ile-X)										
Ile	Pro	cyclo(Ile/Pro)	20.9	210.2	211.2	MS2 (211): 69.5 (3.9); 70.4 (33.1); 86.3 (72.8); 127 (2.1); 138.1 (35); 155 (38.3); 183 (100); 194 (17.1)	JW0233	82	EIC	
Ile	Aze	cyclo(Ile/Aze)	20.2	196.2	197	MS2 (197): 86.3 (50.6); 96.2 (26); 97.1 (100); 99.1 (25.6); 101.1 (81.2); 113 (59.5); 140 (34.4); 152 (65.5); 169 (73.8); 180 (50.8)	JW0233	82	42	
Ile	34DePro	cyclo(Ile/34DePro)	21.8	208.2	209	MS2 (209): 68.4 (100); 72.4 (5.2); 80.3 (4.5); 82.3 (4.1); 86.3 (25.7); 114.1 (58); 136 (27.6); 163 (15.4); 166 (64.9); 181 (8.5)	JW0233	82	50	
Ile	t 4FPro	cyclo(Ile/t4FPro)	21.7	228.2	229	No fragmentation	JW0233	82	EIC	

Table S4 (continued)

Cyclodipeptides containing proline analogs

AA1	AA2	Cyclodipeptide	Retention time (min)	Theoretical MW	m/z (+ mode)	MS fragmentation	Best conditions of production		
							Auxotrophy used	CDPS	UV peak area
cyclo(Ala-X)									
Ala	Pro	cyclo(Ala/Pro)	15.5	168.1	169	MS2 (169): 68.4 (0.5); 70.4 (100); 72.3 (0.4); 96.2 (4.1); 98.1 (2); 113.1 (3); 124 (2.6); 125.1 (3.9); 141 (19.6); 151.9 (8)	JW0233	89	1895
Ala	Aze	cyclo(Ala/Aze)	14.1	154.1	154.9	MS2 (155): 56.4 (34); 72.4 (12.1); 83.3 (9.1); 98.1 (18.1); 99.1 (73.8); 110.1 (11.3); 111.1 (9.4); 113.1 (16.7); 127 (100); 137.9 (20.5)	JW0233	89	3717
Ala	34DePro	cyclo(Ala/34DePro)	15.9	166.1	166.9	MS2 (167): 68.4 (100); 72.4 (18.6); 94.1 (1.8); 96.2 (3.9); 106.1 (2.5); 107.1 (1.6); 121.1 (18.4); 124.1 (23.8); 139 (7.1); 148.9 (1.4)	JW0233	89	542
Ala	c 4FPro	cyclo(Ala/c4FPro)	15.4	186.1	187	MS2 (187): 68.5 (9.9); 88.2 (100); 111.1 (30.9); 139 (33.5); 159 (8)	JW0233	89	507
Ala	t 4FPro	cyclo(Ala/t4FPro)	16	186.1	186.9	MS2 (187): 88.2 (100); 94.2 (11.1); 139 (5.5); 159 (28.5); 169.9 (8.3)	JW0233	89	2291
Ala	t4HOPro	cyclo(Ala/t4HOPro)	14	184.1	184.9	MS2 (185): 68.4 (9.6); 86.2 (100); 94.1 (11.1); 111.1 (16.1); 129 (4.6); 139 (10.5); 156.9 (20.9); 167.9 (8.5)	JW0233	89	200*
cyclo(Pro-X)									
Pro	Pro	cyclo(Pro/Pro)	16.9	194.1	195	MS2 (195): 68.4 (0.4); 70.4 (100); 98.2 (5.9); 134 (0.3); 151 (1.5)	JW0233	25	3657
Pro	Aze	cyclo(Pro/Aze)	15.9	180.1	181	MS2 (181): 56.4 (5.7); 69.7 (2.1); 70.4 (70.2); 72.4 (4); 82.3 (2.2); 112.1 (2.7); 125 (19.7); 139 (100); 151 (13); 152.9 (93.2)	JW0233	100	192
Pro	34DePro	cyclo(Pro/34DePro)	17.3	192.1	192.9	MS2 (193): 68.5 (7.8); 70.4 (100); 96.1 (1.2); 98.2 (12.8); 100.3 (1); 125.1 (0.7); 137 (0.8); 147 (8.9); 149 (1.2); 164.9 (3.8)	JW0233	100	EIC
Pro	c 4FPro	cyclo(Pro/c4FPro)	17.5	212.1	212.9	MS2 (213): 68.5 (1.5); 70.4 (40.3); 88.2 (18.9); 98.1 (4.7); 125.9 (1.1); 137 (13.1); 149 (1.2); 165 (100); 192.9 (18.5); 195 (2.3)	JW0233	100	EIC
Pro	t 4FPro	cyclo(Pro/t4FPro)	17.5	212.1	213	MS2 (213): 68.5 (3.4); 70.4 (100); 88.2 (59.2); 95.2 (0.3); 98.1 (16.5); 116.1 (5.5); 149 (0.6); 168.9 (2); 177.1 (0.3); 195 (0.6)	JW0233	100	102
Pro	t4HOPro	cyclo(Pro/t4HOPro)	15.7	210.1	211	MS2 (211): 68.4 (11.6); 70.4 (95); 86.2 (100); 98.1 (7.3); 114 (10.1); 137 (2.7); 164.9 (4.6); 166.9 (5); 191.9 (1); 193 (1)	JW0233	25	62
cyclo(Met-X)									
Met	Pro	cyclo(Met/Pro)	19.7	228.2	229.1	MS2 (229.1): 180.9 (100) MS3 (229.1->180.9): 70.4 (8.5); 73.9 (0.5); 80.3 (0.8); 82.4 (2.9); 84.3 (8.4); 97.2 (3.3); 125 (100); 153 (42.8)	JW0233	10	817
Met	Aze	cyclo(Met/Aze)	19	214.2	214.9	MS2 (214.9): 167 (100) MS3 (214.9->167): 56.4 (1.7); 74.4 (2.6); 84.3 (20); 85.2 (1.3); 96.2 (3.7); 111.1 (100); 124 (0.7); 138.9 (43.6)	JW0233	10	397
Met	34DePro	cyclo(Met/34DePro)	20.6	226.2	227.1	MS2 (227.1): 178.9 (100) MS3 (227.1->178.9): 68.4 (15.4); 80.2 (5.1); 84.2 (40.1); 95.2 (100); 111 (67.5); 123.1 (75.5); 133.9 (36.4); 150.9 (58.9)	JW0233	98	80
Met	c 4FPro	cyclo(Met/c4FPro)	20.6	246.2	247.5	MS2 (247.5): 199 (100) MS3(247->199): 68.4 (1.9); 84.3 (8.9); 88.3 (34.3); 115.1 (4); 123 (7.5); 143 (100); 151 (26.1); 171 (23.2)	JW0233	10	81
Met	t 4FPro	cyclo(Met/t4FPro)	20.6	246.2	247	MS2 (247): 199 (100) MS3 (247->199): 74.5 (1.7); 84.3 (9.3); 88.2 (27.9); 115 (12.4); 116 (4.3); 123 (9.5); 134 (3.4); 142.9 (100); 151 (21.7); 170.9 (43.3)	JW0233	98	74
cyclo(Val-X)									
Val	Pro	cyclo(Val/Pro)	18.5	196.1	215	MS2 (215): 70.5 (39.9); 72.4 (52.8); 98.1 (1.5); 113.1 (1.7); 124.1 (30.1); 141.1 (42.8); 152 (3.9); 153.1 (1.6); 169 (100); 180 (15.9)	JW0233	82	125
Val	Aze	cyclo(Val/Aze)	17.6	182.1	182.9	MS2 (183): 56.4 (24); 72.4 (56); 83.3 (59.7); 98.2 (23.7); 99.1 (31.4); 101.1 (40.5); 113.1 (26.9); 126.1 (30.6); 127 (39.8); 138 (40); 155 (100)	JW0233	82	358
Val	34DePro	cyclo(Val/34DePro)	19.5	194.1	195	MS2 (195): 68.4 (34.3); 72.4 (100); 80.3 (2.9); 82.3 (1.2); 100.1 (14.3); 122 (54.9); 134 (1.5); 149 (4.8); 152 (11.1); 167 (12.7)	JW0233	82	140
Val	c 4FPro	cyclo(Val/c4FPro)	19.7	214.1	215.1	MS2 (215): 68.5 (11.1); 70.5 (16.3); 72.4 (93.3); 88.2 (91); 100.1 (13); 122.1 (39.5); 139 (80.8); 167 (100); 187 (82.5); 197.9 (11.9)	JW0233	82	88
Val	t 4FPro	cyclo(Val/t4FPro)	19.3	214.1	215	MS2 (215): 72.4 (100); 88.3 (17.1); 122.1 (28.8); 139 (29.1); 159 (6.2); 167 (3.3); 187 (55); 197.9 (7.9)	JW0233	82	EIC
Val	t4HOPro	cyclo(Val/t4HOPro)	17.1	212.1	213	MS2 (213): 68.4 (6.3); 72.4 (73.4); 86.2 (36.7); 122 (40); 139 (35.1); 157 (19.8); 167 (10); 167.9 (5.4); 184.9 (100); 195.9 (9.9)	JW0233	82	EIC

Table S4 (continued)

Cyclodipeptides containing analogs of phenylalanine/tyrosine							Best conditions of production		
AA1	AA2	cyclo(AA1/AA2)	Retention time (min)	Theoretical MW	m/z (+ mode)	MS fragmentation	Strain	CDPS	UV peak area
cyclo(X-X)									
Tyr	Tyr	cyclo(Tyr/Tyr)	22.1	326.4	327.1	MS2 (327): 107 (3.2); 119 (2.7); 135.9 (37.4); 193 (1.2); 220.9 (12); 254 (6.7); 282 (4.5); 299.1 (100)	JW2580	94	>25000
3FTyr	3FTyr	cyclo(3FTyr/3FTyr)	25.2	362.3	363.1	MS2 (363.1): 106.1 (0.7); 109 (5.8); 133.9 (0.9); 136.9 (2.5); 153.9 (48.6); 238.9 (5); 290 (7.3); 300 (1.2); 318.1 (7.9); 335.1 (100)	JW2581	94	11018
DOPA	DOPA	cyclo(DOPA/DOPA)	20.6	358.3	359.1	MS2 (359.1): 115 (4.9); 123 (3.2); 151.9 (15.6); 161.9 (11.9); 206.9 (14.9); 208.9 (19.4); 236.9 (61.7); 286 (95.1); 314.1 (10.9); 331.1 (100)	JW2581	94	10706
Phe	Phe	cyclo(Phe/Phe)	27.7	294.2	295	MS2 (295): 93.2 (1.2); 103.1 (3.3); 120 (100); 130 (0.7); 222 (3.7); 232 (1.1); 250.1 (4.9); 267 (69.8)	JW2326	62	2426
4FPhe	4FPhe	cyclo(4FPhe/4FPhe)	29.4	330.3	331.1	MS2 (331.1): 109 (0.5); 117 (0.8); 118 (13.5); 120.9 (0.4); 137.9 (87.4); 258 (8.1); 268 (0.3); 286 (4.7); 303 (100); 314 (0.5)	JW2580	91	1300*
2FPhe	2FPhe	cyclo(2FPhe/2FPhe)	28.1	330.3	331.1	MS2 (331.1): 117 (0.3); 118 (5.1); 137.9 (46.4); 147.9 (0.5); 175.9 (0.4); 258 (17.8); 286 (12.9); 302 (0.4); 303 (100); 311 (1.1)	JW2580	91	4918
3ThienA	3ThienA	cyclo(3ThienA/3ThienA)	27.1	306.3	306.9	MS2 (306.9): 97.1 (5); 99.1 (27.7); 125.9 (11.7); 135.9 (14.5); 136.9 (20); 180.9 (100); 208.9 (7.9); 233.9 (23); 261.9 (32.6); 278.9 (60.8); 288.9 (3.4); 289.9 (7.5)	JW2326	AlbC	337
3PyrA	3PyrA	cyclo(3PyrA/3PyrA)	13.6	296.2	297.1	93.1 (71.6); 121 (26); 146.9 (13); 161.9 (21.4); 204.9 (100); 234 (12.8); 251 (11.2); 279 (19.3)	JW2326	62	+/- 100*
4PyrA	4PyrA	cyclo(4PyrA/4PyrA)	8.5	296.2	297.1	93.1 (76.4); 94.1 (67.3); 104 (23.9); 148.9 (100); 161.9 (15); 203.9 (35.9); 204.9 (31.3); 252 (22.2)	JW2326	AlbC	495
cyclo(Phe-X)									
Phe	Tyr	cyclo(Phe/Tyr)	23.5	310.3	311.1	MS2 (311.1): 91.2 (2.4); 118.9 (2.5); 120 (25.3); 135.9 (39.1); 204.9 (9); 238 (5.4); 266 (6.2); 283 (100)	JW2326	94	2688
Phe	3FTyr	cyclo(Phe/3FTyr)	24.6	328.3	329.1	MS2 (329.1): 103.1 (2.3); 109.1 (0.8); 120 (80.5); 136.9 (0.5); 153.9 (16.8); 204.9 (1.2); 256 (6.1); 266 (1.3); 284 (6.1); 301 (100)	JW2581	91	3374
Phe	DOPA	cyclo(Phe/DOPA)	25.5	326.3	327.1	MS2 (327.1): 107.1 (1.3); 120 (20.6); 134.9 (1.8); 151.9 (24.4); 165 (1.5); 176.9 (2.2); 204.9 (22.3); 254 (48.7); 282.1 (6.5); 299.1 (100)	JW2581	94	2247
Phe	Phe	cyclo(Phe/Phe)	27.7	294.2	295	MS2 (295): 93.2 (1.2); 103.1 (3.3); 120 (100); 130 (0.7); 222 (3.7); 232 (1.1); 250.1 (4.9); 267 (69.8)	JW2326	62	2426
Phe	2FPhe	cyclo(Phe/2FPhe)	28.2	312.3	313.1	MS2 (313.1): 91.2 (1.8); 93.2 (2.6); 103.1 (7.7); 118 (3); 120 (61.8); 137.9 (29.5); 240 (8); 268 (8.5); 285.1 (100); 296 (1)	JW2326	62	1810
Phe	4FPhe	cyclo(Phe/4FPhe)	29	312.3	313.1	MS2 (313.1): 91.2 (1.8); 93.2 (3.1); 103.1 (10.4); 118 (2.3); 120 (77.9); 137.9 (13.1); 240 (7.2); 250 (0.7); 268 (7.1); 285 (100)	JW2326	62	459
Phe	3ThienA	cyclo(Phe/3ThienA)	27.1	300.3	301	MS2 (301.1): 99.1 (10.4); 120 (27.5); 126 (5.3); 130 (6.8); 136.9 (4.5); 175 (15.7); 202.9 (6); 228 (37.7); 256 (14.1); 273 (100)	JW2326	62	604
Phe	3PyrA	cyclo(Phe/3PyrA)	18.7	295.3	296.1	MS2 (296.1): 93.1 (73.7); 94.1 (15); 120 (23.7); 121 (100); 148.9 (13.8); 204.9 (33.3); 223 (12.5); 232.9 (16.9); 250 (23.7); 268 (17.1)	JW2326	62	774
Phe	4PyrA	cyclo(Phe/4PyrA)	18.6	295.3	296.1	MS2 (296.1): 93.1 (55.9); 94.1 (100); 120 (15.7); 121 (18.3); 131.9 (6.8); 161.9 (3.5); 204.9 (9.1); 222.9 (2.6); 251 (19.3); 268 (17.8)	JW2326	62	1429
cyclo(Tyr-X)									
Tyr	Tyr	cyclo(Tyr/Tyr)	22.1	326.4	327.1	MS2 (327): 107 (3.2); 119 (2.7); 135.9 (37.4); 193 (1.2); 220.9 (12); 254 (6.7); 282 (4.5); 299.1 (100)	JW2580	94	>25000
Tyr	3FTyr	cyclo(Tyr/3FTyr)	23.2	344.3	345.1	MS2 (345.1): 135.9 (34.1); 153.9 (6.1); 238.9 (11.3); 272 (7.3); 300.1 (5); 317.1 (100)	JW2580	94	6000*
Tyr	DOPA	cyclo(Tyr/DOPA)	21.3	342.3	343.1	MS2 (343.1): 107 (2.4); 136 (13.5); 151.9 (10.4); 192.9 (2.2); 208.9 (1.8); 220.9 (15); 236.9 (8.5); 270 (41.7); 298 (4.6); 315.1 (100)	JW2580	94	8300
Tyr	Phe	cyclo(Tyr/Phe)	23.5	310.3	311.1	MS2 (311.1): 91.2 (2.4); 118.9 (2.5); 120 (25.3); 135.9 (39.1); 204.9 (9); 238 (5.4); 266 (6.2); 283 (100)	JW2326	94	2688
Tyr	2FPhe	cyclo(Tyr/2FPhe)	23.7	328.3	329.1	MS2 (329.1): 107.1 (3.7); 118 (2); 119 (3.7); 135.9 (47.5); 137.9 (14.1); 146.9 (1.2); 222.9 (12.9); 256 (6.8); 284 (4.7); 301 (100)	JW2580	91	2855
Tyr	3ThienA	cyclo(Tyr/3ThienA)	22.7	316.3	317.1	MS2 (317.1): 107.1 (9.7); 113 (7.3); 135.9 (20.7); 136.9 (3.4); 190.9 (8.8); 210.9 (10.9); 218.9 (3.6); 243.9 (32.5); 272 (21.2); 289 (100)	JW2580	91	1355
Tyr	3PyrA	cyclo(Tyr/3PyrA)	15.8	311.3	312.1	MS2 (312.1): 93.1 (47.5); 94.1 (8.2); 107 (10.1); 121 (64.1); 135.9 (12.4); 148.9 (8.5); 204.9 (18.9); 205.9 (100); 266 (7.5); 284.1 (12.4)	JW2580	91	2662
Tyr	4PyrA	cyclo(Tyr/4PyrA)	17.9	311.3	312.1	MS2 (312.1): 93.1 (68.3); 94.1 (100); 107 (4.9); 121 (24.5); 135.9 (13.4); 161.9 (6); 204.9 (14.6); 205.9 (11.1); 267 (28.4); 284 (24.7)	JW2580	94	1975
cyclo(Leu-X)									
Leu	Tyr	cyclo(Leu/Tyr)	21.7	276.2	277	MS2 (277): 86.3 (12.7); 91.1 (5.8); 107.1 (5.1); 119 (4.4); 135.9 (90.4); 171 (8.7); 204 (6.3); 221 (1.4); 232 (11.7); 249 (100)	JW2326	AlbC	900*
Leu	3FTyr	cyclo(Leu/3FTyr)	24	294.3	295.1	MS2 (295.1): 86.3 (31.7); 109 (1.3); 136.9 (0.7); 153.9 (33.8); 164.9 (0.6); 171 (2.4); 222 (4.1); 239 (1); 250 (27.2); 267 (100); 278 (0.8)	JW2326	AlbC	9000*
Leu	DOPA	cyclo(Leu/DOPA)	21.4	292.3	293.1	MS2 (293.1): 86.3 (4.8); 123 (1.2); 134.9 (1.6); 151.9 (35.5); 163.9 (1.3); 171 (23.7); 220 (27.2); 248 (7.9); 265 (100); 275 (1.2)	JW2326	AlbC	978
Leu	Phe	cyclo(Leu/Phe)	27.2	260.2	261	MS2 (261): 86.3 (27.5); 103.1 (2.3); 120 (100); 188 (4.1); 197.9 (0.5); 204.9 (0.7); 216 (23); 233 (62.4)	JW2326	AlbC	8461
Leu	2FPhe	cyclo(Leu/2FPhe)	27.5	278.3	279	MS2 (279): 86.3 (17.2); 91.2 (1.6); 118 (5.3); 137.9 (58.9); 149.9 (1.6); 195 (2.3); 206 (27.1); 223 (2.6); 234 (95.3); 251 (100)	JW2326	AlbC	5279
Leu	3-ThienA	cyclo(Leu/3-ThienA)	26.4	266.3	267	MS2 (267): 86.3 (17.6); 99.1 (15.8); 113 (5.7); 126 (12.1); 136.9 (3.5); 141 (15.4); 169 (2.9); 193.9 (23.1); 221.9 (15.3); 238.9 (100)	JW2326	AlbC	2345
Leu	3PyrA	cyclo(Leu/3PyrA)	17.5	261.3	262	MS2 (262): 93.1 (100); 94.1 (6); 121 (11.9); 148.9 (12.9); 188.9 (4); 198.9 (9.7); 216 (10.1); 217 (6); 234 (2.9); 245 (4.8)	JW2326	AlbC	3168
Leu	4PyrA	cyclo(Leu/4PyrA)	17.3	261.3	262	MS2 (262): 78.3 (1.8); 86.3 (3.1); 92.2 (2.3); 93.1 (83.2); 94.1 (100); 106 (0.8); 121 (4.8); 131.9 (5.3); 216.9 (9.7); 234 (5.1)	JW2326	AlbC	1735
cyclo(Ala-X)									
Ala	Tyr	cyclo(Ala/Tyr)	17.2	234.2	235	MS2 (235): 107.1 (29.6); 119 (1.2); 129 (13.4); 135.9 (100); 146.9 (3); 162 (3.1); 163.9 (1.9); 179 (1.3); 189.9 (4.7); 206.9 (86.5)	JW2326	22	3074
Ala	3FTyr	cyclo(Ala/3FTyr)	18.2	252.2	253	MS2 (253): 109.1 (0.7); 125 (1.6); 129 (6.7); 136.9 (0.6); 153.9 (100); 164.9 (1.7); 180 (4.6); 196.9 (0.6); 207.9 (6.8); 224.9 (71.2)	JW2326	22	1036
Ala	DOPA	cyclo(Ala/DOPA)	18.2	250.2	251	MS2 (251): 87.2 (2.3); 89.2 (2.5); 123 (7); 129 (51.3); 151.9 (50.9); 177.9 (21.6); 203.9 (2.7); 204.9 (4.2); 205.9 (4); 222.9 (100)	JW2580	94	5790
Ala	Phe	cyclo(Ala/Phe)	21	218.2	219.1	MS2 (219.1): 120(100); 190.9(43.6)	JW2326	62	2492
Ala	2FPhe	cyclo(Ala/2FPhe)	22.2	236.2	236.9	MS2 (236.9): 91.2 (0.4); 116.9 (0.4); 118 (4); 137.9 (100); 164 (6.1); 180.9 (0.4); 191.9 (27.8); 208.9 (58.4); 216.9 (2.1); 219.9 (0.4)	JW2326	AlbC	300
Ala	4FPhe	cyclo(Ala/4FPhe)	22.4	236.2	236.9	MS2 (236.9): 118.1 (2.2); 137.9 (100); 163.9 (3.3); 166 (0.4); 180.9 (0.4); 191.9 (12.6); 208.9 (47.5); 219.9 (0.4)	JW2326	22	722
Ala	3ThienA	cyclo(Ala/3ThienA)	19.8	224.2	224.9	MS2 (224.9): 71.4 (3.9); 89.2 (3.2); 99.1 (100); 126 (21.4); 127 (2.8); 136.9 (11.7); 151.9 (22); 179.9 (10.9); 196.9 (39.2); 207.9 (5.9)	JW2326	62	304
Ala	3PyrA	cyclo(Ala/3PyrA)	8.9	219.2	220.2	MS2 (220): 93.1 (100); 121 (5); 148.9 (7.5); 156.9 (4.4); 173.9 (4.7); 202.9 (2.2)	JW2326	22	153
Ala	4PyrA	cyclo(Ala/4PyrA)	8	219.2	220.2	MS2 (220): 92.2 (1); 93.1 (100); 94.1 (46.7); 121 (1.7); 131.9 (5); 174.9 (9.4); 191.9 (3.6); 202.9 (1.2)	JW2326	22	804

Table S4 (continued)

Cyclodipeptides containing analogs of phenylalanine/tyrosine

Cyclodipeptides containing analogs of phenylalanine/tyrosine							Best conditions of production		
AA1	AA2	cyclo(AA1/AA2)	Retention time (min)	Theoretical MW	m/z (+ mode)	MS fragmentation	Strain	CDPS	UV peak area
cyclo(Pro-X)									
Pro	Tyr	cyclo(Pro/Tyr)	19.3	260.2	261	MS2 (261): 91.2 (5.2); 107.1 (3.2); 119 (4.9); 135.9 (100); 146.9 (11.4); 154.9 (8.3); 216 (4.4); 233 (40.6)	JW2326	22	11950
Pro	3FTyr	cyclo(Pro/3FTyr)	20.5	278.2	279	MS2 (279): 109.1 (4.7); 137 (3.3); 153.9 (85.4); 154.9 (6.4); 164.9 (7); 205.9 (3.7); 233.9 (4.1); 251 (100)	JW2326	22	5218
Pro	DOPA	cyclo(Pro/DOPA)	18	276.2	277	MS2 (277): 107.1 (3.2); 113.1 (7); 115.1 (3.1); 134.9 (4.3); 151.9 (100); 154.9 (21.8); 161.9 (2.5); 203.9 (3); 232 (2.4); 249 (34.9)	JW2326	22	4662
Pro	Phe	cyclo(Pro/Phe)	23.5	244.2	245	MS2 (245): 70.4 (49.6); 120 (100); 130.9 (2.1); 172 (2); 199.9 (2.1); 217 (41.9)	JW2326	22	1735
Pro	2FPhe	cyclo(Pro/2FPhe)	24.1	262.2	263	MS2 (263): 91.2 (2.1); 98.1 (3); 118 (8.7); 137.9 (100); 153.9 (8.5); 190 (18.4); 207 (2.5); 235 (74.4); 243 (4.6); 245.9 (5.2)	JW2326	22	1369
Pro	3ThienA	cyclo(Pro/3ThienA)	22.1	250.2	250.9	MS2 (250.9): 70.4 (17.3); 97.1 (2.3); 99.1 (100); 125 (87); 125.9 (51.5); 136.9 (8.6); 152.9 (23); 153.9 (3.1); 205.9 (12.2); 222.9 (91.5)	JW2326	22	212
Pro	3PyrA	cyclo(Pro/3PyrA)	15.1	245.2	246	MS2 (246): 70.4 (6.2); 93.2 (83.2); 94.1 (9.3); 121 (29.8); 149 (30.4); 152.9 (4); 153.9 (11.8); 182.9 (9.5); 199.9 (90.9); 227.9 (100)	JW2326	22	394
Pro	4PyrA	cyclo(Pro/4PyrA)	14.8	245.2	246	MS2 (246): 70.4 (2.7); 78.3 (1.3); 93.2 (35); 94.1 (100); 96.1 (0.9); 121 (2.8); 125 (1.8); 131.9 (1.9); 152.9 (1.5); 200.9 (5.9)	JW2326	22	5086
cyclo(Met-X)									
Met	Tyr	cyclo(Met/Tyr)	19.4	294.2	295	MS2 (295.3): 230.9 (100); 246.9 (6.4); 277 (3.9) MS3 (295->231): 84.3 (3.6); 91.2 (21.7); 130 (41.9); 147 (4.1); 156 (4); 156.9 (5.6); 158 (12.6); 174.9 (41.1); 185.9 (49.4); 202.9 (100)	JW2326	62	554
Met	DOPA	cyclo(Met/DOPA)	21.3	310.3	311.1	MS2 (311.1): 107 (5.7); 132.1 (2); 134.9 (6.7); 151.9 (100); 248 (17.1); 265 (27.7); 266 (4.6); 276 (2); 293.1 (3.3); 294.1 (32.2)	JW2326	AlbC	200*
Met	Phe	cyclo(Met/Phe)	24.6	278.2	279	MS2 (279): 203 (0.7); 230.9 (100); 251 (0.3) MS3 (279->231): 84.3 (5.2); 91.2 (19.8); 130 (43.7); 147 (3.7); 156 (3.8); 157 (8.2); 158 (12.8); 175 (39.3); 185.9 (49.7); 202.9 (100)	JW2326	62	1266
Met	2FPhe	cyclo(Met/2FPhe)	24.9	296.3	297	MS2 (297): 248.9 (100) MS3 (297.2->248.9): 84.2 (3.2); 109.1 (17.9); 138 (3.4); 147.9 (17.4); 165 (4.5); 174.9 (3.1); 176 (5.4); 192.9 (64.9); 204 (46.6); 220.9 (100)	JW2326	62	444
Met	4FPhe	cyclo(Met/4FPhe)	26	296.3	297.1	MS2 (297): 248.9 (100) MS3 (297.2->248.9): 84.3 (3.7); 109.1 (33.4); 113.1 (3.6); 148 (22.8); 165 (10.1); 173 (7.3); 175.9 (15.4); 192.9 (45.7); 203.9 (47.4); 220.9 (100)	JW2326	62	458
Met	3ThienA	cyclo(Met/3ThienA)	23.6	284.3	285	MS2 (285): 236.9 (100) MS3 (285->236.9): 97.1 (10.5); 101.1 (35.7); 135.9 (43.9); 136.9 (50.1); 152.9 (11.1); 161.9 (25); 163.9 (13.8); 180.9 (21.4); 191.9 (100); 208.9 (78.5)	JW2326	62	209
Met	3PyrA	cyclo(Met/3PyrA)	16.1	279.3	280	MS2 (280): 86.3 (1.3); 93.1 (1); 94.1 (3.8); 104.1 (2.5); 106.1 (1.4); 121 (100); 166.9 (11.6); 231.9 (1.4); 234 (29.5); 263 (3.6)	JW2326	AlbC	209
Met	4PyrA	cyclo(Met/4PyrA)	15.6	279.3	280.1	MS2 (280.1): 93.2 (3.6); 94.2 (6.8); 104.1 (1.3); 132 (1.7); 149 (1.9); 176 (6.1); 186.9 (7.2); 204 (10.9); 207 (1.5); 231.9 (100); 234.9 (33); 263 (12.7)	JW2326	62	89
cyclo(Val-X)									
Val	Tyr	cyclo(Val/Tyr)	19.3	260.2	261	MS2 (261): 91.2 (5.2); 107.1 (3.2); 113 (2.3); 119 (4.9); 135.9 (100); 146.9 (11.4); 154.9 (8.3); 187.9 (2); 216 (4.4); 233 (40.7)	JW2326	22	5597
Val	3FTyr	cyclo(Val/3FTyr)	21.4	280.2	281	MS2 (281): 100.1 (1); 109.1 (1.4); 137 (0.8); 153.9 (38.6); 156.9 (0.9); 208 (15.7); 236 (14.2); 253 (100)	JW2326	22	1334
Val	DOPA	cyclo(Val/DOPA)	19.2	278.2	279	MS2 (279): 107.1 (0.9); 123 (0.9); 134.9 (1.6); 151.9 (45.9); 157 (10.8); 188.9 (1.5); 206 (40); 234 (5.7); 251 (100); 262 (0.8)	JW2326	22	1065
Val	Phe	cyclo(Val/Phe)	25	246.1	247	MS2 (247): 72.4 (7.8); 100.1 (0.8); 103.1 (1); 120 (100); 174 (10); 184 (0.8); 202 (12.4); 219 (64.8)	JW2326	22	424
Val	2FPhe	cyclo(Val/2FPhe)	25.3	264.2	265	MS2 (265): 91.2 (0.9); 118 (2.5); 138 (51.6); 192 (37.9); 209 (0.5); 220 (33.9); 237 (100); 245 (1)	JW2326	22	226
Val	4FPhe	cyclo(Val/4FPhe)	26	264.2	265	MS2 (265): 138 (41.4); 192 (24.3); 220 (25.1); 237 (100)	JW2326	22	974
Val	3-ThienA	cyclo(Val/3-ThienA)	23.8	252.2	253	MS2 (253): 72.4 (4.4); 99.1 (15.9); 126 (9.2); 127 (22.2); 136.9 (2.6); 154.9 (2.7); 162.9 (2.1); 179.9 (25.4); 207.9 (9.3); 225 (100)	JW2326	22	EIC*
Val	3PyrA	cyclo(Val/3PyrA)	15.4	247.2	248	MS2 (248): 92.2 (3.5); 93.1 (100); 94.1 (4.6); 121 (8.6); 148.9 (14.8); 174.9 (12.1); 184.9 (6); 202 (7.8); 203 (4); 230.9 (3.4)	JW2326	22	55
Val	4PyrA	cyclo(Val/4PyrA)	15.2	247.2	248	MS2 (248): 72.4 (0.7); 78.3 (0.9); 92.2 (1.9); 93.2 (100); 94.1 (98.6); 121 (3.4); 131.9 (3.1); 174.9 (2.2); 202.9 (11.5); 220 (3.9)	JW2326	22	514
cyclo(Thr-X)									
Thr	Tyr	cyclo(Thr/Tyr)	19.4			No fragmentation available	JW2326	22	Unseparated
Thr	3FTyr	cyclo(Thr/3FTyr)	17.4	282.2	283	MS2 (283): 113.1 (20.5); 125 (5.1); 141 (19.9); 191.9 (7.6); 194 (6.1); 208.9 (13.5); 237 (33.5); 239 (6.4); 255 (4.3); 265 (100) MS3 (283->265): 113.1 (100); 125 (19.2); 153.9 (32.3); 179.9 (26.6); 191.9 (39.9); 193.9 (24.3); 219.9 (24.2); 236.9 (71.3)	JW2326	22	1709
Thr	DOPA	cyclo(Thr/DOPA)	15.4	280.2	281	MS2 (281): 85.2 (5.3); 113.1 (10.7); 125 (5.3); 141 (28.1); 190 (16.1); 207 (8.5); 235 (20.1); 237 (8.9); 253 (5.2); 263 (100) MS3 (281->263): 97.1 (26); 113.1 (33.6); 125 (100); 140.9 (72.2); 177.9 (30.5); 189.9 (95.2); 217.9 (43.1); 234.9 (76.2)	JW2326	22	2830
Thr	Phe	cyclo(Thr/Phe)	20.6			No fragmentation available	JW2326	22	Unseparated
Thr	2FPhe	cyclo(Thr/2FPhe)	19.8	266.2	267	MS2 (267): 109.1 (8.5); 138 (2.6); 178 (11.2); 192.9 (18.2); 195 (2.3); 204 (1.4); 221 (64.8); 223 (6.7); 239 (5.4); 248.9 (100)	JW2326	22	296
Thr	4FPhe	cyclo(Thr/4FPhe)	20.8	266.2	267	MS2 (267): 109 (38.3); 176 (3.4); 178 (9.3); 192.9 (16.7); 221 (51.6); 222.9 (6.7); 239 (3.3); 249 (100) MS3 (267->249): 109.1 (35.5); 138 (26.4); 148 (2.6); 163.9 (11.7); 176 (11.5); 178 (33.4); 192.9 (20); 201 (3.5); 204 (12.3); 220.9 (100)	JW2326	22	1272
Thr	3-ThienA	cyclo(Thr/3-ThienA)	18.2	254.2	255	MS2 (255): 97.1 (36.1); 99.1 (10.5); 111.1 (13.8); 124.9 (14.6); 151.9 (5.1); 163.9 (10.3); 180.9 (13.7); 208.9 (20.6); 210.9 (7.7); 236.9 (100)	JW2326	22	EIC
cyclo(His-X)									
His	Phe	cyclo(His/Phe)	17.9	284.2	285	MS2 (285): 93.1 (5.5); 110.1 (97.3); 120 (11.5); 137.9 (9); 165.9 (37); 212 (88.1); 224 (6.8); 240 (22.3); 257 (71.9); 268 (100)	JW2326	17	791
His	2FPhe	cyclo(His/2FPhe)	17.9	302.2	303.1	MS2 (303): 93.1 (4.5); 110.1 (66.6); 137.9 (9.8); 165.9 (25.7); 230 (69.1); 242 (3.7); 258 (24.5); 275 (70.7); 285 (2.3); 286 (100)	JW2326	17	EIC *
His	4FPhe	cyclo(His/4FPhe)	18.6	302.2	303	MS2 (303): 93.1 (5.4); 110 (75.1); 137.9 (8); 165.9 (28.6); 229.9 (76.1); 241.9 (3.8); 258 (22.8); 275 (79.2); 285.1 (2.6); 286 (100)	JW2326	17	126
His	3-ThienA	cyclo(His/3-ThienA)	16.6	290.2	291	MS2 (291): 110 (63.4); 135.8 (11.1); 164.9 (11.9); 165.8 (65.8); 180.9 (16); 217.9 (73.1); 246 (17); 263 (48.8); 273 (9.7); 273.9 (100)	JW2326	17	EIC
His	3FTyr	cyclo(His/3FTyr)	16.2	318.2	319.1	MS2 (319): 110 (110.4); 165.9 (33.9); 245.9 (71.2); 258 (16.3); 273.9 (12.6); 282.7 (10); 290.5 (10.4); 291.1 (100); 302.1 (65.3)	JW2326	17	EIC

Table S4 (continued)

Cyclodipeptides containing analogs of phenylalanine/tyrosine

AA1	AA2	cyclo(AA1/AA2)	Retention time (min)	Theoretical MW	m/z (+ mode)	MS fragmentation	Best conditions of production		
							Strain	CDPS	UV peak area
cyclo(Asn-X)									
Asn	Phe	cyclo(Asn/Phe)	18.5	261.2	262	MS2 (262) : 244.9 (100) MS3 (262->244.9) : 91.1 (0.6); 130 (18.5); 130.9 (0.6); 146 (0.6); 146.9 (0.6); 156.9 (5.4); 157.9 (2.6); 174.9 (7.4); 189 (3); 202.9 (100)	JW2326	20	832
Asn	3FTyr	cyclo(Asn/3FTyr)	17.1	295.2	296	MS2 (296.1) : 279 (100) MS3 (296->279) : 99.2 (5.6); 163.9 (46.5); 180.9 (31); 190.9 (7.4); 191.9 (7.3); 208.9 (100); 222.9 (9.4); 236.9 (85.8)	JW2326	20	EIC
Asn	2FPhe	cyclo(Asn/2FPhe)	18.8	279.2	280	MS2 (280.2) : 263 (100) MS3 (280->263) : 109 (3.7); 121 (0.5); 147.9 (11); 164 (0.6); 164.9 (1.2); 174.9 (4.4); 175.8 (1.1); 192.9 (30.1); 206.9 (2.3); 220.9 (100)	JW2326	20	44
Asn	3-ThienA	cyclo(Asn/3-ThienA)	24.7	267.2	267.9	MS2 (268.2) : 97.1 (2.5); 111 (2.4); 183.9 (2.7); 208.9 (2.7); 221.9 (19.6); 250.9 (100) MS3 (267.9->250.9) : 97.1 (12.5); 123 (3.2); 124.9 (2.4); 148.9 (3); 166.9 (7.1); 170.9 (2); 172.9 (10.9); 204.9 (13.3); 208.8 (100); 216.8 (2)	JW2326	20	EIC
Asn	4FPhe	cyclo(Asn/4FPhe)	19.8	279.2	280	MS2 (280.2) : 263 (100) MS3 (280->263) : 109 (7); 147.9 (22); 165 (20.1); 174.9 (2); 175.9 (5.2); 192.9 (100); 206.9 (6.6); 220.9 (55.5)	JW2326	20	469

Table S4 (continued)

Cyclodipeptides containing analogs of leucine/isoleucine/valine/methionine							Best conditions of production		
AA1	AA2	cyclo(AA1/AA2)	Retention time (min)	Theoretical MW	m/z (+ mode)	MS fragmentation	Strain	CDPS	UV peak area
cyclo(X-X)									
Leu	Leu	cyclo(Leu/Leu)	25.6	226.2	227	MS2 (227): 86.3 (58.3); 98.2 (3.4); 126 (1.9); 143 (1.7); 154.1 (7.4); 171 (3.3); 182 (100); 199 (33.3)	JW5807	54	3420
Met	Met	cyclo(Met/Met)	21.9	262.2	263.1	MS2 (263): 111 (0.2); 167 (10.7); 214 (0.7); 214.9 (100) MS2 (263->214.9): 84.3 (0.4); 111.1 (3.5); 139 (3.3); 166.9 (100); 171.9 (0.9); 186.9 (0.4)	JW3973	36	700*
TriFLeu	TriFLeu	cyclo(TriFLeu/TriFLeu)	31	334.3	335.1	MS2 (335): 140 (23.9); 239.1 (5.6); 240 (18); 247 (31.2); 262 (5.7); 267.1 (14.1); 287.1 (64.8); 290.1 (45.3); 307.1 (15.1); 315.1 (100)	JW5807	54	300*
Aha	Aha	cyclo(Aha/Aha)	20.8	252.3	253	MS2 (253): 80.3 (35.4); 81.2 (21.7); 82.2 (55.1); 99 (19.2); 107 (55.1); 108 (34.7); 111 (15.6); 123 (16.7); 134.9 (100); 224.9 (35.5)	JW3973	54	100*
gMeLeu	gMeLeu	cyclo(gMeLeu/gMeLeu)	29.1	254.3	255.1	MS2 (255): 100.1 (100); 126.1 (2.2); 154 (18.3); 171 (7.3); 182 (11.5); 199 (20.7); 209 (2.8); 210 (95.7); 227.1 (32.5); 237.1 (2)	JW5807	54	EIC
cyclo(Trp-X)									
Trp	Leu	cyclo(Trp/Leu)	25.5	299.2	300.1	MS2 (300.1): 86.2 (0.9); 129.9 (100); 131.9 (22.4); 168.9 (1); 169.9 (4.7); 254.9 (1.8); 272 (1); 283 (3.4)	JW2326	68	>20000
Trp	TriFLeu	cyclo(Trp/TriFLeu)	28.2	353.3	354.1	MS2 (354): 103.1 (0.9); 117 (0.7); 128 (0.9); 130 (100); 132 (5.6); 139.9 (1.3); 158.9 (0.4); 169.9 (2.1); 326.1 (0.3); 337.1 (1.4)	JW5605	68	11811
Trp	Aha	cyclo(Trp/Aha)	24.6	312.3	313.1	MS2 (313): 130 (26.2); 144 (12.6); 155.9 (17.1); 169 (14.2); 195 (10.6); 202.1 (13.1); 222.9 (100); 242 (13.9); 256.1 (14.8); 285 (11.7)	JW5605	68	460
Trp	Norleu	cyclo(Trp/Norleu)	26.3	299.3	300.1	MS2 (300): 86.3 (0.9); 117 (0.9); 130 (100); 132 (23.8); 141 (0.9); 169.9 (4.9); 255 (2.6); 283 (4.1)	JW5605	68	1806
Trp	Norval	cyclo(Trp/Norval)	24.5	285.2	286.1	MS2 (286): 103.1 (0.6); 115.5 (0.7); 117 (2.1); 126.9 (0.8); 130 (100); 132 (21.5); 154.8 (1.4); 169.9 (4.8); 240.9 (2.3); 269 (2.4)	JW5605	68	2993
cyclo(Tyr-X)									
Tyr	Leu	cyclo(Tyr/Leu)	21.7	276.2	277	MS2 (277): 86.3 (12.7); 91.1 (5.8); 107.1 (5.1); 119 (4.4); 135.9 (90.4); 171 (8.7); 204 (6.3); 221 (1.4); 232 (11.7); 249 (100)	JW2326	AlbC	900*
Tyr	V	cyclo(Tyr/V)	19.3	260.2	261	MS2 (261): 91.2 (5.2); 107.1 (3.2); 113 (2.3); 119 (4.9); 135.9 (100); 146.9 (11.4); 154.9 (8.3); 187.9 (2); 216 (4.4); 233 (40.7)	JW2326	22	5597
Tyr	Met	cyclo(Tyr/Met)	19.4	294.2	295	MS2 (295): 104.1 (7.2); 247 (22.5); 247.9 (2.7); 263 (83.3); 264 (12.7); 265 (4.7); 295 (100); 296 (17.3); 297 (5); 317 (8) MS3 (295->247): 85.3 (34.2); 107.1 (15.6); 113.1 (34.1); 141 (27.5); 146 (59.6); 163 (16.1); 174 (23.8); 190.9 (23.7); 201.9 (81.9); 218.9 (100)	JW2326	AlbC	3503
Tyr	TriFLeu	cyclo(Tyr/TriFLeu)	23	330.3	315.1	MS2 (315): 107.1 (15.8); 119 (4.5); 136 (46.3); 140 (2.4); 146.9 (1.2); 225 (8.2); 258.1 (2.7); 286.1 (3.9); 303.1 (100); 311.1 (3)	JW5605	AlbC	694
Tyr	Aha	cyclo(Tyr/Aha)	19.3	289.3	290	MS2 (290): 136 (11.8); 159.9 (6.6); 171.9 (15.4); 199.9 (100); 204.9 (33.2); 216.9 (7.4); 218.9 (11.6); 232.9 (14.7); 233.9 (5.7); 262 (22.2)	JW3973	54	2631
Tyr	Norleu	cyclo(Tyr/Norleu)	21.3	276.3	277.1	MS2 (277): 86.3 (7.7); 91.2 (5.5); 107.1 (3.9); 119 (4.4); 136 (91.5); 146.9 (2); 171 (11.7); 204.1 (4.9); 232.1 (11.7); 249.1 (100)	JW5605	AlbC	469
Tyr	Eth	cyclo(Tyr/Eth)	21.1	308.3	309.1	MS2 (309): 247 (100) MS3 (309.1->247): 85.2 (29); 107.1 (15.6); 113.1 (30.1); 141 (26.5); 145.9 (63.7); 163 (13); 173.9 (20.6); 190.9 (25.7); 201.9 (82.4); 218.9 (100)	JW3973	54	1104
Tyr	MeEGlu	cyclo(Tyr/MeEGlu)	18.6	306.3	307.1	MS2 (307): 275 (100) MS3 (207.1->275): 84.2 (0.6); 91.2 (0.8); 107.1 (9.7); 125 (1.7); 136 (20.8); 152.9 (6); 201.9 (6.6); 218.9 (1.8); 229.9 (0.4); 246.9 (100)	JW3973	54	189
Tyr	gMeLeu	cyclo(Tyr/gMeLeu)	23.2	290.3	291	MS2 (291): 100.2 (22.5); 107.1 (3.5); 119 (2.7); 136 (59.7); 146.9 (1.8); 185 (6.7); 218 (2.1); 246.1 (7.9); 263.1 (100); 274 (1)	JW5605	AlbC	EIC
cyclo(Phe-X)									
Phe	Leu	cyclo(Phe/Leu)	27.2	260.2	261	MS2 (261): 86.3 (27.5); 103.1 (2.3); 120 (100); 188 (4.1); 197.9 (0.5); 204.9 (0.7); 216 (23); 233 (62.4)	JW2326	AlbC	8461
Phe	Val	cyclo(Phe/Val)	25	246.1	247	MS2 (247): 72.4 (7.8); 100.1 (0.8); 103.1 (1); 120 (100); 174 (10); 184 (0.8); 202 (12.4); 219 (64.8)	JW2326	22	424
Phe	Met	cyclo(Phe/Met)	24.6	278.2	279	MS2 (279): 203 (0.7); 230.9 (100); 251 (0.3) MS3 (279->231): 84.3 (5.2); 91.2 (19.8); 130 (43.7); 147 (3.7); 156 (3.8); 157 (8.2); 158 (12.8); 175 (39.3); 185.9 (49.7); 202.9 (100)	JW2326	62	1266
Phe	TriFLeu	cyclo(Phe/TriFLeu)	28.8	314.3	331.1	MS2 (331): 91.2 (1.1); 93.2 (1.1); 103.1 (4.7); 120.1 (100); 140 (5.4); 242 (5.1); 267.1 (1.2); 270.1 (9.2); 287.1 (93.8); 295.1 (5.6)	JW5605	AlbC	5391
Phe	Aha	cyclo(Phe/Aha)	24.6	273.3	274	MS2 (274): 120 (20); 144 (8.6); 156 (28.9); 175 (4.5); 183.9 (100); 188.9 (38.5); 200.9 (6); 202.9 (10.8); 216.9 (22.2); 246 (19)	JW5807	54	2342
Phe	Norleu	cyclo(Phe/Norleu)	26.7	260.3	261.1	MS2 (261): 86.3 (19.9); 91.3 (0.6); 93.2 (0.6); 103.1 (1.7); 120.1 (100); 188 (5.3); 198 (0.7); 216 (24.5); 231.7 (0.4); 233 (59.5)	JW5605	AlbC	2214
Phe	Norval	cyclo(Phe/Norval)	24.6	246.2	247.1	MS2 (247): 72.5 (5.1); 91.3 (0.3); 93.2 (0.4); 103.1 (1.5); 120 (100); 174 (5.4); 202 (16); 218.2 (0.4); 219 (45.9); 230 (0.3)	JW5807	54	296
Phe	Eth	cyclo(Phe/Eth)	26.2	292.3	293.1	MS2 (293): 230.9 (100) MS3 (293.1->230.9): 84.3 (5.7); 91.2 (19.3); 120 (4.1); 130 (43.5); 147 (5); 157 (5.9); 157.9 (12.8); 174.9 (35.4); 185.9 (44.6); 202.9 (100)	JW3973	62	300*
Phe	MeEGlu	cyclo(Phe/MeEGlu)	22.8	290.3	291.1	MS2 (291): 231(5.9); 258.9(100) MS3 (291.1->258.9): 103.1 (0.6); 120 (39.5); 186 (9.7); 203 (2.8); 230.9 (100); 241 (0.8)	JW3973	62	510
Phe	gMeLeu	cyclo(Phe/gMeLeu)	28.5	274.3	275.1	MS2 (275.1): 93.4 (0.8); 100.2 (55.5); 103.1 (2.8); 120 (100); 174 (4.1); 191 (2.1); 202 (5.1); 219 (3.5); 230.1 (18.6); 247.1 (64.4)	JW5605	AlbC	100*

Table S4 (continued)

Cyclodipeptides containing analogs of leucine/isoleucine/valine/methionine								Best conditions of production		
AA1	AA2	cyclo(AA1/AA2)	Retention time (min)	Theoretical MW	m/z (+ mode)	MS fragmentation	Strain	CDPS	UV peak area	
cyclo(Leu-X)										
Leu	Leu	cyclo(Leu/Leu)	25.6	226.2	227	MS2 (227): 86.3 (58.3); 98.2 (3.4); 126 (1.9); 143 (1.7); 154.1 (7.4); 171 (3.3); 182 (100); 199 (33.3)	JW5807	54	3420	
Leu	Ile	cyclo(Leu/Ile)	25.1	226.2	227	MS2 (227): 69.5 (3); 86.3 (96); 98.2 (5.4); 114.1 (2.7); 126.1 (1); 143 (1.1); 154.1 (23.6); 171 (2.4); 182 (100); 199 (77.6)	JW3745	85	1541	
Leu	Val	cyclo(Leu/Val)	23.1	212.2	213	MS2 (213): 72.4 (43.7); 86.3 (40.2); 140.1 (33.5); 168 (100); 185 (88.8)	JW5807	85	560	
Leu	Met	cyclo(Leu/Met)	23.7	244.2	245	MS2 (245): 169.1 (0.7); 196.9 (100); 217 (0.3) MS3 (245->197): 84.3 (3.5); 85.3 (10.9); 101.1 (3.2); 113.1 (19.3); 124.1 (2.5); 141 (23.9); 152 (16.1); 168.9 (100)	JW5807	AlbC	4000*	
Leu	TriFLeu	cyclo(Leu/TriFLeu)	28.6	280.3	281.1	MS2 (281): 86.3 (11.5); 114 (6.3); 140 (22.2); 193 (5.3); 205 (11.3); 208 (17.6); 233.1 (3.5); 236 (100); 253 (34); 261.1 (8.4)	JW5807	54	258	
Leu	Aha	cyclo(Leu/Aha)	23.4	239.3	240	MS2 (240): 86.3 (63.8); 96.1 (21.9); 127 (15.2); 139 (24.2); 141 (15); 150 (100); 154.9 (82.8); 168.9 (13.7); 182.9 (36); 211.9 (48.3)	JW3973	54	654	
Leu	Norval	cyclo(Leu/Norval)	23.6	212.2	213	MS2 (213): 72.4 (20.2); 84.3 (1.4); 86.3 (29.8); 98.2 (1.2); 112.1 (0.9); 129 (2.7); 140 (6.9); 157 (2.6); 168 (100); 185 (29.8)	JW5807	54	EIC	
Leu	Eth	cyclo(Leu/Eth)	25.9	258.3	259	MS2 (259): 196.9 (100) MS3 (259->296.9) : 84.3 (4.4); 85.2 (12.1); 86.3 (2.2); 101.1 (3.4); 113 (26.7); 124.1 (4); 140 (1.8); 141 (22.6); 152 (17); 168.9 (100)	JW3973	54	100*	
Leu	MeEGlu	cyclo(Leu/MeEGlu)	22.1	256.3	257	MS2 (257): 128.9 (1.2); 197 (10.7); 225 (100) MS3 (257->225) : 84.1 (4); 86.3 (19.4); 113 (8.8); 128.9 (29.9); 140.9 (21.7); 197 (100)	JW3973	54	100*	
Leu	gMeLeu	cyclo(Leu/gMeLeu)	27.4	240.3	241.1	MS2 (241): 86.3 (32.2); 100.2 (38.7); 114.1 (2.4); 129 (2.6); 140 (11.2); 157 (4); 168.1 (11); 185 (13.3); 196 (100); 213.1 (32.6)	JW5807	54	EIC	
cyclo(Ile-X)										
Ile	Leu	cyclo(Ile/Leu)	25.1	226.2	227	MS2 (227): 69.5 (3); 86.3 (96); 98.2 (5.4); 114.1 (2.7); 126.1 (1); 143 (1.1); 154.1 (23.6); 171 (2.4); 182 (100); 199 (77.6)	JW3745	85	1541	
Ile	TriFLeu	cyclo(Ile/TriFLeu)	28.3	280.3	281.1	MS2 (281): 86.3 (30.8); 140 (32.6); 186 (4.4); 193 (5.2); 208 (53); 213 (4.5); 233.1 (9.1); 236 (64.9); 253 (100); 261.1 (11)	JW5807	85	689	
Ile	Aha	cyclo(Ile/Aha)	23	239.3	240	MS2 (240): 84.3 (5.5); 85.3 (6.9); 86.3 (100); 87.2 (7.7); 96.2 (6.5); 98.2 (16.1); 127.1 (18.7); 139 (12.1); 141 (25.7); 150 (24.6); 155 (55.4); 167 (12.4); 169 (10.9); 183 (18.8); 212 (35.8)	JW5807	85	461	
Ile	Norleu	cyclo(Ile/Norleu)	25.6	226.3	227.1	MS2 (227): 69.5 (3); 86.3 (94); 98.2 (2.3); 114.1 (3.1); 126.1 (0.5); 131.1 (0.6); 154.1 (26); 171 (0.5); 182 (100); 199 (90.1)	JW5807	85	146	
Ile	Norval	cyclo(Ile/Norval)	23.1	212.2	213	MS2 (213): 69.4 (3.4); 72.4 (16.7); 86.3 (76.4); 98.2 (2.5); 100.2 (3.7); 140.1 (26.1); 157 (0.7); 168 (100); 185 (88.4); 196 (0.7)	JW5807	54	EIC	
Ile	gMeLeu	cyclo(Ile/gMeLeu)	27	240.3	241.1	MS2 (241): 69.5 (2); 86.3 (60.2); 100.2 (47.8); 112.2 (3.5); 140.1 (9.8); 157.1 (4.6); 168 (41); 185 (9.2); 196.1 (78.8); 213 (100)	JW5807	54	EIC	
cyclo(Ala-X)										
Ala	Leu	cyclo(Ala/Leu)	19.8	184.2	185	MS2 (185): 86.3 (88.2); 101.1 (3.3); 112.1 (4.1); 114.1 (1.9); 129 (6.4); 139.1 (1.4); 140 (100); 157 (25.2)	JW5807	85	678	
Ala	Ile	cyclo(Ala/Ile)	19.2	184.2	185	MS2 (185): 69.4 (2.2); 86.2 (100); 112.1 (5.6); 129 (1.3); 140 (30.5); 157 (32.3)	JW3745	85	233	
Ala	Met	cyclo(Ala/Met)	17.8	202.2	203	MS2 (203): 127(1); 154.9(100) MS3 (203->155): 71.5 (0.7); 80.3 (0.8); 84.3 (3); 98.1 (1.9); 99.1 (26.5); 109.1 (1.5); 110.1 (10.9); 126.9 (100)	JW3973	30	234	
Ala	TriFLeu	cyclo(Ala/TriFLeu)	23	238.2	239	MS2 (239): 100.2 (3.4); 140 (100); 143 (6); 151 (14.6); 166 (9.9); 171 (10.5); 191 (21.8); 193.9 (60.5); 211 (22.2); 219 (19.2)	JW5807	85	376	
Ala	Aha	cyclo(Ala/Aha)	16.8	197.2	197.9	MS2 (198): 58.6 (10.3); 71.4 (7.7); 85.3 (23.1); 97.1 (20.1); 99.1 (25.6); 108.1 (32.7); 113 (100); 124.9 (14.5); 140.9 (52.6); 169.9 (19.8)	JW3973	98	EIC	
Ala	Norleu	cyclo(Ala/Norleu)	20.1	184.2	185	MS2 (185): 69.5 (1.5); 86.3 (89.1); 95.1 (0.3); 112.1 (4.7); 114.1 (1.3); 129 (0.7); 139.2 (0.4); 140 (100); 157 (28.8); 167.9 (0.3)	JW3973	30	267	
Ala	Norval	cyclo(Ala/Norval)	17.6	170.1	171	MS2 (171): 72.4 (63.9); 98.2 (7.2); 100.2 (1.9); 107.3 (0.4); 108.1 (0.5); 115.1 (1.1); 125.1 (0.6); 126 (100); 143 (35); 153 (2.1)	JW5807	85	92	
Ala	Eth	cyclo(Ala/Eth)	20.1	216.2	216.9	154.9 (100) MS3 (154.9) : 84.3 (2.2); 99.1 (26.6); 109.2 (1.9); 110 (9.7); 127 (100)	JW3973	98	EIC	
Ala	MeEGlu	cyclo(Ala/MeEGlu)	16.5	214.2	214.9	No fragmentation possible (coeluted compound ionization is too strong)	JW3973	30	100*	
cyclo(Pro-X)										
Pro	Leu	cyclo(Pro/Leu)	21.5	210.2	211	MS2 (211): 70.5 (100); 86.3 (40.3); 114.1 (4.4); 127 (9.5); 138.1 (36.6); 155.1 (14.4); 166 (5.6); 167.1 (4.2); 183 (65.1); 194 (25.5)	JW0233	82	2069	
Pro	Ile	cyclo(Pro/Ile)	20.9	210.2	211.2	MS2 (211): 69.5 (3.9); 70.4 (33.1); 86.3 (72.8); 127 (2.1); 138.1 (35); 155 (38.3); 183 (100); 194 (17.1)	JW0233	82	EIC	
Pro	Met	cyclo(Pro/Met)	19.7	228.2	229.1	MS2 (229.1): 180.9 (100) MS3 (229.1->180.9): 70.4 (8.5); 73.9 (0.5); 80.3 (0.8); 82.4 (2.9); 84.3 (8.4); 97.2 (3.3); 125 (100); 153 (42.8)	JW0233	10	817	
Pro	TriFLeu	cyclo(Pro/TriFLeu)	25.1	264.2	265.1	MS2 (265): 98.2 (51.3); 100.1 (3.2); 120.1 (3.3); 140 (90.2); 177 (1.7); 197 (10.2); 217 (51); 219 (3); 237 (49.4); 245 (100)	JW5605	82	573	
Pro	Aha	cyclo(Pro/Aha)	19.3	223.2	223.9	MS2 (224): 70.4 (31); 96.1 (9.9); 115 (41.1); 125 (53.5); 138.9 (100); 150.9 (17.9); 152.9 (21.6); 164.9 (9.6); 166.9 (32.7); 195.9 (39.3)	JW3973	98	696	
Pro	Norleu	cyclo(Pro/Norleu)	21.8	210.2	211	MS2 (211): 70.4 (100); 86.3 (50.7); 114.1 (2.5); 127 (2.7); 138 (47); 155.1 (16.3); 166 (5.1); 167 (4.1); 183 (74.1); 194 (24.6)	JW3973	10	1000*	
Pro	Norval	cyclo(Pro/Norval)	19.2	196.1	197	MS2 70.4 (100); 72.4 (46.4); 98.1 (2.4); 100.2 (2); 124.1 (28); 141 (16.3); 152 (5); 153 (3.7); 169 (65.7); 179.9 (19.8)	JW5605	82	50	
Pro	Eth	cyclo(Pro/Eth)	22	242.2	243	MS2 (243): 180.9 (100) MS3(243->180.9) : 70.4 (10.1); 82.3 (1.9); 84.3 (5.9); 97.2 (4); 125 (100); 152.9 (46.5)	JW3973	10	500*	
Pro	MeEGlu	cyclo(Pro/MeEGlu)	18.5	240.2	241	MS2 (241): 70.4 (1.4); 181 (3.4); 208.9 (100) MS3 (280.9) : 70.4 (100); 84.3 (8.6); 152.9 (1.3); 179.1 (0.4); 180.9 (81)	JW3973	10	EIC	

Table S4 (continued)

Cyclodipeptides containing analogs of leucine/isoleucine/valine/methionine

Cyclodipeptides containing analogs of leucine/isoleucine/valine/methionine							Best conditions of production		
AA1	AA2	cyclo(AA1/AA2)	Retention time (min)	Theoretical MW	m/z (+ mode)	MS fragmentation	Strain	CDPS	UV peak area
cyclo(Met-X)									
Met	Leu	cyclo(Met/Leu)	23.7	244.2	245	MS2 (245): 169.1 (0.7); 196.9 (100); 217 (0.3) MS3 (245->197): 84.3 (3.5); 85.3 (10.9); 101.1 (3.2); 113.1 (19.3); 124.1 (2.5); 141 (23.9); 152 (16.1); 168.9 (100)	JW5807	AlbC	4000*
Met	TriFLeu	cyclo(Met/TriFLeu)	26.8	298.3	299.1	MS2 (299): 251 (100) MS3 (299->251): 84.3 (5.6); 140 (4.2); 166.9 (6.4); 175 (5); 183 (8.9); 194.9 (41.2); 203 (3.4); 205.9 (4.2); 223 (100); 231 (20.3)	JW5807	54	95
Met	Aha	cyclo(Met/Aha)	19.1	257.3	258	MS2 (258): 84.3 (62.2); 85.2 (24.2); 86.3 (75.5); 113.1 (26.1); 132 (79.3); 139.9 (56.3); 141 (65); 150 (17.4); 184 (100); 230 (19.7)	JW3973	54	EIC*
Met	Norleu	cyclo(Met/Norleu)	24.1	244.3	245.1	MS2 (245): 196.9 (100) MS3 (245->197): 84.3 (5.1); 86.3 (1.9); 96.2 (4.1); 113.1 (13.2); 124.1 (4); 140 (2.4); 141 (47.4); 151 (2.3); 152 (17.1); 168.9 (100)	JW5807	54	50
Met	Norval	cyclo(Met/Norval)	21.8	230.2	231	MS2 (231): 182.9 (100) MS3 (231.1->182.9): 82.4 (1.8); 84.3 (2.8); 95.1 (2.4); 99.1 (12.3); 110.1 (3); 127 (40.3); 138 (12.7); 155 (100)	JW5807	54	EIC
cyclo(Val-X)									
Val	Leu	cyclo(Val/Leu)	23.1	212.2	213	MS2 (213): 72.4 (43.7); 86.3 (40.2); 140.1 (33.5); 168 (100); 185 (88.8)	JW5807	85	80
Val	TriFLeu	cyclo(Val/TriFLeu)	26.3	266.2	267	MS2 (267): 100.1 (2.8); 140 (31.1); 172 (2.8); 179 (7.3); 194 (46.1); 199 (4.7); 219 (13.3); 222 (55.9); 239 (100); 247 (13.8)	JW5807	85	243
Val	Norleu	cyclo(Val/Norleu)	23.1	212.2	213	MS2 (213): 72.4 (37.8); 84.3 (3.5); 86.3 (29.7); 112.1 (0.8); 114.1 (0.8); 129.1 (1.4); 140.1 (27.1); 157 (3.3); 168 (100); 185 (82.9)	JW5807	16	574
Val	Norval	cyclo(Val/Norval)	21.2	198.1	199	MS2 (199): 72.4 (66.3); 84.3 (1.6); 100.2 (1.7); 107.1 (0.8); 126.1 (23); 143 (1.2); 154 (100); 171 (85.5); 175.8 (1); 180.9 (1)	JW5807	85	EIC*
cyclo(Gly-X)									
Gly	Met	cyclo(Gly/Met)	15.9	188.2	189	MS2 (189): 113(0.8); 140.9(100) MS3 (189->141): 84.3 (1.8); 85.2 (31.7); 113 (100)	JW3973	30	168
Gly	Norleu	cyclo(Gly/Norleu)	18.3	170.2	170.9	MS2 (171): 69.5 (2.4); 75.4 (2.6); 86.3 (100); 97.2 (4.8); 98.2 (28); 114.1 (8.8); 115.1 (5.9); 126 (91.4); 143 (48.5); 153.9 (5)	JW3973	30	157
Gly	Eth	cyclo(Gly/Eth)	18.3	202.2	202.9	MS2 (203): 140.9 (100) MS3 (202.9->140.9): 84.4 (3.1); 85.3 (34.6); 113 (100)	JW3973	30	127
Gly	MeEGlu	cyclo(Gly/MeEGlu)	14.8	200.2	200.9	168.9 (100) MS3 (200.9->168.9): 84.2 (100); 96.1 (12); 111.7 (0.5); 113 (70.3); 124 (3.5); 140.9 (55.8); 151 (4.1)	JW3973	30	56
cyclo(Glu-X)									
Glu	Leu	cyclo(Glu/Leu)	20.2	242.2	243	MS2 (243): 141 (1.6); 197 (12.7); 225 (100) MS3 (243->225): 84.3 (9.3); 86.3 (12.2); 113.1 (6.4); 129 (30.9); 141 (28.8); 197 (100)	JW5807	44	6428
Glu	TriFLeu	cyclo(Glu/TriFLeu)	22.3	296.2	297.1	MS2 (297): 251.1(3.1); 279(100) MS3 (279): 84.3 (2); 100.1 (1.6); 120 (1.1); 139.9 (56.9); 206 (12.4); 223 (3.5); 231 (0.4); 251 (100)	JW5605	44	1499
Glu	Norleu	cyclo(Glu/Norleu)	19.6	242.2	243	MS2 (243): 129 (1); 141 (0.5); 197 (14.2); 224.9 (100) MS3 (243.2->211.2): 84.2 (4.4); 86.3 (14.1); 96.1 (0.2); 113.1 (3.3); 129 (11.2); 141 (11.9); 152 (1.5); 169 (0.3); 196.9 (100); 207 (0.2)	JW5807	36	1592
Glu	Norval	cyclo(Glu/Norval)	17.3	228.1	228.9	MS2 (229): 72.5 (0.4); 183 (11.9); 210.9 (100) MS3 (228.9->210.9): 72.4 (11.5); 84.2 (5); 113 (3); 129 (11.4); 138 (3.4); 141 (11.1); 155 (0.7); 182.9 (100); 327.3 (0.2); 438.1 (0.2)	JW5807	36	617

Table S4 (continued)

Cyclodipeptides containing tryptophan analogs

							Best conditions of production		
AA1	AA2	cyclo(AA1/AA2)	Retention time (min)	Theoretical MW	m/z (+ mode)	MS fragmentation	Strain	CDPS	UV peak area
cyclo(X-X)									
Trp	Trp	cyclo(Trp/Trp)	27.3	372.4	373.1	MS2 (373.1): 129.9 (100); 131.9 (6.3); 168.9 (11.5); 169.9 (10.8); 186.9 (7.4); 213.9 (16.9); 241.9 (92.3); 243.9 (6.3); 255.9 (14.3); 356 (33.3)	JW2326	64	>15000
4FTrp	4FTrp	cyclo(4FTrp/4FTrp)	28.6	408.3	409.1	MS2(409.1): 147.8 (100); 149.9 (7.9); 186.9 (25.4); 203.9 (13.2); 231.9 (37.4); 259.9 (51.2); 261.9 (22); 337.1 (4.4); 364 (8); 381.1 (17.6)	JW2326	63	21800
5FTrp	5FTrp	cyclo(5FTrp/5FTrp)	29.4	408.3	409.1	MS2 (409.1): 147.8 (100); 186.9 (13.1); 187.8 (2.3); 203.9 (4.4); 231.9 (12.6); 259.9 (58.2); 261.9 (8.3); 273.9 (6.2); 364.1 (2.1); 392 (9.3)	JW2326	63	16178
cyclo(Trp-X)									
Trp	Trp	cyclo(Trp/Trp)	27.3	372.4	373.1	MS2 (373.1): 129.9 (100); 131.9 (6.3); 168.9 (11.5); 169.9 (10.8); 186.9 (7.4); 213.9 (16.9); 241.9 (92.3); 243.9 (6.3); 255.9 (14.3); 356 (33.3)	JW2326	64	>15000
Trp	4FTrp	cyclo(Trp/4FTrp)	28	390.3	391.1	MS2 (391.1): 129.9 (58.6); 147.8 (10.5); 176.9 (11.7); 203.9 (14.1); 231.9 (21.1); 241.9 (13.7); 255.8 (20.1); 259.9 (100); 363.1 (11.1); 373.9 (13)	JW2326	63	1429
Trp	5FTrp	cyclo(Trp/5FTrp)	28.5	390.3	391.1	MS2 (391.1): 129.9 (70.5); 147.8 (13.6); 186.8 (19.6); 242 (13.8); 259.9 (100); 279.9 (20.8); 374 (32.6)	JW2326	63	1067
cyclo(Tyr-X)									
Tyr	Trp	cyclo(Tyr/Trp)	23.8	349.4	350.1	MS2 (350.1): 116.9 (3.9); 129.9 (100); 131.9 (27.7); 135.9 (11); 169.8 (13.8); 180.9 (2.5); 190.8 (2.1); 218.8 (3); 305 (8.3); 333 (15.1)	JW2326	69	>15000
Tyr	4FTrp	cyclo(Tyr/4FTrp)	24.6	367.3	368.1	MS2 (368.1): 135.9 (3.8); 145.9 (11.7); 147.9 (100); 149.9 (21.6); 187.8 (7.7); 190.8 (9.7); 220.9 (5.8); 323 (9.2); 340 (15.6); 351.1 (10.5)	JW1254	69	7062
Tyr	5FTrp	cyclo(Tyr/5FTrp)	25.1	367.3	368.1	MS2 (368.1): 135.9 (5.5); 145.9 (6.9); 147.8 (100); 149.9 (16); 187.8 (5.5); 190.9 (6.5); 220.9 (6.7); 323.1 (7.5); 340.1 (4.7); 351 (8.2)	JW1254	69	>13000
Tyr	5MeTrp	cyclo(Tyr/5MeTrp)	25.5	363.35	364.1	MS2 (364.1): 114.9 (2); 130.9 (4); 135.9 (6.2); 142.9 (4.5); 143.9 (100); 145.9 (18.1); 183.8 (9.7); 232.9 (3.4); 319.1 (3.1); 347.1 (12.6)	JW1254	69	567
cyclo(Phe-X)									
Phe	Trp	cyclo(Phe/Trp)	27.3	333.3	334.1	MS2 (334.1): 117 (3.6); 120 (11.2); 129.9 (100); 131.9 (31.5); 169.8 (10.2); 202.9 (5.3); 289 (7.1); 317.1 (9.9)	JW2326	68	555
Phe	4FTrp	cyclo(Phe/4FTrp)	28	351.3	352.1	MS2 (352.1): 119.9 (6.2); 129.9 (5.5); 147.9 (100); 149.9 (27.2); 187.7 (4.6); 202.9 (8.8); 307 (10.9); 324 (14.5); 329.2 (9.2); 335.1 (13)	JW2326	68	305
Phe	5FTrp	cyclo(Phe/5FTrp)	28.6	351.3	352.1	MS2 (352.1): 119.9 (5.6); 129.9 (7); 147.9 (100); 149.9 (27.7); 165 (3.8); 187.7 (8); 202.9 (6.2); 205 (14.2); 329.2 (5.4); 335.1 (16.8)	JW2326	68	302
cyclo(Leu-X)									
Leu	Trp	cyclo(Leu/Trp)	25.5	299.2	300.1	MS2 (300.1): 86.2 (0.9); 129.9 (100); 131.9 (22.4); 168.9 (1); 169.9 (4.7); 254.9 (1.8); 272 (1); 283 (3.4)	JW2326	68	>20000
Leu	4FTrp	cyclo(Leu/4FTrp)	26.3	317.3	318.1	MS2 (318.1): 112.9 (1.7); 140.9 (5.1); 147.8 (100); 149.9 (22.2); 168.9 (4.2); 170.9 (1.8); 187.8 (4.6); 273 (4.2); 290 (10.9); 301 (2.3)	JW2326	68	17554
Leu	5FTrp	cyclo(Leu/5FTrp)	26.9	317.3	318.1	MS2 (318.1): 112.9 (0.9); 140.9 (0.6); 147.8 (100); 149.9 (16.3); 168.9 (2.9); 170.9 (1.6); 187.8 (3.2); 272.9 (2.6); 290 (2.3); 301 (2.6)	JW2326	68	14849
Leu	5MeTrp	cyclo(Leu/5MeTrp)	27.4	313.35	314.1	MS2 (314.1): 115 (0.4); 130.9 (1.2); 131.9 (0.5); 141.9 (0.5); 142.9 (0.6); 143.9 (100); 145.9 (14.8); 183.9 (5); 269 (1.8); 297 (2.8)	JW2326	68	593
Leu	5HOTrp	cyclo(Leu/5HOTrp)	21.9	315.3	316.1	MS2 (316.1): 117.9 (1.1); 133.9 (1); 145.9 (100); 147.9 (3.2); 185.8 (11.9); 215.8 (2.5); 243.9 (1.3); 270.9 (1.1); 298.1 (4); 299 (4)	JW2326	68	130
cyclo(Ala-X)									
Ala	Trp	cyclo(Ala/Trp)	21.3	257.2	258	MS2 (258): 89.1 (3.8); 97.1 (2.9); 103 (11.4); 117.9 (2.9); 129.9 (100); 131.9 (10.9); 169.8 (7.2); 240.9 (4.7)	JW2326	63	922
Ala	4FTrp	cyclo(Ala/4FTrp)	22.7	275.2	276	MS2 (276): 99 (2.4); 126.9 (0.3); 147.9 (100); 149.9 (19.5); 187.8 (4.3); 230.8 (1.1); 247.9 (3.3); 256 (0.3); 258.9 (1.6)	JW2326	63	800*
Ala	5FTrp	cyclo(Ala/5FTrp)	23	275.2	275.9	MS2 (275.9): 89.1 (0.3); 91.1 (0.5); 127.9 (0.4); 128.9 (0.3); 147.8 (100); 149.9 (14.4); 187.8 (3); 230.9 (0.7); 247.9 (0.8); 258.9 (1.5)	JW2326	63	877
Ala	5MeTrp	cyclo(Ala/5MeTrp)	23.6	271.25	272	MS2 (272): 142.9 (0.9); 143.9 (100); 145.9 (7.1); 166.8 (0.6); 183.9 (4.4); 226 (0.7); 254.9 (1.3)	JW2326	63	129
cyclo(Pro-X)									
Pro	Trp	cyclo(Pro/Trp)	24.1	283.3	284	MS2 (284): 115 (5.8); 116.9 (1.7); 118 (2.5); 124.9 (1.7); 129.9 (100); 131.9 (26.4); 152.9 (5.8); 169.8 (23.9); 238.9 (8); 266.9 (6.1)	JW2326	74	>20000
Pro	4FTrp	cyclo(Pro/4FTrp)	25.6	301.3	302	MS2 (302): 124.9 (8.6); 147.9 (100); 149.9 (18.7); 152.9 (7.7); 154.9 (3.7); 176.9 (15); 187.8 (12.6); 256.9 (10.2); 274 (7.9); 284.9 (3.4)	JW2326	74	>20000
Pro	5FTrp	cyclo(Pro/5FTrp)	25.6	301.3	302	MS2 (302): 115 (3.1); 124.9 (7); 147.9 (100); 149.9 (14.7); 152.9 (7.9); 154.9 (4.2); 187.8 (10.2); 256.9 (8); 274 (2.5); 284.9 (3.7)	JW2326	74	>20000
Pro	5MeTrp	cyclo(Pro/5MeTrp)	26.2	297.3	298	MS2 (298): 115 (3.7); 131 (1); 131.9 (1.1); 143.9 (100); 145.9 (20.8); 152.9 (3.5); 154.9 (0.8); 183.9 (15.4); 252.9 (3.6); 281 (5.9)	JW2326	74	463
Pro	5HOTrp	cyclo(Pro/5HOTrp)	19.7	299.3	300	MS2 (300): 115 (4.6); 118 (1); 133.9 (1); 145.8 (100); 147.9 (8.8); 152.9 (4); 154.9 (1); 185.8 (11.9); 254.9 (3.6); 282.9 (3.1)	JW2326	74	789
cyclo(Ser-X)									
Ser	Trp	cyclo(Ser/Trp)	20.2	274.2	274	MS2 (374): 105 (0.6); 129.9 (100); 131.9 (12.4); 169.8 (11.2); 186.8 (0.7); 245.9 (1); 255.9 (0.9); 256.9 (2.2)	JW2326	70	750
Ser	4FTrp	cyclo(Ser/4FTrp)	21.6	291.3	292	MS2 (292): 114.9 (1.3); 142.9 (2.8); 144.9 (1.2); 147.8 (100); 149.9 (18.2); 187.8 (5.4); 263.9 (7.3); 271.9 (1.1); 273.9 (1.1); 274.9 (2)	JW2326	63	1134
Ser	5FTrp	cyclo(Ser/5FTrp)	21.8	291.3	292	MS2 (292): 85.1 (0.7); 114.9 (0.6); 142.8 (0.9); 144.9 (1.2); 147.8 (100); 149.9 (12.5); 187.8 (5.9); 263.9 (1.3); 273.9 (2.3); 274.9 (2.9)	JW2326	63	865
Ser	5HOTrp	cyclo(Ser/5HOTrp)	17.2	289.3	290	MS2 (290): 145.9 (21.9); 146.9 (14.2); 174.9 (100); 216.6 (3.2); 217.9 (8.3); 244.9 (5.2); 246.8 (2.8); 263.1 (4.2); 272.9 (4.1); 273.9 (12.9)	JW2326	63	195

Table S5: Fragmentation patterns by residues, observed by comparison between cyclodipeptides
 Bold and italic types indicate respectively strong and weak signals.

Underlined type indicates fragmentation with one single daughter-ion (MS2), for which MS3 is necessary to get further identification data.

Residue	Immonium ion [NH-CHR+H] ⁺	[CO-CR] ⁺	[-(NH-CHR-CO)- +H] ⁺	Related ions	Characteristic loss in cyclodipeptide scaffold
Proline (1)	70		98		
Az.-carboxylic acid (2)	56				
3,4-Dehydroproline (3)	68		96		
<i>cis</i> -4-Fluoroproline (4)	88		116	68	
<i>trans</i> -4-Fluoroproline (5)	88			68	
<i>trans</i> -4-Hydroxyproline (6)	86		114		
Tyrosine (10)	136	147		91, 107, 119	
3-Fluorotyrosine (11)	154	165		109	
3,4-DOPA (12)	152				
Phenylalanine (13)	120	131	148	93, 103	
4-Fluorophenylalanine (14)	138				
2-Fluorophenylalanine (15)	138				
2-Thienylalanine (16)	126			97, 99	
3-(3-Pyridyl)-alanine (17)	121			93 , 94	
3-(4-Pyridyl)-alanine (18)	121			92, 93, 94	
Tryptophan (19)	159	170		117, 130 , 132	-17 (-NH ₃)
4-Fluorotryptophan (20)	177	188		148 , 150	
5-Fluorotryptophan (21)		188		148 , 150	
5-Methyltryptophan (22)		184		144 , 146	
5-Hydroxytryptophan (23)	175	186		146 , 148	
Leucine (24)	86		114	72	
Isoleucine (25)	86		114	69	
Valine (26)	72		100	55	
Methionine (27)	104				<u>- 48 (-CH₃SH)</u>
Trifluoroisoleucine (28)	140				
Azidohomoalanine (29)					-57, -71, -73, -85, -90
Norleucine (30)	86*			69	
Norvaline (31)	72*				
Ethionine (32)					<u>-68 (-C₂H₅SH)</u>
Glutamate γ -methyl ester (33)					<u>-60 (-CH₂COOH), -32 (-CH₃OH)</u>
γ -Methylleucine (34)	100				

*Norleucine is isomassic with leucine and isoleucine. Norvaline is isomassic with valine. We could not detect significant differences in the fragmentation patterns of the non-canonical cyclodipeptides containing norleucine or norvaline, compared to their canonical counterparts. The different retention times of the *de novo* appeared product and the corresponding canonical cyclodipeptides allowed us to assign peaks to the non-canonical cyclodipeptides.

III) Incorporation of non-canonical amino acids by CDPs

a. Production of canonical and non-canonical cyclodipeptides by CDPs

Figure S5 indicates the UV peak area of the cyclodipeptides detected in the different cultures.

For cyclodipeptides available in our laboratory, whose calibration curves were previously established, an order of magnitude of the massic concentration is given (calibration curves are given in Figure S6).

A

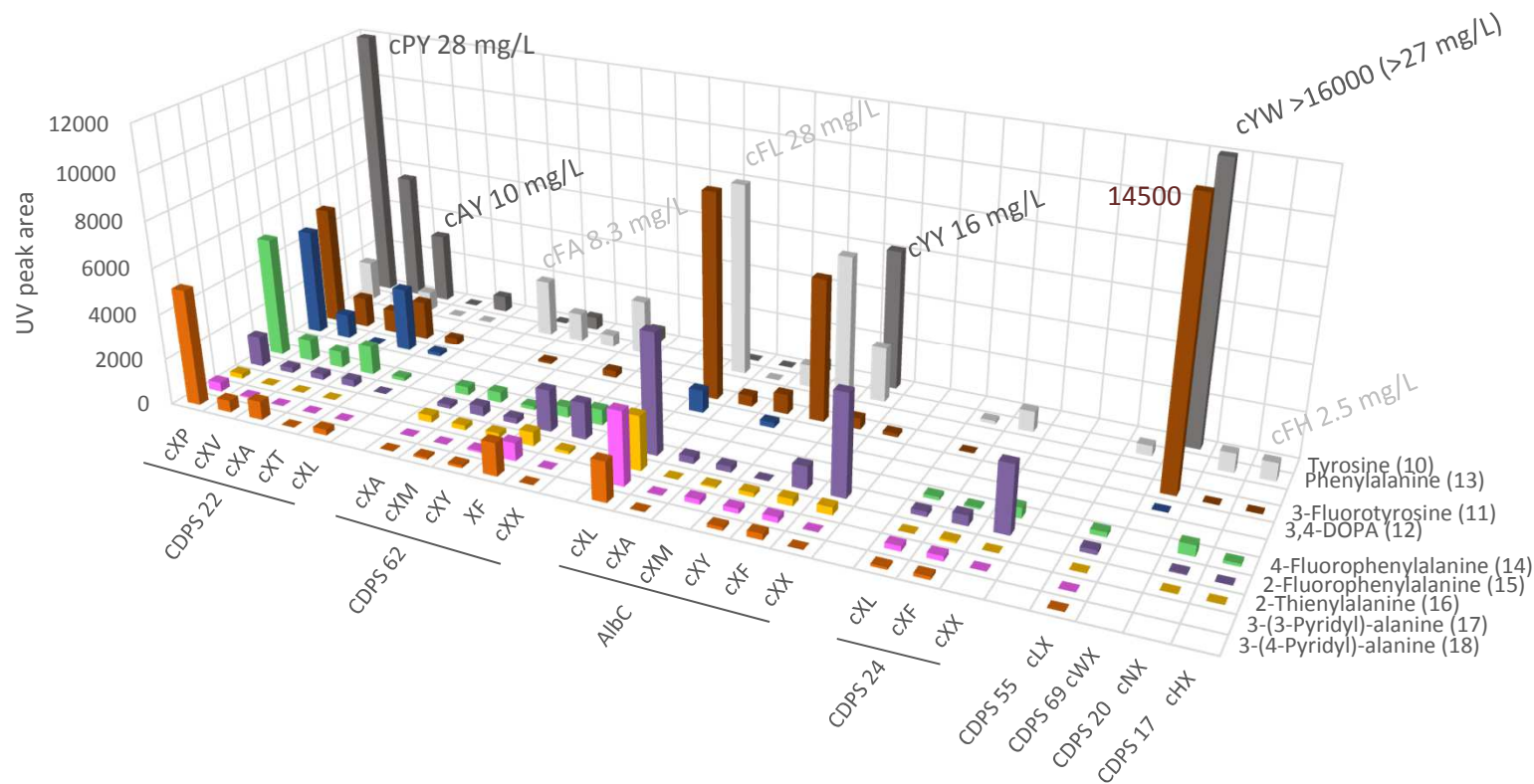


Figure S5 Histograms showing the production of canonical and non-canonical cyclodipeptides by CDPs

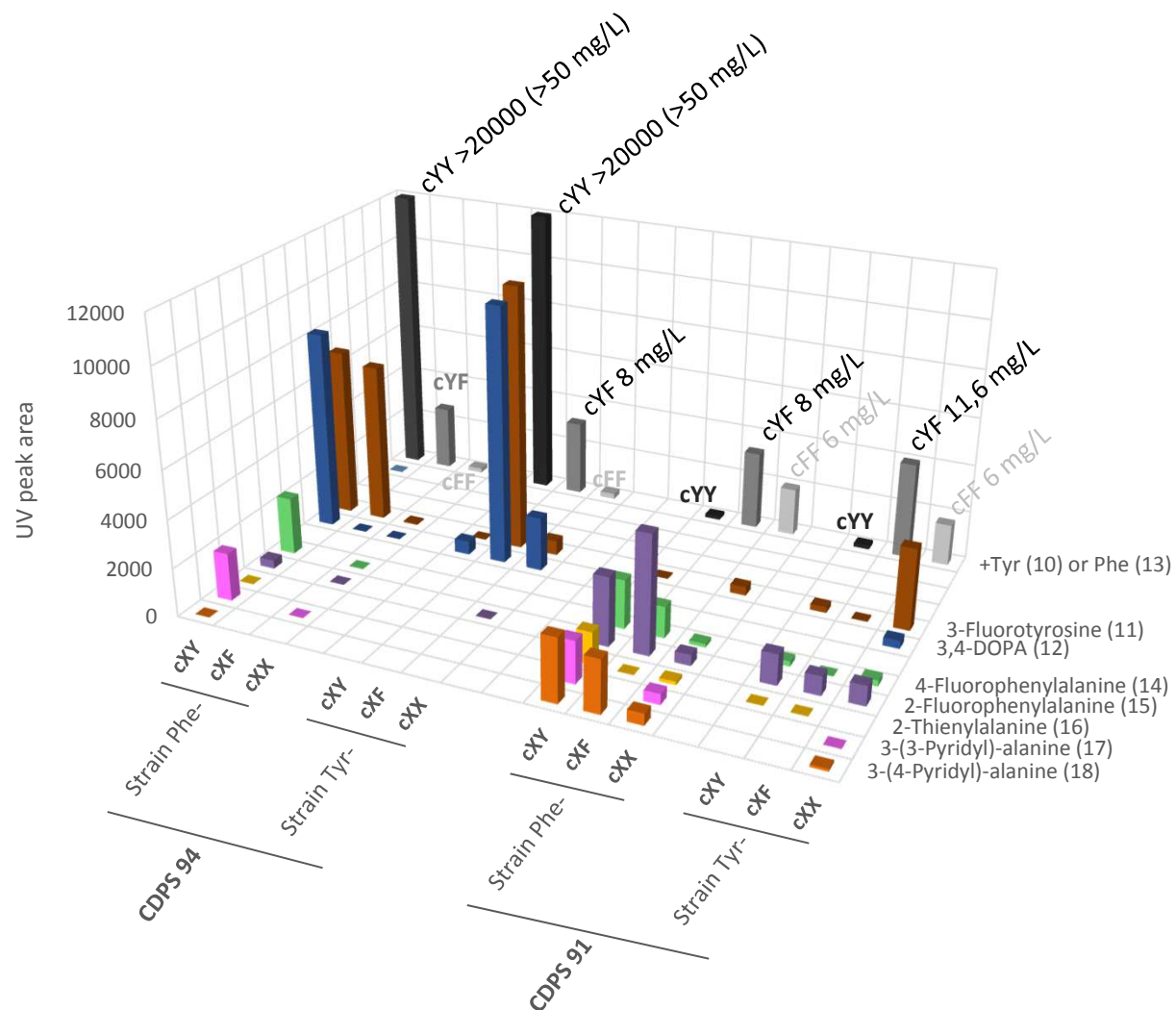
A) Non-canonical cyclodipeptides produced by replacement of phenylalanine and tyrosine

CDPSs were expressed in strain JW2326, auxotrophic for phenylalanine, tyrosine and tryptophan.

Positive standards complemented with phenylalanine and tyrosine (50mg/L each) are shown. The expected canonical cyclodipeptides were produced.

In order to mimic an auxotrophy for phenylalanine and tyrosine only, cultures expressing CDPS 69 (producing cyclo(Trp-Tyr)) were complemented with 50mg/L tryptophan.

For better clarity, UV peak area above 12000 were not fully drawn on the histogram. In these cases, measured UV peak area is given next to the corresponding bars.

B**Figure S5 (continued)**

B) Non-canonical cyclodipeptides produced by replacement of phenylalanine and tyrosine

CDPS 91 and 94 produce cyclo(Phe-Tyr) as one of their product. Depleting the production of Phe and Tyr in the same experiment could have limited the possibility of incorporation of ncAAs (for example if one ncAA is a good substituent for Phe but not for Tyr). Thus, in order to replace specifically Phe or Tyr, CDPS 91 and 94 were expressed separately in strain JW2580, auxotrophic for Phe and in strain JW2581, auxotrophic for Tyr.

Positive standard complemented either in phenylalanine or tyrosine (50mg/L each), depending of the auxotrophy used, are shown. The expected canonical cyclodipeptides were produced.

For better clarity, UV peak area above 12000 were not fully drawn on the histogram. In these cases, measured UV peak area is given next to the corresponding bars.

C

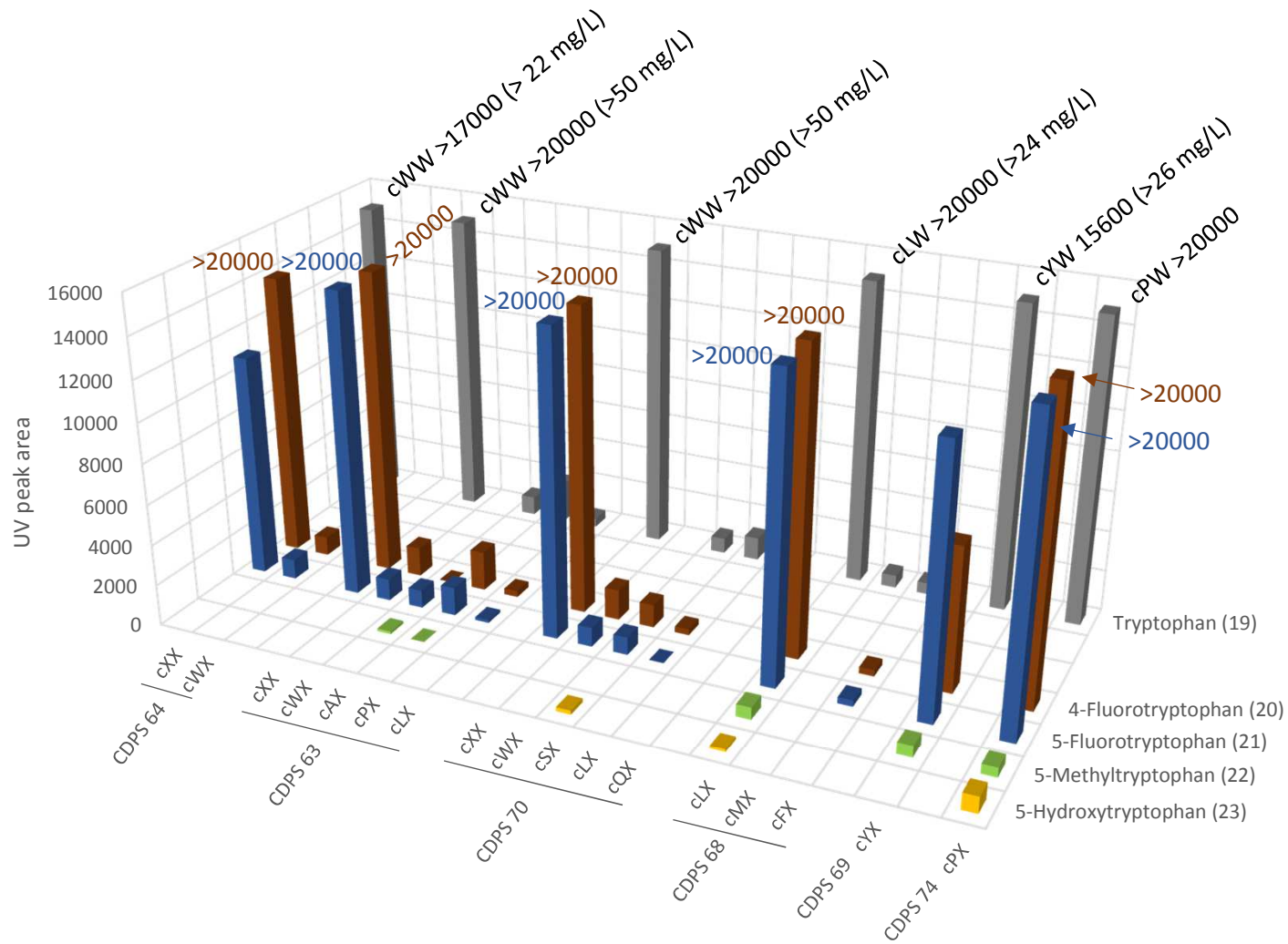
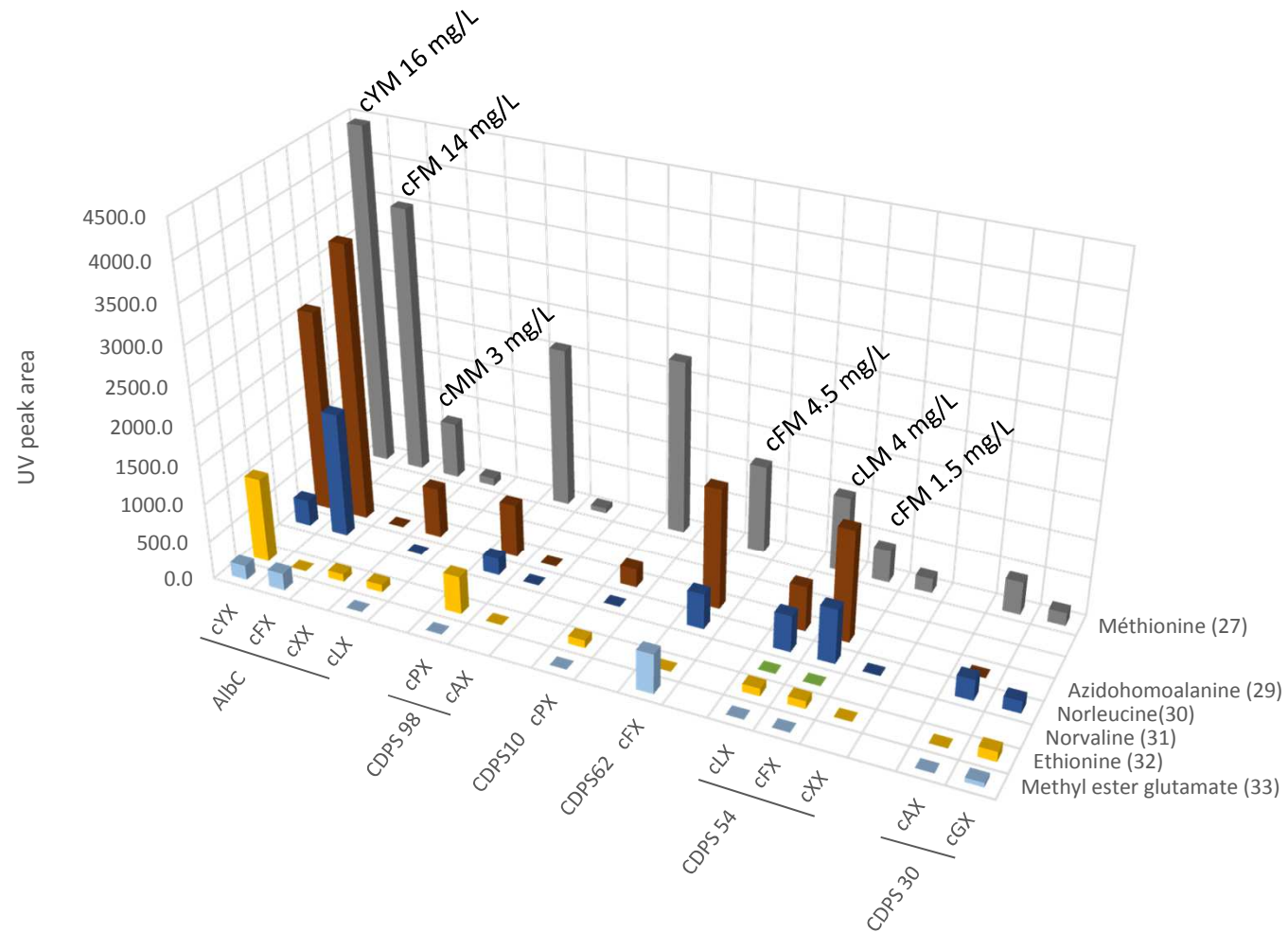


Figure S5 (continued)

C) Non-canonical cyclodipeptides produced by replacement of tryptophan

CDPSs were expressed in strain JW1254, auxotrophic for tryptophan. Positive standards complemented with tryptophan (50mg/L) are shown. The expected canonical cyclodipeptides were produced.

UV peak area above 15000 are above the saturation level of the detector. UV peak area reaching this threshold are indicated on the histogram.

D**Figure S5 (continued)**

D) Non-canonical cyclodipeptides produced by replacement of methionine

CDPSs were expressed in strain JW3973, auxotrophic for methionine. Positive standards complemented with methionine (50mg/L) are shown. The expected canonical cyclodipeptides were produced.

E

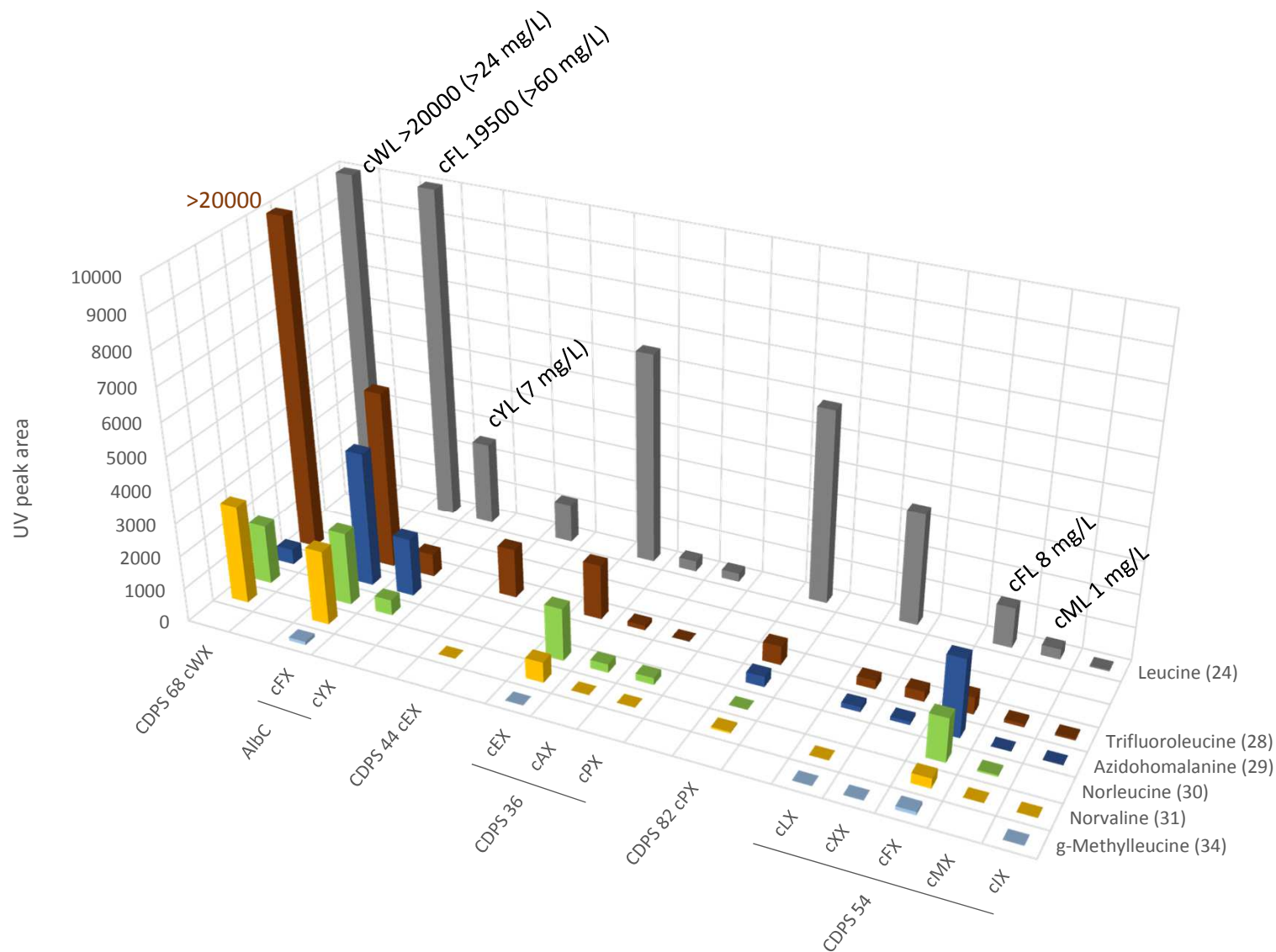


Figure S5 (continued)

E) Non-canonical cyclodipeptides produced by replacement of leucine

CDPSs were expressed in strain JW5605, auxotrophic for leucine, isoleucine and valine. Positive standards complemented with leucine (50mg/L) are shown. The expected canonical cyclodipeptides were produced.

UV peak area above 15000 are above the saturation level of the detector. UV peak area reaching this threshold are indicated on the histogram.

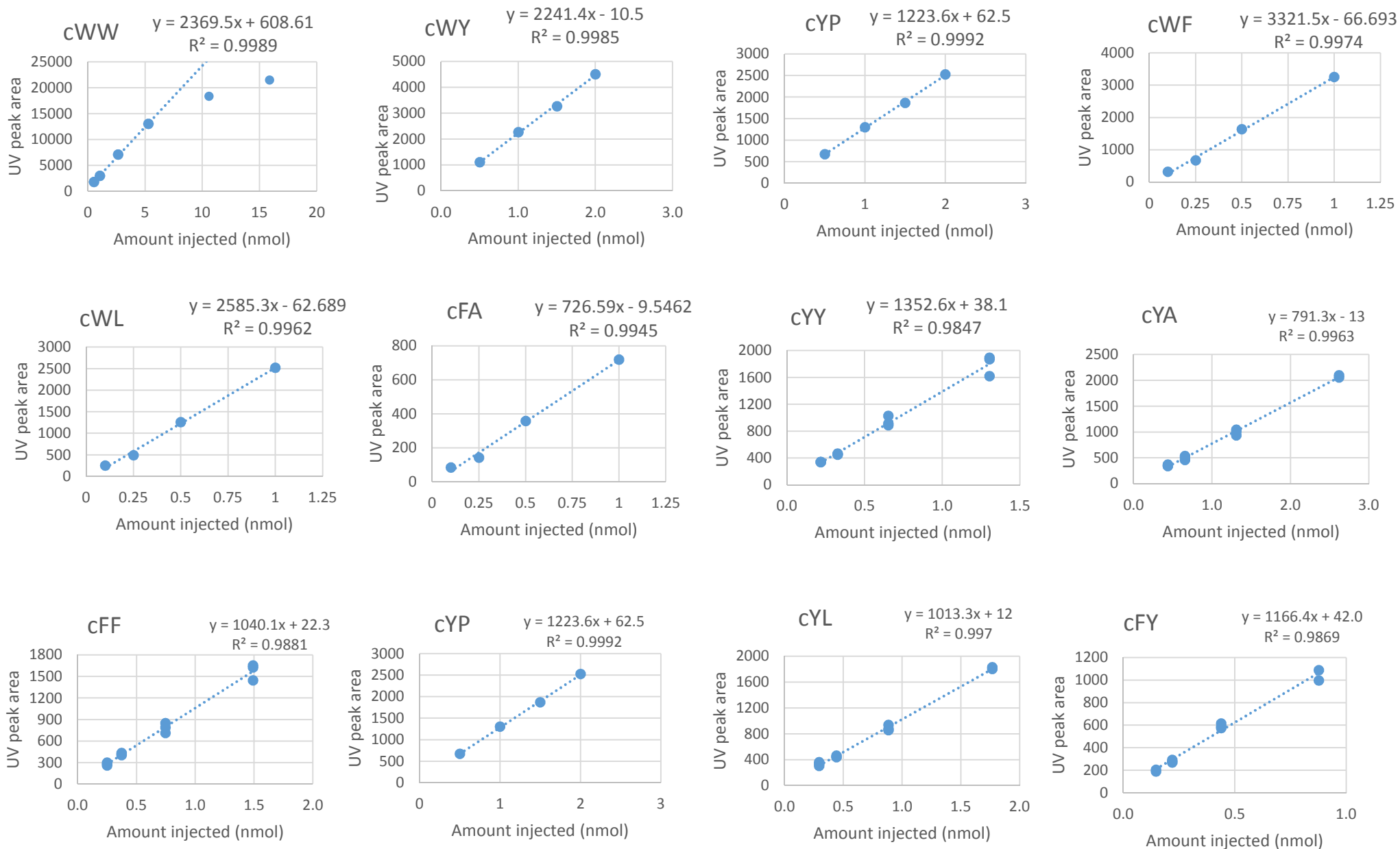


Figure S6 Calibration curves of cyclodipeptides

Saturation of UV detector arise for UV area above 14000 (cf. cWW curve). For UV peak area above this saturation threshold, a lower limit estimation is given in Figure S5.

Origin of compounds: cWW, cWY, cYP, CWF, cFF, cWL: Commercial compounds obtained from Bachem ; cYY, cYA, cYL, cFY: Gondry *et al* (2009) Nat Chem Biol 5, 414-420 ; cFA: Jacques *et al* (2015) Nat Chem Biol 11, 721-727

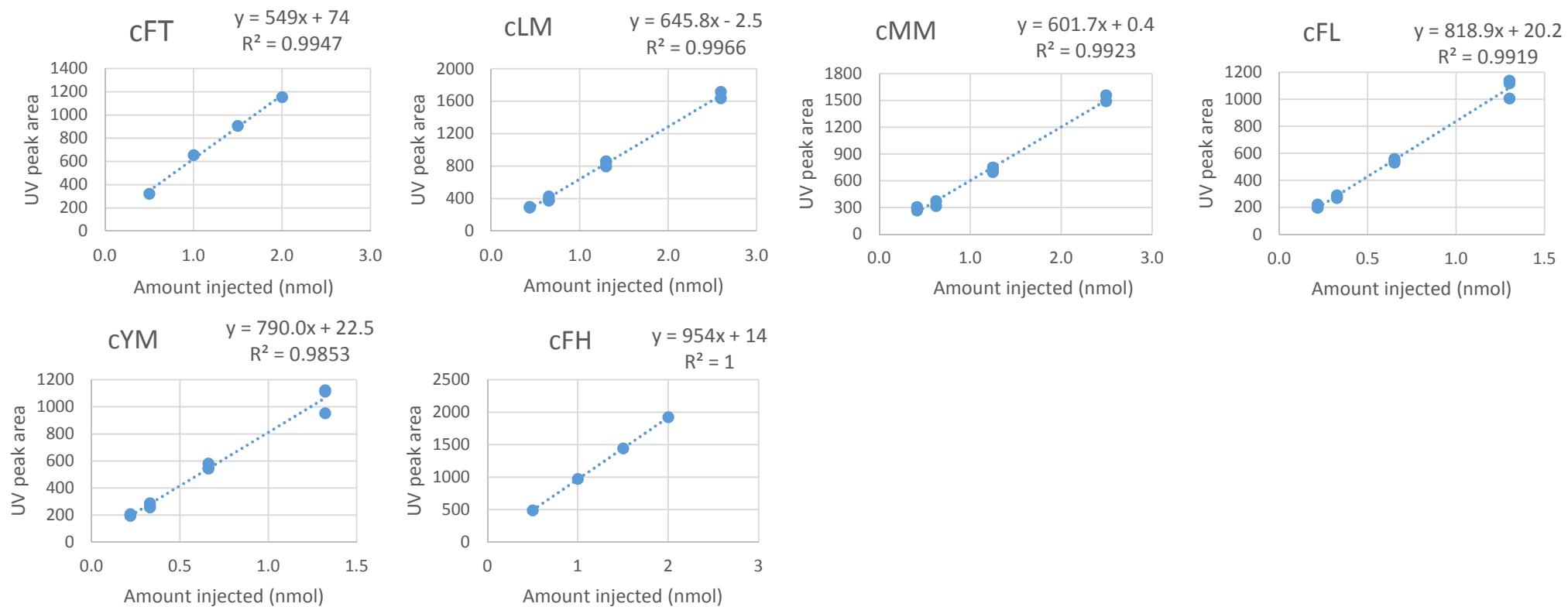


Figure S6 (continued)

Origin of compounds : cMM, cFL, cFH: Commercial compounds from Bachem ; cLM, cYM: Gondry *et al* (2009) Nat Chem Biol 5, 414-420 ; cFT: Jacques *et al* (2015) Nat Chem Biol 11, 721-727

b. Influence of an hyperosmotic stress on the incorporation of *trans*-4-hydroxyproline

We performed a comparative assay to test the effect of high NaCl concentration on the incorporation of *trans*-4-hydroxyproline (Hyp) by CDPSs. The set of CDPSs producing proline-containing cyclodipeptides was expressed in the Pro- strain using the protocol described in Figure S1. After medium replacement, each phase II culture was divided into two. One half was supplemented with NaCl (600 mM final) while the other half was left in standard conditions. Cultures were complemented with 200 mg/L *trans*-4-Hydroxy-L-proline, transferred to 24-well plates and grown overnight. Supernatants were collected and analyzed by LC-MS/MS. Results are given in Figure S6.

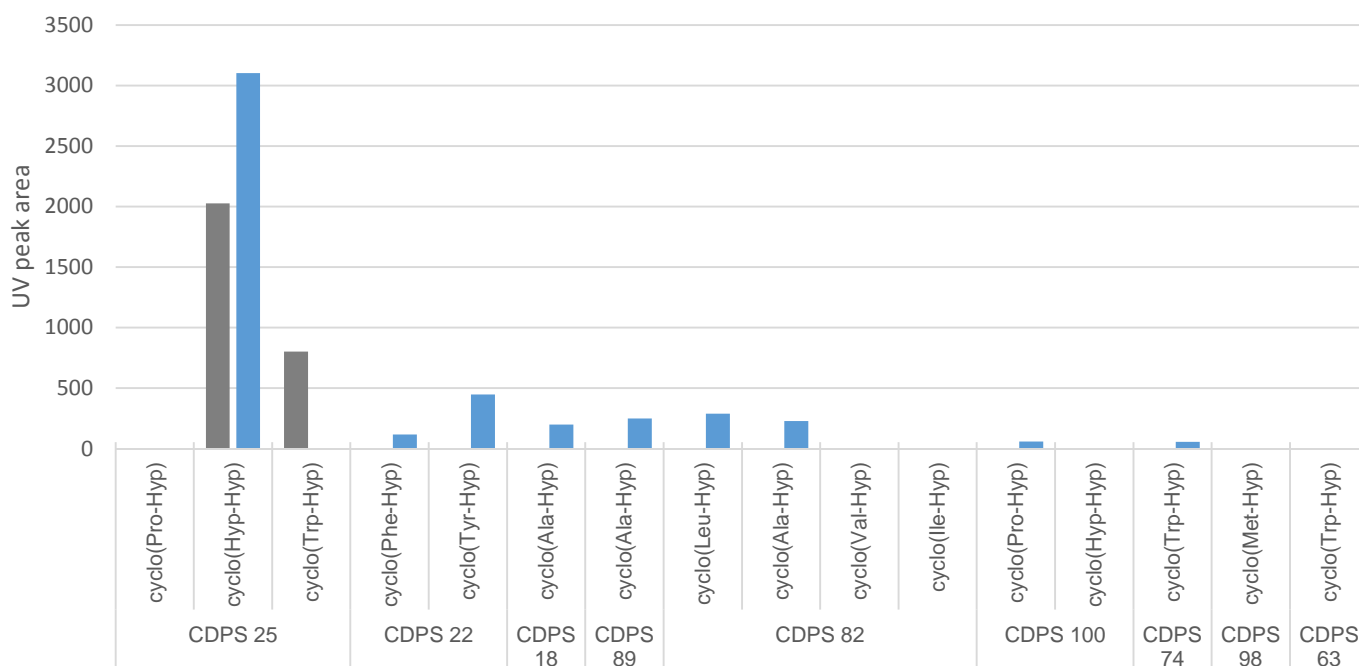










Figure S7 Comparison of hydroxyproline incorporation with and without osmotic stress. Cyclodipeptides containing *trans*-4-Hydroxy-L-proline were identified by LC-MS/MS. UV peak area are displayed for cultures resuspended in classical medium (grey) or in hyperosmotic medium (blue).

c. Comparative overview of loading efficiencies of ncAAs by AARSs and their incorporation by CDPSs

In order to check whether poor loading efficiencies could explain the absence of incorporation by CDPSs for some ncAAs, we gathered the information available in the literature on loading efficiencies by AARSs and we looked for correlation with the results of our CDPS-dependent incorporation assay. These information are given in Table S6.

Table S6 Comparative overview of loading efficiencies of ncAAs by AARSs and their incorporation by CDPSs

Proline analogs								
Name	# in this study	Short form	Formula	Loading efficiencies by AARS				Number of CDPSs incorporating the ncAA over number of CDPSs tested
				Reference	AARS	Direct estimation of loading efficiency by <i>in vitro</i> assay (relative efficiency compared to proteinogenic amino acid)	Indirect estimation by <i>in vivo</i> methods	
L-Azetidine-2-carboxylic acid	2	Aze		Josephson et al., 2005 Hartman et al., 2006	ProRS	Less than 5%		10/10
3,4-Dehydro-L-proline	3	34DePro		Hartman et al., 2006	ProRS	100%		9/10
<i>cis</i> -4-Fluoro-L-proline	4	c4FPro		Oldach et al., 2012	ProRS		High level of incorporation*	9/10
<i>trans</i> -4-Fluoro-L-proline	5	t4FPro		Oldach et al., 2012	ProRS		High level of incorporation*	9/10
<i>trans</i> -4-Hydroxy-L-proline	6	t4HOPro		Hartman et al., 2006 Oldach et al., 2012 Buetcher et al., 2003	ProRS	25%	High level of incorporation*	7/9
α -Methyl-L-proline	7	α MePro		Hartman et al., 2006	ProRS	Less than 5%		1/10
L-2-Thiazolidinecarboxylic acid	8	2ThPro		Hartman et al., 2006 Oldach et al., 2012	ProRS	230%	High level of incorporation*	1/10
L-4-Thiazolidinecarboxylic acid	9	4ThPro		Hartman et al., 2006 Oldach et al., 2012 Budisa et al., 1998	ProRS	250%	High level of incorporation*	1/10

* Residue-specific incorporation of proline analogs in ribosomal peptides using an *E. coli* Pro auxotroph strain.

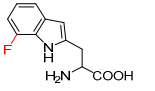
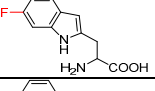
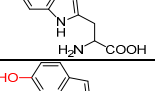
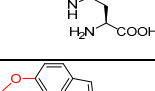
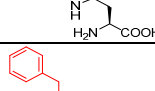
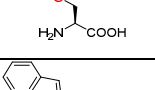
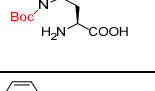
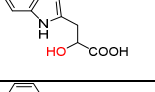
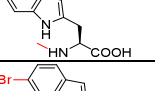
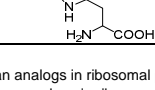
Table S6 (continued)

Phenylalanine/tyrosine analogs				Loading efficiencies by AARS			Number of CDPs incorporating the ncAA over number of CDPs tested
Name	# in this study	Short form	Formula	Reference	AARS	Direct estimation of loading efficiency by <i>in vitro</i> assay (relative efficiency compared to proteinogenic amino acid)	
3-Fluoro-DL-tyrosine	11	3FTyr		Josephson et al., 2005 Hartman et al., 2006	TyrRS	100% (compared to β -Homotyrosine)	9/10
3,4-Dihydroxy-L-phenylalanine	12	DOPA		Budisa et al., 2006	TyrRS	No quantification available	8/10
4-Fluoro-DL-phenylalanine	14	4FPhe		Wang et al., 2002	PheRS	k_{cat}/K_m 4FPhe diminished about 10-fold when compared to Phe	10/10
2-Fluoro-DL-phenylalanine	15	2FPhe		Hartman et al., 2006	PheRS	140%	9/10
3-(2-Thienyl)-DL-alanine	16	3-ThienA		Hartman et al., 2006	PheRS	No quantification available	10/10
3-(3-Pyridyl)-L-alanine	17	3PyrA		Kirshenbaum et al., 2002	PheRS	90% efficiency of <i>in vivo</i> substitution of Phe by 3PyrA*	7/10
3-(4-Pyridyl)-L-alanine	18	4PyrA		Kirshenbaum et al., 2002	PheRS	85% efficiency of <i>in vivo</i> substitution of Phe by 4PyrA*	8/10
β -(2-Thienyl)-DL-serine				Hartman et al., 2006	PheRS	150%	0/10
3,4,5-Trifluoro-L-phenylalanine				Fan et al., 2014	TyrRS	Level of protein production comparable with a standard completed with tyrosine**	0/10
β -Methyl-L-phenylalanine hydrochloride				Hartman et al., 2006	PheRS	65%	0/10
O-Acetyl-L-tyrosine				Fan et al., 2014	TyrRS	Level of protein production comparable with a standard completed with tyrosine**	0/10
3-Nitro-L-tyrosine				Hartman et al., 2006	TyrRS	30% (compared to β -Homotyrosine)	0/10
4-Nitro-L-phenylalanine monohydrate				Fan et al., 2014	TyrRS	Level of protein production comparable with a standard completed with tyrosine**	0/10
3,5-Dibromo-L-tyrosine monohydrate				Fan et al., 2014	TyrRS	Level of protein production comparable with a standard completed with tyrosine**	0/10
3,4-Dichloro-L-phenylalanine				Fan et al., 2014	TyrRS	Level of protein production comparable with a standard completed with tyrosine**	0/10
DL- β -Homophenylalanine				Hartman et al., 2006	PheRS/ TyrRS	25% (compared to Phe)	0/10
L- β -Homotyrosine hydrochloride				Hartman et al., 2006	PheRS/ TyrRS	More than 100% (compared to Tyr)	0/10

* Residue-specific incorporation of phenylalanine analogs in ribosomally translated proteins using an *E. coli* Phe auxotroph strain. Numbers refers to the relative amount of proteins produced with the analogs, compared to a standard complemented with phenylalanine.

** Incorporation of Tyr analogs in proteins by TyrRS-dependent loading of the analogs on a suppressor tRNA necessary for protein translation.

Table S6 (continued)

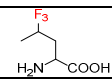
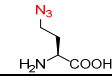
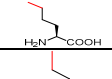
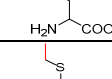
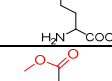
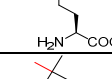
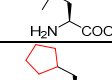
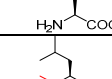
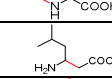
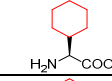
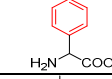
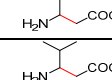
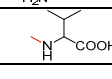
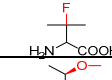
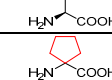
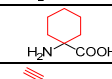
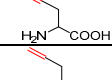
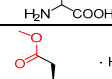
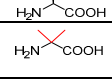
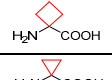
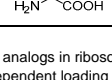
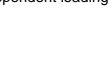



Tryptophan analogs				Loading efficiencies by AARS				Number of CDPs incorporating the ncAA over number of CDPs tested
Name	# in this study	Short form	Formula	Reference	AARS	Direct estimation of loading efficiency by <i>in vitro</i> assay (relative efficiency compared to proteinogenic amino acid)	Indirect estimation by <i>in vivo</i> methods	
4-Fluoro-DL-tryptophan	20	4FTrp		Hartman et al., 2006 Oldach et al., 2012	TrpRS	75%	High level of incorporation*	6/6
5-Fluoro-DL-tryptophan	21	5FTrp		Hartman et al., 2006	TrpRS	50%		6/6
5-Methyl-DL-tryptophan	22	5MeTrp		Zhou et al., 2016 (TrpRS from <i>L. lactis</i>)	TrpRS		Low level of incorporation**	4/6
5-Hydroxy-L-tryptophan	23	5HOTrp		Hartman et al., 2006 Oldach et al., 2012	TrpRS	40%	High level of incorporation*	3/6
5-Methoxy-L-tryptophan				Hartman et al., 2006	TrpRS	30%		0/6
S-benzyl-L-Cysteine				Fan et al., 2014	TrpRS		Detectable level of protein production***	0/6
Nin-Boc-L-tryptophan				Fan et al., 2014	TrpRS/TyrRS		Detectable level of protein production***	0/6
DL-Indole-3-lactic acid				Fan et al., 2014	TrpRS		Higher protein production compared to a standard complemented with tryptophan***	0/6
N-Methyl-L-tryptophan				Hartman et al., 2006	TrpRS	25%		0/6
5-Bromo-DL-tryptophan				Hartman et al., 2006	TrpRS	25%		0/6

* Residue-specific incorporation of tryptophan analogs in ribosomal peptides using an *E. coli* Trp auxotroph strain.

** Residue-specific incorporation of tryptophan analogs in ribosomal peptides using a *L. lactis* Trp auxotroph strain.

*** Incorporation of Trp analogs in proteins by TrpRS-dependent loading of the analogs on a suppressor tRNA necessary for protein translation.

Table S6 (continued)

Leucine/isoleucine/valine/methionine analogs				Loading efficiencies by AARS			Number of CDPSs incorporating the ncAA over number of CDPSs tested	
Name	# in this study	Short form	Formula	Reference	AARS	Direct estimation of loading efficiency by <i>in vitro</i> assay (relative efficiency compared to proteinogenic amino acid)		Indirect estimation by <i>in vivo</i> methods
5,5,5-Trifluoro-DL-leucine	28	TriFLeu		Josephson et al., 2005 Hartman et al., 2006	LeuRS	80%		10/10 (tested as a Leu/Ile/Val analog)
4-Azido-L-homoalanine	29	Aha		Hartman et al., 2006 Oldach et al., 2012 Connor et al., 2008	MetRS	150%	High level of incorporation*	9/12 (tested as a Met analog and as Leu/Ile/Val analog)
L-Norleucine	30	Norleu		Hartman et al., 2006 Oldach et al., 2012	MetRS	No quantification available	High level of incorporation*	10/12 (tested as a Met analog and as Leu/Ile/Val analog)
DL-Norvaline	31	Norval		Hartman et al., 2006	MetRS	75%		8/12 (tested as a Leu/Ile/Val analog and a Met analog)
DL-Ethionine	32	Eth		Hartman et al., 2006 Oldach et al., 2012	MetRS	150%		6/6 (tested as a Met analog)
L-Glutamic acid methyl ester	33	MeEGlu		Hartman et al., 2006	MetRS	120%		6/6 (tested as a Met analog), overall low efficiency
γ-Methyl-L-leucine	34	γMeLeu		Hartman et al., 2006	LeuRS	75%		5/10 (tested as a Leu/Ile/Val analog)
3-Cyclopentylalanine				Hartman et al., 2006	LeuRS	45%		0/10 (tested as a Leu/Ile/Val analog)
N-Methyl-L-leucine				Hartman et al., 2006	LeuRS	75%		0/10 (tested as a Leu/Ile/Val analog)
DL-β-Homoleucine				Hartman et al., 2006	LeuRS	Less than 5%		0/10 (tested as a Leu/Ile/Val analog)
L-Cyclohexylglycine				Hartman et al., 2006	IleRS	75%		0/10 (tested as a Leu/Ile/Val analog)
DL-Phenylglycine				Hartman et al., 2006	IleRS	75%		0/10 (tested as a Leu/Ile/Val analog)
DL-3-aminobutyric acid				Hartman et al., 2006	ValRS	55%		0/10 (tested as a Leu/Ile/Val analog)
DL-β-Homovaline				Hartman et al., 2006	ValRS	20%		0/10 (tested as a Leu/Ile/Val analog)
N-Methyl-DL-valine				Hartman et al., 2006	ValRS	Less than 5%		0/10 (tested as a Leu/Ile/Val analog)
3-fluoro-DL-valine				Hartman et al., 2006	ValRS	80%		0/10, but important contamination of the product by non-fluorated valine
O-Methyl-L-threonine				Hartman et al., 2006	ValRS	100%		0/10 (tested as a Leu/Ile/Val analog)
Cycloleucine				Hartman et al., 2006	LeuRS/ ValRS	50% (compared to Ile), 10% (compared to Val)		0/10 (tested as a Leu/Ile/Val analog)
Homocycloleucine				Hartman et al., 2006	LeuRS/ ValRS	50% (compared to Ile), 25% (compared to Val)		0/10 (tested as a Leu/Ile/Val analog)
DL-Propargylglycine				Hartman et al., 2006 Connor et al., 2008	MetRS	150%	84% of yield compared to the same experiment with Met**	0/10 (tested as a Leu/Ile/Val analog)
DL-2-Allylglycine				Hartman et al., 2006	MetRS	100%		0/10 (tested as a Leu/Ile/Val analog)
L-Aspartic acid methyl ester hydrochloride				Hartman et al., 2006	MetRS	140%		0/10 (tested as a Leu/Ile/Val analog)
2-Aminoisobutyric acid				Hartman et al., 2006	AlaRS	80%		0/10 (tested as a Leu/Ile/Val analog)
1-Amino-1-cyclobutanecarboxylic acid				Hartman et al., 2006	AlaRS	60%		0/10 (tested as a Leu/Ile/Val analog)
1-Aminocyclopropane-carboxylic acid				Hartman et al., 2006	AlaRS	10%		0/10 (tested as a Leu/Ile/Val analog)

* Residue-specific incorporation of methionine analogs in ribosomal peptides using an *E. coli* Met auxotroph strain.

** L/F transferase assay coupled to MetRS-dependent loading of Propargylglycine on tRNAs.

IV) Isolation of selected cyclodipeptides

a. *In vivo* cyclodipeptide production

In order to produce cyclo(5FTrp-Pro), *E. coli* strain JW2326, auxotrophic for the three proteinogenic aromatic amino acids, was transformed with the plasmid pIJ196-CDPS74. A preculture was grown in M9 minimum medium (same composition as in the general method, see p.SXX), supplemented with the 20 proteinogenic amino acids (50mg/L each) and containing ampicillin (200µg/ml) and kanamycin (50µg/ml) and was used to inoculate 5 flasks of 200mL of the same minimum medium (OD at inoculation 0.05). The flasks were incubated at 200rpm and 37°C until an OD600 of 0.6 was reached, at which point protein expression was induced with 1mM IPTG. After 4 hours incubation at 20°C, cells were pelleted by centrifugation and resuspended in 200mL M9 minimum medium containing no vitamins, no antibiotics and no proteinogenic amino acids but supplemented with 200mg/L 5-fluoro-DL-tryptophan (**21**). After 20 hour incubation at 20°C, the supernatant was removed and concentrated to 60mL under vacuum. Cyclodipeptides were extracted with chloroform (3x20mL). The organic phase was washed twice with 20mL K₂CO₃, once with 20mL saturated brine and dried over anhydrous magnesium sulfate. The solvent was removed by rotary evaporation and the residue was taken up in 25% acetonitrile/75% water, filtered and purified by semipreparative HPLC on a LiChrospher® 100 RP-18 C18 column (250x10mm, 5µm) (Merck Millipore) in an isocratic gradient of 25% mobile phase B for 5 min followed by a gradient of 25 to 35% mobile phase B over 20min (mobile phase A : water + 0.1% formic acid; mobile phase B : acetonitrile + 10% water + 0.1% formic acid) to yield approximately 6mg of cyclo(5FTrp-Pro) (purity above 99%, see NMR characterization).

A similar protocol was used for the production of cyclo(Aha-Phe) and cyclo(Aha-Leu) with the following modifications: *E. coli* strain JW5807, auxotrophic for leucine was transformed with the plasmid pIJ196-CDPS54. Cyclodipeptide production was performed with supplementation of 100mg/L L-azidohomoalanine (**29**) and the cyclodipeptides were purified in an isocratic gradient of 20% mobile phase B for 5 min followed by a gradient of 20 to 30% mobile phase B over 20min (mobile phase A : water + 0.1% formic acid; mobile phase B : acetonitrile + 10% water + 0.1% formic acid). Around 1.5mg of cyclo(Aha-Phe) was obtained (purity around 85% estimated by analytical HPLC and NMR, contamination with cyclo(Phe-Met)). About 0.4mg of cyclo(Aha-Leu) was obtained (purity above 99%, see NMR characterization).

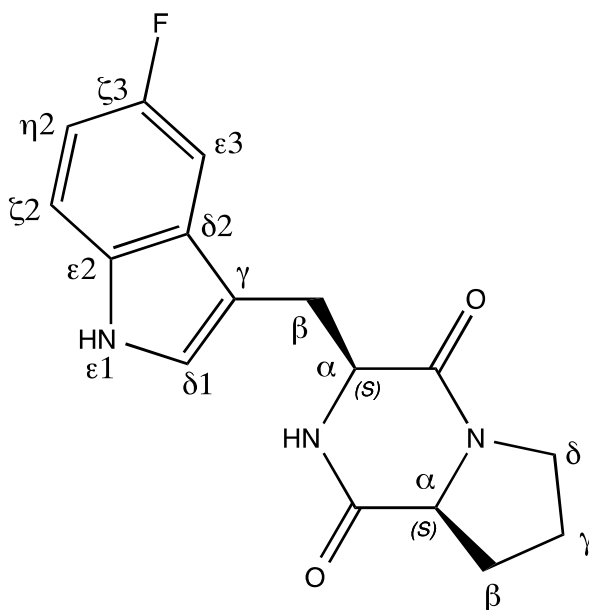
Notably, in the three cases, a unique peak exhibited the m/z ratio of the compound of interest and was purified.

b. NMR analysis

1) NMR spectroscopy

^1H , ^{13}C and ^{15}N NMR experiments were recorded on a Bruker Avance III spectrometer equipped with a 5-mm TCI inverse z-gradient cryoprobe and operating at a ^1H frequency of 500 MHz. One-dimensional ^{19}F NMR experiment was recorded on a Bruker Nanobay Avance III spectrometer equipped with a 5-mm BBFO direct probe operating at a ^1H frequency of 300 MHz. Spectra were recorded at 298.6 K. All data were processed and analyzed with Bruker TOPSPIN software. ^1H and ^{13}C resonances were assigned via the analysis of one-dimensional ^1H , one-dimensional ^{13}C DEPTQ (Distortionless Enhancement by Polarization Transfer), two-dimensional ^1H - ^1H COSY, two-dimensional ^1H - ^1H TOCSY (Total Correlation Spectroscopy, mixing time of 66 ms), two-dimensional ^1H - ^1H ROESY (Rotating frame Overhauser Enhancement Spectroscopy, mixing time of 200 or 600 ms), two-dimensional ^1H - ^{13}C HSQC (Heteronuclear Single Quantum Correlation), two-dimensional ^1H - ^{13}C HMBC (Heteronuclear Multiple-Bond Correlation, optimized for $J = 8$ Hz). ^1H and ^{13}C chemical shifts were referenced to the DMSO solvent signal (δ 2.50 ppm and 39.5, respectively). ^{15}N resonances were assigned via the analysis of two-dimensional ^1H - ^{15}N HMBC (optimized for $J = 4$ or 6 Hz). ^{19}F resonance was assigned on one-dimensional ^{19}F NMR spectrum. ^{15}N and ^{19}F chemical shifts were referenced indirectly to liquid ammonia and trichloro-fluoro-methane, respectively, by using the chemical shift of the lock solvent.

2) NMR assignments of cyclo(5-Fluoro-L-Tryptophan-L-Proline)



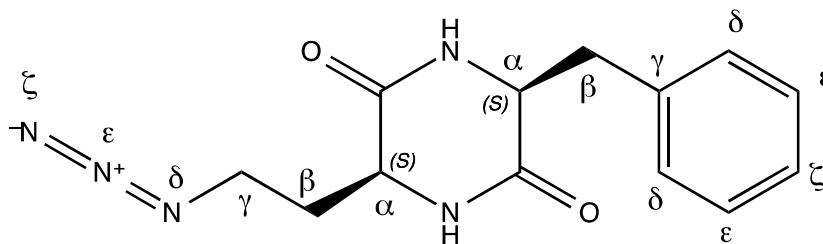
^1H NMR (500.3 MHz, DMSO): δ 10.95 (d, $J = 2.1$ Hz, 1H, H ϵ_1 Trp), 7.83 (s, 1H, H N Trp), 7.34 (dd, $J = 10.4, 2.6$ Hz, 1H, H ϵ_3 Trp), 7.30 (dd, $J = 8.8, 4.6$ Hz, 1H, H ζ_2 Trp), 7.23 (d, $J = 2.4$ Hz, 1H, H δ_1 Trp), 6.88 (td, $J = 9.1, 2.6$ Hz, 1H, H η_2 Trp), 4.30 (t, $J = 5$ Hz, 1H, H α Trp), 4.04 (ddd, $J = 10.0, 7.0, 1.3$ Hz, 1H, H α Pro), 3.37 (dt, $J = 11.6, 8.0$ Hz, 1H, H δ Pro), 3.25 (ddd, $J = 11.7, 8.9, 3.9$ Hz, 1H, H δ' Pro), 3.15 (ABX, $J = 14.8, 4.9$ Hz, 1H, H β Trp), 3.09 (ABX, $J = 14.8, 5.1$ Hz, 1H, H β' Trp), 1.96 (m, 1H, H β Pro), 1.69 (m, 1H, H γ Pro), 1.61 (m, 1H, H γ' Pro), 1.30 (m, 1H, H β' Pro).

^{13}C NMR (125.8 MHz, DMSO): δ 168.8 (C' Pro), 165.4 (C' Trp), 156.6 (d, $^1J_{CF} = 231$ Hz, C ζ_3 Trp), 132.5 (C ϵ_2 Trp), 127.7 (d, $^3J_{CF} = 10$ Hz, C δ_2 Trp), 126.6 (C δ_1 Trp), 112.0 (d, $^3J_{CF} = 10$ Hz, C ζ_2 Trp), 109.5 (d, $^4J_{CF} = 5$ Hz, C γ Trp), 108.9 (d, $^2J_{CF} = 26$ Hz, C η_2 Trp), 103.5 (d, $^2J_{CF} = 23$ Hz, C ϵ_3 Trp), 58.3 (C α Pro), 55.3 (C α Trp), 44.5 (C δ Pro), 27.7 (C β Pro), 25.9 (C β Trp), 21.7 (C γ Pro).

^{15}N NMR (50.7 MHz, DMSO): δ 132.2 (N ϵ_1 Trp), 125.9 (N Pro), 117.6 (N Trp)

^{19}F NMR (282.4 MHz, DMSO): δ -125.66 (td, $J_{FH} = 10, 4.6$ Hz, F Trp)

3) NMR assignments of cyclo (L-Azidohomoalanine-L-Phenylalanine)

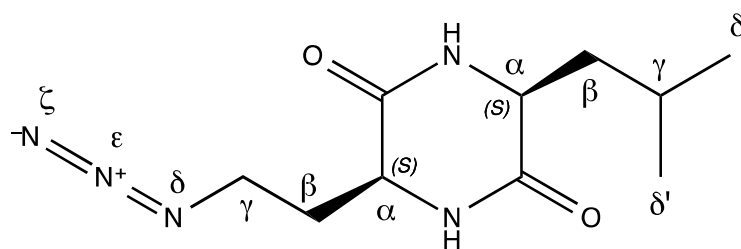


^1H NMR (500.3 MHz, DMSO): δ 8.26 (d, $J = 2.0$ Hz, 1H, H^{N} Phe), 8.12 (d, $J = 2.1$ Hz, 1H, H^{N} Aha), 7.29 (m, 2H, H_{ϵ} Phe), 7.23 (m, 1H, H_{ζ} Phe), 7.16 (m, 2H, H_{δ} Phe), 4.20 (m, 1H, H_{α} Phe), 3.67 (m, 1H, H_{α} Aha), 3.15 (dd, $J = 13.4, 3.6$ Hz, 1H, H_{β} Phe), 2.85 (dd, $J = 13.4, 4.9$ Hz, 1H, H_{β} Phe), 2.80 (m, 2H, H_{γ} Aha), 1.17 (m, 1H, H_{β} Aha), 0.85 (m, 1H, $\text{H}_{\beta'}$ Aha).

^{13}C NMR (125.8 MHz, DMSO): δ 166.3 (C' Aha), 166.1 (C' Phe), 136.0 (C_{γ} Phe), 130.4 (C_{δ} Phe), 128.0 (C_{ϵ} Phe), 126.8 (C_{ζ} Phe), 55.3 (C_{α} Phe), 51.4 (C_{α} Aha), 46.1 (C_{γ} Aha), 38.1 (C_{β} Phe), 32.5 (C_{β} Aha).

^{15}N NMR (50.7 MHz, DMSO): δ 119.1 (N Aha), 117.2 (N Phe), 71.1 (N_{δ} Aha), 248.4 (N_{ϵ} Aha).

4) NMR assignments of cyclo (L-Azidohomoalanine-L-Leucine)



^1H NMR (500.3 MHz, DMSO): δ 8.24 (d, J = 1.9 Hz, 1H, H^{N} Leu), 8.21 (d, J = 1.8 Hz, 1H, H^{N} Aha), 3.87 (m, 1H, H_{α} Aha), 3.77 (m, 1H, H_{α} Leu), 3.47 (m, 2H, H_{γ} Aha), 1.98 (m, 1H, H_{β} Aha), 1.84 (m, 1H, $\text{H}_{\beta'}$ Aha), 1.81 (m, 1H, H_{γ} Leu), 1.60 (ddd, J = 13.7, 8.5, 4.9 Hz, 1H, H_{β} Leu), 1.46 (ddd, J = 13.7, 8.3, 5.7 Hz, $\text{H}_{\beta'}$ Leu), 0.88 (d, J = 6.6 Hz, 3H, CH_3 δ Leu), 0.86 (d, J = 6.6 Hz, 3H, CH_3 δ' Leu).

^{13}C NMR (125.8 MHz, DMSO): δ 168.6 (C' Leu), 167.6 (C' Aha), 52.5 (C_{α} Leu), 51.6 (C_{α} Aha), 46.9 (C_{γ} Aha), 42.8 (C_{β} Leu), 32.5 (C_{β} Aha), 23.6 (C_{γ} Leu), 22.9 (C_{δ} Leu), 21.8 ($\text{C}_{\delta'}$ Leu).

^{15}N NMR (50.7 MHz, DMSO): δ 117.9 (N Leu), 115.5 (N Aha), 71.1 (N_{δ} Aha), 248.3 (N_{ϵ} Aha).

5) Conformational and configurational analysis

The analysis of $^3J_{H\alpha-H\beta}$ coupling constants for both cyclo(5-Fluoro-L-Tryptophan-L-Proline) and cyclo(L-Azidohomoalanine-L-Phenylalanine) revealed that the χ_1 gauche+ ($\approx +60^\circ$) conformation of the aromatic side chain is the major rotamer ($\approx 80\%$ population). This indicates a privileged orientation of the aromatic moiety above the DKP ring, as already described for other DKP compounds.^[37] Such positioning is also confirmed by chemical shift analysis, as strong upfield shifts were observed for some side chain protons due to the anisotropic magnetic susceptibility of aromatic rings. Indeed, an upfield shift of ≈ 1 ppm is observed for $H_{\beta'}$ proton of Pro (with respect to random coil value) and for H_{β} and H_{γ} protons of homoalanine side chain (with respect to corresponding chemical shift values observed in cyclo(L-Azidohomoalanine-L-Leucine) peptide). These large chemical shift variations are only possible if both side chains lie on the same face of the DKP ring.

The ROE analysis unambiguously proved the homochirality of aminoacids in the different DKP compounds. Importantly, ROE correlations were observed between H_{α} protons for the three compounds, indicating a *syn* relationship of these protons. In the case of cyclo(5-Fluoro-L-Tryptophan-L-Proline), ROEs were observed between one H_{δ} proton of Pro and $H_{\delta 1}$ and $H_{\epsilon 3}$ protons of Trp, confirming the *syn* stereochemistry.

6) NMR spectra of cyclo(5-Fluoro-L-Tryptophane-L-Proline) in DMSO (298.6 K)

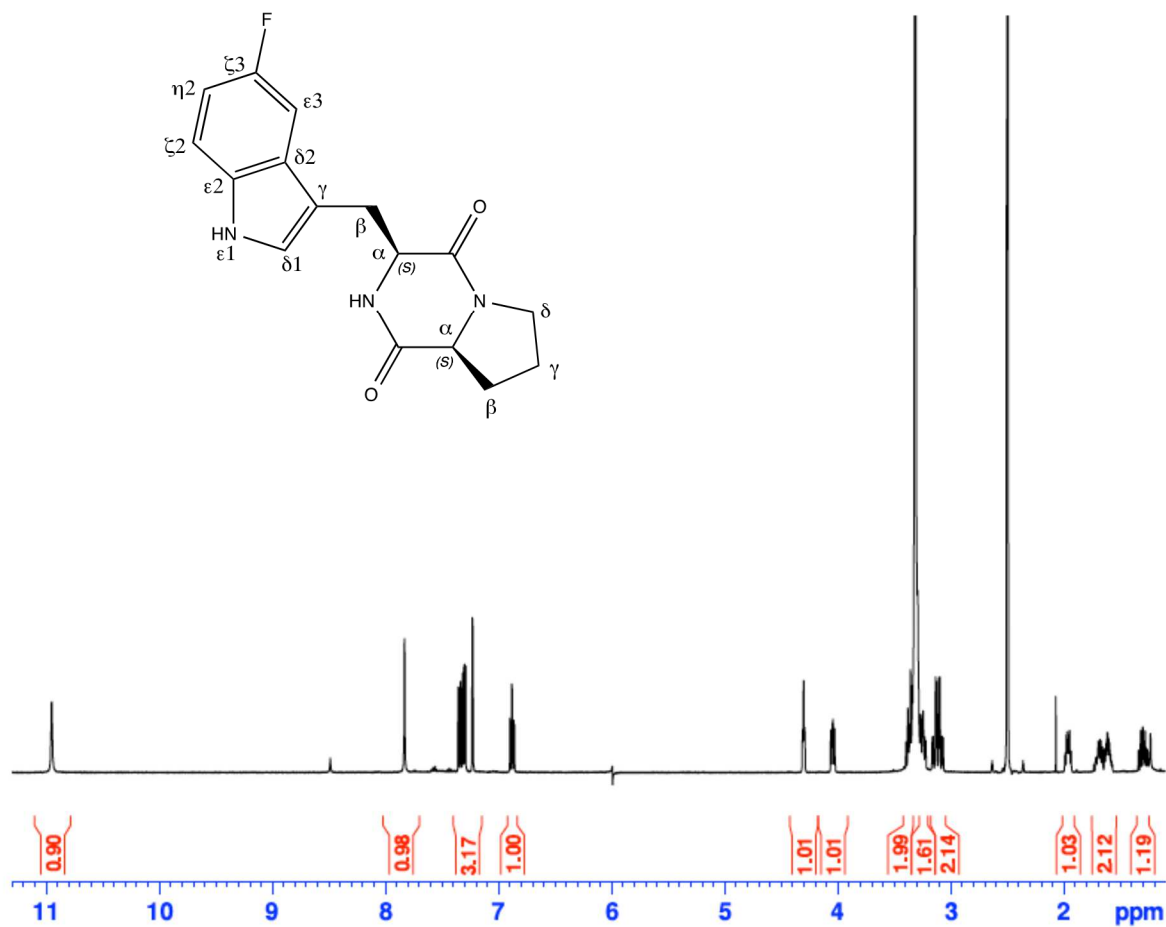


Figure S8 1D ^1H spectrum of cyclo(5-Fluoro-L-Tryptophane-L-Proline) in DMSO

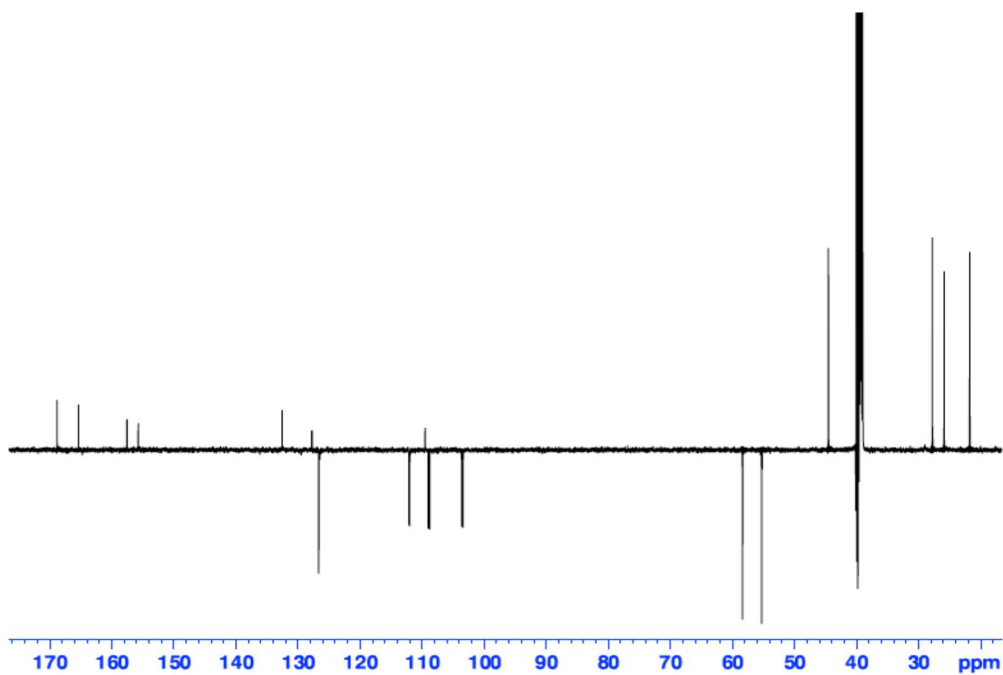


Figure S9 1D ^{13}C DEPTQ spectrum of cyclo(5-Fluoro-L-Tryptophane-L-Proline) in DMSO

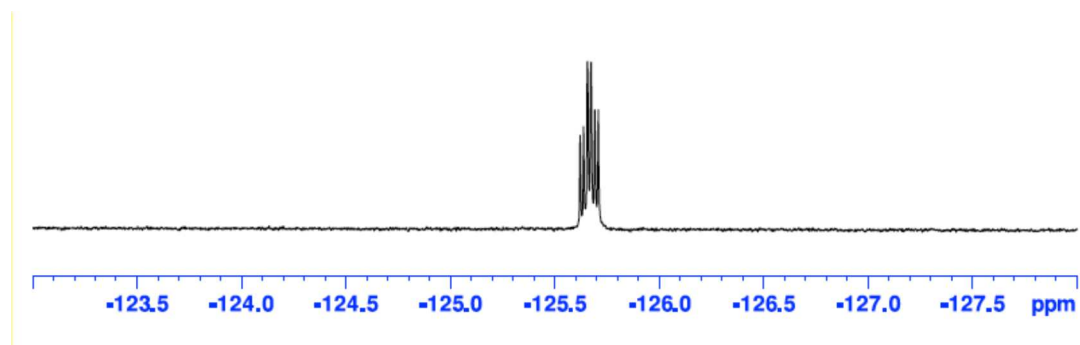


Figure S10 1D ^{19}F spectrum of cyclo(5-Fluoro-L-Tryptophane-L-Proline) in DMSO

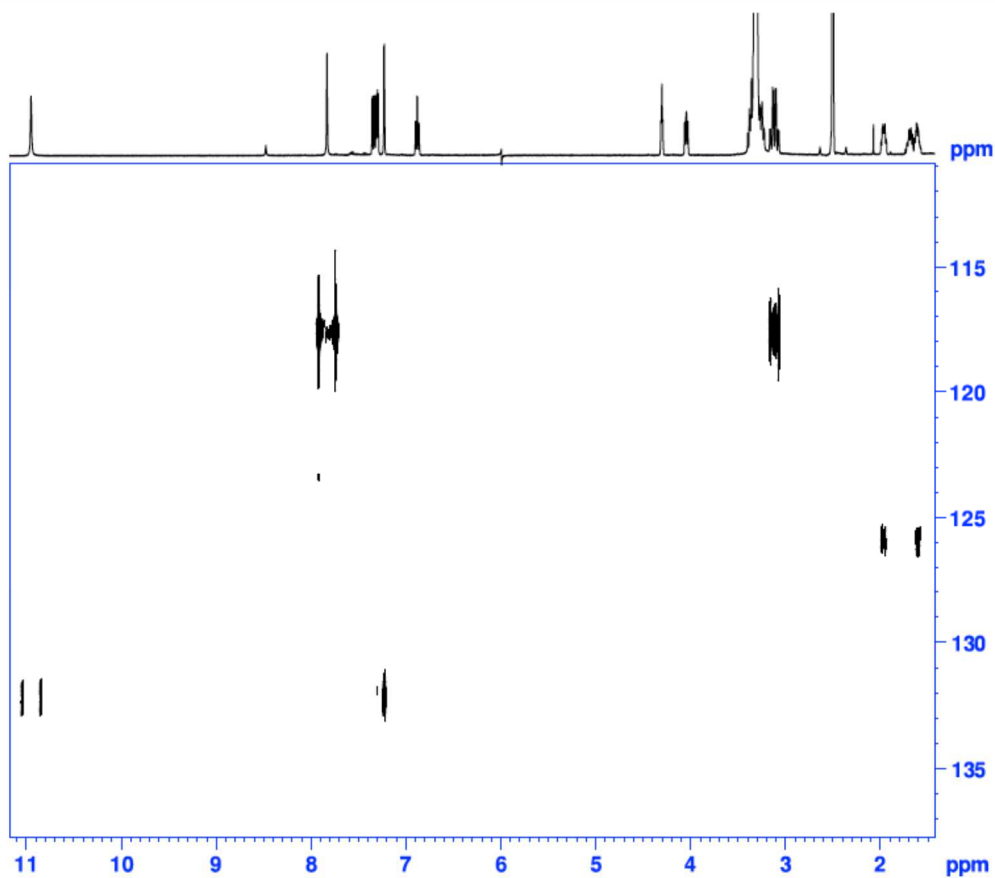


Figure S11 2D ^{15}N - ^1H HMBC spectrum of cyclo(5-Fluoro-L-Tryptophane-L-Proline) in DMSO

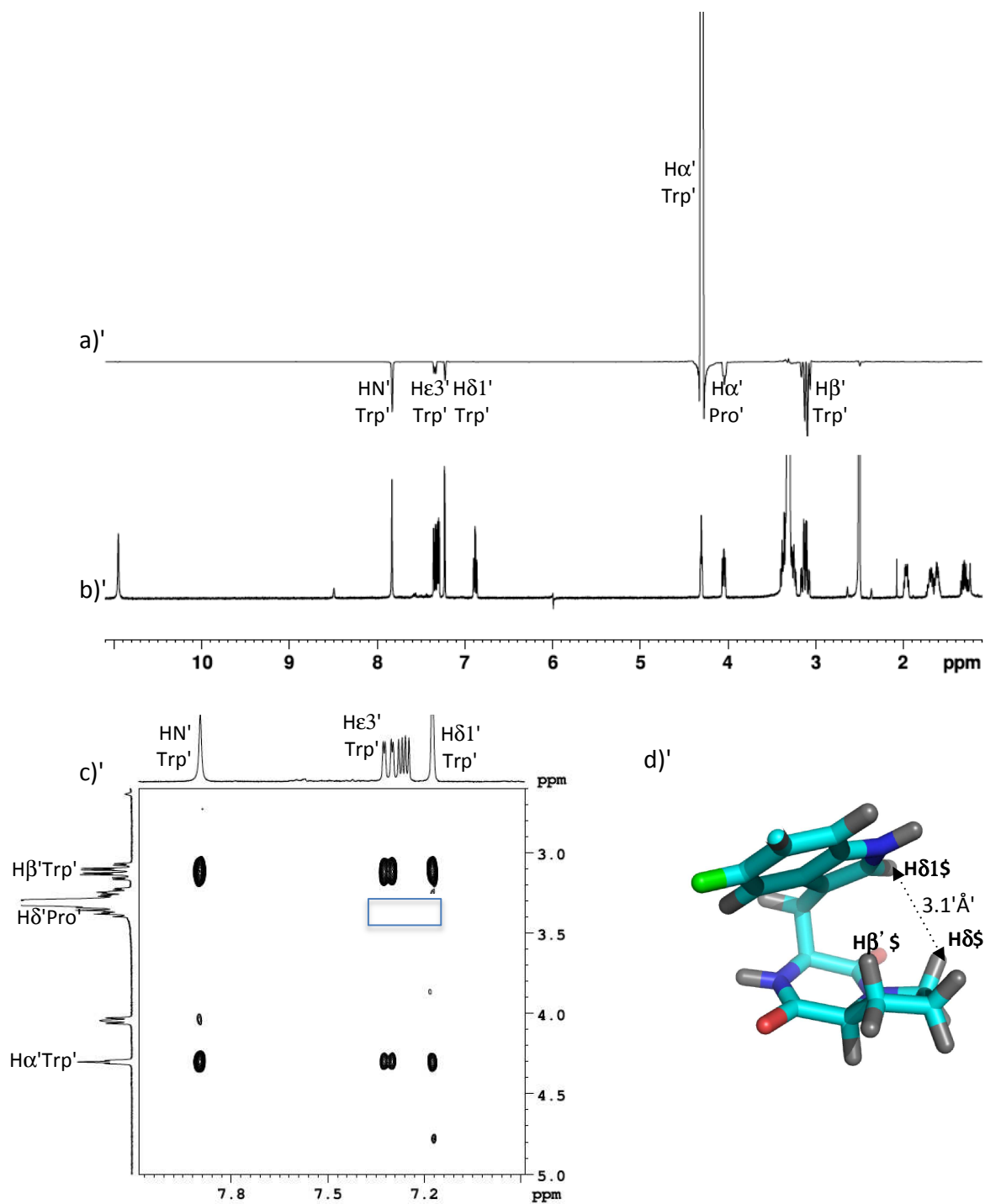


Figure S12 Stereochemical analysis of cyclo(5-Fluoro-L-Tryptophan-L-Proline) in DMSO.

(a) 1D row (F1 $\delta=4.30$ ppm) extracted from the 2D ^1H - ^1H ROESY spectrum (200 ms mixing time), showing the spatial proximity of the α protons of Trp and Pro residues. (b) 1D ^1H NMR spectrum. (c) Region of the 2D ^1H - ^1H ROESY spectrum, showing the spatial proximity of H δ 1 and H ϵ 3 aromatic protons of Trp with H δ proton of Pro. (d) Selected low energy conformer (χ_1 \square +60°, χ_2 \square +90° for Trp) showing the position of one H β proton in the shielding cone of indole ring and the close proximity between H δ Pro proton and Trp.

7) NMR spectra of cyclo(L-Azidohomoalanine-L-Phenylalanine) in DMSO (298.6 K)

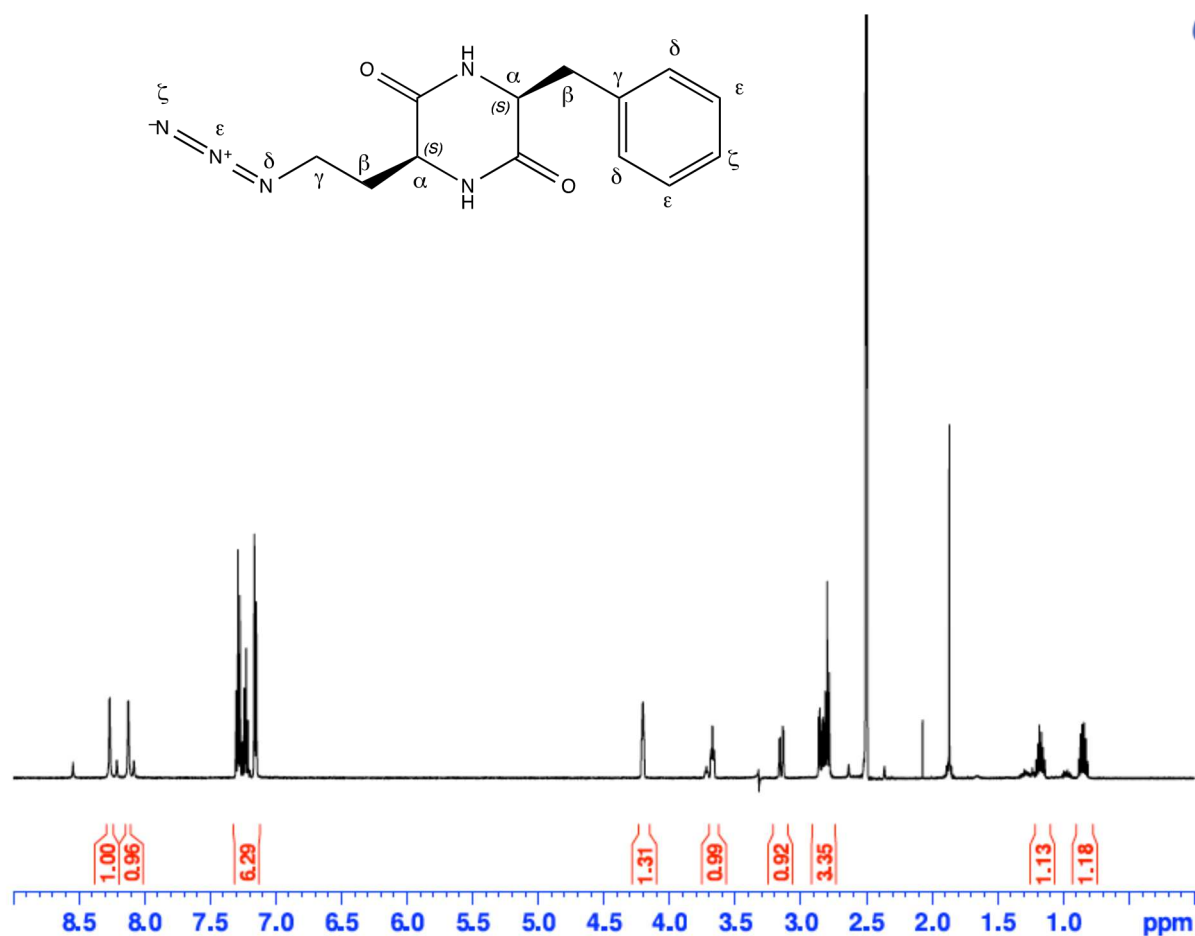


Figure S13 1D ¹H spectrum of cyclo(L-Azidohomoalanine-L-Phenylalanine) in DMSO. A weak presaturation was applied to suppress the water signal at 3.3 ppm. The spectrum contains 15% impurity corresponding to cyclo(L-Phe-L-Met).

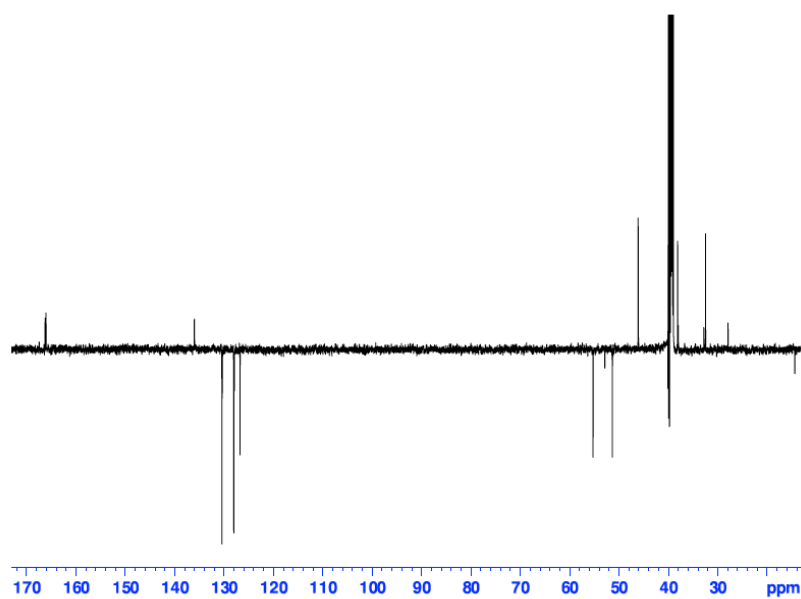


Figure S14 1D ^{13}C DEPTQ spectrum of cyclo(L-Azidohomoalanine-L-Phenylalanine) in DMSO

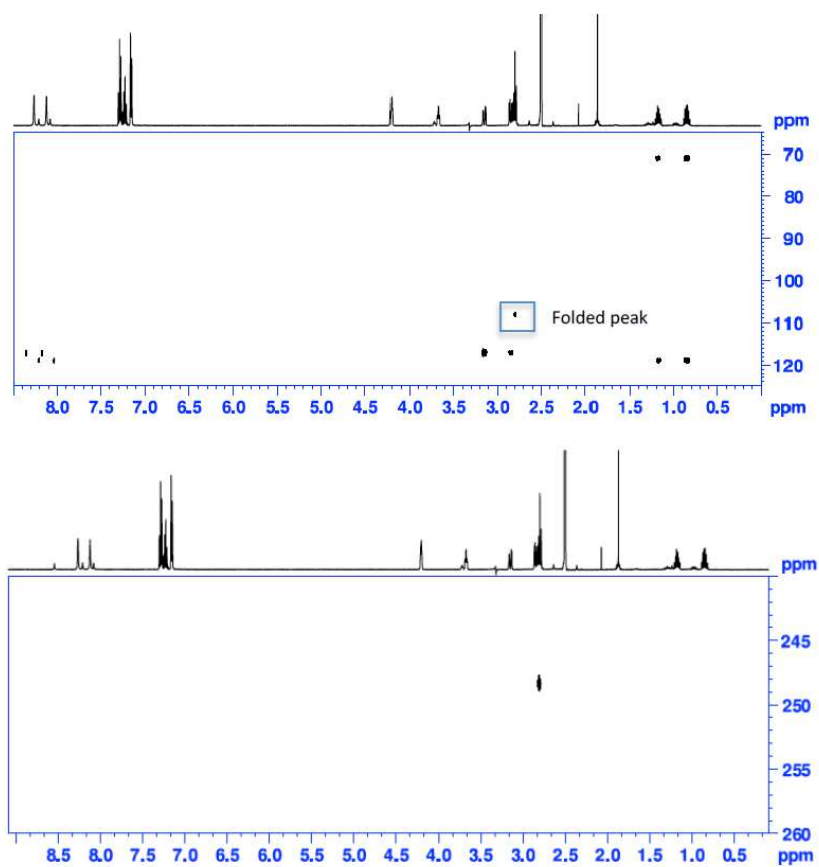


Figure S15 2D ^{15}N - ^1H HMBC spectra of cyclo(L-Azidohomoalanine-L-Phenylalanine) in DMSO

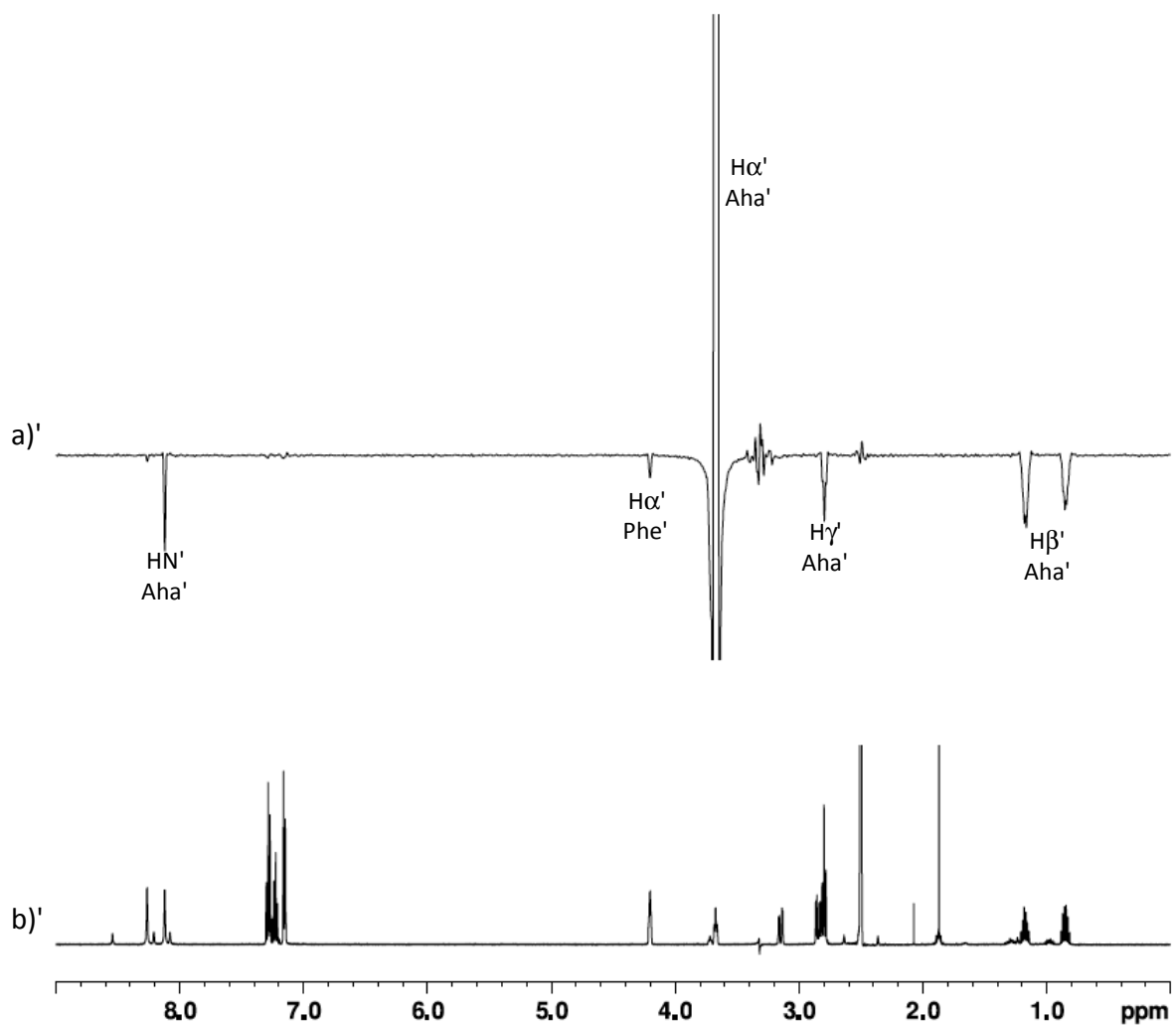


Figure S16 Stereochemical analysis of cyclo(L-Azidohomoalanine-L-Phenylalanine) in DMSO. (a) 1D row (F1 $\delta=3.67$ ppm) extracted from the 2D ¹H-¹H ROESY spectrum (600 ms mixing time), showing the spatial proximity of the H_α protons of Phe and Aha residues. (b) 1D ¹H NMR spectrum.

8) NMR spectra of cyclo(L-Azidohomoalanine-L-Leucine) in DMSO (298.6 K)

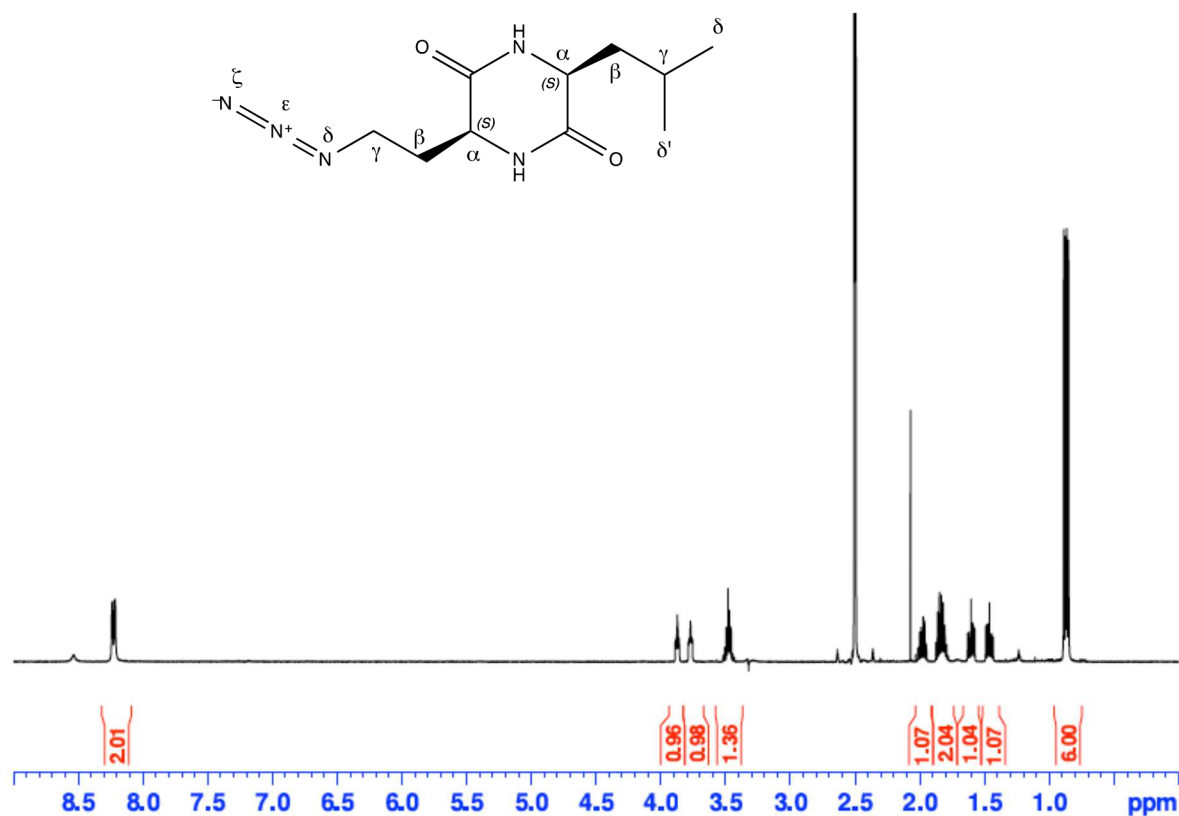


Figure S17 1D ¹H spectrum of cyclo(L-Azidohomoalanine-L-Leucine) in DMSO. A weak presaturation was applied to suppress the water signal at 3.3 ppm.

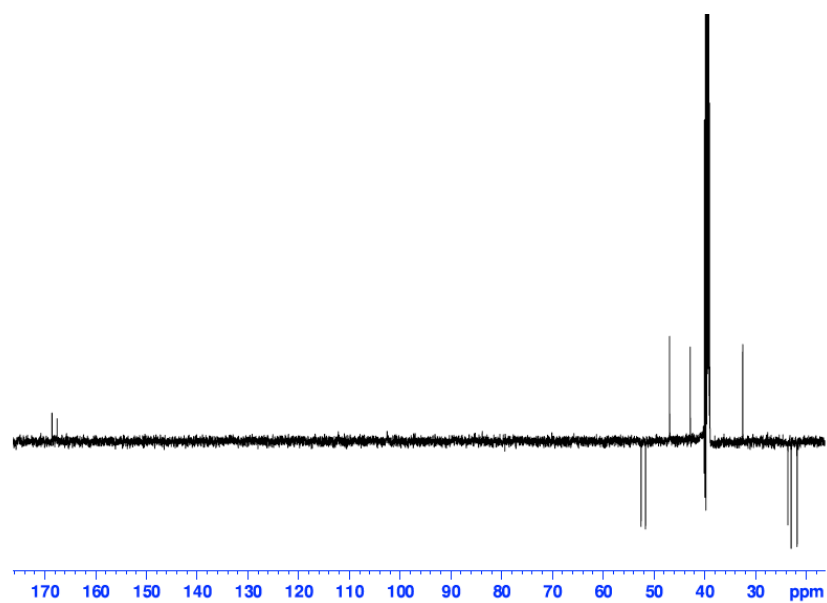


Figure S18 1D ^{13}C DEPTQ spectrum of cyclo(L-Azidohomoalanine-L-Leucine) in DMSO

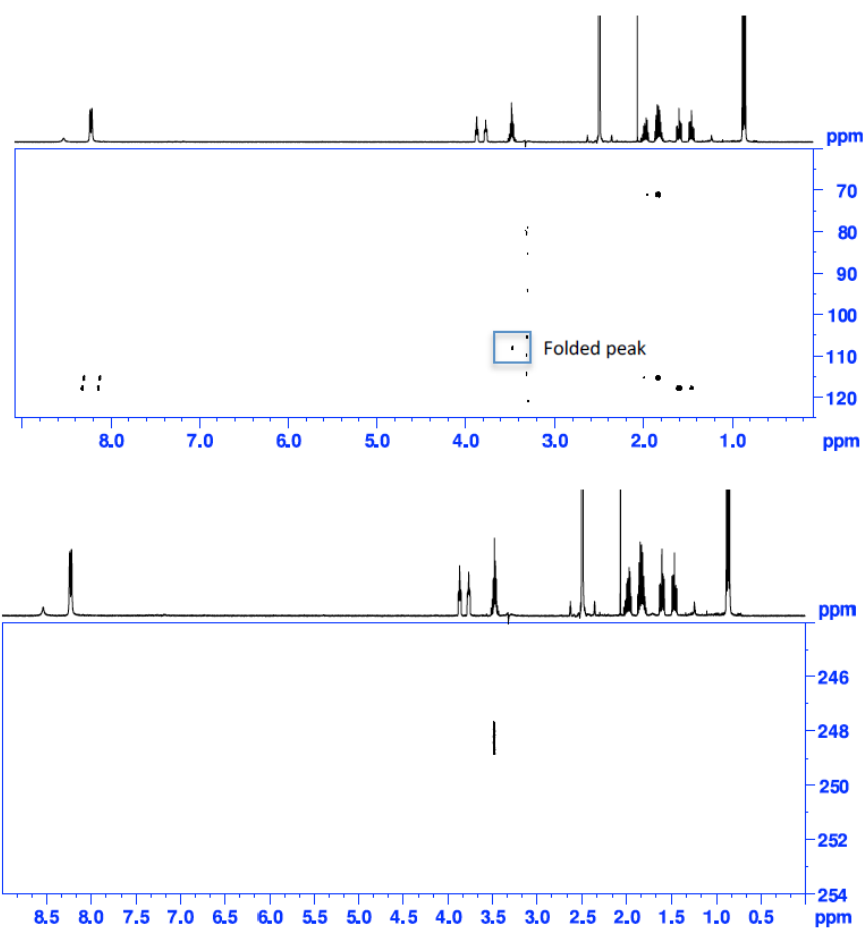


Figure S19 2D ^{15}N - ^1H HMBC spectra of cyclo(L-Azidohomoalanine-L-Leucine) in DMSO

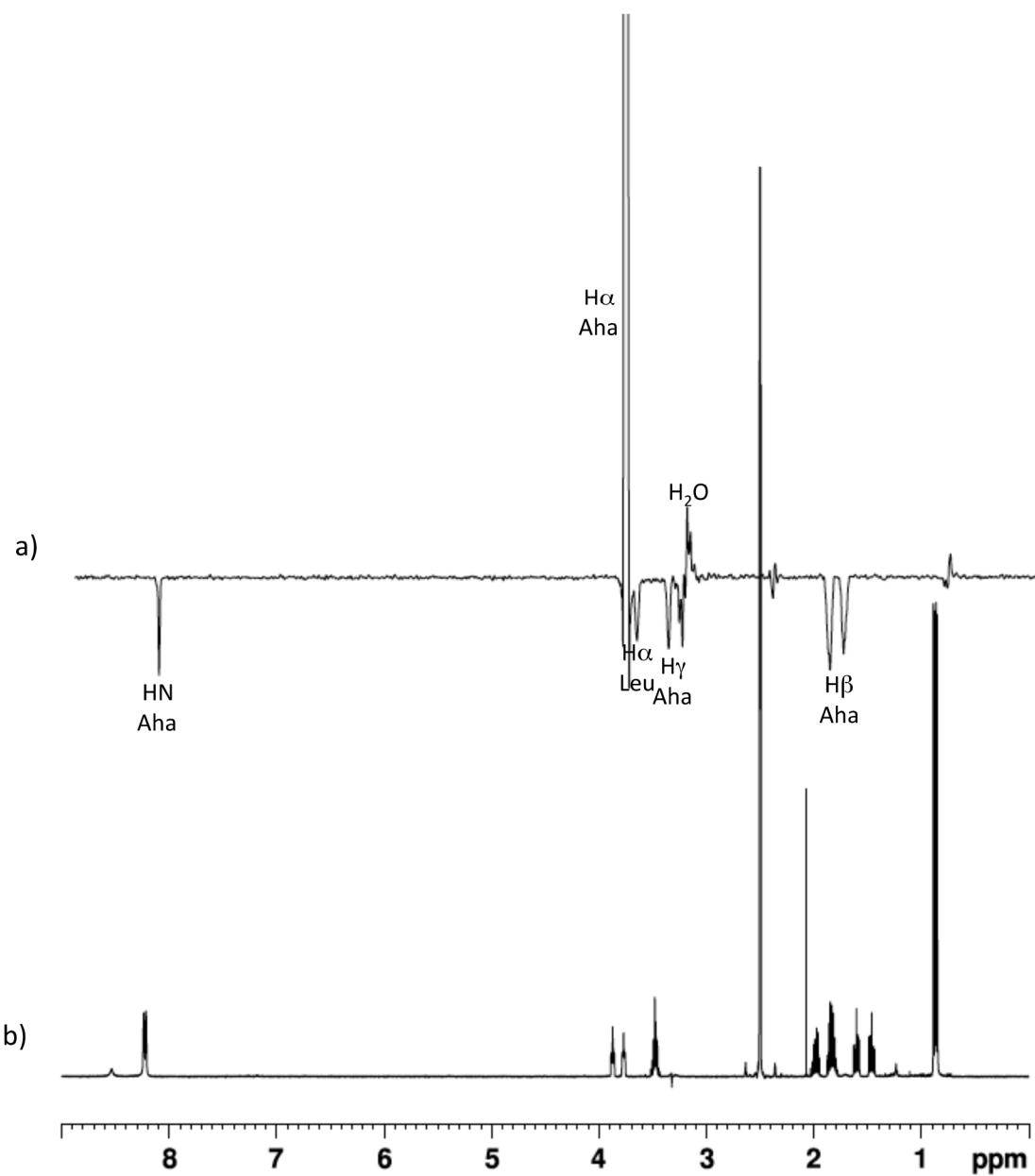


Figure S20 Stereochemical analysis of cyclo(L-Azidohomoalanine-L-Leucine) in DMSO. (a) 1D row (F1 δ =3.87 ppm) extracted from the 2D ¹H-¹H ROESY spectrum (600 ms mixing time), showing the spatial proximity of the H α protons of Aha and Leu residues. (b) 1D ¹H NMR spectrum.

References

- [1] M. Gondry, L. Sauguet, P. Belin, R. Thai, R. Amouroux, C. Tellier, K. Tuphile, M. Jacquet, S. Braud, M. Courçon, et al., *Nat. Chem. Biol.* **2009**, *5*, 414–420.
- [2] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, H. Mori, *Mol. Syst. Biol.* **2006**, *2*, 1–11.
- [3] F. Bertels, H. Merker, C. Kost, *PLoS One* **2012**, *7*, e41349.
- [4] I. B. Jacques, M. Moutiez, J. Witwinowski, E. Darbon, C. Martel, J. Seguin, E. Favry, R. Thai, A. Lecoq, S. Dubois, et al., *Nat. Chem. Biol.* **2015**, *11*, 721–731.
- [5] S. Lautru, M. Gondry, R. Genet, J.-L. Pernodet, *Chem. Biol.* **2002**, *9*, 1355–1364.
- [6] M. Gondry, I. B. Jacques, R. Thai, M. Babin, N. Canu, J. Seguin, P. Belin, J.-L. Pernodet, M. Moutiez, *Front. Microbiol.* **2018**, DOI 10.3389/fmicb.2018.00046.
- [7] K. Josephson, M. C. T. Hartman, J. W. Szostak, *J. Am. Chem. Soc.* **2005**, *127*, 11727–11735.
- [8] M. C. T. Hartman, K. Josephson, J. W. Szostak, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 4356–4361.
- [9] F. Oldach, R. Altoma, A. Kuthning, T. Caetano, S. Mendo, N. Budisa, R. D. Süßmuth, *Angew. Chemie - Int. Ed.* **2012**, *51*, 415–418.
- [10] D. D. Buechter, D. N. Paoella, B. S. Leslie, M. S. Brown, K. A. Mehos, E. A. Gruskin, *J. Biol. Chem.* **2003**, *278*, 645–650.
- [11] N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann, R. Huber, *Eur. J. Biochem.* **1995**, *230*, 788–796.
- [12] P. Wang, N. Vaidehi, D. A. Tirrell, W. A. Goddard, *J. Am. Chem. Soc.* **2002**, *124*, 14442–14449.
- [13] C. Fan, J. M. L. Ho, N. Chirathivat, D. Söll, Y. S. Wang, *ChemBioChem* **2014**, *15*, 1805–1809.
- [14] K. Kirshenbaum, I. S. Carrico, D. A. Tirrell, *ChemBioChem* **2002**, *3*, 235–237.
- [15] N. Budisa, C. Minks, F. J. Medrano, J. Lutz, R. Huber, L. Moroder, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 455–459.
- [16] N. Budisa, in *Eng. Genet. Code Expand. Amin. Acid Repert. Des. Nov. Proteins*, **2006**, pp. 68–141.
- [17] J. Achenbach, *Through the Mirror : Translation with D-Amino Acids*, **2015**.
- [18] J. Soutourina, P. Plateau, F. Delort, A. Peirottes, S. Blanquet, *J. Biol. Chem.* **1999**, *274*, 19109–19114.
- [19] J. Soutourina, P. Plateau, S. Blanquet, *J. Biol. Chem.* **2000**, *275*, 32535–32542.
- [20] O. Soutourina, J. Soutourina, S. Blanquet, P. Plateau, *J. Biol. Chem.* **2004**, *279*, 42560–42565.
- [21] T. W. Giessen, F. Altegoer, A. J. Nebel, R. M. Steinbach, G. Bange, M. A. Marahiel, *Angew. Chemie - Int. Ed.* **2015**, *54*, 2492–2496.
- [22] J. C. MacDonald, *Biochem. J.* **1965**, *96*, 533–8.
- [23] T. W. Giessen, A. M. Von Tesmar, M. A. Marahiel, *Chem. Biol.* **2013**, *20*, 828–838.
- [24] T. W. Giessen, A. M. Von Tesmar, M. A. Marahiel, *Biochemistry* **2013**, *52*, 4274–4283.
- [25] J. B. Patteson, W. Cai, R. A. Johnson, K. C. Santa Maria, B. Li, *Biochemistry* **2017**, acs.biochem.7b00943.
- [26] M. W. Vetting, S. S. Hegde, J. S. Blanchard, *Nat. Chem. Biol.* **2010**, *6*, 797–799.
- [27] L. Bonnefond, T. Arai, Y. Sakaguchi, T. Suzuki, R. Ishitani, O. Nureki, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3912–7.
- [28] L. Sauguet, M. Moutiez, Y. Li, P. Belin, J. Seguin, M. H. Le Du, R. Thai, C. Masson, M. Fonvielle, J. L. Pernodet, et al., *Nucleic Acids Res.* **2011**, *39*, 4475–4489.
- [29] M. Moutiez, E. Schmitt, J. Seguin, R. Thai, E. Favry, P. Belin, Y. Mechulam, M. Gondry, *Nat. Commun.* **2014**, *5*, 1–7.
- [30] M. Moutiez, P. Belin, M. Gondry, *Chem. Rev.* **2017**, *117*, 5578–5618.
- [31] I. A. Papayannopoulos, *Mass Spectrom. Rev.* **1995**, *14*, 49–73.
- [32] Y. H. Chen, S. E. Liou, C. C. Chen, *Eur. Food Res. Technol.* **2004**, *218*, 589–597.
- [33] J. Xing, Z. Yang, B. Lv, L. Xiang, *Rapid Commun. mass Spectrom.* **2008**, *22*, 1415–1422.
- [34] T. Stark, T. Hofmann, *J. Agric. Food Chem.* **2005**, *53*, 7222–7231.

- [35] A. J. C. Furtado, R. Vessecchi, J. C. Tomaz, S. Galembeck, J. Bastos, N. Lopes, A. Crotti, *J. mass Spectrom.* **2007**, *42*, 1279–1286.
- [36] X. Guo, X. Liu, J. Pan, H. Yang, *Sci. Rep.* **2015**, *5*, 1–12.
- [37] M. Fonvielle, M. H. Le Du, O. Lequin, A. Lecoq, M. Jacquet, R. Thai, S. Dubois, G. Grach, M. Gondry, P. Belin, *J. Biol. Chem.* **2013**, *288*, 17347–17359.

IV) Discussion and perspectives

The approach presented in this study was rather straightforward, but its high throughput allowed us to assess an important number of ncAAs and of CDPSs and thus to deliver interesting results. However, several questions remain unanswered and interesting follow-up studies could be considered.

Incorporation by CDPSs of ncAAs with modifications of the amino acid backbones

As noted in the manuscript, we did not observe incorporation for 36 out of the 60 ncAAs tested. In particular, no incorporation was observed for any of the tested ncAAs with modifications of the amino acid backbones (β -amino acids such as β -homophenylalanine, β -homotyrosine, β -homoleucine, β -homovaline, N-methyl amino acids such as N-methyl-tryptophan, N-methyl-leucine, N-methyl-valine, or α -hydroxy acid such as indole-3-lactic acid), whereas these ncAAs have all been reported to be substrates of their cognate AARSs. The utilisation of such backbone-modified amino acids by CDPSs would be of great interest for the expansion of the diversity accessible with these enzymes because it would result in molecules with interesting scaffolds, such as rings with seven atoms¹⁶⁶, N-methylated 2,5-DKPs or morpholine-2,5-diones¹⁶⁷. Since no control was implemented in our experiment to ensure that ncAAs were actually loaded on tRNAs and that subsequent ncAA-tRNAs were present in the cells at significant concentrations, we cannot deduce from our results that the absence of incorporation observed is due to the lack of CDPS activity on the corresponding ncAA-tRNAs. Further studies would be required, for example *in vitro* assays which would include a control of the actual acylation of ncAAs on tRNAs.

Several strategies to further increase the diversity of non-canonical 2,5-DKPs

First, about 40 ncAAs that have been shown to be substrate of *E. coli* AARSs were not tested in our study (see Table 1 below). Some of them are analogues of amino acids that are rare and/or poor substrates of CDPSs characterized so far (Arg, Lys, His, Asp, Gln). Others were not included because they were not commercially available or at prohibitive costs.

Besides the ncAAs listed in Table 1, another class of ncAAs was not tested in our assay: D-amino acids. Some *E. coli* AARSs (HisRS, LysRS, TrpRS, TyrRS, ValRS, AspRS) have been shown to

load the D-stereoisomer of their cognate amino acid on their cognate tRNAs^{168,169}. However, in order to avoid detrimental incorporation of D-amino acids into proteins by ribosomes, the misacylated D-AA-tRNAs are specifically recognized and deacylated by an enzyme, called D-Tyr-tRNA^{Tyr} deacylase (Dtd)¹⁷⁰. Therefore, in order to test the ability of CDPSs to use D-AA-tRNAs as substrates, it would be required to delete the gene coding for the Dtd in the genomes of the *E. coli* strains used. Double mutants bearing auxotrophy-causing gene-deletion and Dtd-deletion could be generated using phage transductions between two different Keio strains.














Another attractive approach to increase the chemical diversity accessible with CDPSs would be to combine our assay with the expression of modified AARSs having increased promiscuity (see section II.B.ii. of the introduction and Figure 12). The two variants of *E. coli* AARSs, A294G PheRS¹⁰⁹ and T222P ValRS¹¹², could be co-expressed with CDPSs and the new ncAAs they are able to load on their cognate tRNAs could be tested. Similarly, systems allowing the expression of AARS/tRNA orthogonal pairs could be combined with CDPS expression in auxotrophic strains. Given the diversity of CDPS specificity patterns involving aromatic amino acids, the TyrRS/tRNA^{Tyr} pair from *Methanococcus jannaschii* and the PylRS/tRNA^{Pyl} pair from *Methanosarcina barkeri* would be of particular interest, because they allow the loading on tRNAs of several dozens of aromatic ncAAs with diverse chemical properties¹²⁰.

Finally, a really interesting approach would be to combine our assay with the expression of 2,5-DKP tailoring enzymes. Indeed, if the chemical synthesis of most “simple” cyclodipeptides is not very challenging for organic chemists, the total synthesis of some complex 2,5-DKPs has proven to be difficult and only multi-steps and low yield synthetic routes have been proposed^{171,172}. Therefore, biosynthetic approaches have a real potential for the production of these molecules. The field of 2,5-DKP tailoring enzymes has been very active in the past few years and many new fascinating activities have been reported^{26,75}.

Table 1: NCAAs substrates of AARSs that were not included in the study

Name	Formula	AARS	Reference	Name	Formula	AARS	Reference
L-Canavanine		ArgRS	Hartman et al., 2006	L-Albizziine		GlnRS	Hartman et al., 2006
N-Methyl-L-lysine		LysRS	Hartman et al., 2006	L-Theanine		GlnRS	Hartman et al., 2006
<i>trans</i> -4,5-Dehydro-L-lysine		LysRS	Hartman et al., 2006	L-Glutamine-γ-hydrazide		GlnRS	Hartman et al., 2006
L-β-Homolysine		LysRS	Hartman et al., 2006	Methylether-L-threonine		ValRS	Hartman et al., 2006
N-Methyl-L-histidine		HisRS	Hartman et al., 2006	4-Thia-L-isoleucine		ValRS/ IleRS	Hartman et al., 2006
β-(2-Thiazolyl)-L-alanine		HisRS	Hartman et al., 2006	2-L-Thienylglycine		IleRS	Hartman et al., 2006
β-(1,2,4-Triazol-3-yl)-L-alanine		HisRS	Hartman et al., 2006	Acivicin		IleRS	Hartman et al., 2006
N-Methyl-L-aspartic acid		AspRS	Hartman et al., 2006	β-tert-Butyl-L-alanine		LeuRS	Hartman et al., 2006
4-Methyl-L-glutamate		GluRS	Hartman et al., 2006	Azaleucine		LeuRS	Hartman et al., 2006
4-Fluoro-L-glutamate		GluRS	Hartman et al., 2006	2-Hydroxy-L-leucine		LeuRS	Fan et al., 2014
4,4-Difluoro-L-glutamate		GluRS	Hartman et al., 2006	2-Amino-4-methyl-L-pentenoic acid		LeuRS	Connor et al., 2008 Fan et al., 2014
Quisqualic acid		GluRS	Hartman et al., 2006	S-2-Aminoethyl cysteine		LysRS	Hartman et al., 2006
Ibotenic acid		GluRS	Hartman et al., 2006	Homopropargylglycine		MetRS	Hartman et al., 2006 Oldach et al., 2012
Isoasparagine		AsnRS	Hartman et al., 2006	Trifluoronorleucine		MetRS	Hartman et al., 2006

Table S1 (continued)

Name	Formula	AARS	Source bibliographique
Trifluoronorvaline		MetRS	Hartman et al., 2006
Crotylglycine		MetRS	Hartman et al., 2006
L-b-Homométhionine		MetRS	Hartman et al., 2006
β -Cyclopropyl-L-alanine		MetRS	Hartman et al., 2006
3-(2-Pyridyl)-L-alanine		PheRS	Kirshenbaum et al., 2002
3-Amino-2-benzylpropanoic acid		PheRS	Hartman et al., 2006
β -Thienyl-L-serine		PheRS /SerRS	Hartman et al., 2006
3-Hydroxy-L-leucine		SerRS	Fan et al., 2014
β -Hydroxy-L-norvaline		ThrRS	Hartman et al., 2006 Minajigi et al., 2010
L-7-Azatriptophan		TrpRS	Hartman et al., 2006 Oldach et al., 2012
3-(Thianaphten-3-yl)-L-alanine		TrpRS	Hartman et al., 2006
2-amino-3-(1-formylindol-3-yl)-L-propanoic acid		TrpRS/ TyrRS	Fan et al., 2014
Methylether-L-threonine		TyrRS	Fan et al., 2014

CHAPTER 2: TOWARDS A BETTER UNDERSTANDING OF RNA RECOGNITION BY CYCLODIPEPTIDE SYNTHASES

I) Introduction

In the course of better understanding how CDPSs recognize the tRNA moieties of their substrates, AA-tRNA analogues are required but the diversity of tRNA mutants and tRNA analogues that can be aminoacylated using AARSs is limited. Thanks to their infinite versatility, flexizymes have been applied with a lot of success to the ribosomal incorporation of ncAAs into peptides, but they have not been yet used for the study of AA-tRNA-utilizing enzymes⁷⁶. We therefore decided to apply the flexizyme technology to the study of CDPSs.

This chapter is presented as a draft article that will be submitted for publication to *Nucleic Acids Research*. Note that, for reasons of time, several aminoacylated shortened tRNAs we produced could not have been tested as substrates in the final assays. These molecules are shaded in the manuscript. Missing experiments will be performed shortly to complete the manuscript.

Author contributions:

Muriel Gondry and I designed the project and the experimental strategy.

Carine Tellier-Lebègue produced *E. coli* AlaRS, T7 RNA polymerase and *Nbra*-CDPS. Jérôme Seguin produced *E. coli* GluRS. Carine Tellier-Lebègue, Morgan Babin and Mireille Moutiez performed the cloning of tRNA sequences into pBSTNAV. Carine Tellier-Lebègue, Jérôme Seguin and I produced *E. coli* tRNAs and performed tRNA aminoacylation by AARSs.

Chemical synthesis of activated amino acids and stable isotope-labelled cyclodipeptides was done under the supervision of Jean-Christophe Cintrat with advice from Robin Vinck. Olivier Cinquin optimised the synthesis protocol of cAE and cAA. I performed the chemical synthesis of activated amino acids and of stable isotope-labelled cAE and cAA.

I implemented and optimized the *in vitro* transcription of tRNA microhelices and flexizymes. I implemented the purification of transcripts and of aminoacylation reaction by anion-exchange chromatography, with assistance from Carine Tellier-Lebègue, Morgan Babin and Jérôme Seguin. Ines Ajel and Alexandre Couëtoux performed aminoacylation kinetics using flexizymes. Morgan Babin and I performed deacylation assays. Carine Tellier-Lebègue, Morgan Babin, Ines Ajel and I produced flexizymes, microhelices and aminoacylated shortened tRNAs. Robert Thai and I optimized and performed the analyses of RNAs by MALDI-TOF-MS.

I optimised the enzymatic assay for *Nbra*-CDPS, with advice of Robert Thai for the optimisation of LC-MS detection of cyclodipeptides. I performed all initial velocities determination. Morgan Babin and I performed the comparative assays of cyclodipeptide production for minimal substrates.

The manuscript was written by Muriel Gondry and me.

II) Article

Flexizyme-aminoacylated shortened tRNAs demonstrate that only acceptor arms of the two substrates are required for cyclodipeptide synthase activity

Nicolas Canu¹, Carine Tellier¹, , Morgan Babin¹, Robert Thai², Inès Ajel¹, Jérôme Seguin¹, Olivier Cinquin^{1,3}, Robin Vinck^{2,3}, Mireille Moutiez¹, Pascal Belin¹, Jean-Christophe Cintrat³ and Muriel Gondry^{1,*}

¹Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91198 Gif-sur-Yvette cedex, France, ²SIMOPRO, CEA-Saclay, 91198 Gif-sur-Yvette cedex, France, ³SCBM, CEA-Saclay, 91198 Gif-sur-Yvette cedex, France

*To whom correspondence should be addressed. Tel: +33 169 08 76 47; Email: muriel.gondry@i2bc.paris-saclay.fr

Present address: Jérôme Seguin, CEA, DEN, Centre de Marcoule, 30207 Bagnols-sur-Cèze, France

INTRODUCTION

Cyclodipeptide synthases (CDPSs) constitute a family of tRNA-dependent enzymes involved in the biosynthesis of 2,5-diketopiperazines (DKPs) (1), a class of microbial natural products with therapeutically promising bioactivities, such as antibacterial, antifungal or antitumoral (2, 3). CDPSs use two aminoacyl-tRNAs (AA-tRNAs), diverted from their canonical role in ribosomal protein synthesis, to catalyse the formation of various cyclodipeptides, the simplest representatives of 2,5- DKPs (4).

CDPSs, which are phylogenetically partitioned between two subfamilies named NYH and XYP (5, 6), have been extensively studied (for reviews, see (1, 7–9)). The catalytic mechanism used by AlbC from *Streptomyces noursei*, the first NYH CDPS member to be identified, has been elucidated (10–12). This mechanism begins with the binding of the first substrate, Phe-tRNA^{Phe}, followed by the transfer and the covalent attachment of its phenylalanyl moiety to a conserved serine residue, thus resulting in the formation of a phenylalanyl-enzyme intermediate. The second substrate, Leu-tRNA^{Leu}, interacts with the phenylalanyl-enzyme, on which its leucyl moiety is transferred, leading to a dipeptidyl-enzyme intermediate. Finally, the dipeptidyl moiety undergoes a cyclisation, which involves a conserved tyrosine residue acting as a proton relay, yielding the final cyclodipeptide cyclo(L-

Phe-L-Leu) (cFL). The crystallographic structures of AlbC alone or in complex with a dipeptidyl intermediate analogue and those of six other NYH or XYP CDPSs have been solved (10, 11, 13–15). Despite the fact that these CDPSs may display very low amino acid sequence identities (down to less than 10%), they share a common architecture with a monomer built around a Rossmann fold domain, which displays high structural similarity to the catalytic domain of homodimeric class-Ic aminoacyl-tRNA synthetases (AARSs). The catalytic residues conserved throughout the CDPS family have the same positioning in all structures, strongly suggesting that CDPSs, whatever their subfamily membership, employ similar mechanisms.

The two AA-tRNAs used successively as substrates by CDPSs are accommodated in different enzyme regions. CDPSs contain two solvent-accessible pockets, each accommodating the aminoacyl moiety of one of the substrates (11). One pocket, named P1, is structurally analogous to the aminoacyl binding pocket in class-Ic AARSs and the other, named P2, has no equivalent in these enzymes. Moreover, biochemical data with variants of NYH CDPSs indicated that the tRNA moiety of the first substrate interact with one patch of basic residues at the surface of a conserved helix (10, 13, 14, 16) while that of the second substrate may interact with few residues from another region whose equivalent is involved in tRNA binding in class-Ic AARSs (14, 16). This remains to be strengthened by additional data, in particular for XYP CDPSs, which do not contain the basic patch involved in the binding of the first tRNA moiety in NYH CDPSs (15).

Enzymological studies were previously conducted with two NYH CDPSs, Rv2275 from *Mycobacterium tuberculosis* and AlbC. However, they did not allow the estimation of catalytic efficiencies and affinities for the two substrates of these enzymes. Indeed, Rv2275 uses two identical AA-tRNAs to produce a cyclodipeptide composed of two identical amino acids, thus preventing the discriminative study of each of the two substrates (13). This drawback could have been overcome by the study of AlbC but the enzyme was shown during *in vitro* experiments to use more efficiently Phe-tRNA^{Phe} than any of the Leu-tRNA^{Leu} isoacceptors as the second substrate to produce mostly cFF (16). An attractive candidate to overcome the issue of CDPSs producing homocyclodipeptides is the recently identified XYP CDPS from *Nocardia brasiliensis*, *Nbra*-CDPS, because it uses Ala-tRNA^{Ala} and Glu-tRNA^{Glu} as substrates for the first and the second pocket respectively, thus synthesizing cAE as major product (5).

Nbra-CDPS is also an interesting candidate to determine, for each of the two substrates, the regions of tRNAs moieties which are essential for catalysis, a question that remains to be documented. In class-Ic AARSs, which have a conserved homodimeric quaternary structure, each monomer contains an anticodon-binding domain, involved in the recognition of the

anticodon loop of a tRNA that will be acylated by the catalytic domain of the other monomer (17). As CDPSs do not contain these anticodon-binding domains, are almost all monomeric (1) and use amino acids attached to the acceptor arm of tRNAs, they may not interact with the anticodon loop or other distal part of tRNAs. A previous study on AlbC using mutated tRNAs suggested that sequence motifs in the acceptor arm (particularly the first base pair N₁-N₇₂ of the second substrate) are required for CDPS activity (16). In order to further investigate the tRNA regions necessary for CDPS activity, one possible approach is the use of shortened tRNA analogues (minitrnAs), which have already been valuable molecules for the study of AARSs (18, 19). An attractive strategy to perform aminoacylation on diverse minitrnAs is to use flexizymes (20). Flexizymes are short AARS-like ribozymes that recognise their RNA acceptor by base-pairing with only three bases of their conserved 3' termini. Thus, they are able to accept a wide range of RNA acceptors as substrates: entire tRNAs, shortened tRNAs such as microhelices and even tetramer oligonucleotides mimicking the 3'-terminus tail of tRNA (20–22).

Here, we report an enzymological study of *Nbra*-CDPS towards diverse substrates. We implement an *in vitro* assay for *Nbra*-CDPS which allows to study the complex behaviour of the enzyme towards its two AA-tRNA substrates. We then introduce a new method for the production of aminoacylated minitrnAs (AA-minitrnAs), by combining the use of flexizymes with separation of the molecules by anion-exchange chromatography and report the production of several shortened analogues of AA-tRNAs. Finally, we study the activity of *Nbra*-CDPS on these shortened analogues and demonstrate that the acceptor arms of the two substrates are the only part of tRNA required for CDPS activity.

MATERIAL AND METHODS

Protein production

The his-tagged versions of T7 RNA polymerase, *E. coli* AlaRS, *E. coli* GluRS and *Nbra*-CDPs were expressed and purified according to previously published protocols (respectively (23), (24), (25) and (4)), which have been slightly modified (see detailed protocols in section I.A. of SI). Purified proteins were quantified by UV spectrophotometry using a DS-11 spectrophotometer (Denovix). Theoretical extinction coefficients were used for T7 RNA polymerase, *E. coli* AlaRS, *E. coli* GluRS. Experimental extinction coefficient of *Nbra*-CDPS was determined by acidic hydrolysis and amino acid analysis (26).

Production of *Escherichia coli* tRNAs

Plasmids allowing the constitutive expression of tRNA^{Ala}_{TGC}, tRNA^{Ala}_{GGC} and tRNA^{Glu} (see Table S1 for a summary of sequences of all RNAs used in the study) were constructed by cloning tRNA sequences into pBSTNAV2, a kind gift of Dr. Yves Mechulam (section I.B.ii. of SI and Table S2). tRNA production was performed using a protocol from the team of Dr. Yves Mechulam, which is detailed in (27). Briefly, cells overexpressing tRNAs were harvested by centrifugation and RNAs were phenol extracted. tRNAs were deacylated at pH 8.0 and 37°C for 90 min, precipitated with ethanol and purified on a 5 mL HiTrap Q HP anion-exchange chromatography column (GE Healthcare) with a linear gradient of 0.2 M to 1M NaCl. Fractions containing tRNAs were precipitated with ethanol and stored at -20°C. tRNAs and all other RNAs used in this study were quantified by UV spectrophotometry using a DS-11 spectrophotometer (Denovix), assuming that OD_{260nm} unit is equal to 40 ng/μL.

Aminoacylation of tRNAs

tRNAs were aminoacylated using the following conditions: 50 mM HEPES-KOH pH7.5, 20 mM KCl (only for tRNA^{Ala}), 15 mM MgCl₂, 1 mM or 10 mM DTT for tRNA^{Glu} or tRNA^{Ala} respectively, 4 mM ATP, 15 μM tRNAs, 2 μM corresponding *E. coli* AARS and 250 μM corresponding L-amino acid. Reactions were incubated at 37°C for 20 min and stopped by adding sodium acetate pH 5.0 at 100 mM final concentration. AA-tRNAs together with remaining non acylated tRNAs, were purified on a 5mL HiTrap Q HP anion-exchange chromatography column (GE Healthcare) using a linear gradient of 0 to 1 M NaCl in 10 mM sodium acetate pH 5.0. Fractions containing tRNAs were precipitated and stored as dried pellet at -20°C. Proportion of aminoacylated tRNAs were determined by enzymatic assays using high concentrations (500 nM) of *Nbra*-CDPS (section I.B.iii. of SI).

Production of flexizymes dFx and minitRNAs

Microhelices mimicking the acceptor arms of *E. coli* tRNAs are abbreviated as miHx^{Yyy-X}. Yyy refers to the isoacceptors which are mimicked (for example miHx^{Ala} reproduces the acceptor arm of tRNA^{Ala} isoacceptors) and X is the number of base pairs that is kept in the stem. MiHx^{Ala_4}, MiHx^{Ala_3}, miHx^{Glu_7}, miHx^{Glu_6}, miHx^{Glu_5}, miHx^{Glu_4}, miHx^{Glu_3} and tetramer oligos ACCA and GCCA were synthesized chemically by Eurogentec. MiHx^{Ala_7}, miHx^{Ala_6}, miHx^{Ala_5} and flexizymes (dFxs) were synthesized by *in vitro* transcription and then purified by anion-exchange chromatography. Their syntheses were carried out using a protocol from the team of Pr. Suga (28), with small modifications to avoid the formation of template-independent

transcripts (29). DNA matrices were obtained by mixing T7_F primer and the corresponding reverse primer (Figure S1 and Table S3) at 10 μ M in 35 mM KCl, heating at 95°C for 3 min then letting cooling to room temperature for 5 min. *In vitro* transcription of dFxs was performed in 5 mL reactions containing 40 mM Tris-HCl pH 8.0, 1 mM spermidine, 0.01% (v/v) Triton-X100, 10 mM DTT, 75 mM NaCl, 30 mM MgCl₂, 30 mM KOH, 5 mM of each NTP (New England Biolabs), 250 nM DNA matrix and 10 μ g/mL T7 RNA polymerase purified in our laboratory. The same protocol was used for transcription of miHxs, except that MgCl₂ concentration was 22.5 mM and that 10 mM GMP was added, in order to obtain 5' mono-phosphate ends that mimic native tRNAs (30). After incubation at 37°C for 14 h, DNA matrices were digested by addition of 10 U/mL DNase 1 (New England Biolabs) and 2 mM MnCl₂ and incubation at 37°C for 1 h. Transcripts were phenol extracted, precipitated with ethanol and solubilized in 10 mM MgCl₂. Unincorporated nucleotides and small aborted transcripts were removed by washing the transcripts three times with 10 mM MgCl₂ over Vivaspin centrifugation filters (3kDa cut-off, GE Healthcare) with monitoring of the OD₂₆₀ of the flow-through. Transcripts were renatured by heating to 80°C for 5 min then rapidly decreasing temperature to 4°C before purification by HPLC on an Elite LaChrom HPLC system (VWR). Samples were loaded onto a DNAPAC-PA100 semi-preparative column (9 mm x 250 mm, 13.5 μ m, ThermoFisher Scientific). Separations were performed under non-denaturing conditions with increasing linear gradients made from buffer A (25 mM ammonium acetate pH 5.2, 0.5% acetonitrile) and buffer B (2.5 M ammonium acetate pH 5.2, 0.5% acetonitrile) at a flow rate of 5 mL/min (31). Different gradients were optimised according to the transcripts to be purified (Table S4). Fractions of interest were desalted to 25 mM ammonium acetate pH 5.2 using a Hi-Trap 26-10 desalting column (GE Healthcare), lyophilised and stored at -20°C. Purified transcripts were heat-renatured using the same protocol as described above and analysed using a DNAPAC-PA100 analytical column (4 mm x 250 mm, 13.5 μ m, ThermoFisher Scientific) with the same gradients at a flow rate of 1 ml/min.

Production of AA-minitrRNAs using flexizymes

MinitRNAs were aminoacylated by dFxs using 3,5-dinitrobenzyl esters (DBE) of L-alanine or L-glutamate (Figure S2). These chemically-activated amino acids were synthesized according to standard procedures (21, 32). Standard aminoacylation conditions were used (28): 25 μ M minitRNA and 37.5 μ M of its corresponding dFx were mixed in 50 mM HEPES-KOH pH 7.5 buffer and were heated at 95°C for 2 min, then slowly cooled down at room temperature for 5 min. 600 mM MgCl₂ was added, solution was incubated at room temperature for 5 min then

on ice for 3 min. Reaction was initiated by addition of 5 mM of the corresponding DBE-activated amino acid and incubated on ice for 1h45 for DBE-activated alanine and 3h for DBE-activated glutamate. Reactions were quenched by sodium acetate pH 5.2, ethanol precipitated and stored as a pellet. Pellets were dissolved in 20 mM sodium acetate pH 5.2 and RNAs were renatured by applying a gentle heat/quick cool protocol: heating to 60°C for 30 s then rapidly decreasing temperature to 5°C. All samples were purified and analysed on DNAPAC-PA100 columns as described above, using gradients reported in Table S4.

Enzymatic assays

Substrates to be tested, AA-tRNAs and AA-minitrRNAs, in dried pellets were solubilized in 20 mM sodium acetate pH 5.2 shortly before use and kept on ice during all experiments to limit deacylation. Enzymatic assays were performed in 100 mM phosphate-sodium buffer pH 7.5, 50 mM KCl, 15 mM MgCl₂ and 0.1 mM β-mercaptoethanol. Substrates were added to the solution, incubated 1min at 20°C and reactions were initiated by addition of 5nM *Nbra*-CDPS. Aliquots were withdrawn after specific times (see below) acidified with 2% TFA to stop the reactions and mixed with known concentrations of stable isotope internal standards to quantify the cAE and cAA produced by LC-MS/MS analyses, as previously described (4, 16) (see section I.C. of SI). Stable isotope-labelled standards cAE and cAA, each containing one ¹³C₃-¹⁵N-L-alanine, were chemically synthesized (see section I.C.i. of SI). LC-MS analyses were performed on a Elute SP HPLC chain (Bruker Daltonik GmbH) coupled *via* a split system to an amazon SL ion trap mass spectrometer (Bruker Daltonik GmbH) set in positive mode. Samples were injected onto an Hypercarb column (2 mm × 150 mm, 3 μm, 100 Å, ACE), at 0.2 mL/min, with a linear gradient from 2 to 42% solvent B at 0.2 mL/min over 20 min (solvent A: 0.1% (v/v) formic acid in H₂O; solvent B: 0.1% (v/v) formic acid in acetonitrile/H₂O (90/10)) and concentrations of cyclodipeptides were calculated using the ratio of the EIC peak area of the analyte and of the standard (section I.C.ii. of SI). For initial velocities determination, three aliquots were taken in the first two minutes of incubation and initial velocities were calculated by plotting the concentrations obtained at the three different sampling times. For endpoint assays, the cyclodipeptides were quantified after 30 min-incubation with the appropriate substrates.

RESULTS

Implementation of an *in vitro* assay

Choice of enzyme and substrates More than 100 CDPSs have been characterized for the cyclodipeptides they produce when heterologously expressed in *E. coli* (7). Most CDPSs produce cyclodipeptides composed of two hydrophobic residues and often produce significant amount of homocyclodipeptides, which prevents the discrimination of the use of substrates at the two binding sites. *Nbra*-CDPS stands out from other CDPSs because it incorporates the hydrophobic Ala and the acidic Glu, producing cAE, only about 2% cAA and no cEE (5). Protein sequence analyses have established that the aminoacyl moiety of Ala-tRNA^{Ala} is accommodated in P1 whereas that of Glu-tRNA^{Glu} is accommodated in the second pocket P2 (5, 6). Despite the fact that Ala-tRNA^{Ala} is also poorly used as a second substrate, as demonstrated by the low amount of cAA produced during *in vivo* experiments in *E. coli*, *Nbra*-CDPS appears to be one of the most appropriate candidates to perform the discriminative study of the use of the two substrates.

Since *Nbra*-CDPS is active in *E. coli*, *E. coli* tRNAs were used as substrates for *in vitro* cyclodipeptide-production assays, as previously done for other CDPSs (4, 13, 16). *E. coli* genome encodes one tRNA^{Glu} sequence and two tRNA^{Ala} isoacceptors, tRNA^{Ala}_{TGC} and tRNA^{Ala}_{GGC}, with only six nucleotides mutations between both alanine isoacceptors (Figure 1).

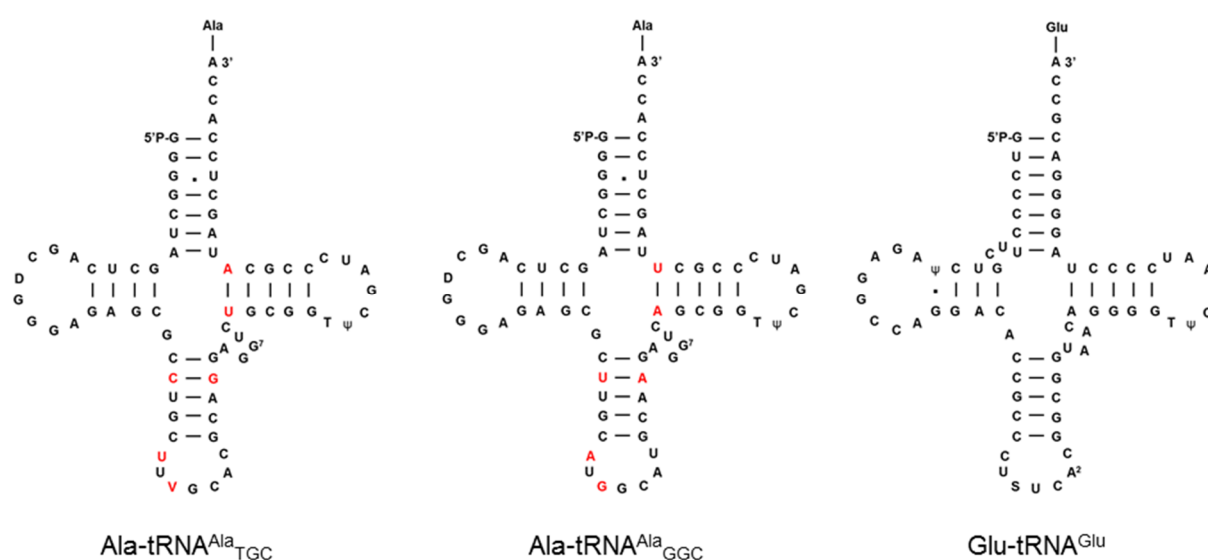


Figure 1: Secondary structures of *E. coli* aminoacylated tRNAs^{Ala} and tRNA^{Glu}. Differences between the two tRNA^{Ala} isoacceptors are in red. Post-transcriptional modifications were taken from the MODOMICS database (33) and are abbreviated as follow: D: dihydrouridine; G⁷: 7-methylguanosine; V: uridine 5-oxyacetic acid; T: 5-methyluridine; ψ : pseudouridine; S: 5-methylaminomethyl-2-thiouridine; A²: 2-methyladenosine.

The three corresponding *E. coli* AA-tRNAs were assayed as substrates for *Nbra*-CDPS. Since preliminary results suggested that *Nbra*-CDPS had similar activities whatever the tRNA^{Ala} isoacceptor used, we optimized the *in vitro* assay with Ala-tRNA^{Ala}_{TGC} before performing accurate analyses with the two isoacceptors.

Principle and optimisation of the in vitro assay Previous enzymological studies of CDPSs used CDPS-AARS coupled assays, in which *E. coli* AA-tRNAs are generated *in situ* by *E. coli* AARSs (10, 16). However, such approach cannot be applied for the test of shortened tRNA analogues which, for most of them, are not substrates of AARSs. We therefore developed a direct assay, in which tRNAs and their shortened analogues are aminoacylated either by AARSs or flexizymes and purified before being tested as substrates for cyclodipeptide production by *NBra*-CDPS. Careful optimization of the different components and parameters of the assay was necessary in order to maximize the cyclodipeptide-synthesizing activity of *Nbra*-CDPS and to ensure that the spontaneous hydrolysis of the aminoacyl-linkage to tRNAs of the substrates could be neglected. Nature of buffer was shown to strongly influence *Nbra*-CDPS activity. The two buffers previously used in enzymatic assays, Tris-HCl and HEPES-KOH (10, 16), strongly inhibit *Nbra*-CDPS activity (Figure S6). Note that the binding of HEPES into the P1 pocket of *Nbra*-CDPS, as previously observed for CHES and CAPSO in the crystal structure of the CDPS YvmC (14), might explain the inhibition observed. Sodium phosphate was chosen as a more suitable buffer. *Nbra*-CDPS activity significantly increased with increasing pH values from 7.00 to 7.75 (Figure S7A). However, in order to limit the spontaneous deacylation of the substrates, which is favoured with increasing pH values (34), pH of the assay was set at 7.50. Since AA-tRNA deacylation is also favoured by increasing temperature (35) and *Nbra*-CDPS had similar activities between 20°C and 30°C (Figure S7B), temperature was set at 20°C. Finally, a concentration of 5 nM *Nbra*-CDPS was shown to allow to conveniently measure initial velocities (Figure S8). Maximal duration of the assay for initial velocities determination was 2 min. Under these conditions, deacylation of the substrates was estimated to be at most 2.3% and 1% for Ala and Glu, respectively (Table S7).

Discriminative study of the two substrates of *Nbra*-CDPS

We first intended to characterise the behaviour of the first substrate of *Nbra*-CDPS, Ala-tRNA^{Ala}, by measuring initial velocities at a fixed concentration of Glu-tRNA^{Glu} and varying concentrations of Ala-tRNA^{Ala}. We obtained similar results for both Ala-tRNA^{Ala} isoacceptors (Figure S9), which suggests that the six nucleotides differing in the T ψ C stem and the anticodon stem-loop (Figure 1) are not essential for substrate recognition. Subsequent experiments were made using Ala-tRNA^{Ala}_{TGC}. We measured cyclodipeptide production using 600 nM Glu-tRNA^{Glu} and Ala-tRNA^{Ala}_{TGC} concentrations ranging from 300 to 12000 nM (Figure 2A and Figure S10). Up to about 4800 nM Ala-tRNA^{Ala}_{TGC}, initial velocities of cAE production increased with increasing substrate concentrations and almost none cAA production was detected. At higher Ala-tRNA^{Ala}_{TGC} concentrations, saturable velocities for cAE production seemed to be obtained before decreasing with increasing substrate concentrations. Concomitantly, low amount of cAA production were detected and increased with increasing substrate concentrations.

Similar experiments were performed to characterize the second substrate, with Ala-tRNA^{Ala}_{TGC} as the constant substrate and Glu-tRNA^{Glu} as the variable one, but saturable velocities for cAE production were observed, whatever the Glu-tRNA^{Glu} concentration used. We therefore tested lower concentrations of Glu-tRNA^{Glu} and increased the fixed concentration of Ala-tRNA^{Ala}_{TGC} to 1200 nM to increase cyclodipeptide production and

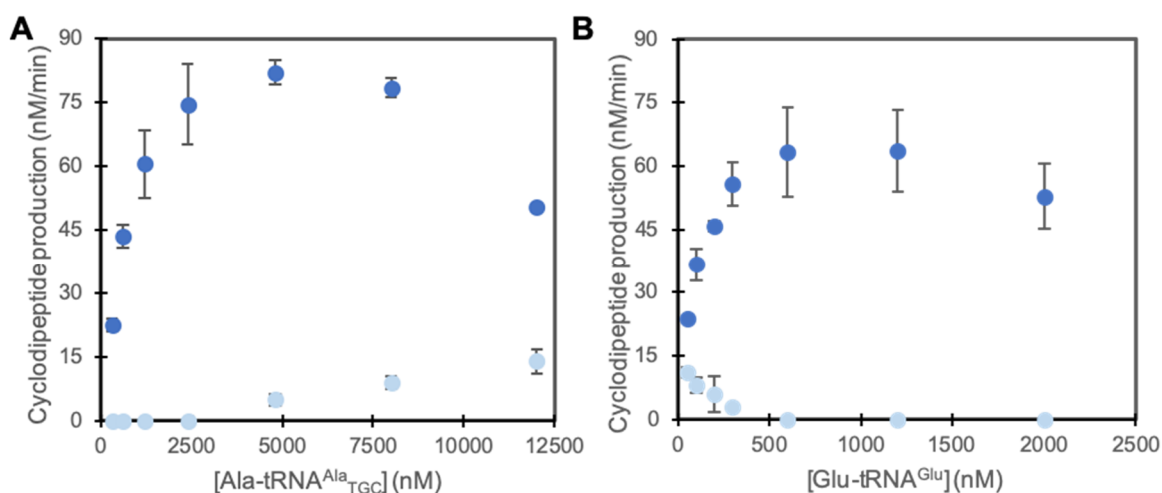


Figure 2: Kinetics study of *Nbra*-CDPS towards its AA-tRNA substrates. Initial velocities of cAE (dark blue) and cAA (light blue) production were measured with A) 600 nM Glu-tRNA^{Glu} and varying concentrations of Ala-tRNA^{Ala}_{TGC} and B) 1200 nM Ala-tRNA^{Ala}_{TGC} and varying concentrations of Glu-tRNA^{Glu}. Error bars correspond to the standard errors between duplicates. Cyclodipeptide titration curves used for initial velocities determination are given in Figure S10.

facilitate the detection (Figure 2B and Figure S10). Up to about 600 nM Glu-tRNA^{Glu}, initial velocities for cAE production increased and those for cAA production decreased. At higher Glu-tRNA^{Glu} concentrations, saturable then decreasing velocities were obtained for cAE production and cAA production was no longer detected.

Taken together, the results clearly demonstrate that competition and/or inhibition occur at the two binding sites. When Ala-tRNA^{Ala} is present in large excess, it is used instead of Glu-tRNA^{Glu} as second substrate. However, cAA production requires a large molar excess of Ala-tRNA^{Ala} compared to Glu-tRNA^{Glu} (at least 4-fold) and remains very low compared to cAE production (Figure 2), which confirms that Glu-tRNA^{Glu} is highly preferred by *Nbra*-CDPS as second substrate. When Glu-tRNA^{Glu} is present in large excess, a decrease in cAE production is observed, suggesting that Glu-tRNA^{Glu} might act as an inhibitor at the first or its own binding site (Figure 2B). These results reflect the complexity of the reaction catalysed by *Nbra*-CDPS, a multisubstrate and multiproduct reaction that prevents any accurate determination of kinetic parameters. However, the results seem to indicate that the affinity of *Nbra*-CDPS towards its two substrates differ significantly: maximal initial velocities are reached at about 5000 nM for Ala-tRNA^{Ala}_{TGC} and about 600 nM for Glu-tRNA^{Glu}. In both cases, the “pre-saturation” increase of initial velocities could be fitted in a Michaelis-Menten equation, yielding apparent K_M values of 805 ± 120 nM for Ala-tRNA^{Ala}_{TGC} and 110 ± 10 nM for Glu-tRNA^{Glu} (Figure S11). This suggests that the affinity of the second binding site towards Glu-tRNA^{Glu} is higher than that of the first binding site towards Ala-tRNA^{Ala}_{TGC}.

This study using entire AA-tRNAs confirms that *Nbra*-CDPS is a relevant model for the discriminative study of the use of its two substrates and sets a reference for the comparative study of shortened substrates.

Production of aminoacylated shortened tRNA analogues using flexizymes

In order to identify the regions of tRNA^{Ala} and tRNA^{Glu} required for *Nbra*-CDPS activity, we designed a set of shortened tRNA analogues (minitrRNAs) (Figure 3). This set contains tRNA miHxs, stem-loop RNAs built with stems corresponding to the acceptor arms of the tRNAs closed by the tetramer loop 5'-UUCG-3', which is known to favour stability of short stems (36, 37). Note that since the two tRNA^{Ala} isoacceptors of *E. coli* have similar acceptor arms they were mimicked by the same miHxs^{Ala}. The set contains miHxs with stems of different lengths, from 7bp down to 3 bp. However, miHx^{Ala-3} was excluded from the study because the corresponding hairpin conformation was not predicted to be stable (Table S5). The set also

contains RNA tetramers mimicking the 3' NCCA tail of the tRNAs. All minitRNAs were obtained either by *in vitro* transcription or chemical synthesis (Table S1).

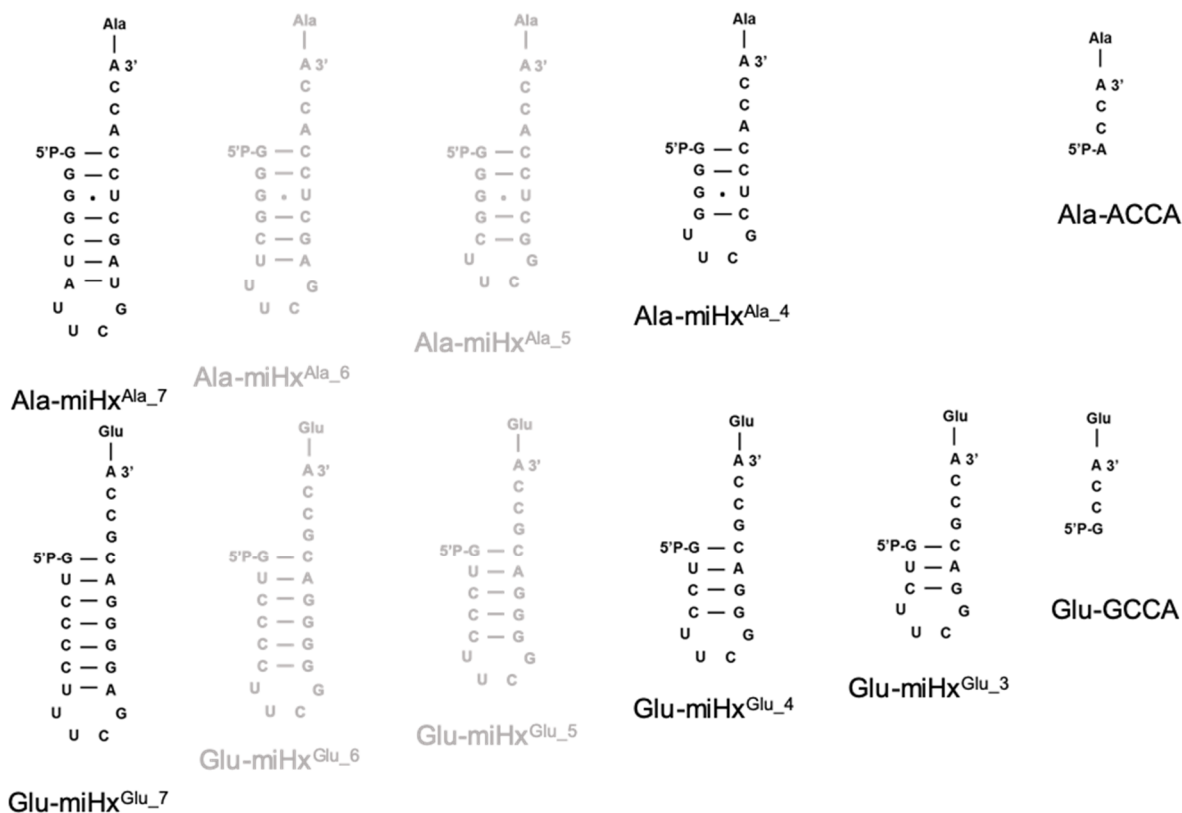


Figure 3: Secondary structures of minitRNAs used in the study. *Shaded miHxs remain to be tested.*

The whole procedure for production of aminoacylated minitRNAs was developed with Ala-miHx^{Ala_7}. MiHx^{Ala_7} was synthesized by *in vitro* transcription, purified by anion-exchange chromatography using a semi-preparative DNAPAC column, desalted and lyophilised (see the Material and Methods section). Anion-exchange chromatography was performed under non-denaturing conditions and miHx^{Ala_7}, as many small RNA hairpins, can fold into different intra- and inter-molecular conformations (Table S5). In order to favour the most stable intramolecular conformation, we applied heat and quick cool renaturation protocol to miHx^{Ala_7} samples before purification and analysis (38). This renaturation protocol had a dramatic effect on HPLC profiles: for example, the number of peaks observed after injection of purified miHx^{Ala_7} was reduced from two to one upon renaturation (Figure S13). Alanylation of miHx^{Ala_7} was performed using dFx under standard conditions and yielded about 60% alanylation, as demonstrated by acid-PAGE analysis (Figure S14A). To separate and purify Ala-miHx^{Ala_7} from residual miHx^{Ala_7}, we developed a new procedure based on the use of the DNAPAC column. We assumed that the shift in the charge upon RNA aminoacylation (the amine function of amino acid backbones is positively charged at neutral or slightly acidic pH)

could induce a different behaviour during anion exchange chromatography. By comparing chromatogram obtained for the alanylation reaction of miHx^{Ala-7} to the chromatogram obtained for purified miHx^{Ala-7}, we observed a new peak with an earlier retention time (Figure S14B). The corresponding eluted compound was purified and its characterization by acid-PAGE analysis and MALDI-TOF-MS demonstrated that it was Ala-miHx^{Ala-7} (Figure S14A,C). Figure 4 sums up the different purification steps from raw *in vitro* transcription reaction to purified Ala-miHx^{Ala-7}. Deacylation of Ala-miHx^{Ala-7} occurred during its purification but could be limited to less than 20%, as shown by the ratio between the area of the peaks corresponding to Ala-miHx^{Ala-7} and miHx^{Ala-7} in purified Ala-miHx^{Ala-7} (Figure 4D). Aminoacylation reaction and purification could easily be scaled-up and allowed to obtain several dozens of μg of Ala-miHx^{Ala-7}. Interestingly, the protocol of purification also allowed the separation and purification of flexizymes from the reaction (Figure S14B). Flexizymes could be purified and reused several times with no loss of efficiency.

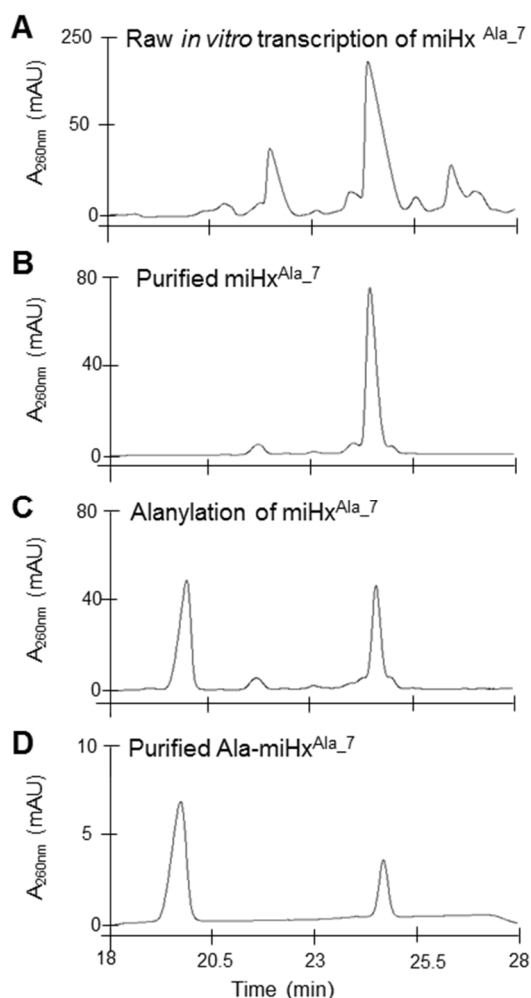


Figure 4: Purification of Ala-miHx^{Ala-7}. Chromatograms corresponding to the injection on DNAPAC of A) 20 μg raw *in vitro* transcription of miHx^{Ala-7}, B) 1 μg purified miHx^{Ala-7}, C) alanylation of 1 μg miHx^{Ala-7} by dFx and D) 0.25 μg purified Ala-miHx^{Ala-7}.

We performed a similar experiment for the production of Glu-miHx^{Glu-7}. Chemically synthesized miHx^{Glu-7} was renatured then glutamylated using dFx. Aminoacylation reaction was analysed onto the DNAPAC column. The chromatogram was similar to the one previously obtained for Ala-miHx^{Ala-7}, with the presence of an additional peak compared to the chromatogram of miHx^{Glu-7}, with an earlier retention time than the peak corresponding to miHx^{Glu-7} (Figure S15A). The eluted compound was confirmed to be Glu-miHx^{Glu-7} by MALDI-TOF (Figure S15B). Albeit the separation was done using similar gradients, glutamylation of miHx^{Glu-7} resulted in a lower shift in retention time (about 2.8 min) than alanylation of miHx^{Ala-7} (about 5 min). This might be due to the negative charge of the side chain of glutamate at pH 5.5 which partly compensates the effect of the additional positive charge brought by the backbone amine group.

Similar experiments were performed with the remaining minitRNAs, which were all shown to be retained on the DNAPAC column. The shorter a minitRNA was, the earlier was its retention times (Figures S16), except for miHx^{Ala-3} which exhibited an unexpected behaviour most certainly due to intermolecular pairings shown to be resistant to several renaturation protocols. Anion-exchange chromatography was used to analyse the flexizyme aminoacylation reactions for all minitRNAs. Additional peaks corresponding to AA-minitRNAs were observed for all minitRNAs, except for miHx^{Ala-3}. Aminoacylation yields were determined easily by calculating the ratios of peak areas of acylated and non-acylated forms and were about 60% for minitRNAs^{Ala} and 45% for minitRNAs^{Glu}, consistent with the results of Suga and coll. (21, 22). Scale-up of the reactions allowed us to obtain several nmoles of AA-minitRNAs and identity of the products was confirmed by MALDI-TOF (Figure S17). Deacylation during the purification process did not exceed 20% for alanylated minitRNAs and 10% for glutamylated minitRNAs.

Anion-exchange chromatographic analysis of AA-minitRNAs also allowed us to perform time courses of aminoacylation by dFxs and determine the optimum reaction times of 1 h 45 and 3 h for minitRNA^{Ala} and minitRNA^{Glu}, respectively (Figure S18). We also applied it to study the spontaneous deacylation of Ala-miHx^{Ala-7} and Glu-miHx^{Glu-7} in the conditions of the cyclodipeptide-production assay and determined the half-life of these products to be 60 and 145 min, respectively (Figure S19 and Table S7), which is in line with previous reports showing that alanylated-tRNAs are among the least stable AA-tRNAs (32). Based on these experiments, we estimated that spontaneous deacylation for respectively Ala-minitRNA^{Ala} and Glu-minitRNA^{Glu} was limited to 2.3% and 1.0% after 2 min (maximal duration of assays for which

initial velocities were calculated) but can reach 29.3% and 13.3% after 30 min (duration of the assay for characterization of poor substrates) (Table S7).

Aminoacylated minitRNAs as substrates for *Nbra*-CDPS

We first tested the use by *Nbra*-CDPS of Ala-miHx^{Ala-7} and Glu-miHx^{Glu-7}, which mimic the entire acceptor stems of tRNA^{Ala} and tRNA^{Glu}, by using a similar approach as for AA-tRNA substrates (*i.e.* by measuring initial velocities of cyclodipeptide production at increasing concentrations of Ala-miHx^{Ala-7} in the presence of a fixed concentration of 600 nM Glu-tRNA^{Glu} or at increasing concentrations of Glu-miHx^{Glu-7} in the presence of a fixed concentration of 1200 nM Ala-tRNA^{Ala-TGC}). The results were compared to those obtained with the corresponding AA-tRNAs (Figure 5A and 5B). With Ala-miHx^{Ala-7}, they were very close to the ones obtained with Ala-tRNA^{Ala-TGC} (Figure 5A), indicating that Ala-miHx^{Ala-7} is as good a substrate as Ala-tRNA^{Ala}. With Glu-miHx^{Glu-7}, they were also close to those observed with Glu-tRNA^{Glu}, except that the production rates of cAE seemed slightly faster (Figure 5B). This suggests that Glu-miHx^{Glu-7} is as good, or even better substrate than Glu-tRNA^{Glu} at the P2 pocket. When used at high concentrations, Glu-miHx^{Glu-7} inhibits the cAE production even more than Glu-tRNA^{Glu}, suggesting that Glu-miHx^{Glu-7} might be a better inhibitor of Glu-tRNA^{Glu} at the first or its own binding site.

We then performed similar experiments except that AA-miHxs were replaced with aminoacylated tetramer oligonucleotides mimicking the 3' NCCA of tRNAs (Figure 3). No cAE production was detected either with Ala-ACCA and Glu-tRNA^{Glu} or with Glu-GCCA and Ala-tRNA^{Ala-TGC} (Figure 5A and 5B). These results clearly showed that the 3' tails of the tRNAs are not sufficient to be used as substrates by *Nbra*-CDPS. However, the *in vitro* assay is not suitable to characterize very poor substrates because of the detection threshold of cyclodipeptide production (section I.C.ii of SI). We therefore used an endpoint assay, in which the cyclodipeptide production was quantified after a 30 min-incubation, using 600 nM of the minitRNAs to be tested and 600 nM Glu-tRNA^{Glu} or 1200 nM Ala-tRNA^{Ala-TGC}. No cAE production was detected with the minitRNAs Ala-ACCA or Glu-GCCA, suggesting that they are not substrates of *Nbra*-CDPS (Figure 5C and 5D).

Finally, we used the endpoint assay to evaluate and compare the activity of *Nbra*-CDPS towards miHxs with decreasing stem lengths. Ala-miHx^{Ala-4} was shown to be substrate of

Nbra-CDPS but was by far a less good substrate than Ala-miHx^{Ala_7} (Figure 5C). [*Ala-miHx*^{Ala_5} and *Ala-miHx*^{Ala_6} remained to be tested during the writing of the manuscript.] Both Glu-miHx^{Glu_3} and Glu-miHx^{Glu_4} were shown to be substrates of *Nbra*-CDPS, but cAE production was lower than with Glu-miHx^{Glu_7} (Figure 5D). The amount of cAE produced with Glu-miHx^{Glu_4} was higher than with Glu-miHx^{Glu_3}, which suggests that Glu-miHx^{Glu_3} is a poorer

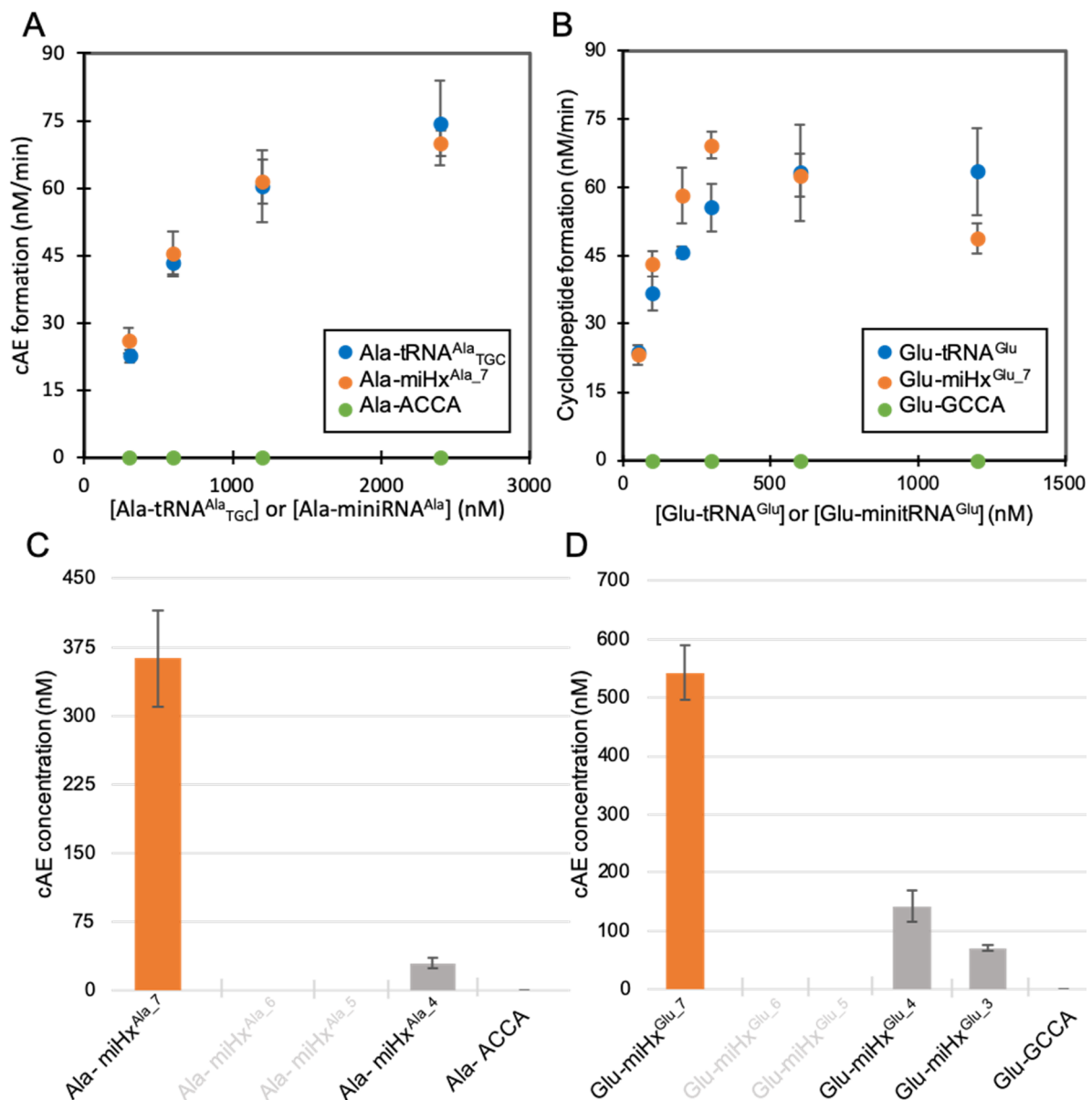


Figure 5: Kinetics study of *Nbra*-CDPS towards AA-minitRNA substrates. Initial velocities of cAE were measured with A) 600 nM Glu-tRNA^{Glu} and varying concentrations of Ala-minitRNA^{Ala} and B) 1200 nM Ala-tRNA^{Ala}_{TGC} and varying concentrations of Glu-minitRNA^{Glu}. End-point assays were performed with C) 600 nM Glu-tRNA^{Glu} and varying concentrations of Ala-minitRNA^{Ala} and D) 1200 nM Ala-tRNA^{Ala}_{TGC} and varying concentrations of Glu-minitRNA^{Glu}. Error bars correspond to the standard errors between duplicates. For clarity reasons, cAA production were not indicated on Figure 5. Data with cAA are presented on Figure S12.

substrate than Glu-miHx^{Glu-4} for *Nbra*-CDPS. [*Glu-miHx^{Glu-5}* and *Glu-miHx^{Glu-6}* remain to be tested].

DISCUSSION

Here, we showed that *Nbra*-CDPS is a relevant model for the discriminative study of the use of the two substrates at the two binding sites of CDPSs. Accurate determination of kinetic parameters was not possible due to the apparent cross-inhibition or cross-competition of the substrates and the probable conformational changes of the CDPS between its free, aminoacyl and dipeptidyl states. However, our data suggest that the affinities of *Nbra*-CDPS for its two substrates are significantly different: the affinity of the first binding site towards Ala-tRNA^{Ala} appears to be lower than that of the second binding site towards Glu-tRNA^{Glu}.

Direct *in vitro* assays offer the possibility to investigate the use of various substrates for each of the two binding sites. We used it to identify the regions of the two tRNAs required for CDPS activity. Our primary interest was to determine the role of the acceptor arms of tRNAs, as we previously showed that sequence motifs in these regions could be essential to retain the activity of the CDPS AlbC (16). We demonstrated that AA-miHxs, which mimic the entire 7 bp stems of tRNA^{Ala} and tRNA^{Glu} acceptor arms, are as good substrates of *Nbra*-CDPS as the corresponding AA-tRNAs. On the other hand, no activity was observed when the substrates were reduced to the aminoacylated 3'-terminus tails. This shows that tRNA acceptor arms are required at the two binding sites for CDPS activity whereas the three stem-loops of tRNAs are fully dispensable. In order to further explore the minimal substrates of CDPSs, we then planned to test AA-miHxs with shorter stems. The first experiments showed that the smallest stable microhelices, Ala-miHx^{Ala-4} and Glu-miHx^{Glu-3}, were poorer substrates than the corresponding entire AA-miHxs, which suggests that *Nbra*-CDPS interact with distal nucleotides of the acceptor arm. [*The complete set of data to be obtained, including those with Ala-miHx^{Ala-5}, Ala-miHx^{Ala-6}, Glu-miHx^{Glu-5} and Glu-miHx^{Glu-6}, will allow a more precise insight into the size of the stems required to retain full activity for the two different substrates.*] One concern about these shorter miHxs is that the tetramer loops introduced to close the miHxs might hamper the interaction with the binding sites of the enzymes by causing artificial steric hindrances. This should be considered in the interpretation of the results.

Besides CDPSs, other families of enzymes have evolved the ability to hijack AA-tRNAs from their canonical role to perform different classes of reactions (1). This can be illustrated with three families of enzymes which have been extensively studied and for which the mode of interaction with their AA-tRNA substrates have been investigated: class I lanthipeptide dehydratases (LanBs) that use Glu-tRNA^{Glu} for the successive glutamylation and dehydrogenation of serine and threonine residues in ribosomally synthesized and posttranslationally modified peptides (39), Fem transferases X (FemX) that transfer the alanyl moiety of Ala-tRNA^{Ala} to UDP-MurNAc-pentapeptide during peptidoglycan biosynthesis (40) and L/F transferases that transfer the leucyl moiety of Leu-tRNA^{Leu} (and in a lesser extend the phenylalanyl moiety of Phe-tRNA^{Phe}) to the N-terminus of proteins as a label for targeted degradation (41). Interestingly, despite their differences in substrate specificities and functions, LanBs, FemX and L/F transferases were all shown by various experimental strategies to essentially interact with the acceptor arms of the tRNA moieties of their substrates.

A first approach to investigate tRNA regioselectivity of AA-tRNA utilizing enzymes rely on the use of mutants tRNAs. Mutations are introduced into tRNAs and assayed as substrates in order to study whether mutations have an impact on the activity, which demonstrate that the mutated nucleotides have a role in the recognition by the enzyme. Using such approaches, it was shown that L/F transferases and FemX interact respectively with up to the first five and four base pairs of the acceptor arms of their substrates (42, 43). Similar studies have also demonstrated that the LanB-like MibB specifically interacts with the discriminator base A₇₃ and the nucleotide U₇₂ within the Glu-tRNA^{Glu} (44) and the CDPS AlbC interacts with the first base pair of its second AA-tRNA substrate (16). However, methods relying on AARSs for aminoacylation are intrinsically limited in terms of the range of RNA acceptors that can be tested because the specificity determinants required for aminoacylation must be conserved (45). An interesting approach to circumvent this limitation is to synthesize AA-tRNA analogues by chemoenzymatic synthesis, through enzymatic ligation of aminoacylated dinucleotide to tRNA analogues lacking two nucleotides at their 3' terminus (46, 47). This strategy allows to aminoacylate a much broader range of tRNA-like molecules than AARS-based methods, including shortened tRNA analogs or tRNAs with mutations in AARS specificity determinants. It was used to generate various analogues of Ala-tRNA^{Ala} for the study of FemX, including shortened analogues such as Ala-miHx^{Ala}, which was shown to be a good substrate of FemX,

and various mutated miHxs, which were used to elucidate the role of nucleotide sequence for specificity (31). This strategy would be particularly interesting to apply to other AA-tRNA-utilizing enzymes. However, it requires to include a 2'-deoxycytidine as a penultimate nucleotide and preliminary results with such molecules (kindly prepared by our collaborators M. Fonvielle, M. Ethève-Quellejeu and M. Arthur) on the CDPS AlbC suggested that this modification might induce a significant decrease of CDPS activity. Thus, the flexizyme-based production pipeline described in this study constitute a valuable alternative. Thanks to the infinite diversity of RNA acceptors that can be aminoacylated using flexizymes and to the ability to use any amino acids, AA-minitRNAs of any given sequence could be produced. This would allow to dissect the sequence-dependent recognition of tRNAs at the two binding sites of *Nbra*-CDPS or of other CDPSs.

The 3D-structures of members of the previously presented families of AA-tRNA-utilizing enzymes are available (39, 48, 49, 15). They exhibit different structural folds and only CDPSs have a Rossmann-fold domain similar to that of class I AARSs, indicating that different structural folds have evolved the ability to interact with AA-tRNAs. Structures in complex with AA-tRNA analogues have also been reported for these enzyme families but the size, intrinsic flexibility of tRNAs and the relatively low affinities of the enzymes towards their substrates have prevented obtaining high-resolution structures with analogues of full-size substrates. Therefore, minimal substrates analogues had to be used. We previously reported the structure of AlbC in complex with a dipeptidyl intermediate, which allowed to better understand aminoacyl binding by CDPSs (11). In the case of LanBs and L/F transferases, structures have been reported with synthetic analogues of puromycin, which mimic the 3'-terminal aminoacylated adenosines of AA-tRNAs and in which the labile ester bond between the nucleotide and the amino acid is replaced by a stable amide bond (48, 50). By using peptidyl-RNA conjugates produced through a solid-phase synthesis/click chemistry approach, Fonvielle et al. reported the structure of FemX in complex with a peptidyl-RNA miHx with two base pairs (43). The flexizyme-based production of various aminoacylated RNAs appears to be of great interest for obtaining structures of CDPSs or other AA-tRNA-utilizing enzymes in complex with such analogues. Indeed, Suga and coll. have recently reported that flexizymes are active on RNA with final adenosine bearing the 3'-deoxy-3'-amino modification, allowing the production of analogues in which the labile ester bond of the aminoacyl linkage is replaced by the stable amide bond (51).

The diverse methods to obtain AA-tRNA analogues have different advantages and drawbacks. Methods based on AARs are simple and efficient, but the spectrum of amino acids and of RNA acceptors that can be used is restricted by the natural specificity of these enzymes. Chemoenzymatic synthesis of aminoacylated- or peptidylated-tRNA analogues gives access to a broad range of AA-tRNA analogues bearing diverse modifications. However, this approach requires to use 2'-deoxycytidine as a penultimate nucleotide, which may be an issue if the 2' hydroxyl of this C₇₅ is involved in the interaction with the enzymes. Above all, the multi-steps synthesis routes are not easily reproducible in a typical biology laboratory. Comparatively, flexizymes represent a convenient aminoacylation strategy since they allow to obtain good aminoacylation yields in one step at mild conditions with an almost infinite versatility of amino acids and RNA acceptors. The combination of flexizymes with anion-exchange chromatography presented in this study represents a valuable tool for the study of AA-tRNAs and the associated mechanisms, since it allows to easily generate a very broad range of aminoacylated RNAs. The application of anion-exchange chromatography to the study and the manipulation of small aminoacylated RNAs also allows to characterize with precision the aminoacylated proportion of an RNA sample by HPLC. This represents an interesting addition to the experimental toolbox used for these molecules and the accuracy of this method appears to be much higher than the conventional method which rely densitometry on acid-PAGE densitometry. We envision that this method could greatly benefit to the community studying the growing number of AA-tRNA utilizing enzymes.

ACKNOWLEDGEMENT

The work was supported the French National Research Agency (ANR-14-CE09-0021 and ANR-19-CE44-0012). N.C. and R.V. hold PhD scholarships from the Commissariat à l'Énergie Atomique et aux Energies Alternatives (CEA). We thank Magali Frugier, Joëlle Rudinger-Thirion and Matthieu Fonvielle for helpful advices concerning RNA handling and Alexandre Couëtoux for experimental assistance.

REFERENCES

1. Moutiez, M., Belin, P. and Gondry, M. (2017) Aminoacyl-tRNA-Utilizing Enzymes in Natural Product

- Biosynthesis. *Chem. Rev.*, **117**, 5578–5618.
- Borthwick, A.D. (2012) 2,5-diketopiperazines: Synthesis, reactions, medicinal chemistry, and bioactive natural products. *Chem. Rev.*, **112**, 3641–3716.
 - Wang, X., Li, Y., Zhang, X., Lai, D. and Zhou, L. (2017) Structural Diversity and Biological Activities of the Cyclodipeptides from Fungi. *Molecules*, **22**, 2026.
 - 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.
 - Jacques, I.B., Moutiez, M., Witwinowski, J., Darbon, E., Martel, C., Seguin, J., Favry, E., Thai, R., Lecoq, A., Dubois, S., *et al.* (2015) Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity. *Nat. Chem. Biol.*, **11**, 721–731.
 - Gondry, M., Jacques, I., Thai, R., Babin, M., Canu, N., Seguin, J., Belin, P., Pernodet, J.-L. and Moutiez, M. (2018) A comprehensive overview of the cyclodipeptide synthase family enriched with the characterization of 32 new enzymes. *Front. Microbiol.*, **9**, 1–14.
 - Canu, N., Moutiez, M., Belin, P. and Gondry, M. (2019) Cyclodipeptide synthases: a promising biotechnological tool for the synthesis of diverse 2,5-diketopiperazines. *Nat. Prod. Rep.*, **to be publ.**
 - Giessen, T. and Marahiel, M. (2014) The tRNA-Dependent Biosynthesis of Modified Cyclic Dipeptides. *Int. J. Mol. Sci.*, **15**, 14610–14631.
 - 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.*, **29**, 961–979.
 - 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.
 - Moutiez, M., Schmitt, E., Seguin, J., Thai, R., Favry, E., Belin, P., Mechulam, Y. and Gondry, M. (2014) Unravelling the mechanism of non-ribosomal peptide synthesis by cyclodipeptide synthases. *Nat. Commun.*, **5**, 1–7.
 - Schmitt, E., Bourgeois, G., Gondry, M. and Aleksandrov, A. (2018) Cyclization Reaction Catalyzed by Cyclodipeptide Synthases Relies on a Conserved Tyrosine Residue. *Sci. Rep.*, **8**, 1–11.
 - 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.
 - 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**, 3912–7.
 - Bourgeois, G., Seguin, J., Babin, M., Belin, P., Moutiez, M., Mechulam, Y., Gondry, M. and Schmitt, E. (2018) Structural basis for partition of the cyclodipeptide synthases into two subfamilies. *J. Struct. Biol.*, **203**, 17–26.
 - Moutiez, M., Fonvielle, M., Belin, P., Favry, E., Arthur, M. and Gondry, M. (2014) Specificity

- determinants for the two tRNA substrates of the cyclodipeptide synthase AlbC from *Streptomyces noursei*. *Nucleic Acids Res.*, **42**, 7247–7258.
17. Kobayashi, T., Nurekil, 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.
 18. Francklyn, C. and Schimmel, P. (1989) Aminoacylation of RNA minihelices with alanine. *Nature*, **337**, 478–481.
 19. Frugier, M., Florentz, C. and Giegé, R. (1994) Efficient aminoacylation of resected RNA helices by class II aspartyl-tRNA synthetase dependent on a single nucleotide. *EMBO J.*, **13**, 2218–2226.
 20. Passioura, T. and Suga, H. (2014) Flexizymes, their evolutionary history and diverse utilities. *Top Curr Chem*, **11**, 331–346.
 21. Murakami, H., Ohta, A., Ashigai, H. and Suga, H. (2006) A highly flexible tRNA acylation method for non-natural polypeptide synthesis. (SI). *Nat. Methods*.
 22. Fujino, T., Kondo, T., Suga, H. and Murakami, H. (2019) Exploring of minimal RNA substrate of flexizymes. *ChemBioChem*, **8603**, 1–8.
 23. Ellinger, T. and Ehricht, R. (1998) Single-step purification of T7 RNA polymerase with a 6-histidine tag. *Biotechniques*, **24**, 718–720.
 24. Beuning, P.J., Yang, F., Schimmel, P. and Musier-Forsyth, K. (2002) Specific atomic groups and RNA helix geometry in acceptor stem recognition by a tRNA synthetase. *Proc. Natl. Acad. Sci.*, **94**, 10150–10154.
 25. Dubois, D.Y., Blais, S.P., Huot, J.L. and Lapointe, J. (2009) AC-truncated glutamyl-tRNA synthetase specific for tRNA^{Glu} is stimulated by its free complementary distal domain: Mechanistic and evolutionary implications. *Biochemistry*, **48**, 6012–6021.
 26. 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. U. S. A.*, **106**, 7426–7431.
 27. Mechulam, Y., Guillon, L., Yatime, L., Blanquet, S. and Schmitt, E. (2007) Protection-Based Assays to Measure Aminoacyl-tRNA Binding to Translation Initiation Factors First Edit. Elsevier Inc.
 28. Goto, Y., Katoh, T. and Suga, H. (2011) Flexizymes for genetic code reprogramming. *Nat. Protoc.*, **6**, 779–790.
 29. Konarska, M.M. and Sharp, P.A. (1990) Structure of RNAs replicated by the DNA-dependent T7 RNA polymerase L-amino acids. *Cell*, **63**, 609–618.
 30. Betat, H., Long, Y., Jackman, J.E. and Mörl, M. (2014) From end to end: tRNA editing at 5'- and 3'-terminal positions. *Int. J. Mol. Sci.*, **15**, 23975–23998.
 31. Fonvielle, M., Chemama, M., Villet, R., Lecerf, M., Bouhss, A., Valéry, J.M., Ethève-Quelejeu, M. and Arthur, M. (2009) Aminoacyl-tRNA recognition by the FemXWv transferase for bacterial cell wall synthesis. *Nucleic Acids Res.*, **37**, 1589–1601.
 32. Peacock, J.R., Walvoord, R.R., Chang, A.Y. and Kozlowski, M.C. (2014) Amino acid – dependent stability of the acyl linkage in aminoacyl-tRNA Amino acid – dependent stability of the acyl

- linkage in aminoacyl-tRNA. *RNA*, **20**, 758–764.
33. Boccaletto, P., MacHnicka, M.A.A., Purta, E., Pitkowski, P., Baginski, B., Wirecki, T.K.K., De Crécy-Lagard, V., Ross, R., Limbach, P.A.A., Kotter, A., *et al.* (2018) MODOMICS: A database of RNA modification pathways. 2017 update. *Nucleic Acids Res.*, **46**, D303–D307.
 34. Schuber, F. and Pinck, M. (1974) On the chemical reactivity of aminoacyl-tRNA ester bond. I - Influence of pH and nature of the acyl group on the rate of hydrolysis. *Biochimie*, **56**, 383–390.
 35. Stepanov, V.G. and Nyborg, J. (2002) Thermal stability of aminoacyl-tRNAs in aqueous solutions.
 36. Tuerk, C., Gauss, P., Thermes, C., Groebe, D.R., Gayle, M., Guild, N., Stormo, G., D'Aubenton-Carafa, Y., Uhlenbeck, O.C. and Tinoco, I. (1988) CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc. Natl. Acad. Sci.*, **85**, 1364–1368.
 37. Antao, V.P., Lai, S.Y. and Tinoco, I. (1991) A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res.*, **19**, 5901–5905.
 38. Sun, X., Li, J.M. and Wartell, R.M. (2007) Conversion of stable RNA hairpin to a metastable dimer in frozen solution. *Rna*, **13**, 2277–2286.
 39. Ortega, M.A., Hao, Y., Zhang, Q., Walker, M.C., van der Donk, W.A. and Nair, S.K. (2014) Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. *Nature*, **517**, 509–512.
 40. Maillard, A.P., Biarrotte-Sorin, S., Villet, R., Mesnage, S., Bouhss, A., Sougakoff, W., Mayer, C. and Arthur, M. (2005) Structure-based site-directed mutagenesis of the UDP-MurNAc-pentapeptide-binding cavity of the FemX alanyl transferase from *Weissella viridescens*. *J. Bacteriol.*, **187**, 3833–3838.
 41. Leibowitz, M.J. and Soffer, R.L. (1969) No Title. *Biochem. Biophys. Res. Commun.*, **36**, 47–53.
 42. Fung, A.W.S., Leung, C.C.Y. and Fahlman, R.P. (2014) The determination of tRNA Leu recognition nucleotides for *Escherichia coli* L/F transferase. *RNA*, **20**, 1210–1222.
 43. Fonvielle, M., Li De La Sierra-Gallay, I., El-Sagheer, A.H., Lecerf, M., Patin, D., Mellal, D., Mayer, C., Blanot, D., Gale, N., Brown, T., *et al.* (2013) The structure of FemXWv in complex with a peptidyl-RNA conjugate: Mechanism of aminoacyl transfer from Ala-tRNA^{Ala} to peptidoglycan precursors. *Angew. Chemie - Int. Ed.*, **52**, 7278–7281.
 44. Ortega, M.A., Hao, Y., Walker, M.C., Donadio, S., Sosio, M., Nair, S.K. and Van Der Donk, W.A. (2016) Structure and tRNA Specificity of MibB, a Lantibiotic Dehydratase from Actinobacteria Involved in NAI-107 Biosynthesis. *Cell Chem. Biol.*, **23**, 370–380.
 45. Giegé, R., Sissler, M. and Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.*, **26**, 5017–35.
 46. Heckler, T.G., Chang, L.H., Zama, Y., Naka, T., Chorghade, M.S. and Hecht, S.M. (1984) T4 RNA Ligase Mediated Preparation of Novel “Chemically Misacylated” tRNAs. *Biochemistry*, **23**, 1468–1473.
 47. Robertson, S.A., Ellman, J.A. and Schultz, P.G. (1991) A General and Efficient Route for Chemical Aminoacylation of Transfer RNAs. *J. Am. Chem. Soc.*, **113**, 2722–2729.
 48. 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.
49. 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.
50. Bothwell, I.R., Cogan, D.P., Kim, T., Reinhardt, C.J., van der Donk, W.A. and Nair, S.K. (2019) Characterization of glutamyl-tRNA-dependent dehydratases using nonreactive substrate mimics. *Proc. Natl. Acad. Sci.*, **116**, 17245–17250.
51. Katoh, T. and Suga, H. (2019) Flexizyme-catalyzed synthesis of 3'-aminoacyl-NH-tRNAs. *Nucleic Acids Res.*, 10.1093/nar/gkz143.

III) Supplementary Information

I)	Additional methods.....	3
A)	Protein production.....	3
i.	<i>E. coli</i> AlaRS.....	3
ii.	<i>E. coli</i> GluRS	3
iii.	T7 RNA Polymerase	4
iv.	<i>Nbra</i> -CDPS	5
B)	RNA production.....	6
i.	RNA sequences and production methods	6
	Table S1: RNA molecules used in this study.	7
ii.	Cloning of tRNAs into pBSTNAV2.....	8
	Table S2: DNA primers used for the cloning of tRNAs into pBSTNAV2.....	8
iii.	Determination of the proportion of aminoacylated tRNAs.....	9
iv.	Synthesis of RNAs by <i>in vitro</i> transcription.....	10
	Figure S1: Strategy of DNA matrix production by oligo annealing.	10
	Table S3: DNA oligos used for <i>in vitro</i> transcription.	10
v.	Aminoacylation by flexizymes	10
	Figure S2: Hybridization between dF _x and their corresponding RNA substrates. ..	10
vi.	Analysis and purification of RNAs by anion-exchange chromatography	11
	Table S4: Gradients used for purification and analysis of RNAs and AA-RNAs.	11
vii.	Acid-PAGE electrophoresis of RNA and AA-miHX.....	12
viii.	MALDI-TOF characterization of minitRNAs and AA-minitRNAs	12
C)	Quantification of cAE and cAA	13
i.	Chemical syntheses of labelled standards of cAE and cAA.....	13
	Scheme S1: Synthesis of cAE*.....	14
	Figure S3: ¹ H NMR spectra of cAE*.....	14
	Scheme S2: Synthesis of cAA*.....	15
	Figure S4: ¹ H NMR spectra of cAA*.....	15
ii.	LC-MS quantification of cAE and cAA	16

	Figure S5: LC-MS chromatograms used for cyclodipeptide quantification.....	16
II)	Additional results	17
A)	Optimization of the <i>in vitro</i> enzymatic assay for <i>Nbra</i> -CDPS.....	17
	Figure S6: Influence of the nature of buffers on cyclodipeptide production.	17
	Figure S7: Influence of pH and temperature on cAE production.....	18
	Figure S8: Influence of the enzyme concentration on cAE production.	19
B)	Additional kinetics using AA-tRNA substrates.....	20
	Figure S9: Comparison of <i>Nbra</i> -CDPS activity on the two tRNA ^{Ala} isoacceptors.....	20
	Figure S10: Cyclodipeptides concentrations curves for initial velocities determination	21
	Figure S11: Fitting of cAE production to Michaelis-Menten equation.....	22
	Figure S12: Kinetics study of <i>Nbra</i> -CDPS towards AA-minitRNA substrates (with cAA production).	23
C)	Prediction of conformation stabilities of miHxs with decreasing stem sizes	24
	Table S5: Sequences and calculated Gibson free energies of miHxs with decreasing stem sizes.....	24
	Comment on Table S5.....	24
D)	Production of aminoacylated minitRNAs.....	25
	Figure S13: Effect of heat and quick cool-renaturation on the HPLC profile of miHxAla ₇	25
	Figure S14: Analysis of miHx ^{Ala₇} alanylation reaction.	26
	Comment on Figure S14.....	27
	Figure S15: Analysis of miHx ^{Glu₇} glutamylation reaction.	28
	Figure S16: Comparison of HPLC profiles for all minitRNAs.....	29
	Figure S17: MALDI-TOF spectra of minitRNAs and AA-minitRNAs.....	31
	Figure S18: Acylation kinetics of minitRNAs by flexizymes.....	33
	Figure S19: Deacylation kinetics of AA-minitRNAs.	34
	Table S7: Predicted deacylation of AA-minitRNAs.....	35
	References	36

l) Additional methods

A) Protein production

i. *E. coli* AlaRS

M15(pREP4) *E. coli* cells were transformed with plasmid pQE875-6His-AlaRS(1) (a kind gift of Dr. Karine Musier-Forsyth) and were used to inoculate one 10 mL LB preculture supplemented with ampicillin (100 µg/mL, Sigma-Aldrich) and kanamycin (30 µg/mL, Sigma-Aldrich). After an incubation at 37°C under agitation, this preculture was used to inoculate two 1 L LB culture, supplemented with ampicillin and kanamycin at the same concentration, at OD₆₀₀ 0.05. When OD₆₀₀ reached 0.6, cultures were induced with 0.5 mM isopropyl-β-D thiogalactopyranoside (IPTG, Euromedex) and incubated overnight at 37°C under agitation. Cells were pelleted, frozen by liquid nitrogen and stored at -80°C. Cells were resuspended in 2x40 mL lysis buffer (100 mM KPO₄ buffer pH 7.8, 300 mM NaCl, 5% glycerol, 2mM β-mercaptoethanol, 1mM phenylmethylsulfonylfluoride (PMSF)) and broken with an Eaton press (Rassant). After addition of benzonase (6 U/mL final, Sigma-Aldrich) and MgCl₂ (10 mM final), the cell-free extracts were incubated 20 min at 4°C with agitation then centrifuged at 35,000 g for 20 min at 4°C. Supernatant containing the soluble fraction was loaded on a 5mL HisTrap Column (GE Healthcare) equilibrated with lysis buffer. Column was washed first with lysis buffer and by a first step at 10% elution buffer (100 mM KPO₄ buffer pH 7.8, 300 mM NaCl, 5% glycerol, 2 mM β-mercaptoethanol, 1 mM PMSF, 1 M imidazole), then 6His-AlaRS was eluted at 30% elution buffer. Fractions containing the protein were pooled and desalted using three 5mL Hi-Trap Desalting columns connected in series in desalting buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 10% glycerol, 1 mM DTT). 6His-AlaRS was concentrated on 5mL Vivaspin (30 kDa cut Off, Sartorius). Glycerol was added to 40% final, the protein was aliquoted at a final concentration of 30.2 µM and stored at -20°C.

ii. *E. coli* GluRS

BL21 (DE3)pLysS *E. coli* cells were transformed with the derived pET-28c-6His-GluRS plasmid(2) (a kind gift of Dr. Jacques Lapointe) and were used to inoculate one 50 mL LB preculture supplemented with chloramphenicol (25 µg/mL, Sigma-Aldrich) and kanamycin (30µg/mL). After an overnight incubation at 37°C under agitation, this starter culture was used to inoculate two 1 L LB culture, supplemented with chloramphenicol and kanamycin at the same concentration, at OD₆₀₀ 0.05. Cultures were incubated at 30°C and, when OD₆₀₀ reached 0.4, they were induced with 0.5 mM IPTG then incubated overnight at 20°C under agitation. Cells were pelleted, frozen by liquid nitrogen and stored at -80°C. Cells were resuspended in 2x40 mL lysis buffer (20 mM Tris-HCl buffer pH 7.9, 1 mM β-mercaptoethanol, 5 mM imidazole, 5% glycerol, 1 mM PMSF) and broken with an Eaton press (Rassant). After addition of benzonase (6U/mL final, Sigma-Aldrich) and MgCl₂ (15 mM final), the cell-free extracts were incubated at 4°C under agitation to decrease the sample viscosity then centrifuged at 35,000g for 30 min at 4°C. Supernatant containing the soluble fraction was loaded on a 5mL HisTrap Column (GE Healthcare) equilibrated with column buffer (20 mM Tris-HCl buffer pH 7.9, 1 mM β-mercaptoethanol). The column was successively washed with column buffer containing 5 and 25mM imidazole. The 6His-GluRS protein was eluted with 1 M imidazole in the same buffer. Fractions containing the protein were pooled and dialyzed against 40 mM HEPES-KOH, pH 7,2, 10% glycerol and 10 mM β-mercaptoethanol using a Spectra/Por membrane (12 to 14 kDa MWCO). 6His-GluRS was concentrated on 5mL Vivaspin (10 kDa cut Off, Sartorius), and stored at -20°C in the storage buffer (20 mM HEPES-KOH, pH 7,2, 45% glycerol and 5 mM β-mercaptoethanol).

iii. T7 RNA Polymerase

BL21(pREP4) *E. coli* cells were transformed with plasmid pQE60- 6His-T7RNAPol (a kind gift of Dr. Matthieu Fonvielle and Dr. Michel Arthur) and were used to inoculate two 20 mL LB precultures supplemented with ampicillin (100µg/mL) and kanamycin (30µg/mL). After an incubation at 37°C under agitation, these precultures were used to inoculate two 800mL LB cultures OD₆₀₀ 0.05, supplemented with ampicillin and kanamycin at the same concentration, incubated at 30°C under agitation. When OD₆₀₀ reached 0.6, cultures were induced with IPTG (0.5 mM final) and incubated overnight at 20°C under agitation. Cells were pelleted, frozen by

liquid nitrogen and stored at -80°C . Cells were resuspended in 2x40 mL lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole, 10% glycerol, 1 mM DTT, 1 mM PMSF) and broken with an Eaton press (Rassant). After addition of benzonase (6.25 U/mL final) and MgCl_2 (5mM final), the cell-free extracts were incubated 20 min at 4°C under agitation then centrifuged at 35,000g for 20 min at 4°C . Supernatant containing the soluble fraction was loaded on a 5 mL HisTrap Column (GE Healthcare) equilibrated with equilibration buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM imidazole, 10% glycerol, 1 mM DTT, 1 mM PMSF). Column was washed with equilibration buffer, then 6His-T7 RNA Polymerase was eluted with a gradient of 0 to 100% elution buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 500mM imidazole, 10% glycerol, 1mM DTT, 1mM PMSF) over 20 column volumes. Fractions containing the protein were pooled and dialyzed 1 h, then overnight at 4°C in dialysis buffer (50mM Tris-HCl pH 8.0, 300 mM NaCl, 10% glycerol, 1 mM DTT, 1 mM EDTA, 1 mM PMSF). 6His-T7 RNA Polymerase was concentrated to 20 mg/mL on 5mL Vivaspin (30 kDa cut Off, Sartorius). Glycerol was added to 50% final, the protein was aliquoted at a final concentration of 10mg/mL and stored at -80°C .

iv. *Nbra*-CDPS

BL21-AI *E. coli* cells were transformed with plasmid PIJ196-CDPS39(3) and were used to inoculate one 5mL LB preculture supplemented with ampicillin (100 $\mu\text{g}/\text{mL}$). After an incubation at 37°C under agitation, this preculture was used to inoculate one 800 mL LB culture, supplemented with ampicillin at the same concentration, at OD_{600} 0.05. When OD_{600} reached 0.6, cultures were induced with IPTG (0.5 mM final) and incubated overnight at 20°C under agitation. Cells were pelleted, frozen by liquid nitrogen and stored at -80°C . Cells were resuspended in 2x40mL lysis buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 5% glycerol, 10 mM β -mercaptoethanol, 1 mM PMSF) and broken with an Eaton press (Rassant). After addition of benzonase (12.5 U/mL final) and MgCl_2 (10 mM final), the cell-free extracts were incubated 40 min at 4°C under agitation then centrifuged at 35,000 g for 25 min at 4°C . Supernatant containing the soluble fraction was loaded on a 5 mL HisTrap Column (GE Healthcare) equilibrated with equilibration buffer (same composition as lysis buffer with 40 mM imidazole). Column was washed with equilibration buffer, then *Nbra*-CDPS was eluted with a gradient of 0 to 100% elution buffer (same composition as lysis buffer with 1 M imidazole) over 25 column volumes. NaCl concentration was lowered to 50 mM by diluting

the sample in a dilution buffer (same composition as lysis buffer with no NaCl) and the protein was loaded on an Heparine 5 mL column (GEHealthcare). Column was washed with equilibration buffer (same as lysis buffer but only 50mM NaCl) and protein was eluted with a gradient of 0 to 100% elution buffer (same composition as lysis buffer with 1 M NaCl). Fractions containing the proteins were concentrated on Vivaspin 10 kDa centrifugal concentrator (VWR). Experimental extinction coefficient was determined by amino acid analysis for *Nbra*-CDPS at $38800 \text{ M}^{-1} \cdot \text{cm}^{-1}$.

B) RNA production

- i. RNA sequences and production methods

Table S1: RNA molecules used in this study.

Name	Sequence	Production	Molecular weight
tRNA ^{Ala} _{TGC}	GGGGCUAUAGCUCAGCUGGGAGAGCGCCUGCUUUUGCACGCAGGAGGUCUGCGGUUCGAUCCCGCAUAGCUCCACCA	<i>In vivo</i> production in <i>E. coli</i>	24651
tRNA ^{Ala} _{GGC}	GGGGCUAUAGCUCAGCUGGGAGAGCGCUUGCAUGGCAUGCAAGAGGUCAGCGGUUCGAUCCCGCUUAGCUCCACCA		24699
tRNA ^{Glu}	GUCCCCUUCGUCUAGAGGCCAGGACACCGCCCUUUCACGGCGGUAACAGGGGUUCGAAUCCCUAGGGGACGCCA		24551
dFx _{ACCA} *	GGAUCGAAAGAUUUCGCAUCCCCGAAAGGGUACAUGGCGUUAGGU	<i>In vitro</i> transcription	15061.8
dFx _{GCCA} *	GGAUCGAAAGAUUUCGCAUCCCCGAAAGGGUACAUGGCGUUAGGC		15060.8
miHx ^{Ala} ₇	GGGGCUAUUCGUAGCUCCACCA		7072.8
miHx ^{Ala} ₆	GGGGCUUUCGAGCUCCACCA		6437.4
miHx ^{Ala} ₅	GGGGCUUCGGCUCCACCA		5802
miHx ^{Ala} ₄	GGGGUUCGCUCCACCA		5151.2
miHx ^{Ala} ₃	GGGUUCGUCCACCA	Chemical synthesis	4500.4
ACCA	ACCA		1283.2
miHx ^{Glu} ₇	GUCCCCUUUCGAGGGGACGCCA		7088.2
miHx ^{Glu} ₆	GUCCCCUUCGGGGACGCCA		6452.8
miHx ^{Glu} ₅	GUCCCUUCGGGGACGCCA		5802
miHx ^{Glu} ₄	GUCCUUCGGGACGCCA		5151.2
miHx ^{Glu} ₃	GUCUUCGGACGCCA		4500.4
GCCA	GCCA		1300.2

dFxs and minitRNAs are unmodified, except for the 5' terminus which bears a triphosphate for dFxs and a monophosphate for the different minitRNAs. tRNAs produced *in vivo* contain post-transcriptional modifications. Molecular weights of tRNAs were calculated on the basis of the sequence with a 5' monophosphate terminus and the following modified bases (modified bases contained in *E. coli* tRNAs were taken from the RNA modification database Modomics(4), the mass difference brought by modified bases is indicated in parentheses): for tRNA^{Ala}_{TGC} and tRNA^{Ala}_{GGC} dihydrouridine (+2); uridine 5-oxyacetic acid (+74); 7-methylguanosine (+14); 5-methyluridine (+14); pseudouridine (+0) and for tRNA^{Glu}: 4-thiouridine (+16); two pseudouridine (+0); 5-methylaminomethyl-2-thiouridine (+59); 2-methyladenosine (+14); 5-methyluridine (+14).

* For a comment on the two dFx sequences that were used in this study, see section I.A.v.

ii. Cloning of tRNAs into pBSTNAV2

Genes were amplified from *E. coli* genomes using Phusion High-Fidelity DNA polymerase, *E. coli* cell lysates as matrices and primers indicated in Table S2, designed for the amplification of tRNA sequence from *E. coli* genome and introduction of EcoR1 and PstI restriction sites. Amplicons were digested with EcoRI/PstI and ligated into pBSTNAV2. Constructs were verified by sequencing and tRNA sequences were in accordance with tRNAs sequences of *E. coli* from the literature.

Table S2: DNA primers used for the cloning of tRNAs into pBSTNAV2.

Name	Sequence
EcoR1_tRNA_Ala_F	5' CTTGTAACGCTGAATTCGGGGCTATAGCTCAG 3'
tRNA_Ala_PstI_R	5' CGCTAAGGATCTGCAGTGGTGGAGCTAAGCGG 3'
EcoR1_tRNA_Glu_F	5' CTTGTAACGCTGAATTCGTCCCCTTCGTCTAG 3'
tRNA_Glu_PstI_R	5' CGCTAAGGATCTGCAGTGGCGTCCCCTAGGGG 3'

iii. Determination of the proportion of aminoacylated tRNAs

For all the optimization of the enzymatic assay (section II.A.i. to II.A.iii), proportions of aminoacylated tRNAs were determined using ^{14}C -labelled amino acids in plateau charging experiments, as in (5)(5). Briefly, analytical 50 μL aminoacylation reactions were prepared concomitantly with 5mL preparative reactions using the same components except that they contained the corresponding L-amino acid labelled with ^{14}C at 60 mCi/mmol (2.22 GBq) (Perkin-Elmer). After incubation, AA-tRNAs and remaining non acylated tRNAs of the analytical reactions were precipitated with 5% TCA and 0.5% casamino acids, filtered on Whatman GF/C filters and quantified by liquid scintillation counting.

Such approach does not consider the effect of scaling-up the reaction and the possible deacylation during the purification process of preparative aminoacylation reactions. Therefore, once conditions of enzymatic assays were set, we used a modified enzymatic assay to determine directly the concentration of AA-tRNAs after preparative aminoacylation reactions and purification.

Enzymatic assays were performed in 100 mM Na-Phosphate buffer pH 7.5, 50 mM KCl, 15 mM MgCl_2 , 0.1 mM β -mercaptoethanol. Approximative substrate concentrations were fixed so that one of the substrates was present in limiting amount and could be totally consumed: either 100 nM Ala-tRNA^{Ala} and 2000 nM Glu-tRNA^{Glu} for the determination of the proportion of alanylated tRNA^{Ala} or 100 nM Glu-tRNA^{Glu} and 2000 nM Ala-tRNA^{Ala_{TGC}} for the determination of the proportion of glutamylated tRNA^{Glu}. Assays were started by addition of 500 nM *Nbra*-CDPS, incubated at 20°C and were quenched with 2% TFA before being analysed by LC-MS. Concentrations of cAE were stable after 6 min for assays containing 100 nM Ala-tRNA^{Ala} and 2000 nM Glu-tRNA^{Glu} and 2 min for assays containing 100 nM Glu-tRNA^{Glu} and 2000 nM Ala-tRNA^{Ala_{TGC}}, suggesting that the totality of the assayed AA-tRNA substrate was consumed and incorporated into the cAE product. We calculated the proportion of aminoacylated tRNAs using the final concentrations of cAE (after 3 or 7 min for characterization of Ala-tRNA^{Ala} or Glu-tRNA^{Glu}, respectively). They were calculated to be 53% for Ala-tRNA^{Ala_{TGC}}, 56% for Ala-tRNA^{Ala_{GCC}} and 67% for Glu-tRNA^{Glu}. These values were used for subsequent experiments.

iv. Synthesis of RNAs by *in vitro* transcription

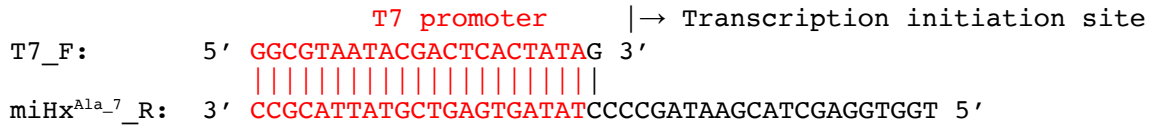


Figure S1: Strategy of DNA matrix production by oligo annealing. Example of the matrix for synthesis of miHx^{Ala_7}.

Table S3: DNA oligos used for *in vitro* transcription.

Name	Sequence
T7_F	GCGTAATACGACTCACTATAG
miHx ^{Ala_7} _R	TGGTGGAGCTACGAATAGCCCCATAGTGAGTCGTATTACGCC
miHx ^{Ala_6} _R	TGGTGGAGCTCGAAAAGCCCCATAGTGAGTCGTATTACGCC
miHx ^{Ala_5} _R	TGGTGGAGCCGAAAGCCCCATAGTGAGTCGTATTACGCC
dFx_ACCA_R	ACCTAACGCCATGTACCCTTTCGGGGATGCGGAAATCTTTCGATCCTATAGTGAGTCGTATTACGCC
dFx_GCCA_R	GCCTAACGCCATGTACCCTTTCGGGGATGCGGAAATCTTTCGATCCTATAGTGAGTCGTATTACGCC

v. Aminoacylation by flexizymes

Two different flexizymes, dFx_ACCA and dFx_GCCA, were produced. The sequences of the two dFx differ by only one nucleotide, located at the 3' terminus (U in dFx^{ACCA} and C in dFx^{GCCA}). This design was chosen in order to ensure optimal hybridization between the flexizymes and RNAs bearing 3' terminus of sequence ACCA (such as tRNA^{Ala} and its truncated analogues) or GCCA (such as tRNA^{Glu} and its shortened analogues), respectively (Figure S2). Indeed, preliminary studies on flexizymes suggested that acylation is favoured by optimal base-pairing

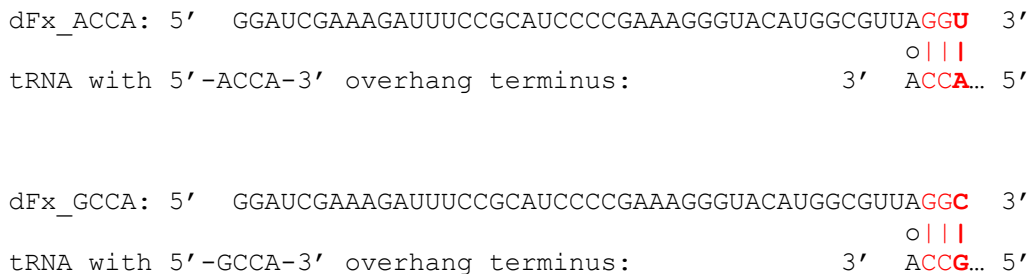


Figure S2: Hybridization between dFx and their corresponding RNA substrates. between the flexizymes and their RNA acceptors (6).

vi. Analysis and purification of RNAs by anion-exchange chromatography

Table S4: Gradients used for purification and analysis of RNAs and AA-RNAs.

Gradient number	RNAs for which the gradient was used	Gradient	
1	Comparative analysis of all minitRNAs	0-5 min	100%A/0%B
		5-75 min	30%A/70%B
2	miHx ^{Ala_7} miHx ^{Ala_6} miHx ^{Ala_5}	0-5 min	65%A/35%B
		5-25 min	65%A/35%B → 55%A/45%B
		25-27 min	55%A/45%B → 100%B
3	dFx	0-5 min	60%A/40%B
		5-20 min	60%A/40%B → 52.5%A/47.5%B
		20-21 min	52.5%A/47.5%B → 40%A/60%B
		21-31 min	40%A/60%B
		31-33 min	40%A/60%B → 100%B
4	Ala-miHx ^{Ala_7} Glu-miHx ^{Glu_7}	0-5 min	68%A/32%B
		5-37 min	68%A/32%B → 52%A/48%B
		37-38 min	52%A/48%B → 40%A/60%B
		38-48 min	40%A/60%B
		48-50 min	40%A/60%B → 100%B
5	Ala-miHx ^{Ala_6} Ala-miHx ^{Ala_5} Glu-miHx ^{Glu_6} Glu-miHx ^{Glu_5}	0-5 min	71%A/ 29%B
		5-37 min	71%A/ 29%B → 55%A/45%B
		37-38 min	55%A/45%B → 40%A/60%B
		38-48 min	40%A/60%B
		48-50 min	40%A/60%B → 100%B
6	Ala-miHx ^{Ala_4} Glu-miHx ^{Glu_4} Glu-miHx ^{Glu_3}	0-5 min	74%A/26%B
		5-41 min	74%A/26%B → 56%A/44%B
		41-42 min	56%A/44%B → 40%A/60%B
		42-52 min	40%A/60%B
		52-54 min	40%A/60%B → 100%B
7	Ala-ACCA Glu-GCCA	0-5 min	100%A/0%B
		5-45 min	100%A/0%B → 80%A/20%B
		45-46 min	80%A/20%B → 40%A/60%B
		46-56 min	40%A/60%B
		56-58 min	40%A/60%B → 100%B

vii. Acid-PAGE electrophoresis of RNA and AA-miHX

For acidic-PAGE analysis, samples were mixed with three times their volume of denaturing dye (93% formamide, 10 mM EDTA, 150 mM sodium acetate pH 5.2). Gels (10 cm x 8 cm) contained 7M urea, 15 % acrylamide (29:1 acrylamide/bisacrylamide ratio), 50mM sodium acetate pH 5.2, 0,04 % TEMED and 0,075 % ammonium persulfate and were run in 50mM sodium acetate pH 5.2 for 2h30 at 4°C using a Mini-PROTEAN electrophoresis chamber (Biorad).

viii. MALDI-TOF characterization of minitRNAs and AA-minitRNAs

MALDI-TOF characterization of minitRNAs and AA-minitRNAs was performed using parameters optimized by Dr. Yang (7). 0.5 µL matrix solution (250 mM 3-hydroxypicolinic acid (Sigma), 10 mM ammonium citrate dibasic, 10% acetonitrile) was spotted on a MALDI plate and allowed to dry for 7-8 min. Then, 0.5 µL 10 µM RNA solution was spotted on top of the matrix dry spot and the sample was allowed to dry for 7-8 min. MALDI-TOF analyses were done in linear/positive mode on a 4800 MALDI TOF MDS Sciex Analyzer (Applied Biosystems), set at a laser energy of 7000. External calibration was done using a mixture of three synthetic DNA oligos at 10 µM (Sigma): 5'-TGGTGGAGCTAGATCG-3' (m/z 4978.2), 5'-TGGTGGAGCTAAGCGGGATC-3' (m/z 6239.1) and 5'-CATTACTGGATCTATCAACAGGAG-3' (m/z 7361.9).

C) Quantification of cAE and cAA

Given the very low concentrations of cyclodipeptides to titrate (in the 10 to 100 nM range) and the low molar extinction coefficient of cyclodipeptides, no UV-based titration is possible. Because of its very high sensitivity and selectivity, mass spectrometry is commonly used for such applications, although quantification by MS is not straightforward. Indeed, the ionization which takes place in the ion source of a mass spectrometer is a stochastic process (only a portion of the molecules is ionized). The ionization efficiency depends mainly on the chemical structure of the compound to analyse but also on several parameters that are difficult to control and which vary on a day-to-day basis, such as the temperature and pressure of the ion source (8). One common strategy to take this variation of ionization efficiency into account is to add an internal standard to the sample, whose concentration is known and which will behave similarly to the analyte in the ion source. Ideally, this internal standard must have exactly the same structure as the analyte, in order to have the same ionization, but a different mass to prevent coverage of the signal of the analyte. In order to combine these two criteria, compounds labelled with stable isotopes are used when possible. In order to prevent any cross-talk between the signals of the analyte and the standard, a mass difference of at least 3 units is required(8). Therefore, in order to titrate cAA and cAE produced by *Nbra*-CDPS, ^{13}C - ^{15}N -labelled cAA and cAE were synthesized.

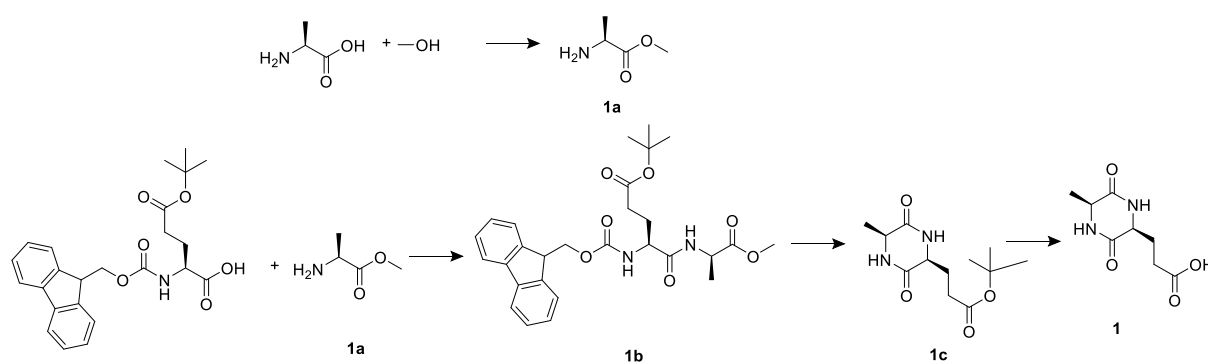
i. Chemical syntheses of labelled standards of cAE and cAA

Synthesis protocol was inspired from Deigin et al. (2016)(9) and Jeedigunta et al. (2000).(10)

Synthesis of $^{13}\text{C}_3$ - ^{15}N -cyclo(Ala-Glu) (cAE*, 1)

$^{13}\text{C}_3$ - ^{15}N -L-alanine (100mg, 1.08mmol, 1eq.) was diluted in 10 mL anhydrous methanol, thionyl chloride (SOCl_2 , 235 μL , 3.24 mmol, 3eq.) was carefully added at 0°C under argon. The mixture was heated to reflux at 70°C for 12 h then dried to vacuum to give **1a** as a white powder with quantitative yield. **1a** (150 mg, 1.05 mmol, 1eq.) was diluted in 10 mL freshly distilled dichloromethane and mixed with N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC, 362 mg, 1.89 mmol, 1.8 eq.), hydroxybenzotriazole hydrate (HOBt, 290 mg, 1.89 mmol, 1.8 eq.) and Fmoc-L-glutamate-O-(tBu) (580 mg, 1.37 mmol, 1.3 eq.). *N,N*-diisopropylethylamine (DIPEA, 143 μL , 0.78 mmol, 3 eq.) was carefully added at 0°C under

argon. The mixture was stirred at 0°C for 1 h then at room temperature for 12 h. The mixture was dried under vacuum, suspended in dichloromethane, washed with water three times, dried over magnesium sulphate and dried under vacuum to afford a white solid. The residue was applied to FC (silica gel, 40 g column, eluted with DCM/MetOH 100:0 → 95:5) to give **1b** as a white powder (324 mg, 60%). **1b** (324 mg, 0.63 mmol) was diluted in 8 mL dichloromethane. 2 mL piperidine was added and the mixture was stirred at 30°C for 12 h, dried under vacuum and applied to FC (silica gel, 80 g column, eluted with DCM/MetOH 10:0 → 9:1) to give **1c** as a white powder (130 mg, 79%). **1c** (130 mg, 51 mmol) was diluted in 6 mL dichloromethane. 6 mL TFA was added and the mixture was stirred at room temperature for 8 h. The mixture was dried under vacuum to give **1** as a white powder (98 mg, 95%).



Scheme S1: Synthesis of cAE*.

¹H NMR spectra of cAE* is given in Figure S3.

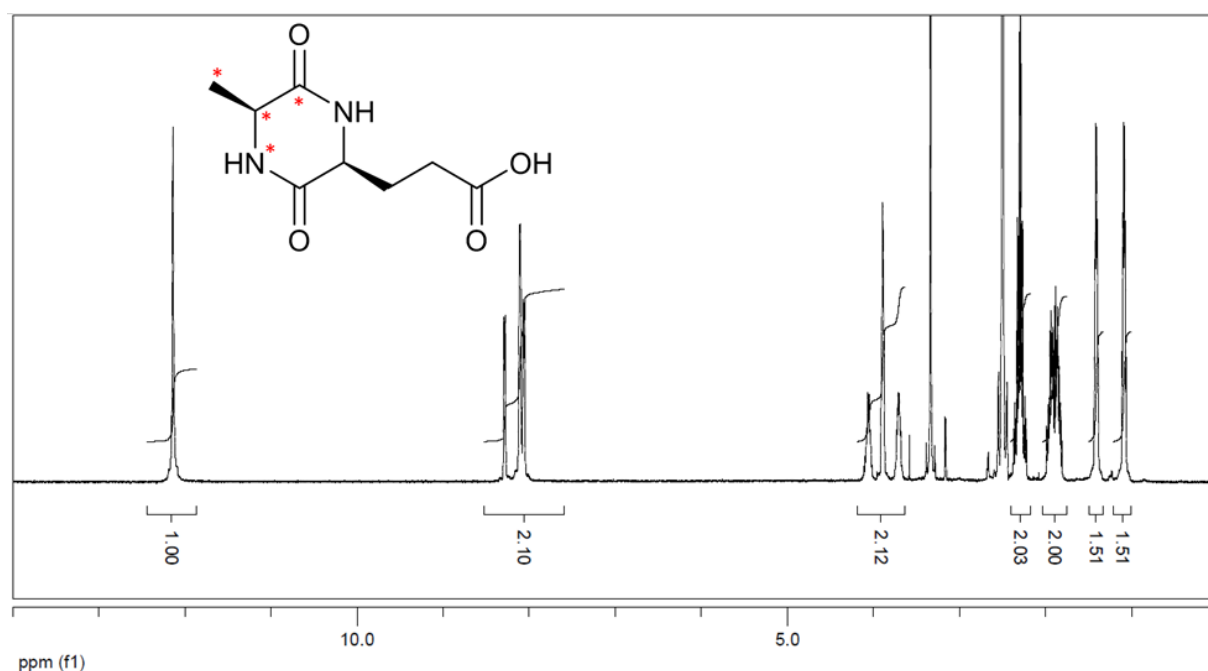
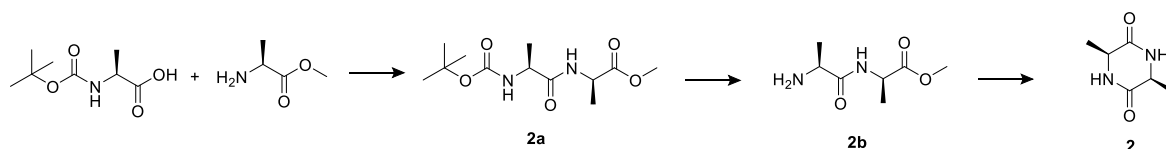


Figure S3: ¹H NMR spectra of cAE*. ¹³C and ¹⁵N atoms are highlighted with a red star.

Synthesis of ^{13}C - ^{15}N -cyclo(Ala-Ala) (cAA*, **2**)

$^{13}\text{C}_3$ - ^{15}N -Boc-L-alanine (50 mg, 0.26 mmol, 1 eq.) was diluted in 5 mL freshly distilled dichloromethane with EDC (100 mg, 0.52 mmol, 2 eq.) and HOBt (80 mg, 0.52 mmol, 2 eq.). L-alanine methyl ester hydrochloride (47.5 mg, 0.34 mmol, 1.3 eq.) and DIPEA (143 μL , 0.78 mmol, 3 eq.) were added under argon and the mixture was stirred at 0°C for 1 h then at room temperature for 12 h. The mixture was dried under vacuum, suspended in dichloromethane, washed with water three times, dried over magnesium sulphate and dried under vacuum to afford a white solid. The residue was applied to FC (silica gel, 12 g column, eluted with cyclohexane/ethyl acetate 9:1 \rightarrow 5:5) to give **2a** as a white powder (47 mg, 50%). **2a** was diluted in 10 mL formic acid and stirred at room temperature for 8 h to yield **2b** (29 mg, quantitative). **2b** was resuspended in 10 mL butan-2-ol and heated to reflux for 24 h to yield **2** (22 mg, quantitative).



Scheme S2: Synthesis of cAA*.

^1H NMR spectra of cAA* is given in Figure S4.

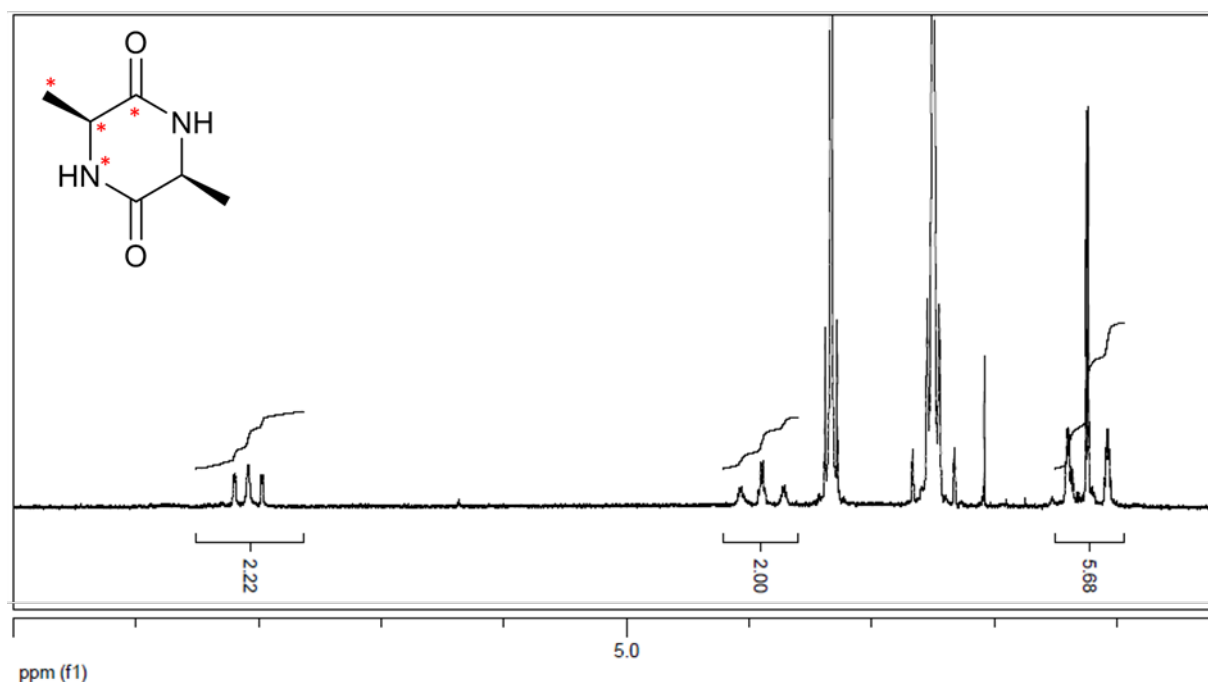


Figure S4: ^1H NMR spectra of cAA*. ^{13}C and ^{15}N atoms are highlighted with a red star.

ii. LC-MS quantification of cAE and cAA

Once LC method was optimized for separation of cAA and cAE, UV-calibration curves was established using commercial cAA and cAE. These calibration curves were used to titrate with accuracy solutions of labelled cAA* and cAE*. These standard solutions were used to add known concentrations of cAA* and cAE* to each sample. Concentrations of cAA and cAE were then determined using the ratio of the EIC peak area of the analyte and the standard (Figure S5). The detection threshold of this method is about 0.3 pmol, which corresponds to a concentration of 3 nM in 100 μ L injected sample and accurate quantification requires at least about 1 pmol, which corresponds to a concentration of 10 nM in 100 μ L sample.

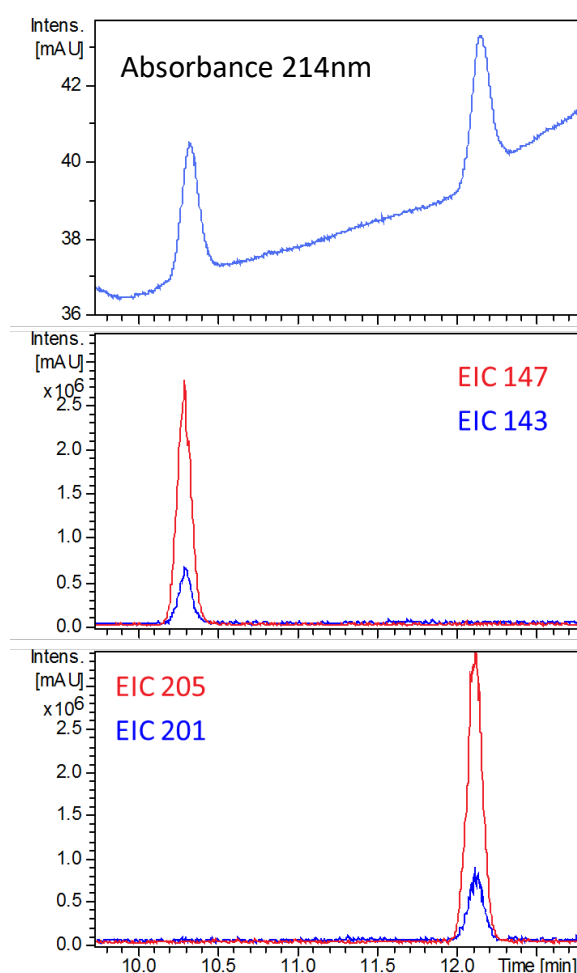


Figure S5: LC-MS chromatograms used for cyclodipeptide quantification. Chromatograms correspond to the injection of 100 μ L of a sample containing 227 nM stable isotope labelled cAA (MW = 146), 223 nM stable isotope labelled cAE (MW = 204), 30nM cAA (MW = 142) and 30 nM cAE (MW = 200).

II) Additional results

A) Optimization of the *in vitro* enzymatic assay for *Nbra*-CDPS

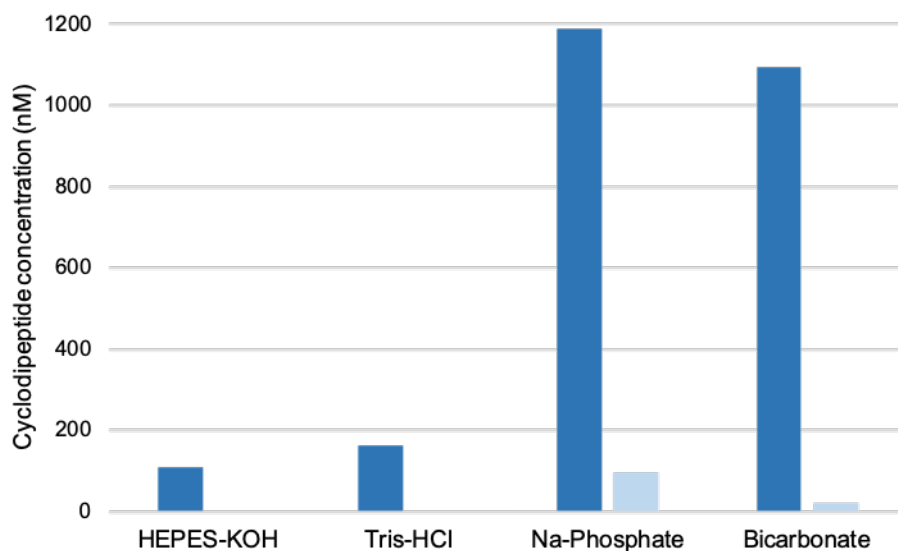


Figure S6: Influence of the nature of buffers on cyclodipeptide production. Enzymatic assays were performed with each buffer at 100 mM and pH 7.0 and 50mM KCl, 15mM MgCl₂, 0.1mM β -mercaptoethanol. Substrates concentrations were 1500 nM Ala-tRNA^{Ala}_{TGC} and 1500 nM Glu-tRNA^{Glu}. Reactions were started by addition of 100 nM *Nbra*-CDPS, incubated for 10 min at 30°C, quenched with 2% TFA before being analysed by LC-MS. cAE and cAA production are in dark and light blue, respectively. This experiment was repeated several times with slightly different conditions and gave similar results.

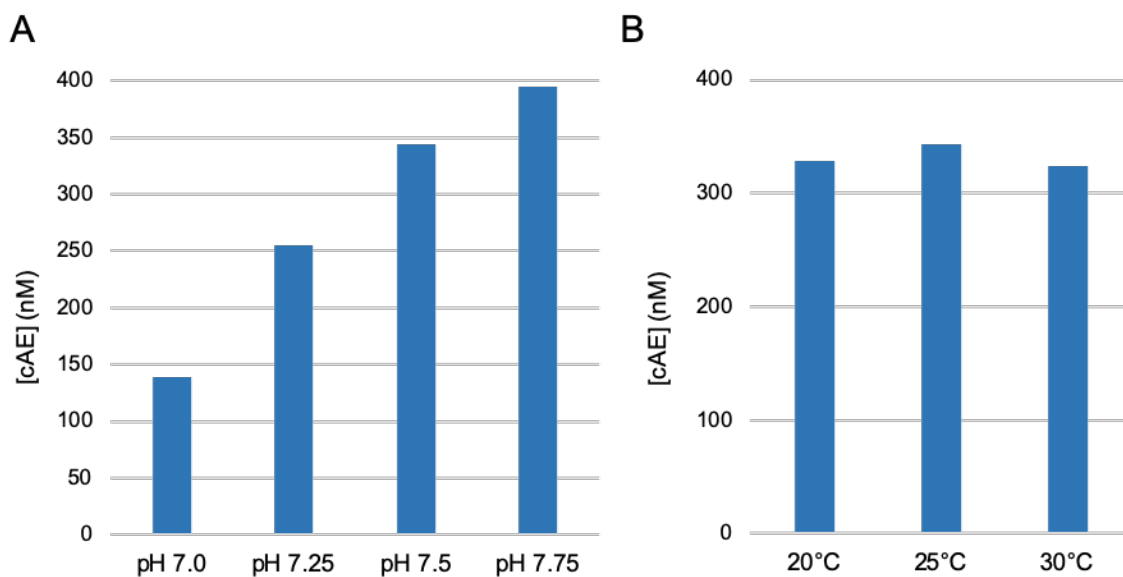


Figure S7: Influence of pH and temperature on cAE production. A) Enzymatic assays were performed in 100 mM Na-Phosphate buffer with pH values varying between 7.00 and 7.75, 50mM KCl, 15mM MgCl₂, 0.1mM β-mercaptoethanol. Substrates concentration were 1000 nM Ala-tRNA^{Ala}_{TGC} and 1000 nM Glu-tRNA^{Glu}. Reactions were started by addition of 10 nM *Nbra*-CDPS, incubated for 20 min at 30°C and were quenched with 2% TFA before being analysed by LC-MS. B) Influence on temperature was tested with similar conditions as in A, except that the pH buffer was set at 7.5 and temperature of incubation varied from 20°C and 30°C. cAA concentration was less than 5nM in all assays.

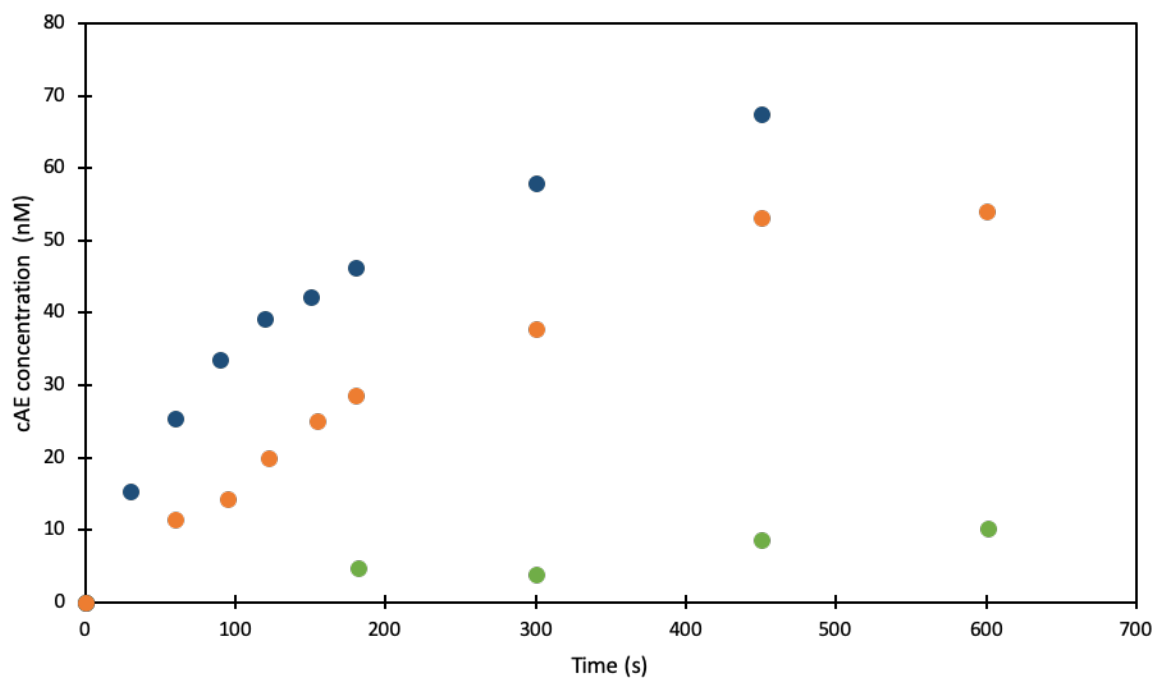


Figure S8: Influence of the enzyme concentration on cAE production. Enzymatic assays were performed in 100 mM Na-Phosphate buffer pH 7.5, 50mM KCl, 15mM MgCl₂, 0.1mM β-mercaptoethanol. Substrates concentrations were 200 nM Ala-tRNA^{Ala}_{TGC} and 200 nM Glu-tRNA^{Glu}. They were started by addition of various concentrations of *Nbra*-CDPS (green, orange and blue correspond to 1 nM, 5 nM and 10 nM *Nbra*-CDPS, respectively), incubated at 20°C and were quenched with 2% TFA before being analysed by LC-MS. No cAA production was observed.

B) Additional kinetics using AA-tRNA substrates

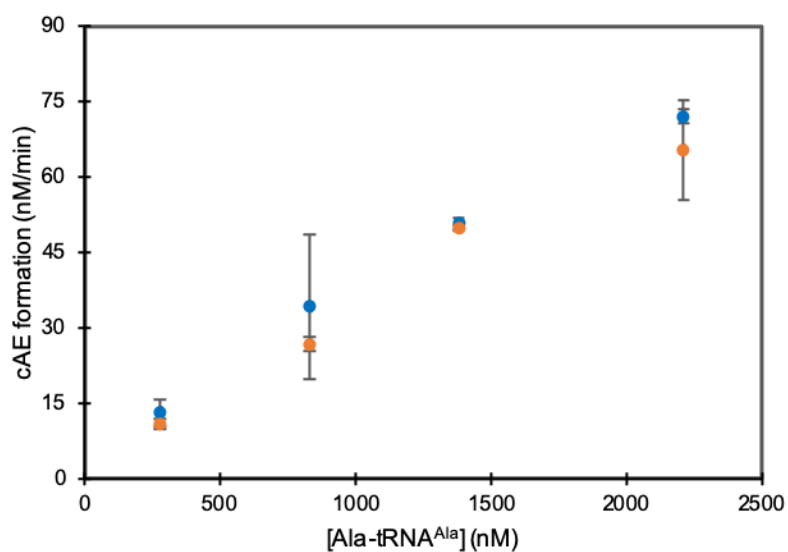


Figure S9: Comparison of *Nbra*-CDPS activity on the two tRNA^{Ala} isoacceptors. Initial velocities of cAE production were measured using 500 nM Glu-tRNA^{Glu} and the indicated concentrations of Ala-tRNA^{Ala}_{TGC} (blue) and Ala-tRNA^{Ala}_{GGC} (orange). Error bars correspond to the standard errors between duplicates.

Figure S10: Cyclodipeptides concentrations curves for initial velocities determination. This figure will be prepared and introduced shortly.

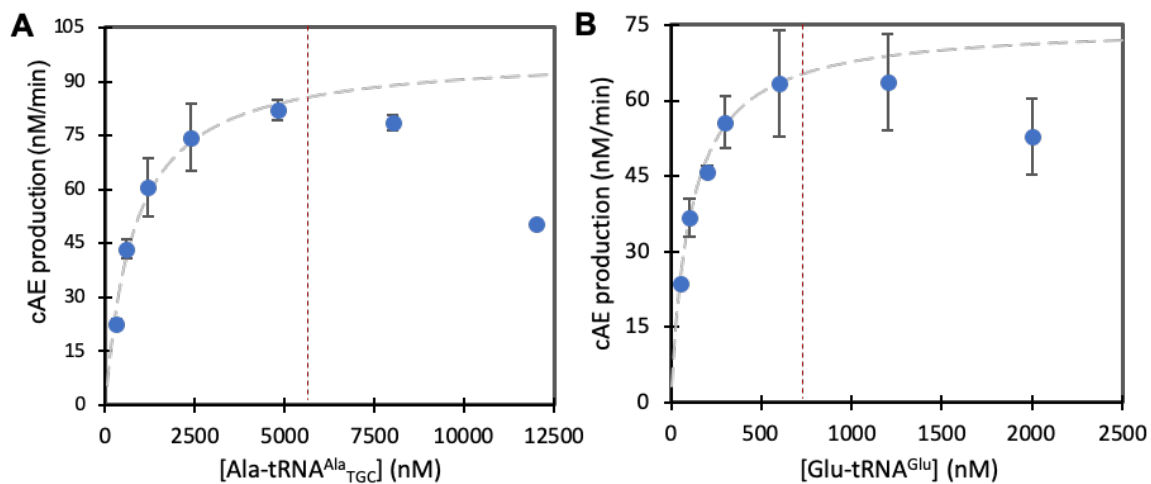


Figure S11: Fitting of cAE production to Michaelis-Menten equation. As in Figure 3, initial velocities are shown for A) increasing concentrations of Ala-tRNA^{Ala}_{TGC} in the presence of 600 nM Glu-tRNA^{Glu} and B) increasing concentrations of Glu-tRNA^{Glu} in the presence of 1200 nM Ala-tRNA^{Ala}_{TGC}. Error bars correspond to the standard errors between duplicates. Pre-saturation initial velocities of cAE production (the 5 points located at the left of the red dotted lines on the graphs in both cases) were fitted to Michael-Menten equation using the nonlinear regression module of the Prism software. Estimated values were $K_M = 805 (\pm 120)$ nM and $V_{max} = 98 \text{ nM}\cdot\text{min}^{-1}$ for Ala-tRNA^{Ala}_{TGC} and $K_M = 110 (\pm 10)$ nM and $V_{max} = 75 \text{ nM}\cdot\text{min}^{-1}$. Corresponding curves are drawn in grey dotted lines.

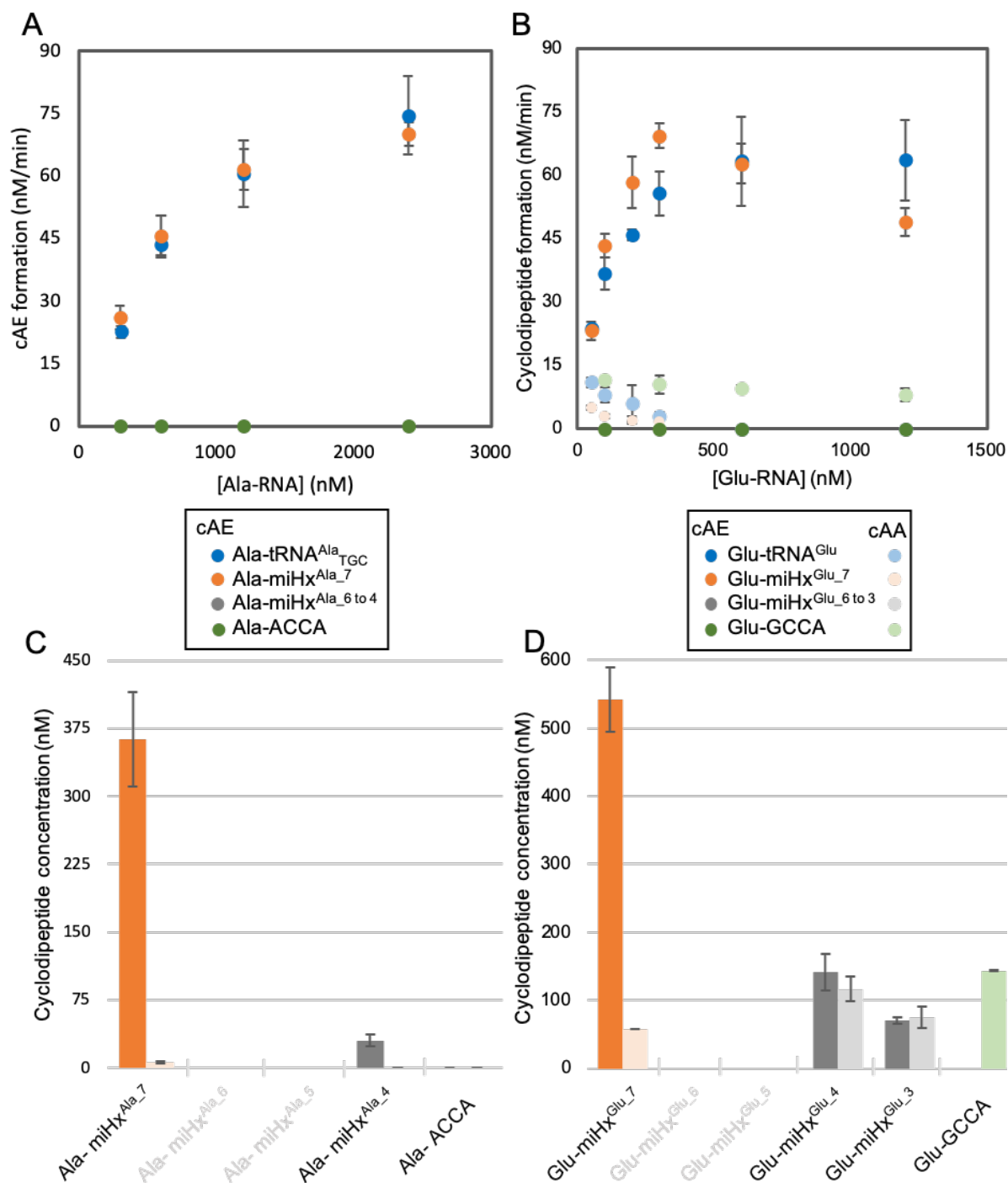


Figure S12: Kinetics study of Nbra-CDPS towards AA-minitRNA substrates (with cAA production). Initial velocities of cAE were measured with A) 600 nM Glu-tRNA^{Glu} and varying concentrations of Ala-minitRNA^{Ala} and B) 1200 nM Ala-tRNA^{Ala}_{TGC} and varying concentrations of Glu-tRNA^{Glu}. End-point assays were performed with C) 600 nM Glu-tRNA^{Glu} and varying concentrations of Ala-minitRNA^{Ala} and D) 1200 nM Ala-tRNA^{Ala}_{TGC} and varying concentrations of Glu-minitRNA^{Glu}. Error bars correspond to the standard errors between duplicates.

C) Prediction of conformation stabilities of miHxs with decreasing stem sizes

Table S5: Sequences and calculated Gibson free energies of miHxs with decreasing stem sizes. ΔG of the two most stable hairpins and of the most stable homodimer were calculated using the OligoAnalyzer tool of IDT (<https://eu.idtdna.com/calc/analyzer>). Lines colored in green indicates sequences for which the most stable predicted hairpin corresponds to the correct conformation (i.e. with a 4-mer NCCA overhang and a 4-mer UUCG loop). Lines in orange indicate sequences for which the most stable hairpin does not correspond to the correct conformation or for which homodimer are predicted to be more stable than the most stable hairpin.

	Sequence	ΔG (kcal.mole ⁻¹)		
		Most stable hairpin	Second most stable hairpin	Most stable homodimer
miHx ^{Ala} ₇	GGGGCUAUUCGUAGCUCCACCA	-14.5	-10.3	-6.34
miHx ^{Ala} ₆	GGGGCUUUCGAGCUCCACCA	-12.6	-8.4	-9.49
miHx ^{Ala} ₅	GGGGCUUCGGCUCCACCA	-12	-7.8	-3.61
miHx ^{Ala} ₄	GGGGUUCGCUCCACCA	-7.4	-4.4	-4.41
miHx ^{Ala} ₃	GGGUUCGUCCACCA	-5.1	-4	-4.41
miHx ^{Ala} ₂	GGUUCGCCACCA	-1.15	-0.48	-4.41
miHx ^{Ala} ₁	GUUCGCACCA	1.41	/	-3.61
miHx ^{Glu} ₇	GUCCCCUUCGAGGGGACGCCA	-16.6	-11.1	-13.72
miHx ^{Glu} ₆	GUCCCCUUCGGGGACGCCA	-15.9	-9.5	-12.12
miHx ^{Glu} ₅	GUCCCUUCGGGGACGCCA	-12.6	/	-9.06
miHx ^{Glu} ₄	GUCCUUCGGGACGCCA	-9.4	/	-5.99
miHx ^{Glu} ₃	GUCUUCGGACGCCA	-6.1	/	-3.61
miHx ^{Glu} ₂	GUUUCGACGCCA	-2.1	/	-6.76
miHx ^{Glu} ₁	GUUCGCGCCA	2.53	/	-10.36

Comment on Table S5.

For miHx^{Ala} with stems sizes of 7 to 4 bp and miHx^{Glu} with stems sizes of 7 to 3 bp, the correct conformation was predicted to be the most stable and the predicted ΔG of the correct conformation was several kcal.mole⁻¹ lower than for other conformations. For miHx^{Ala}₃, the correct conformation was predicted to be the most stable but the predicted ΔG of the 3 different conformations were close (less than 1,1 kcal.mole⁻¹). As miHx^{Ala}₃ exhibited an atypical behaviour in anion-exchange chromatography under non-denaturing conditions (Figure S16), likely to be due to misfolding, it was excluded from the study. For miHx^{Ala} and miHx^{Glu} with stems sizes of 1 and 2 bp, the stability of hairpin was predicted to be very low and homodimers were predicted to be more stable and they were excluded from the study.

D) Production of aminoacylated minitRNAs

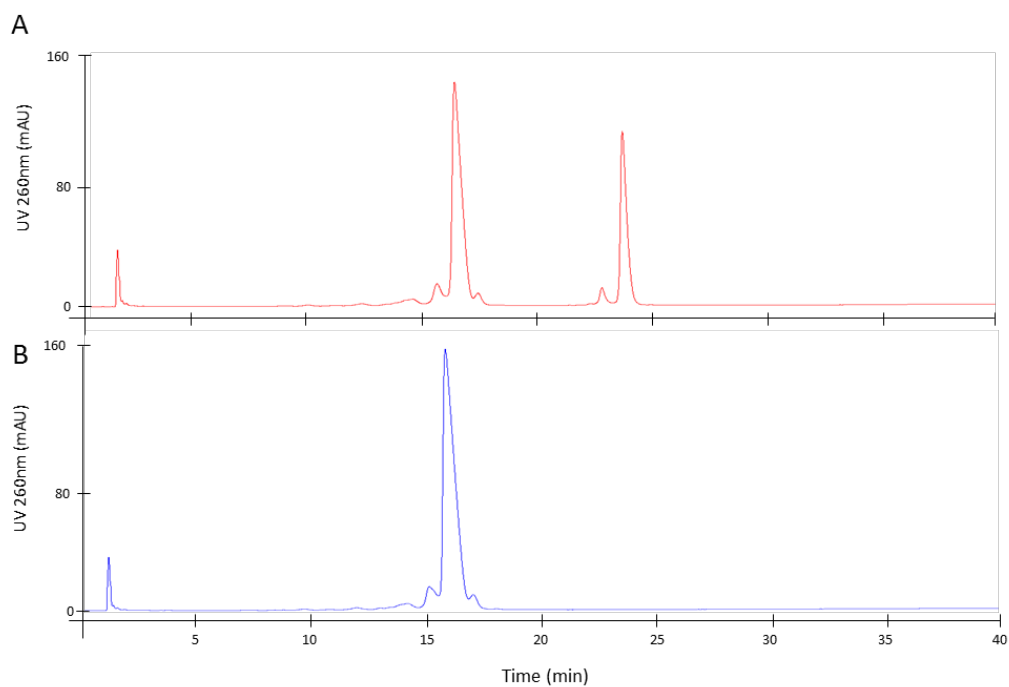


Figure S13: Effect of heat and quick cool-renaturation on the HPLC profile of miHxAla₇. Chromatogram correspond to the injection on analytical DNAPAC column using gradient 2 of 5 μ g purified miHx^{Ala-7} before (A) and after (B) heat and quick-cool renaturation (80°C for 5 min then quickly cooling to 4°C using a PCR thermal cycler). Reverse conversion of one form to the other was commonly observed during freeze/thaw cycles, as has been reported for such molecules (11).

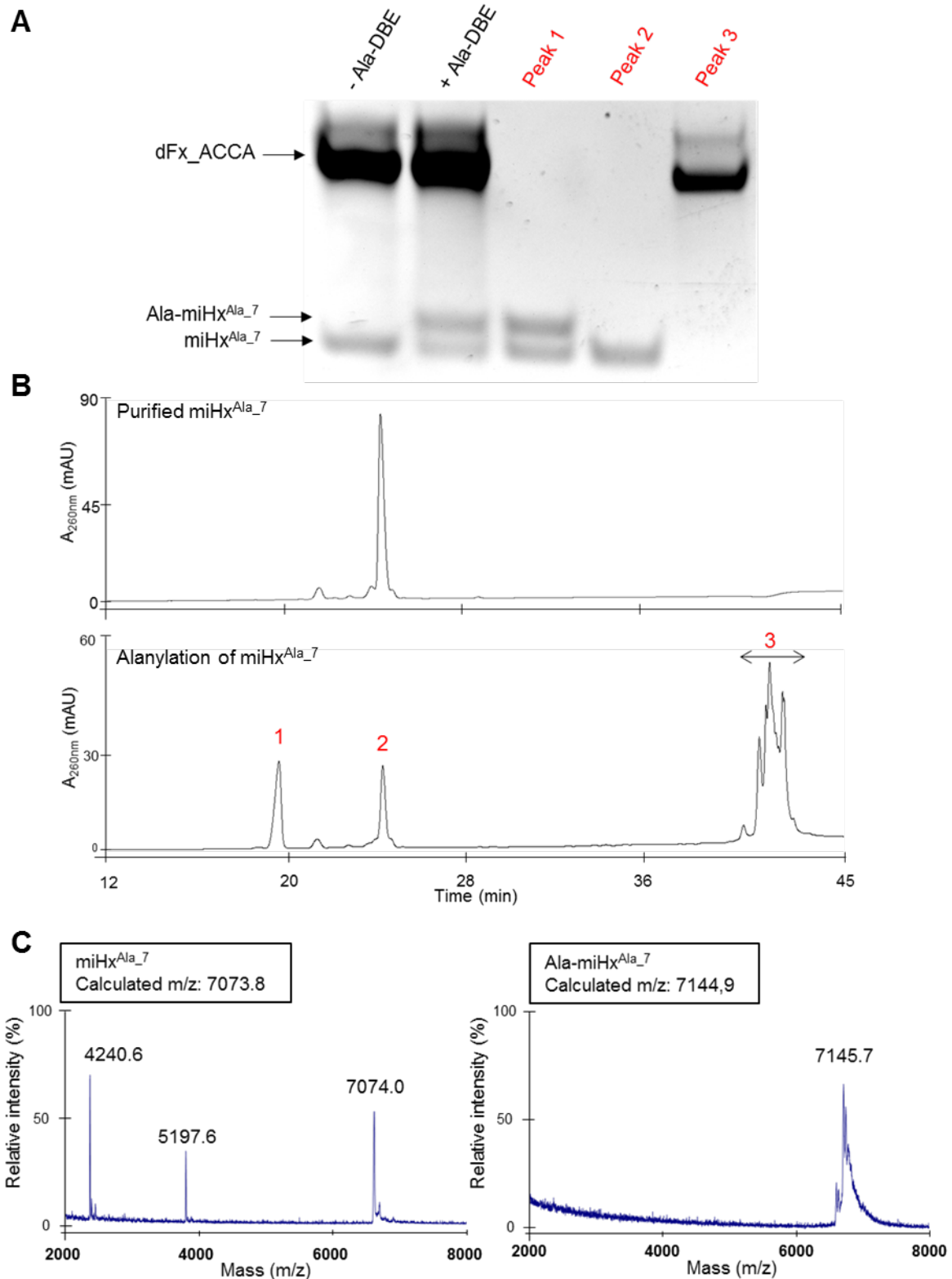


Figure S14: Analysis of miHx^{Ala_7} alanylation reaction. A) Acid-PAGE analysis. Lane 1: Aminoacylation reaction with no Ala-DBE added (250 ng miHx^{Ala_7}). Lane 2: Aminoacylation reaction with Ala-DBE added (250 ng miHx^{Ala_7}). Lane 3 to 5: Compounds contained in peak 1 to 3 from the corresponding chromatogram below, purified and lyophilized (250 ng total RNA in each case). B) Chromatograms corresponding to the injection on analytical DNAPAC column of (from top to bottom) 1 μ g purified miHx^{Ala_7} and alanylation of 1 μ g miHx^{Ala_7} by dFx using gradient 4. C) MALDI-TOF spectra of purified miHx^{Ala_7} and Ala-miHx^{Ala_7}.

Comment on Figure S14.

Figure S14A shows the acid-PAGE-analysis of the alanylation reaction of miHx^{Ala-7} by dFx and its purification using DNAPAC column. Lanes 1 and 2 correspond respectively to the alanylation of miHx^{Ala-7} by dFx without and with the activated amino acid, Ala-DBE. An additional band, corresponding to Ala-miHx^{Ala-7}, was visible when Ala-DBE was present in the reaction. By comparing the intensity of the bands corresponding to Ala-miHx^{Ala-7} and miHx^{Ala-7}, the aminoacylation yield was evaluated to be about 60%.

Figure S14B shows the chromatograms obtained after the injections on the DNAPAC column of the purified miHx^{Ala-7} (up) and the alanylation reaction (down). For the alanylation reaction, three major peaks were observed. The eluted corresponding compounds were purified, lyophilised and analysed by acid-PAGE in Lanes 3 to 5 of Figure S14A. Peak 2 contains only non-acylated miHx^{Ala-7}, which is coherent since its retention time is the same as purified miHx^{Ala-7} (Figure S14B, up). Peak 1 contains about 65% Ala-miHx^{Ala-7} but also 35% deacylated miHx^{Ala-7}. The presence of about 35% deacylated miHx^{Ala-7} results from deacylation during the purification process. Such deacylation was reduced during subsequent purifications by optimizing the purification pipeline (cooling of the purification system, addition of a desalting step before lyophilisation, storage at -80°C after lyophilisation) so that purified RNAs were obtained acylated up to 80% with Ala and 90% with Glu. Peak 3 corresponds to dFx. A minor contaminant, appearing as a faint band above the band of dFx was observed. Indeed, total purification of flexizymes could not be obtained on the DNAPAC column.

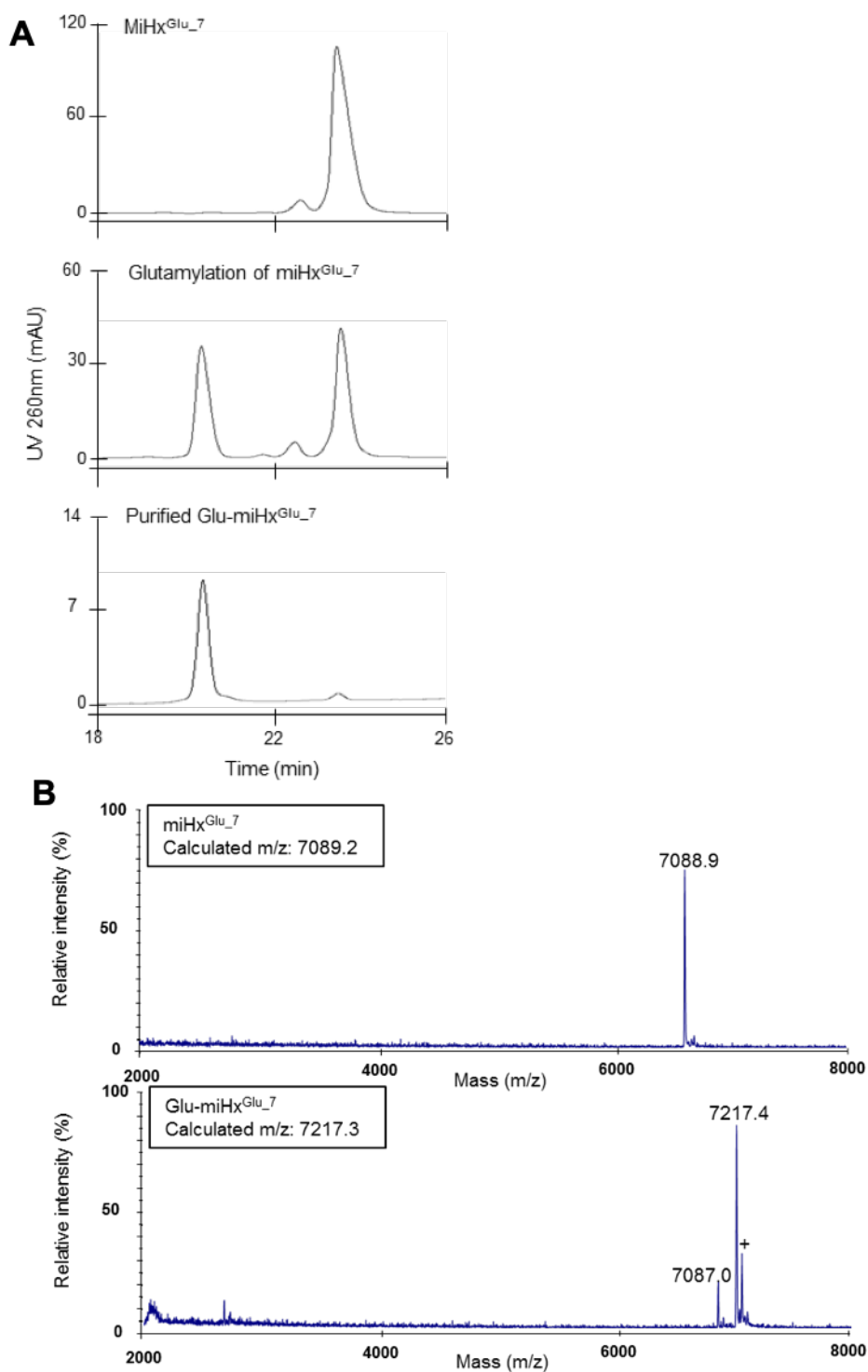


Figure S15: Analysis of miHx^{Glu}₇ glutamylation reaction. A) Chromatograms corresponding to the injection on analytical DNAPAC column of (from top to bottom) 2 µg miHx^{Glu}₇, glutamylation of 1 µg miHx^{Glu}₇ by dFx and 0,25 µg Glu-miHx^{Glu}₇ using gradient 4. Deacylation in the Glu-miHx^{Glu}₇ sample was less than 5%. B) MALDI-TOF spectra of purified miHx^{Glu}₇ and Glu-miHx^{Glu}₇.

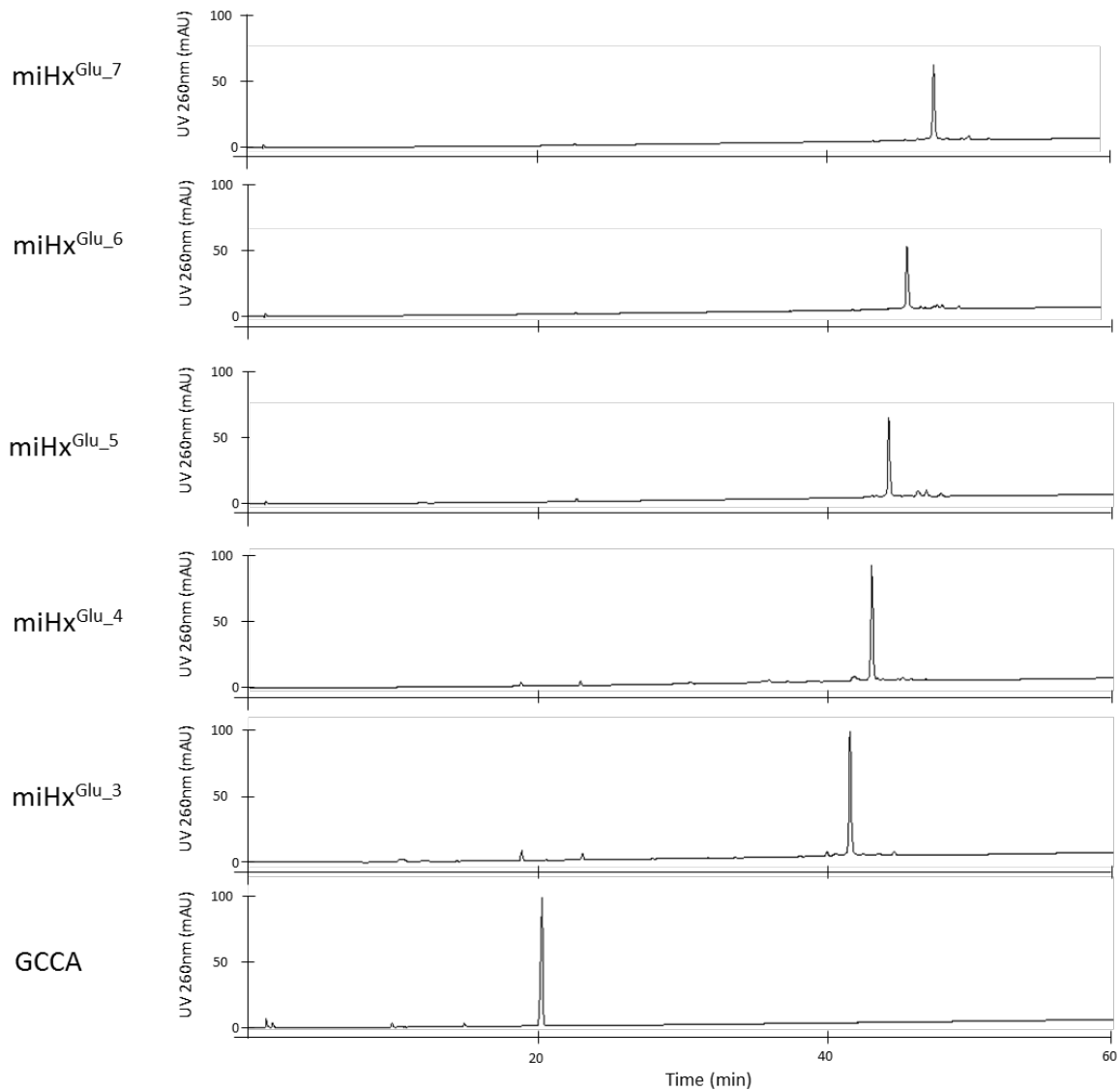


Figure S16: Comparison of HPLC profiles for all minitRNAs. Chromatograms corresponding to the injection on analytical DNAPAC column using gradient 1 of 1 μ g of purified minitRNAs^{Glu} of different sizes, after heat and quick-cool renaturation.

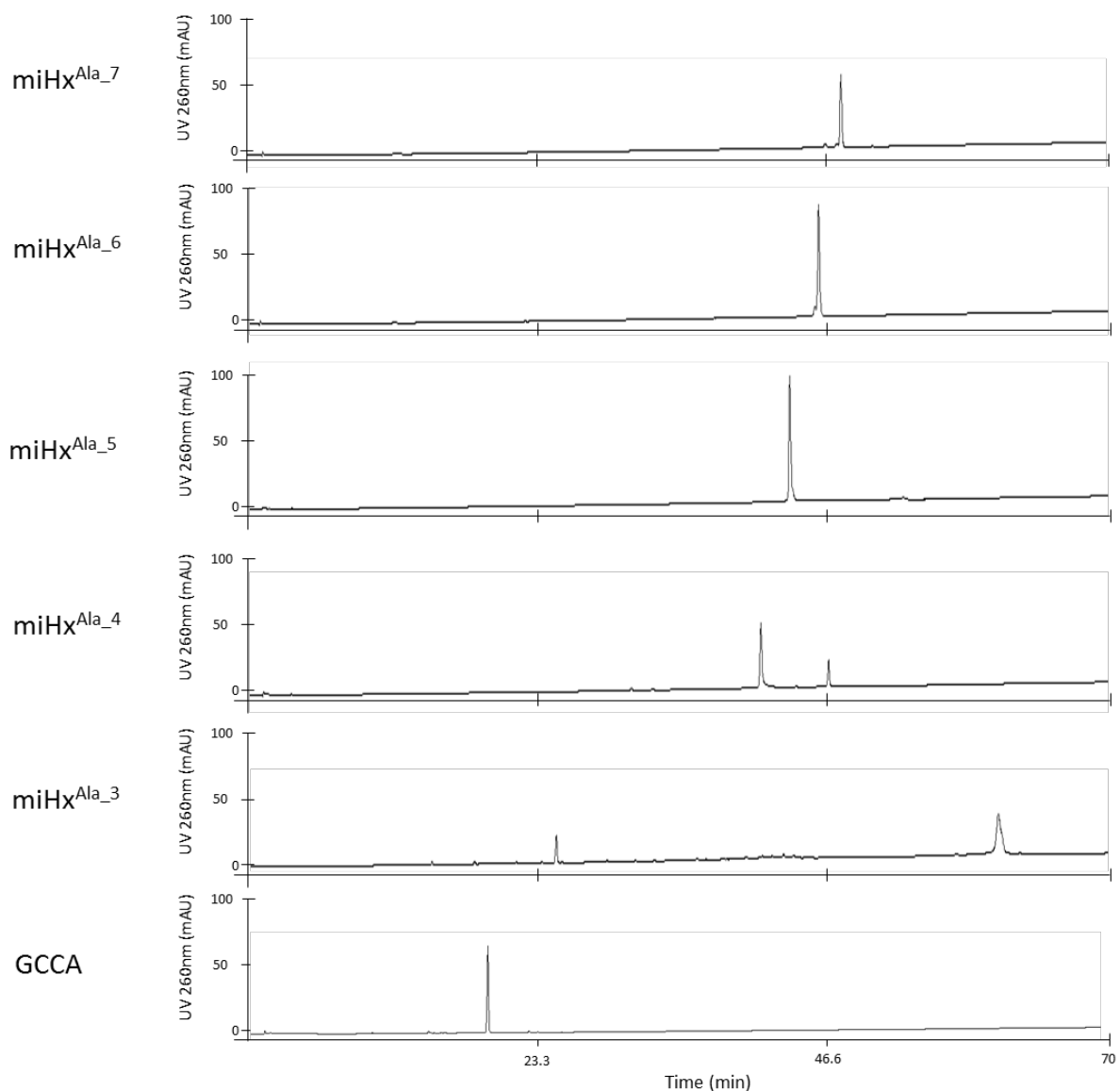


Figure S16 (continued): Chromatograms corresponding to the injection on analytical DNAPAC column using gradient 1 of 1 μ g of purified minitRNAs^{Glu} of different sizes, after heat and quick-cool renaturation. The profile corresponding to the injection of miHx^{Ala_3} is atypical compared to the others, presumably because of misfolding (the peak eluting at a late elution time might correspond to homodimers).

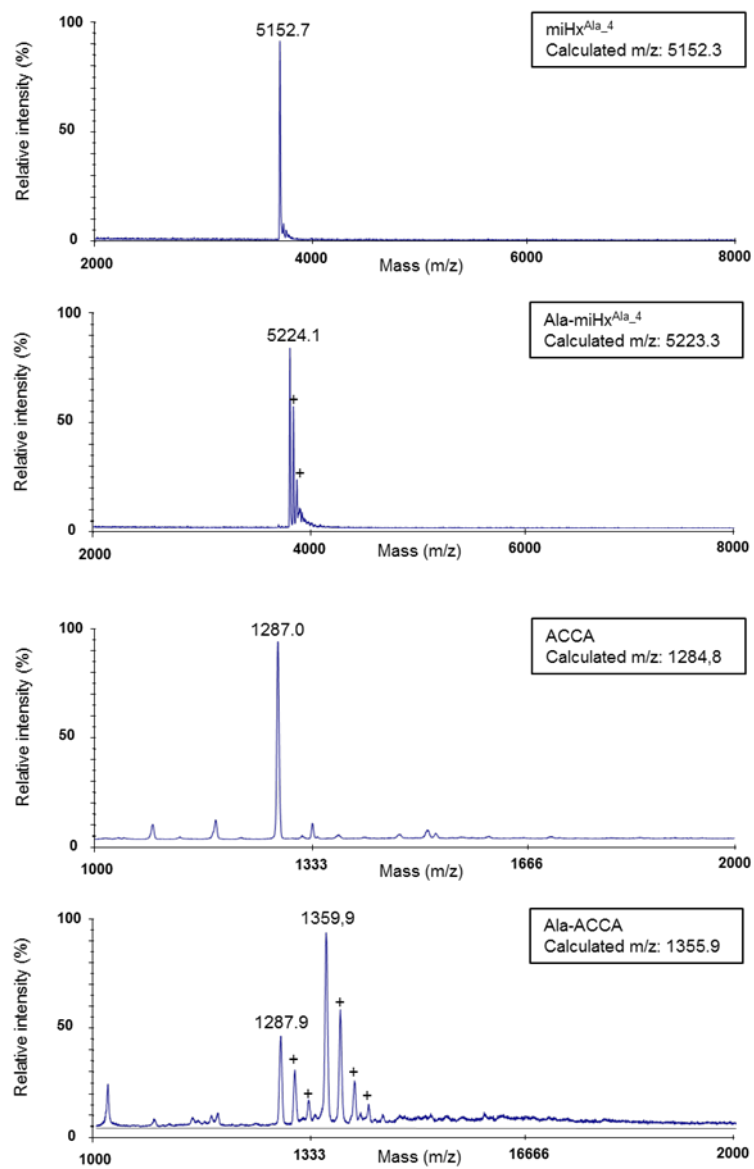


Figure S17: MALDI-TOF spectra of minitRNAs and AA-minitRNAs. Peak corresponding to putative sodium adducts (increment of mass of about 22) are indicated with a +.

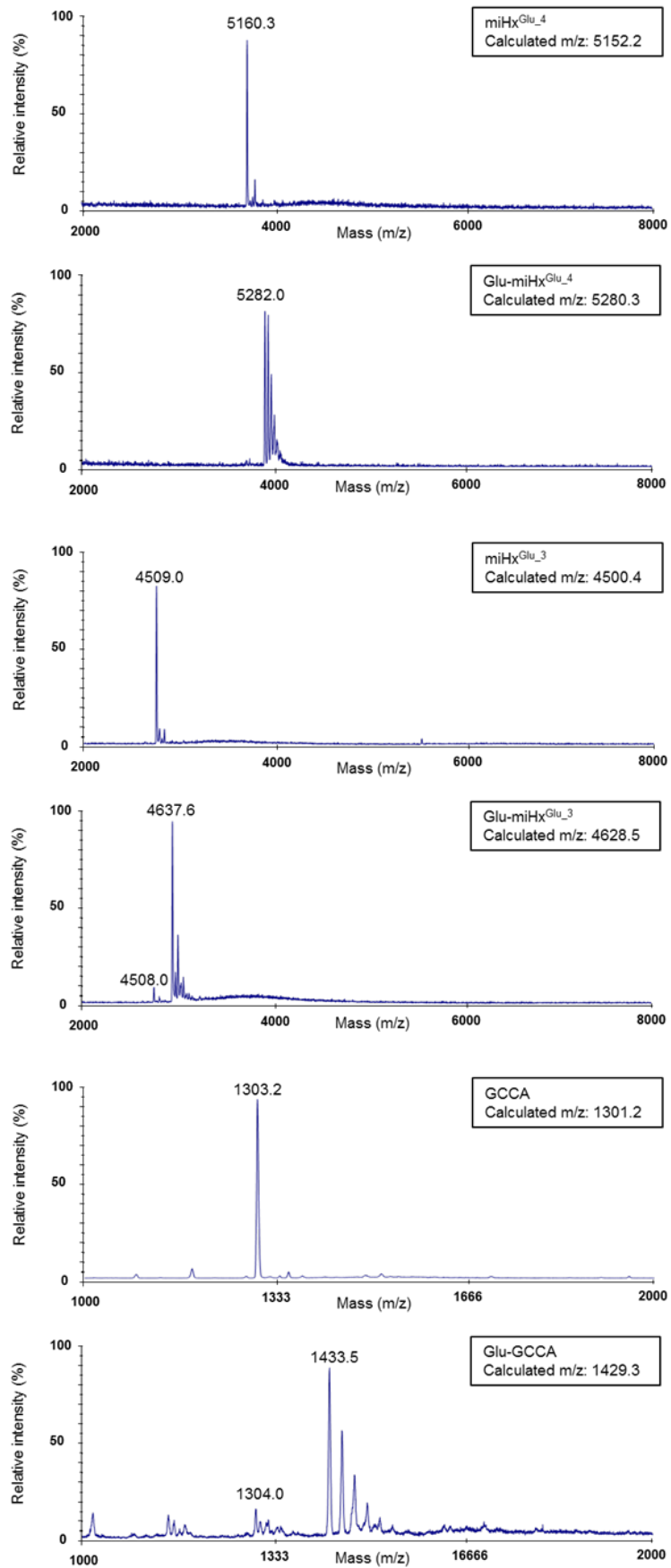


Figure S17 (continued)

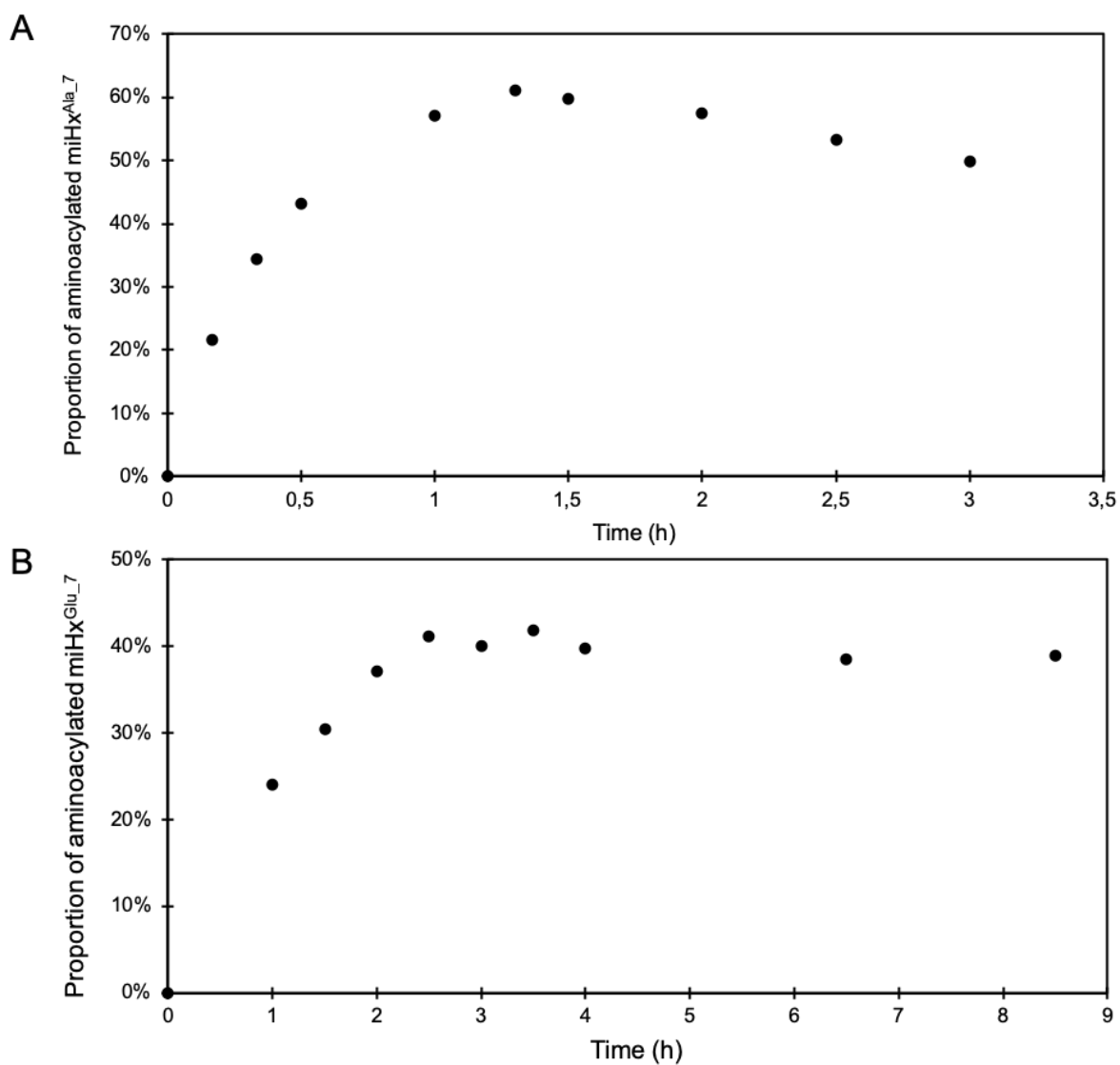


Figure S18: Acylation kinetics of minitRNAs by flexizymes. Flexizyme-mediated alanylation of miHx^{Ala_7} and glutamylation of miHx^{Glu_7} were prepared as indicated in the Material and Methods section. Aliquots of 250 ng were taken at several time points and analysed by anion-exchange chromatography on DNAPAC using gradient number 4. Proportion of aminoacylated miHx were determined using the ratio of the peaks corresponding to AA-miHx and miHx. Note that kinetics of aminoacylation by flexizymes have been shown not to be dependent on the size of the RNA acceptor (12).

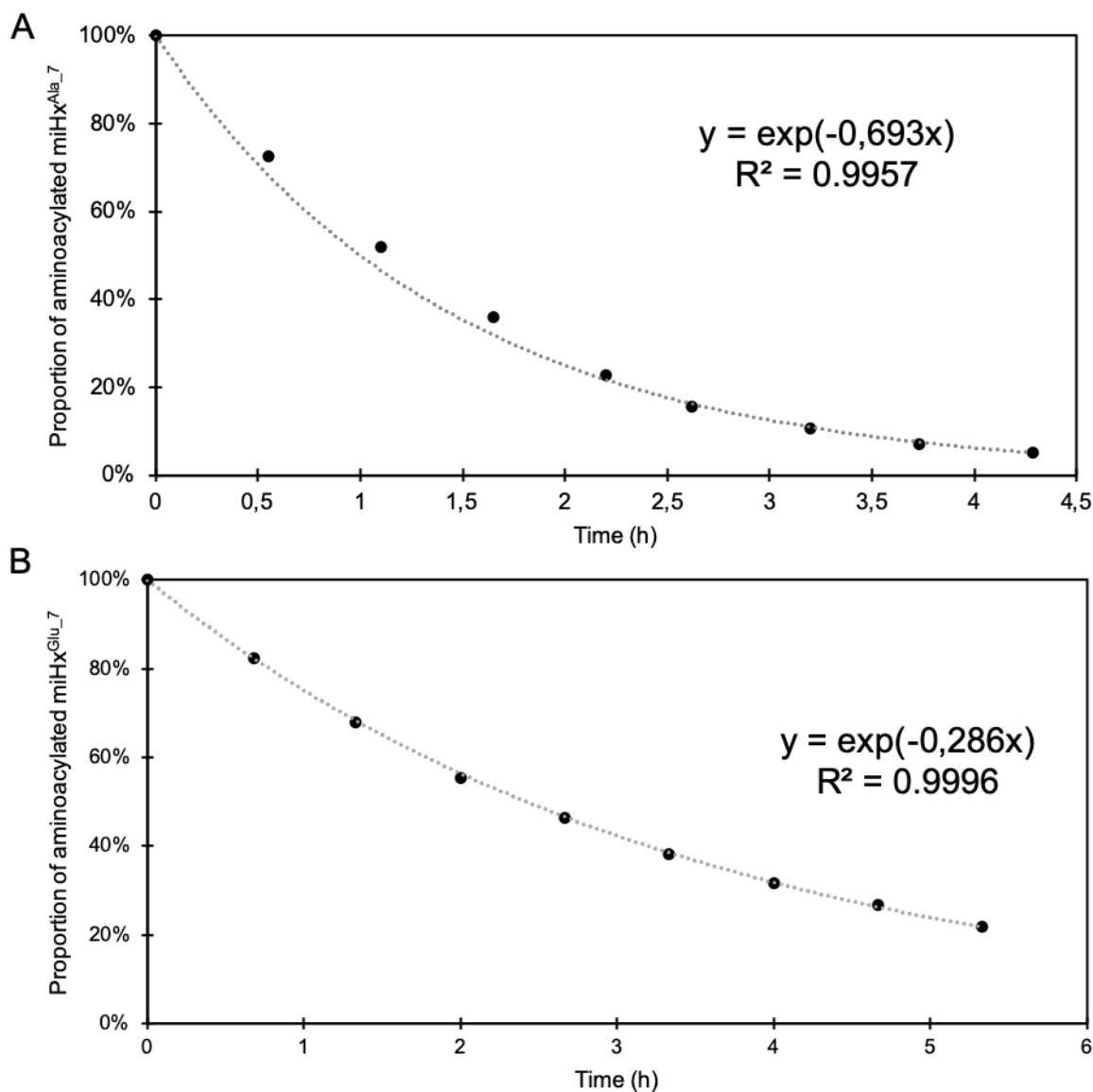


Figure S19: Deacylation kinetics of AA-miRNAs. Starting from purified Ala-miHx^{Ala_7} or Glu-miHx^{Glu_7}, solutions containing 1 μ M AA-miHx were prepared and incubated at 20°C in the conditions of the cyclodipeptide production assay (100 mM phosphate-sodium buffer pH 7.5, 50 mM KCl, 15 mM MgCl₂ and 0.1 mM β -mercaptoethanol). Aliquots of 250 ng were taken at several time points and analysed by anion-exchange chromatography on DNAPAC using gradient number 4. Proportion of aminoacylated miHx were determined using the ratio of the peaks corresponding to AA-miHx and miHx. Proportion of aminoacylated miHx are normalized to the initial value at T₀. Curves were fitted to an exponential decay using Excel.

Table S7: Predicted deacylation of AA-miRNAs. Half-lives and predicted deacylation after 2 and 30 min for Ala-miHx^{Ala_7} and Glu-miHx^{Glu_7} were calculated using the corresponding exponential decay constant.

	Ala-miHx ^{Ala_7}	Glu-miHx ^{Glu_7}
Exponential decay constant (h ⁻¹)	0.693	0.286
Half-life (min)	60	145
Deacylation after 2 min (%)	2.3	1
Deacylation after 30 min (%)	29.3	13.3

References

1. Beuning, P.J., Gulotta, M. and Musier-Forsyth, K. (1997) Atomic group 'mutagenesis' reveals major groove fine interactions of a tRNA synthetase with an RNA helix. *J. Am. Chem. Soc.*, **119**, 8397–8402.
2. Dubois, D.Y., Blais, S.P., Huot, J.L. and Lapointe, J. (2009) AC-truncated glutamyl-tRNA synthetase specific for tRNA^{Glu} is stimulated by its free complementary distal domain: Mechanistic and evolutionary implications. *Biochemistry*, **48**, 6012–6021.
3. Jacques, I.B., Moutiez, M., Witwinowski, J., Darbon, E., Martel, C., Seguin, J., Favry, E., Thai, R., Lecoq, A., Dubois, S., *et al.* (2015) Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity. *Nat. Chem. Biol.*, **11**, 721–731.
4. Boccaletto, P., MacHnicka, M.A.A., Purta, E., Pitkowski, P., Baginski, B., Wirecki, T.K.K., De Crécy-Lagard, V., Ross, R., Limbach, P.A.A., Kotter, A., *et al.* (2018) MODOMICS: A database of RNA modification pathways. 2017 update. *Nucleic Acids Res.*, **46**, D303–D307.
5. Villet, R., Fonvielle, M., Busca, P., Chemama, M., Maillard, A.P., Hugonnet, J.E., Dubost, L., Marie, A., Josseume, N., Mesnage, S., *et al.* (2007) Idiosyncratic features in tRNAs participating in bacterial cell wall synthesis. *Nucleic Acids Res.*, **35**, 6870–6883.
6. Saito, H., Watanabe, K. and Suga, H. (2001) Concurrent molecular recognition of the amino acid and tRNA by a ribozyme. *RNA*, **7**, 1867–78.
7. Yang, C.H. (2010) Detection and Sequencing of MicroRNA using MALDI Time-of-Flight Mass Spectrometry.
8. Stokvis, E., Rosing, H. and Beijnen, J.H. (2005) Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography / mass spectrometry : necessity or not ? *Rapid Commun. Mass Spectrom.*, **19**, 401–407.
9. Deigin, V., Ksenofontova, O., Khrushchev, A., Yatskin, O., Goryacheva, A. and Ivanov, V. (2016) Chemical Platform for the Preparation of Synthetic Orally Active Peptidomimetics with Hemoregulating Activity. *ChemMedChem*, 10.1002/cmdc.201600157.
10. Jeedigunta, S., Krenisky, J.M. and Kerr, R.G. (2000) Diketopiperazines as advanced intermediates in the biosynthesis of ecteinascidins. *Tetrahedron*, **56**, 3303–3307.
11. Sun, X., Li, J.M. and Wartell, R.M. (2007) Conversion of stable RNA hairpin to a metastable dimer in frozen solution. *Rna*, **13**, 2277–2286.
12. Fujino, T., Kondo, T., Suga, H. and Murakami, H. (2019) Exploring of minimal RNA substrate of flexizymes. *ChemBioChem*, **8603**, 1–8.
13. Peacock, J.R., Walvoord, R.R., Chang, A.Y. and Kozlowski, M.C. (2014) Amino acid – dependent stability of the acyl linkage in aminoacyl-tRNA Amino acid – dependent stability of the acyl linkage in aminoacyl-tRNA. *RNA*, **20**, 758–764.
14. Stepanov, V.G. and Nyborg, J. (2002) Thermal stability of aminoacyl-tRNAs in aqueous solutions.
15. Schuber, F. and Pinck, M. (1974) On the chemical reactivity of aminoacyl-tRNA ester bond. I - Influence of pH and nature of the acyl group on the rate of hydrolysis. *Biochimie*, **56**, 383–390.
16. Hentzen, D.L.E., Mandel, P. and Garel, J. (1972) Relation between aminoacyl-tRNA stability and the fixed amino acid. *Biochim. Biophys. Acta - Gen. Subj.*, **281**, 228–232.

IV) Discussion and perspectives

The study presented in the second chapter of this manuscript required an interdisciplinary approach, at the crossroads between synthetic chemistry, molecular biology, biochemistry and analytical chemistry. The use of flexizyme-aminoacylated minitRNAs represents a new application of this technology, which will hopefully be transposed beyond the field of CDPSs. The results obtained allowed us to move forward in our analysis of tRNA regioselectivity. However, many questions remain unanswered and our flexizyme-based approach holds great promise for further studies.

For reasons of time, we could only tackle the issue of the size of the substrates and the role of tRNA sequences for CDPS activity and specificity was not studied in the frame of my PhD. However, once the size of minimal substrates for the binding sites of *Nbra*-CDPS is characterized, flexizymes could be used to easily generate sets of aminoacylated miHx mutants. The enzymological assay that we developed could be used to test the impact of these mutations in the acceptor arms of both substrates on *Nbra*-CDPS activity. One particular lead to investigate is the impact of using *E. coli* tRNAs instead of the host native tRNAs. The sequences of the acceptor arms of Ala isoacceptors of *E. coli* and *Nocardia brasiliensis* are almost similar (two tRNA^{Ala} isoacceptors of *N. brasiliensis* have the exact same acceptor arm sequence as the tRNA^{Ala} isoacceptors *E. coli* and only one has a few variations in the sixth and seventh distal base pairs). On the other hand, several important differences can be observed between the sequences of the acceptor arm of tRNA^{Glu} of *E. coli* and *N. brasiliensis* (the discriminator base, the first two base pairs and the seventh distal base pair is different). The fact that *Nbra*-CDPS appears active *in vivo* in *E. coli* and *in vitro* using *E. coli* tRNAs raises questions about the importance of these sequence motifs for activity and specificity. The methods developed during my PhD could be applied to the study of this particular question.

Once optimal minimal substrates (i.e. minimum size and maximal activity) have been characterized, the production pipeline that we develop for AA-minitRNAs can be scaled up to produce the quantities of molecules required for biophysical approaches such as isothermal titration calorimetry, nuclear magnetic resonance or crystallisation. Stable analogues will be

produced thanks to the recent methods proposed by Prof. Suga (introduction of an amido group instead of the labile ester bond).

The perspective of applying flexizymes for a better understanding of CDPS specificity was the basis for a grant request as a Collaborative Research Project (PRC) to the French National Research Agency (ANR). This project, entitled “FlexPep” was laureate of a PRC grant for the 2019 campaign and will therefore be funded for the next four years. The methods developed during my PhD and the results obtained on *Nbra*-CDPS constituted an important part of the strategy presented in the grant request and I was personally involved in its redaction.

The FlexPep projects aims at better understanding CDPS specificity and in particular the interaction between CDPSs and the tRNA moieties of their two substrates. The core idea of the project is to use flexizyme-aminoacylated minitRNAs in biophysical approaches to better understand CDPS/tRNA interaction. My host group, coordinator of the project, will produce AA-RNA analogues and CDPSs and will contribute to the biochemical and biophysical characterization of “AA-RNA:CDPSs” complexes. Partners groups include those of Jean-Christophe Cintrat (CEA/DRF/Joliot/MTS, Bio-organic Chemistry Laboratory) for chemical synthesis of amino acid or nucleotide analogues, Nelly Morellet (ICSN, CNRS UPR 2301, Structural Biology and Chemistry Laboratory) for NMR analyses and Jean-Baptiste Charbonnier (I2BC, UMR 9198, Nuclear Envelope, Telomeres and DNA Repair Laboratory) for crystallography experiments.

GENERAL DISCUSSION AND PERSPECTIVES

CDPSs constitute a good example of how technical advances have contributed to a renewal of our approach of natural product research. The discovery of CDPSs was driven by biochemical analysis and intensive biochemical characterization have been required to characterize the diversity of this family. With these experimental data on hands, bioinformatics analysis can now be used to perform efficient genome mining and in particular to predict the activity of the vast majority of CDPSs. This is a precious asset for the *in silico* analysis of CDPS-containing gene clusters. The description of CDPS diverse activities now seems well advanced, even if new groups with unpredictable activities continue to emerge. Concomitantly, the number of 2,5-DKP-tailoring enzymes described has boomed in the past few years. Therefore, it seems that the field has come to sufficient maturity to consider the manipulation of CDPS-based pathways for the production of new compounds. One of the approaches that can be undertaken is to extend the range of precursors accepted by these pathways. In this context, the fundamental question of my PhD was to extend our understanding of CDPS specificity towards non-natural substrates. Better understanding how CDPS recognize their AA-tRNA substrates appears a mandatory step before the introduction of non-canonical precursors into 2,5-DKP biosynthetic cascades.

In the first part of this manuscript, I used the natural promiscuity of *E. coli* AARSs to screen ncAAs for their ability to be incorporated into 2,5-DKPs by diverse CDPSs. We discovered that CDPSs tolerate a wide range of chemical modifications in the side chains of amino acids which significantly expands the chemical space accessible using 2,5-DKP biosynthetic gene clusters.

In the second part of my PhD, we investigated the requirement for the different part of the tRNA moieties of the two substrates of one CDPSs. Thanks to the application of the flexizyme technology and by using an interdisciplinary approach, we produced several shortened analogues of AA-tRNAs. By using enzymatic assays, we demonstrated that the acceptor arms of tRNAs are the only part required for CDPS activity. The methods developed during this part of my PhD pave the way for interesting further studies, in the field of CDPSs as well as other AA-tRNA-utilizing enzymes.

Globally speaking, the results obtained are encouraging. Natural CDPSs, without any engineering efforts, are able to use a wide range of ncAAs, provided that they are attached on tRNAs. Only a small part of the tRNA moieties of the substrates appears to be involved in the interaction with the enzymes, which should facilitate the utilization of exogenous tRNAs. Further studies using the methods developed during my PhD and complementary approaches are required in order to fully understand the role of tRNA sequences in CDPS activity and specificity. In order to further increase the diversity that can be produced using these fascinating small enzymes, several strategies can be considered. The use of more promiscuous aminoacyl-tRNA synthetases could be easily implemented and would further expand the range of ncAAs that can be incorporated into 2,5-DKPs by CDPSs. In some cases where CDPS specificity would appear to be the limitation, the engineering of more promiscuous CDPSs could be undertaken. Previous studies have led to a good understanding of the architecture of the aminoacyl-binding pockets. The ongoing project using flexizymes should give new precious information on the residues involved in tRNA binding and on the tRNA sequence motifs that play a role in activity. These results will allow engineering efforts to be targeted more precisely. Finally, the manipulation of CDPSs for the production of non-canonical cyclodipeptides should be fully integrated into the booming field of 2,5-DKP-biosynthetic pathways.

Bibliography

1. Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43**, 155–176 (2016).
2. Rodrigues, T., Reker, D., Schneider, P. & Schneider, G. Counting on natural products for drug design. *Nat. Chem.* **8**, 531–541 (2016).
3. Scherlach, K. & Hertweck, C. Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **7**, 1753–1760 (2009).
4. Shen, B. A New Golden Age of Natural Products Drug Discovery. *Cell* **163**, 1297–1300 (2015).
5. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455–9 (2015).
6. Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chem. Rev.* **117**, 11570–11648 (2017).
7. Robbins, T., Liu, Y. C., Cane, D. E. & Khosla, C. Structure and mechanism of assembly line polyketide synthases. *Curr. Opin. Struct. Biol.* **41**, 10–18 (2016).
8. Süssmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis — Principles and Prospects. *Angew. Chemie - Int. Ed.* **56**, 3770–3821 (2017).
9. Hetrick, K. J. & van der Donk, W. A. Ribosomally synthesized and post-translationally modified peptide natural product discovery in the genomic era. *Curr. Opin. Chem. Biol.* **38**, 36–44 (2017).
10. van der Donk, W. A. Introduction: Unusual Enzymology in Natural Product Synthesis. *Chem. Rev.* 5223–5225 (2017). doi:10.1021/acs.chemrev.7b00124
11. Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **47**, W81–W87 (2019).
12. Walsh, C. T. T. & Fischbach, M. A. A. Natural products version 2.0: Connecting genes to molecules. *J. Am. Chem. Soc.* **132**, 2469–2493 (2010).
13. Baral, B., Akhgari, A. & Metsä-Ketelä, M. Activation of microbial secondary metabolic pathways: Avenues and challenges. *Synth. Syst. Biotechnol.* **3**, 163–178 (2018).
14. Winter, J. M., Behnken, S. & Hertweck, C. Genomics-inspired discovery of natural products. *Curr. Opin. Chem. Biol.* **15**, 22–31 (2011).
15. Huo, L. *et al.* Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* (2019). doi:10.1039/c8np00091c
16. Hover, B. M. M. *et al.* Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens.

- Nat. Microbiol.* **3**, 415–422 (2018).
17. Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528–532 (2013).
 18. Nielsen, J. Cell factory engineering for improved production of natural products. *Nat. Prod. Rep.* (2019). doi:10.1039/c9np00005d
 19. Pearsall, S. M., Rowley, C. N. & Berry, A. Advances in Pathway Engineering for Natural Product Biosynthesis. *ChemCatChem* **7**, 3078–3093 (2015).
 20. Kim, E., Moore, B. S. & Yoon, Y. J. Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nat. Chem. Biol.* **11**, 649–659 (2015).
 21. Borthwick, A. D. 2,5-diketopiperazines: Synthesis, reactions, medicinal chemistry, and bioactive natural products. *Chem. Rev.* **112**, 3641–3716 (2012).
 22. Lautru, S., Gondry, M., Genet, R. & Pernodet, J.-L. The Albonoursin Gene Cluster of *S. noursei*. *Chem. Biol.* **9**, 1355–1364 (2002).
 23. Gondry, M. *et al.* Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. *Nat. Chem. Biol.* **5**, 414–420 (2009).
 24. Belin, P. *et al.* The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways. *Nat. Prod. Rep.* **29**, 961–979 (2012).
 25. Giessen, T. W. & Marahiel, M. A. Rational and combinatorial tailoring of bioactive cyclic dipeptides. *Front. Microbiol.* **6**, 1–11 (2015).
 26. Borgman, P., Lopez, R. D. & Lane, A. L. The Expanding Spectrum of Diketopiperazine Natural Product Biosynthetic Pathways Containing Cyclodipeptide Synthases. *Org. Biomol. Chem.* **17**, 2305–2314 (2019).
 27. González, J. F., Ortín, I., De La Cuesta, E. & Menéndez, J. C. Privileged scaffolds in synthesis: 2,5-piperazinediones as templates for the preparation of structurally diverse heterocycles. *Chem. Soc. Rev.* **41**, 6902–6915 (2012).
 28. Wang, X., Li, Y., Zhang, X., Lai, D. & Zhou, L. Structural Diversity and Biological Activities of the Cyclodipeptides from Fungi. *Molecules* **22**, 2026 (2017).
 29. Ortiz, A. & Sansinenea, E. Cyclic Dipeptides: Secondary Metabolites Isolated from Different Microorganisms with Diverse Biological Activities. *Curr. Med. Chem.* **24**, 2773–2780 (2017).
 30. Borthwick, A. D. & Da Costa, N. C. 2,5-diketopiperazines in food and beverages: Taste and bioactivity. *Crit. Rev. Food Sci. Nutr.* **57**, 718–742 (2017).
 31. Huang, R. M. *et al.* An update on 2,5-Diketopiperazines from marine organisms. *Mar. Drugs* **12**, 6213–6235 (2014).
 32. de Carvalho, M. P. & Abraham, W.-R. Antimicrobial and Biofilm Inhibiting

Diketopiperazines. *Curr. Med. Chem.* **19**, 3564–3577 (2012).

33. Perry, T. L., Richardson, K. S., Hansen, S. & Friesen, A. J. Identification of the diketopiperazine of histidylproline in human urine. *J. Biol. Chem.* **240**, 4540–4542 (1965).
34. Ginz, M. & Engelhardt, U. H. Identification of new diketopiperazines in roasted coffee. *Eur. Food Res. Technol.* **213**, 8–11 (2001).
35. Tommonaro, G., Abbamondi, G. R., Iodice, C., Tait, K. & De Rosa, S. Diketopiperazines Produced by the Halophilic Archaeon, *Haloterrigena hispanica*, Activate AHL Bioreporters. *Microb. Ecol.* **63**, 490–495 (2012).
36. Holden, M. T. G. *et al.* Quorum-sensing cross talk: Isolation and chemical characterization of cyclic dipeptides from *Pseudomonas aeruginosa* and other Gram-negative bacteria. *Mol. Microbiol.* **33**, 1254–1266 (1999).
37. Degrassi, G. *et al.* Plant growth-promoting *Pseudomonas putida* WCS358 produces and secretes four cyclic dipeptides: Cross-talk with quorum sensing bacterial sensors. *Curr. Microbiol.* **45**, 250–254 (2002).
38. Park, D. K. *et al.* Cyclo(Phe-Pro) modulates the expression of ompU in *Vibrio* spp. *J. Bacteriol.* **188**, 2214–2221 (2006).
39. Li, J., Wang, W., Xu, S. X., Magarvey, N. A. & McCormick, J. K. *Lactobacillus reuteri*-produced cyclic dipeptides quench agr-mediated expression of toxic shock syndrome toxin-1 in staphylococci. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 3360–3365 (2011).
40. Bofinger, M. R., De Sousa, L. S., Fontes, J. E. N. & Marsaioli, A. J. Diketopiperazines as Cross-Communication Quorum-Sensing Signals between *Cronobacter sakazakii* and *Bacillus cereus*. *ACS Omega* **2**, 1003–1008 (2017).
41. Prasad, C. Bioactive cyclic dipeptides. *Peptides* **16**, 151–164 (1995).
42. Ortiz-Castro, R. *et al.* Transkingdom signaling based on bacterial cyclodipeptides with auxin activity in plants. *Proc. Natl. Acad. Sci.* **108**, 7253–7258 (2011).
43. Arnaouteli, S. *et al.* Pulcherrimin formation controls growth arrest of the *Bacillus subtilis* biofilm. *Proc. Natl. Acad. Sci.* **116**, 1–48 (2019).
44. Bischoff, V., Cookson, S. J., Wu, S. & Scheible, W. R. Thaxtomin A affects CESA-complex density, expression of cell wall genes, cell wall composition, and causes ectopic lignification in *Arabidopsis thaliana* seedlings. *J. Exp. Bot.* **60**, 955–965 (2009).
45. King, R. R. & Calhoun, L. A. The thaxtomin phytotoxins: Sources, synthesis, biosynthesis, biotransformation and biological activity. *Phytochemistry* **70**, 833–841 (2009).
46. Matsuda, K. Okaramines and other plant fungal products as new insecticide leads. *Curr. Opin. Insect Sci.* **30**, 67–72 (2018).
47. Scoffone, V. C. *et al.* Discovery of new diketopiperazines inhibiting *Burkholderia cenocepacia* quorum sensing in vitro and in vivo. *Sci. Rep.* **6**, 1–11 (2016).

48. Kanoh, K., Kohno, S., Katada, J. & Takahashi, J. Phenylahistin arrests cells in mitosis by inhibiting tubulin polymerization. *J. Antibiot. (Tokyo)*. **52**, 134–141 (1999).
49. Hayashi, Y., Yamazaki-Nakamura, Y. & Yakushiji, F. Medicinal chemistry and chemical biology of diketopiperazine-type antimicrotubule and vascular-disrupting agents. *Chem. Pharm. Bull.* **61**, 889–901 (2013).
50. Cimino, P. *et al.* Plinabulin, an inhibitor of tubulin polymerization, targets KRAS signaling through disruption of endosomal recycling. *Biomed. Reports* **10**, 218–224 (2019).
51. Gomes, N. G. M., Lefranc, F., Kijjoo, A. & Kiss, R. Can some marine-derived fungal metabolites become actual anticancer agents? *Mar. Drugs* **13**, 3950–3991 (2015).
52. Miyoshi, T., Miyairi, N., Aoki, H., Kosaka, M. & Sakai, H. Bicyclomycin, a new antibiotic. I. Taxonomy, isolation and characterization. *J. Antibiot. (Tokyo)*. **25**, 569–75 (1972).
53. Kohn, H. & Widger, W. The molecular basis for the mode of action of bicyclomycin. *Drug Targets-Infectious Disord.* **5**, 273–295 (2005).
54. Harford, P. S., Murray, B. E., DuPont, H. L. & Ericsson, C. D. Bacteriological studies of the enteric flora of patients treated with bicozamycin (CGP 3543/E) for acute nonparasitic diarrhea. *Antimicrob. Agents Chemother.* **23**, 630–633 (1983).
55. Malik, M. *et al.* Lethal synergy involving bicyclomycin: An approach for reviving old antibiotics. *J. Antimicrob. Chemother.* **69**, 3227–3235 (2014).
56. Park, B. S., Widger, W. & Kohn, H. Fluorine-substituted dihydrobicyclomycins: Synthesis and biochemical and biological properties. *Bioorganic Med. Chem.* **14**, 41–61 (2006).
57. Bloudoff, K., Fage, C. D., Marahiel, M. A. & Schmeing, T. M. Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 95–100 (2017).
58. Yu, D., Xu, F., Zhang, S. & Zhan, J. Decoding and reprogramming fungal iterative nonribosomal peptide synthetases. *Nat. Commun.* **8**, 1–11 (2017).
59. Mootz, H. D., Schwarzer, D. & Marahiel, M. A. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem* **3**, 490–504 (2002).
60. Wei, X., Yang, F. & Straney, D. C. Multiple non-ribosomal peptide synthetase genes determine peptaibol synthesis in *Trichoderma virens*. *Can. J. Microbiol.* **51**, 423–429 (2005).
61. Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174–4183 (1997).
62. Röttig, M. *et al.* NRSPredictor2 - A web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res.* **39**, 362–367 (2011).

63. Walsh, C. T., O'Brien, R. V. & Khosla, C. Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds. *Angew. Chemie - Int. Ed.* **52**, 7098–7124 (2013).
64. Pupin, M. *et al.* Norine: A powerful resource for novel nonribosomal peptide discovery. *Synth. Syst. Biotechnol.* **1**, 89–94 (2016).
65. Winn, M., Fyans, J. K., Zhuo, Y. & Micklefield, J. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* 317–347 (2016). doi:10.1039/c5np00099h
66. Brown, A. S., Calcott, M. J., Owen, J. G. & Ackerley, D. F. Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. *Nat. Prod. Rep.* **35**, 1210–1228 (2018).
67. Gu, B., He, S., Yan, X. & Zhang, L. Tentative biosynthetic pathways of some microbial diketopiperazines. *Appl. Microbiol. Biotechnol.* **97**, 8439–8453 (2013).
68. Stachelhaus, T., Mootz, H. D., Bergendahl, V. & Marahiel, M. A. Peptide Bond Formation in Nonribosomal Peptide Biosynthesis. *J. Biol. Chem.* **273**, 22773–22781 (1998).
69. Schultz, A. W. *et al.* Biosynthesis and Structures of Cyclomarins and Cyclomarazines , Prenylated Cyclic Peptides of Marine Actinobacterial Origin. *J. Am. Chem. Soc.* **130**, 4507–4516 (2008).
70. Schroeder, F. C. *et al.* Diketopiperazine Formation in Fungi requires Dedicated Cyclization and Thiolation Domains. *Angew. Chemie Int. Ed.* 1–6 (2019). doi:10.1002/anie.201909052
71. Balk-bindseip, W., Helmke, E. & Weyland, H. Maremycin A and B, New Diketopiperazines from a Marine Streptomyces sp. *Liebigs Ann.* 1291–1294 (1995).
72. Huang, T., Duan, Y., Zou, Y., Deng, Z. & Lin, S. NRPS Protein MarQ Catalyzes Flexible Adenylation and Specific S-Methylation. *ACS Chem. Biol.* **13**, 2387–2391 (2018).
73. Barry, S. M. *et al.* Cytochrome P450 – catalyzed L -tryptophan nitration in thaxtomin phytotoxin biosynthesis. *Nat. Chem. Biol.* **8**, 814–816 (2012).
74. Alkhalaf, L. M. *et al.* Binding of distinct substrate conformations enables hydroxylation of remote sites in thaxtomin D by cytochrome P450 TxtC. *J. Am. Chem. Soc.* jacs.8b08864 (2018). doi:10.1021/jacs.8b08864
75. Canu, N., Moutiez, M., Belin, P. & Gondry, M. Cyclodipeptide synthases: a promising biotechnological tool for the synthesis of diverse 2,5-diketopiperazines. *Nat. Prod. Rep.* (2019). doi:10.1039/c9np00036d
76. Moutiez, M., Belin, P. & Gondry, M. Aminoacyl-tRNA-Utilizing Enzymes in Natural Product Biosynthesis. *Chem. Rev.* **117**, 5578–5618 (2017).
77. Betat, H., Long, Y., Jackman, J. E. & Mörl, M. From end to end: TRNA editing at 5'-and 3'-terminal positions. *Int. J. Mol. Sci.* **15**, 23975–23998 (2014).

78. Boccaletto, P. *et al.* MODOMICS: A database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **46**, D303–D307 (2018).
79. Robertus, J. D. *et al.* Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **250**, 546–551 (1974).
80. Giegé, R. & Frugier, M. *Transfer RNA Structure and Identity*. *Madame Curie Bioscience Database* (2003).
81. Schimmel, P. RNA Processing and Modifications: The emerging complexity of the tRNA world: Mammalian tRNAs beyond protein synthesis. *Nat. Rev. Mol. Cell Biol.* **19**, 45–58 (2018).
82. Taiji, M., Yokoyama, S. & Miyazawa, T. Transacylation Rates of (Aminoacyl)adenosine Moiety at the 3'-Terminus of Aminoacyl Transfer Ribonucleic Acid. *Biochemistry* **22**, 3220–3225 (1983).
83. Stepanov, V. G. & Nyborg, J. Thermal stability of aminoacyl-tRNAs in aqueous solutions. 485–490 (2002).
84. Peacock, J. R., Walvoord, R. R., Chang, A. Y. & Kozlowski, M. C. Amino acid – dependent stability of the acyl linkage in aminoacyl-tRNA Amino acid – dependent stability of the acyl linkage in aminoacyl-tRNA. *RNA* **20**, 758–764 (2014).
85. Blanquet, S., Mechulam, Y. & Schmitt, E. The many routes of bacterial transfer RNAs after aminoacylation. *Curr. Opin. Struct. Biol.* **10**, 95–101 (2000).
86. Asahara, H. & Uhlenbeck, O. C. Predicting the binding affinities of misacylated tRNAs for *Thermus thermophilus* EF-Tu·GTP. *Biochemistry* **44**, 11254–11261 (2005).
87. Bothwell, I. R. *et al.* Characterization of glutamyl-tRNA–dependent dehydratases using nonreactive substrate mimics. *Proc. Natl. Acad. Sci.* **116**, 17245–17250 (2019).
88. Eriani, G., Delarue, M., Poch, O., Gangloff, J. & Moras, D. Partition of tRNA synthetases into two classes based on mutually. *Nature* **347**, 203–206 (1990).
89. Kaiser, F. *et al.* Characterization of Amino Acid Recognition in Aminoacyl-tRNA Synthetases. *bioRxiv* 606459 (2019). doi:10.1101/606459
90. Qin, X. *et al.* Cocrystal structures of glycyl-tRNA synthetase in complex with tRNA suggest multiple conformational states in glycylation. *J. Biol. Chem.* **289**, 20359–20369 (2014).
91. Jakubowski, H. Proofreading in trans by an aminoacyl-tRNA synthetase: A model for single site editing by isoleucyl-tRNA synthetase. *Nucleic Acids Res.* **24**, 2505–2510 (1996).
92. Sankaranarayanan, R. & Moras, D. The fidelity of the translation of the genetic code. *Acta Biochim. Pol.* **48**, 323–335 (2001).
93. Yadavalli, S. S. & Ibba, M. *Quality control in aminoacyl-tRNA synthesis: its role in translational fidelity*. *Advances in protein chemistry and structural biology* **86**, (Elsevier Inc., 2012).

94. Ahel, I., Korencic, D., Ibba, M. & Soll, D. Trans-editing of mischarged tRNAs. *Proc. Natl. Acad. Sci.* **100**, 15422–15427 (2003).
95. Giegé, R., Sissler, M. & Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* **26**, 5017–35 (1998).
96. Danhart, E. M. *et al.* Conformational and chemical selection by a trans -acting editing domain. *Proc. Natl. Acad. Sci.* **114**, E10254–E10254 (2017).
97. Francklyn, C. & Schimmel, P. Aminoacylation of RNA minihelices with alanine. *Nature* **337**, 478–481 (1989).
98. Cochella, L. & Green, R. Fidelity in protein synthesis. *Curr. Biol.* **15**, 536–540 (2005).
99. Cowie, D. B. & Cohen, G. N. Biosynthesis by *Escherichia coli* of active altered proteins containing selenium instead of sulfur. *BBA - Biochim. Biophys. Acta* **26**, 252–261 (1957).
100. Richmond, M. H. The effect of amino acid analogues on growth and protein synthesis in microorganisms. *Bacteriol. Rev.* **26**, 398–420 (1962).
101. Hartman, M. C. T., Josephson, K. & Szostak, J. W. Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4356–4361 (2006).
102. Budisa, N. *et al.* Residue-specific bioincorporation of non-natural, biologically active amino acids into proteins as possible drug carriers: structure and stability of the per-thiaproline mutant of annexin V. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 455–459 (1998).
103. Fan, C., Ho, J. M. L., Chirathivat, N., Söll, D. & Wang, Y. S. Exploring the substrate range of wild-type aminoacyl-tRNA synthetases. *ChemBioChem* **15**, 1805–1809 (2014).
104. Oldach, F. *et al.* Congeneric lantibiotics from ribosomal in vivo peptide synthesis with noncanonical amino acids. *Angew. Chemie - Int. Ed.* **51**, 415–418 (2012).
105. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 1–11 (2006).
106. Bertels, F., Merker, H. & Kost, C. Design and characterization of auxotrophy-based amino acid biosensors. *PLoS One* **7**, e41349 (2012).
107. Kuthning, A. *et al.* Towards Biocontained Cell Factories: An Evolutionarily Adapted *Escherichia coli* Strain Produces a New-to-nature Bioactive Lantibiotic Containing Thienopyrrole-Alanine. *Sci. Rep.* **6**, 1–7 (2016).
108. Hoesl, M. G. *et al.* Chemical Evolution of a Bacterial Proteome. *Angew. Chemie - Int. Ed.* **54**, 10030–10034 (2015).
109. Kast, P. & Hennecke, H. Amino acid substrate specificity of *Escherichia coli* phenylalanyl-tRNA synthetase altered by distinct mutations. *J. Mol. Biol.* **222**, 99–124 (1991).
110. Kirshenbaum, K., Carrico, I. S. & Tirrell, D. A. Biosynthesis of proteins incorporating a versatile set of phenylalanine analogues. *ChemBioChem* **3**, 235–237 (2002).

111. Bentin, T. *et al.* Photoreactive Bicyclic Amino Acids as Substrates for Mutant Escherichia coli Phenylalanyl-tRNA Synthetases. *J. Biol. Chem.* **279**, 19839–19845 (2004).
112. Döring, V. *et al.* Enlarging the Amino Acid Set of Escherichia coli by Infiltration of the Valine Coding Pathway. *Science (80-.)*. **292**, 501–504 (2001).
113. Iqbal, E. S., Dods, K. K. & Hartman, M. C. T. Ribosomal incorporation of backbone modified amino acids via an editing-deficient aminoacyl- tRNA synthetase. *Org. Biomol. Chem.* 1073–1078 (2018). doi:10.1039/c7ob02931d
114. Young, D. D. & Schultz, P. G. Playing with the Molecules of Life. *ACS Chem. Biol.* **13**, 854–870 (2018).
115. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science (80-.)*. **292**, 498–500 (2001).
116. Santoro, S. W., Wang, L. & Herberich, B. An efficient system for the evolution of aminoacyl- tRNA synthetase specificity. *Nat. Biotechnol.* **20**, 1044–1048 (2002).
117. Zhang, Y. A. N., Wang, L. E. I., Schultz, P. G. & Wilson, I. A. 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 (2005).
118. Wan, W., Tharp, J. M. & Liu, W. R. Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta - Proteins Proteomics* **1844**, 1059–1070 (2014).
119. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413 (2010).
120. Dumas, A., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical codon reassignment- Expanding the chemistry in biology. *Chem. Sci.* **6**, 50–69 (2015).
121. Melo Czekster, C., Robertson, W. E., Walker, A. S., Söll, D. & Schepartz, A. In Vivo Biosynthesis of a β -Amino Acid-Containing Protein. *J. Am. Chem. Soc.* **138**, 5194–5197 (2016).
122. Guo, J., Wang, J., Anderson, J. C. & Schultz, P. G. Addition of an alpha -hydroxy acid to the genetic code of bacteria. *Angew. Chemie - Int. Ed.* **47**, 722–725 (2008).
123. Santarem, M. *et al.* Synthesis of 3'-triazoyl-dinucleotides as precursors of stable Phe-tRNAPhe and Leu-tRNA^{Leu} analogues. *Bioorganic Med. Chem. Lett.* **24**, 3231–3233 (2014).
124. Heckler, T. G. *et al.* T4 RNA Ligase Mediated Preparation of Novel “Chemically Misacylated” tRNAPhes. *Biochemistry* **23**, 1468–1473 (1984).
125. Robertson, S. A., Ellman, J. A. & Schultz, P. G. A General and Efficient Route for Chemical Aminoacylation of Transfer RNAs. *J. Am. Chem. Soc.* **113**, 2722–2729 (1991).
126. Fonvielle, M. *et al.* Decoding the logic of the tRNA regiospecificity of nonribosomal femX

- Wv aminoacyl transferase. *Angew. Chemie - Int. Ed.* **49**, 5115–5119 (2010).
127. Iannazzo, L. *et al.* Synthesis of tRNA analogues containing a terminal ribose locked in the South conformation to study tRNA-dependent enzymes. *Org. Biomol. Chem.* **16**, 1903–1911 (2018).
 128. Chemama, M., Fonvielle, M., Arthur, M., Valéry, J. M. & Etheve-Quelquejeu, M. Synthesis of stable aminoacyl-tRNA analogues containing triazole as a bioisoster of esters. *Chem. - A Eur. J.* **15**, 1929–1938 (2009).
 129. Lünse, C. E. *et al.* New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nat. Chem. Biol.* **11**, 606–610 (2016).
 130. Steitz, T. A. & Moore, P. B. RNA, the first macromolecular catalyst: The ribosome is a ribozyme. *Trends Biochem. Sci.* **28**, 411–418 (2003).
 131. Hammann, C., Luptak, A. & Perreault, J. The ubiquitous hammerhead ribozyme. *Rna* **18**, 871–885 (2012).
 132. Webb, C. H. T. & Lupták, A. HDV-like self-cleaving ribozymes. *RNA Biol.* **8**, (2011).
 133. Ellis, J. C. & Brown, J. W. The RNase P family. *RNA Biol.* **6**, (2009).
 134. Müller, S., Appel, B., Balke, D., Hieronymus, R. & Nübel, C. Thirty-five years of research into ribozymes and nucleic acid catalysis: Where do we stand today? *F1000Research* **5(F1000 Fa)**, (2016).
 135. Blind, M. & Blank, M. Aptamer Selection Technology and Recent Advances. *Mol. Ther. - Nucleic Acids* **4**, (2015).
 136. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
 137. Darmostuk, M., Rimpelova, S., Gbelcova, H. & Ruml, T. Current approaches in SELEX: An update to aptamer selection technology. *Biotechnol. Adv.* **33**, 1141–1161 (2014).
 138. Jenne, A. & Famulok, M. A novel ribozyme with ester transferase activity. *Chem. Biol.* **5**, 23–34 (1998).
 139. Lohse, P. A. & Szostak, J. W. Ribozyme-catalysed amino-acid transfer reactions. *Nature* **381**, 442–4 (1996).
 140. Lee, N., Bessho, Y., Wei, K., Szostak, J. W. & Suga, H. Ribozyme-catalyzed {tRNA} aminoacylation. *Nat. Struct. Mol. Biol.* **7**, 28–33 (2000).
 141. Saito, H., Kourouklis, D. & Suga, H. An in vitro evolved precursor tRNA with aminoacylation activity. *EMBO J.* **20**, 1797–1806 (2001).
 142. Passioura, T. & Suga, H. Flexizymes, their evolutionary history and diverse utilities. *Top Curr Chem* **11**, 331–346 (2014).
 143. Saito, H., Watanabe, K. & Suga, H. Concurrent molecular recognition of the amino acid

- and tRNA by a ribozyme. *RNA* **7**, 1867–78 (2001).
144. Xiao, H., Murakami, H., Suga, H. & Ferré-D'Amaré, A. R. Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature* **454**, 358–361 (2008).
 145. Fujino, T., Kondo, T., Suga, H. & Murakami, H. Exploring of minimal RNA substrate of flexizymes. *ChemBioChem* **8603**, 1–8 (2019).
 146. Niwa, N., Yamagishi, Y., Murakami, H. & Suga, H. A flexizyme that selectively charges amino acids activated by a water-friendly leaving group. *Bioorganic Med. Chem. Lett.* **19**, 3892–3894 (2009).
 147. Öjemalm, K. *et al.* Energetics of side-chain snorkeling in transmembrane helices probed by nonproteinogenic amino acids. *Proc. Natl. Acad. Sci. U. S. A.* 4–10 (2016). doi:10.1073/pnas.1606776113
 148. Fujino, T., Goto, Y., Suga, H. & Murakami, H. Reevaluation of the d-amino acid compatibility with the elongation event in translation. *J. Am. Chem. Soc.* **135**, 1830–1837 (2013).
 149. Achenbach, J. *et al.* Outwitting EF-Tu and the ribosome: Translation with D-amino acids. *Nucleic Acids Res.* **43**, 5687–5698 (2015).
 150. Fujino, T., Goto, Y., Suga, H. & Murakami, H. Ribosomal Synthesis of Peptides with Multiple beta-Amino Acids. *J. Am. Chem. Soc.* **138**, 1962–1969 (2016).
 151. Katoh, T. & Suga, H. Ribosomal Incorporation of Consecutive β -Amino Acids. *J. Am. Chem. Soc.* **140**, jacs.8b07247 (2018).
 152. Kawakami, T., Murakami, H. & Suga, H. Messenger RNA-Programmed Incorporation of Multiple N-Methyl-Amino Acids into Linear and Cyclic Peptides. *Chem. Biol.* **15**, 32–42 (2008).
 153. Kawakami, T., Murakami, H. & Suga, H. Ribosomal Synthesis of Polypeptoids and Peptoid–Peptide Hybrids. *J. Am. Chem. Soc.* 16861–16863 (2008). doi:10.1021/ja806998v
 154. Ohta, A., Murakami, H., Higashimura, E. & Suga, H. Synthesis of Polyester by Means of Genetic Code Reprogramming. *Chem. Biol.* **14**, 1315–1322 (2007).
 155. Seefeldt, A. Inhibition of the bacterial ribosome by nascent and antimicrobial peptides. *Biochemistry, Molecular Biology.* Université de Bordeaux. (2018).
 156. Rogers, J. M. *et al.* Ribosomal synthesis and folding of peptide-helical aromatic foldamer hybrids (SI). *Nat. Chem.* **10**, 405–412 (2018).
 157. Katoh, T., Iwane, Y. & Suga, H. Logical engineering of D-arm and T-stem of tRNA that enhances D-amino acid incorporation. *Nucleic Acids Res.* **45**, 12601–12610 (2017).
 158. Katoh, T., Wohlgemuth, I., Nagano, M., Rodnina, M. V. & Suga, H. Essential structural elements in tRNA^{Pro} for EF-P-mediated alleviation of translation stalling. *Nat. Commun.* **7**, 1–12 (2016).

159. Ogawa, A., Namba, Y. & Gakumasawa, M. Rational optimization of amber suppressor tRNAs toward efficient incorporation of a non-natural amino acid into protein in a eukaryotic wheat germ 1. *Org. Biomol. Chem.* **14**, 2671–2678 (2016).
160. Passioura, T. & Suga, H. A RaPID way to discover nonstandard macrocyclic peptide modulators of drug targets. *Chem. Commun.* (2017). doi:10.1039/C6CC06951G
161. Liu, Z. *et al.* Homologous trans -editing factors with broad tRNA specificity prevent mistranslation caused by serine/threonine misactivation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6027–6032 (2015).
162. Blaskovich, M. A. T. Unusual Amino Acids in Medicinal Chemistry. *J. Med. Chem.* **59**, 10807–10836 (2016).
163. Baumann, T. *et al.* Prospects of in vivo incorporation of non-canonical amino acids for the chemical diversification of antimicrobial peptides. *Front. Microbiol.* **8**, 1–9 (2017).
164. Hirose, H., Tsiamantas, C., Katoh, T. & Suga, H. In vitro expression of genetically encoded non-standard peptides consisting of exotic amino acid building blocks. *Curr. Opin. Biotechnol.* **58**, 28–36 (2019).
165. Giessen, T. W. & Marahiel, M. A. Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity. *FEBS Lett.* **586**, 2065–2075 (2012).
166. Niquille, D. L. *et al.* Nonribosomal biosynthesis of backbone-modified peptides. *Nat. Chem.* **10**, (2017).
167. Abdelmelek, M., Nguyen, M., Bradford, V., Iverson, B. & Suggs, L. Synthesis of Morpholine-2,5-dione Monomers for the Preparation of Polydepsipeptides
Polydepsipeptides : Hybrids of Synthetic and Natural Materials Morpholine Diones : A Historical Perspective. (2010).
168. Achenbach, J. Through the mirror : Translation with D-amino acids. (2015).
169. Bhatt, T. K., Soni, R. & Sharma, D. Recent Updates on DTD (D-Tyr-tRNA(Tyr) Deacylase): An Enzyme Essential for Fidelity and Quality of Protein Synthesis. *Front. cell Dev. Biol.* **4**, 32 (2016).
170. Soutourina, J., Plateau, P., Delort, F., Peirottes, A. & Blanquet, S. Functional Characterization of the D-Tyr-tRNA Tyr Deacylase from *Escherichia coli*. *J. Biol. Chem.* **274**, 19109–19114 (1999).
171. Cochrane, J. R., White, J. M., Wille, U. & Hutton, C. A. Total synthesis of mycocyclosin. *Org. Lett.* **14**, 2402–2405 (2012).
172. Amatov, T., Pohl, R., Císařová, I. & Jahn, U. Synthesis of Bridged Diketopiperazines by Using the Persistent Radical Effect and a Formal Synthesis of Bicyclomycin. *Angew. Chemie - Int. Ed.* **54**, 12153–12157 (2015).

RESUME EN FRANCAIS

Les cyclodipeptides constituent, avec leurs dérivés plus complexes les 2,5-dicétopiperazines (2,5-DKP), une importante famille de produits naturels, synthétisés essentiellement par des micro-organismes. Certaines de ces 2,5-DKP naturellement produites présentent des activités pharmacologiques et le spectre de ces activités est large (antibiotiques, anti-tumoraux, anti-viraux, herbicides, ...), ce qui a suscité un vif intérêt en chimie médicinale pour le squelette cyclodipeptidique. Des approches purement chimiques ont permis de générer des analogues de 2,5-DKP avec des propriétés pharmacologiques améliorées. Une approche complémentaire de la chimie organique pour générer une plus grande diversité de 2,5-DKP est l'étude et la manipulation des voies métaboliques de biosynthèse des 2,5-DKP. Deux types de voies métaboliques conduisant aux 2,5-DKP ont été décrits. Ceux-ci reposent respectivement sur les enzymes de synthèse des peptides non-ribosomiques (NRPS, *non-ribosomal peptide synthetase*), des complexes multi-enzymatiques de grande taille, ou sur une famille d'enzymes de petite taille (autour de 25 kDa), dédiées à la synthèse de cyclodipeptides, les synthases de cyclodipeptides (CDPS, *cyclodipeptide synthase*). Les CDPS sont particulièrement intéressantes : ces petites enzymes catalysent la formation de cyclodipeptides en utilisant les acides aminés sous la forme d'ARNt aminoacylés, qu'elles détournent de leur rôle canonique comme substrats du ribosome. D'autre part, les gènes codant les CDPS sont souvent associés dans les génomes avec des gènes codant des enzymes de modification des cyclodipeptides qui introduisent une complexité et une diversité structurale sur ces molécules. De précédents travaux ont montré que les CDPS sont capables d'utiliser la quasi-totalité des acides aminés protéinogènes pour produire un grand nombre de cyclodipeptides. Les caractéristiques des CDPS en font un outil biotechnologique intéressant pour la synthèse de cyclodipeptides. Cependant, afin d'exploiter complètement le potentiel synthétique de ces enzymes, il est nécessaire de mieux comprendre leur spécificité, notamment vis-à-vis d'acides aminés non naturels.

Mon travail de thèse débute par la présentation d'une nouvelle méthode qui permet l'annotation fonctionnelle des CDPS, basée sur l'utilisation des réseaux de spécificité de séquences (SSN, *sequence similarity network*). En nous basant sur les activités des CDPS précédemment caractérisées, nous avons montré que l'activité de plus de 80% des CDPS putatives peut être prédite. Puis, nous avons étendu de manière significative le spectre des produits des CDPS en montrant que ces enzymes sont capables d'incorporer des acides aminés non naturels. Pour cela, nous avons transposé aux CDPS la méthode d'insertion d'acides aminés non naturels dite « residue-spécifique » qui est couramment appliquée à la synthèse ribosomique des protéines. Nous avons tiré profit de la promiscuité des aminoacyl-ARNt synthétases (AARS) d'*Escherichia coli* vis-à-vis d'acides aminés non naturels pour étudier la promiscuité des CDPS *in vivo*, en utilisant des souches d'*E. coli* auxotrophes pour les acides aminés protéinogènes. 26 acides aminés non naturels ont été incorporés, conduisant à la mise en évidence d'environ 200 nouveaux

cyclodipeptides. Enfin, nous avons cherché à mieux comprendre la manière dont les CDPS reconnaissent la partie ARNt de leurs substrats. Nous avons mis au point un nouvel essai enzymatique *in vitro* qui permet d'étudier séparément la reconnaissance des deux différents substrats des CDPS et d'estimer des paramètres cinétiques. En utilisant une méthode d'aminocyclation innovante, basée sur une classe de ribozymes à activité AARS, les flexizymes, nous avons généré des analogues d'ARNt aminoacylés portant des parties ARNt tronquées. Parmi des différents « mini ARNt aminoacylés », nous avons pu montrer que ceux reproduisant l'intégralité des sept paires de bases du bras accepteur des ARNt sont d'aussi bons substrats que les ARNt aminoacylés complets. Cela suggère que les CDPS interagissent principalement avec le bras accepteur des ARNt. La méthode de production de « mini ARNt aminoacylés » développée dans le cadre de cette thèse ouvre d'intéressantes perspectives, notamment pour des études biophysiques et structurales des enzymes de plus en plus nombreuses qui utilisent les ARNt aminoacylés comme substrats et pour lesquelles la synthèse d'analogues de substrats est difficile.

Titre : Vers une meilleure compréhension de la reconnaissance des substrats des synthases de cyclodipeptides

Mots clés : Produits naturels peptidiques, Promiscuité enzymatique, Biosynthèse peptidique non ribosomique, ARNt aminoacylés, Flexizymes,

Les cyclodipeptides constituent, avec leurs dérivés plus complexes les dicétopiperazines (DKP), une importante famille de produits naturels, synthétisés essentiellement par des micro-organismes. Une approche intéressante pour synthétiser une grande diversité de DKP consiste à étudier et manipuler les voies de biosynthèse de ces molécules. Les synthases de cyclodipeptides (CDPS) constituent une famille d'enzymes, dédiées à la production de cyclodipeptides, qui ont la particularité d'utiliser les ARNt aminoacylés (AA-ARNt) comme substrats. Afin d'exploiter complètement le potentiel synthétique de ces enzymes, il est nécessaire de mieux comprendre leur spécificité, notamment vis-à-vis de substrats non naturels.

Dans cette thèse, nous avons tout d'abord

démonstré que les CDPS sont capables d'incorporer des acides aminés non naturels, en utilisant la promiscuité des aminoacyl-ARNt synthétases (AARS) d'*Escherichia coli*. Ceci augmente de manière significative la diversité de cyclodipeptides accessible de manière enzymatique. Puis nous avons amélioré notre compréhension de la reconnaissance de la partie ARNt des substrats par les CDPS. En utilisant les flexizymes, des ribozymes à activité AARS, nous avons généré des analogues d'AA-ARNt avec des parties ARNt tronquées. Nous avons pu montrer que des « mini AA-ARNt » reproduisant les 7 paires de bases du bras accepteur des ARNt sont d'aussi bons substrats que les AA-ARNt complets, ce qui suggère que les CDPS interagissent principalement avec les bras accepteurs de leurs substrats ARNt.

Title: New insights into the recognition of the substrates of cyclodipeptide synthases

Keywords: Peptide natural products, Enzyme promiscuity, Non-ribosomal peptide synthesis Aminoacyl-tRNAs, flexizymes,

Cyclodipeptides and their complex derivatives, the diketopiperazines (DKPs), constitute a large class of natural products with diverse and noteworthy pharmacological activities observed for many naturally occurring DKPs. A promising approach to generate diverse DKPs is to study and manipulate DKP biosynthetic pathways. Cyclodipeptide synthases (CDPSs) constitute an enzyme family dedicated to the synthesis of cyclodipeptides, with the particularity to use aminoacylated-tRNAs (AA-tRNAs) as substrates. In order to unlock the biosynthetic potential of these enzymes, better understanding their specificity, in particular towards non-natural substrates, is required. In this thesis, we first significantly expanded the diversity of cyclodipeptides accessible through

enzymatic synthesis by showing that CDPSs could incorporate non-canonical amino acids, through the use of the promiscuity of *E. coli* aminoacyl-tRNA synthetases. Then, we gave new insights into the recognition by CDPSs of the tRNA moieties of their substrates. By using an innovative RNA acylation strategy based on a class of ribozymes called flexizymes, we generated analogues of AA-tRNAs with truncated RNA moieties. Among these "AA-minitRNAs", we showed that those mimicking the entire 7 bp stems of tRNAs are as good substrates as AA-tRNAs, which suggests that CDPSs interact mainly with the acceptor arms of tRNAs and paves the way for promising biophysical and structural studies.

