



**HAL**  
open science

**Découverte et déchiffrement de nouvelles voies de biosynthèse dépendant des synthèses de cyclodipeptides : les clés d'une diversité accrue de dicétopipérazines potentiellement bioactives**

Isabelle Jacques

► **To cite this version:**

Isabelle Jacques. Découverte et déchiffrement de nouvelles voies de biosynthèse dépendant des synthèses de cyclodipeptides : les clés d'une diversité accrue de dicétopipérazines potentiellement bioactives. Microbiologie et Parasitologie. Université Paris Sud - Paris XI, 2015. Français. NNT : 2015PA114838 . tel-01674165v1

**HAL Id: tel-01674165**

**<https://theses.hal.science/tel-01674165v1>**

Submitted on 2 Jan 2018 (v1), last revised 2 Jan 2018 (v2)

**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.

## UNIVERSITÉ PARIS-SUD 11

### ECOLE DOCTORALE :

INNOVATION THÉRAPEUTIQUE : DU FONDAMENTAL A L'APPLIQUÉ  
PÔLE : INGENIERIE DES PROTEINES ET CIBLES THERAPEUTIQUES

### DISCIPLINE :

STRUCTURE, FONCTION ET INGÉNIERIES DES PROTÉINES

ANNÉE 2012 - 2015

SÉRIE DOCTORAT N°

### THÈSE DE DOCTORAT

soutenue le 23/09/2015

par

**Isabelle JACQUES**

Découverte et déchiffrement de nouvelles voies de biosynthèse dépendant des synthèses de cyclodipeptides : les clés d'une diversité accrue de dicétopipérazines potentiellement bioactives

**Directeur de thèse :** Muriel GONDROY Docteur, HDR (Saclay)  
**Co-encadrant de thèse :** Pascal BELIN Docteur (Saclay)

#### **Composition du jury :**

*Président du jury :* Hermann van TILBEURGH Professeur, HDR (Orsay)  
*Rapporteurs :* Bertrand AIGLE Professeur, HDR (Nancy)  
Marcel SALANOUBAT Docteur, HDR (Evry)  
*Examineurs :* Pierre MONSAN Professeur, thèse d'Etat (Toulouse)  
Séverine ZIRAH Docteur, HDR (Paris)







## **REMERCIEMENTS**

Tout d'abord, je souhaiterais remercier les membres de mon jury pour avoir accepté d'évaluer mes travaux par la lecture du manuscrit puis au moment de la soutenance. Leurs remarques et leurs commentaires lors de ma soutenance m'ont permis d'enrichir ma réflexion et de proposer une version finale améliorée de mon manuscrit de thèse.

Dans un second temps, je tiens à adresser mes remerciements à Vincent DIVE, chef du Service d'Ingénierie Moléculaire des Protéines (SIMOPRO) ainsi qu'à Annick HAREL-BELLAN, chef du Service de Biologie Intégrative et Génétique Moléculaire (SBIGEM) au sein de l'Institut de Biologie et de Technologie de Saclay du Commissariat à l'Energie Atomique, pour m'avoir accueillie dans leur service.

Je souhaite remercier sincèrement Muriel GONDROY de m'avoir accordé sa confiance dès le tout début, de m'avoir offert l'opportunité de rejoindre son équipe et de faire partie de l'aventure des CDPS. Ses grandes qualités scientifiques et humaines ont été durant ces trois années une aide exceptionnelle et seront toujours pour moi un exemple. Je souhaite dire un grand merci à Pascal BELIN d'avoir accepté de me former à la microbiologie et à la biologie moléculaire et de m'avoir encadrée, suivie, guidée et encouragée au quotidien durant ces trois années. Notre collaboration a été pour moi particulièrement enrichissante autant d'un point de vue scientifique, méthodologique et rédactionnel que d'un point de vue relationnel. Sa rigueur et son expertise scientifique ont constitué le socle de ce travail et me permettront, je l'espère, d'aborder à l'avenir une grande diversité de problématiques scientifiques avec la même qualité. Je souhaite remercier également Mireille MOUTIEZ de m'avoir fait bénéficier de ses compétences en analyse chimique et en enzymologie pour mener à bien mon projet et de m'avoir apporté du soutien dans les petits moments de découragement. Merci à Jérôme SEGUIN pour avoir pris part au projet en prenant en charge et en mettant au point les purifications des CDPS et également pour la gentillesse et le souci dont il a fait preuve à mon égard lorsque je partageais son bureau. Un grand merci enfin à Emmanuel FAVRY pour l'aide qu'il m'a apportée pour la réalisation de mes travaux et pour son énergie communicative au sein du laboratoire. Ces trois années passées au sein de cette équipe ont été remplies d'expériences positives et formatrices.

Je souhaite remercier nos collaborateurs Jean-Luc PERNODET, Emmanuelle DARBON, Sylvie LAUTRU et Jerzy WITWINOSKY du département Microbiologie de l'Institut de Biologie Intégrative de la Cellule, Lucrèce MATERON, Emmanuelle SACHON et Gérard BOLBACH de la plateforme de spectrométrie de masse et protéomique de l'Université Pierre et Marie Curie ainsi qu'Isabelle CORREIA et Olivier LEQUIN du Laboratoire de Biomolécules de l'Université Pierre et Marie Curie.

Je tiens également à remercier au sein du SIMOPRO, Robert et Steven pour les nombreuses séries de LC-MS/MS lancées sur les weekends, ainsi que toutes les personnes qui ont rendu ces années sympathiques à l'intérieur et à l'extérieur du CEA : Cécile, Carole M., Marion, Gael, Jennifer, Marjorie,

Caroline, Stéphanie, Thomas, Matthieu, Céline, Aurélien, Claire, Guillaume, Laura, Bertrand, Carole F., Hervé, Elodie, Pascal K. ... et tous ceux auxquels je ne pense pas au moment d'écrire ces lignes mais que je n'ai pas vraiment oubliés !

Enfin, un merci tout particulier à ma famille, à mes parents pour m'avoir donné l'envie et les moyens de faire des études, pour m'avoir soutenue pendant celles-ci et guidée dans mes choix parfois cornéliens, à mes frères pour m'avoir montré deux beaux exemples de réussite (même si cela met un peu la pression) et pour leur soutien et leur regard bienveillant. Pour finir, un immense merci à mon fiancé, Ludovic, pour m'avoir donné la force de continuer durant toutes nos études et notre doctorat par l'exemple de courage et de travail qu'il m'a offert à voir, pour le réconfort et la confiance qu'il me donnait dans les moments de doute (« une chose après l'autre »), pour les petits déjeuners du matin et pour l'équipe que nous avons formée et allons continuer à former.

## RÉSUMÉS

Malgré l'intérêt et la diversité des propriétés pharmacologiques des 2,5-dicétopipérazines (DKP), les voies de biosynthèse de ces molécules d'origine microbienne sont très peu connues. L'objectif de mes travaux de thèse a été *i*) de documenter de nouvelles voies de biosynthèse de DKP qui se caractérisent par la présence d'une synthase de cyclodipeptides (CDPS) travaillant souvent de concert avec une ou plusieurs enzymes de modification des cyclodipeptides et *ii*) d'explorer la diversité chimique codée par ces voies. Dans un premier temps, je me suis intéressée aux CDPS. Après la sélection par bioinformatique de candidats dans les bases de données génomiques, j'ai pu identifier 51 nouvelles CDPS actives et montrer que ces enzymes peuvent incorporer 17 des 20 acides aminés naturels. Par ailleurs, ce travail a permis de mieux caractériser la famille des CDPS, de définir l'existence de plusieurs sous-familles aux signatures fonctionnelles spécifiques et d'établir les premiers éléments d'un code de spécificité pour la synthèse de cyclodipeptides. Dans un second temps, je me suis attachée à caractériser les enzymes de modification associées aux nouvelles CDPS et, en particulier, les dioxygénases dépendant du Fe(II) et du 2-oxoglutarate (OG) qui sont très représentées dans ces voies. J'ai ainsi pu détecter une activité *in vivo* pour 11 OG et poursuivre la caractérisation *in vitro* pour l'une de ces OG, ce qui a permis de caractériser les DKP qu'elle synthétise et d'ainsi montrer la complexité des modifications chimiques introduites. L'ensemble de ces travaux a donc permis d'identifier et de caractériser de nouvelles voies de biosynthèse qui donnent accès à une diversité accrue de DKP.

Despite the interest and diversity of the pharmacological properties of 2,5-diketopiperazines (DKPs), the biosynthetic pathways of these microbial molecules are poorly documented. The aim of my doctoral work was *i*) to identify new DKP biosynthetic pathways that are characterized by the presence of a cyclodipeptide synthase (CDPS) often associated with one or more cyclodipeptide tailoring enzymes and *ii*) to explore the chemical diversity encoded by these pathways. First of all, my study focused on CDPSs. After the bioinformatics based selection of candidates, 51 novel CDPS were characterized, revealing the incorporation of 17 of the 20 proteinogenic amino acids. Moreover, this work has allowed a better characterization of the CDPS family, by showing the existence of several subfamilies with specific functional signatures and laying the foundations of a specificity conferring code for the synthesis of cyclodipeptides. Second, I characterized the tailoring enzymes associated with the newly identified CDPSs and, in particular, the FeII and oxoglutarate dependent dioxygenases (OGs) that are highly represented in these pathways. I detected the *in vivo* activity for 11 OGs and characterized the *in vitro* activity for one of them, showing the complexity of the chemical modifications introduced into the cyclodipeptide. This work has led to identify and characterize novel biosynthetic pathways that provide access to a greater diversity of DKPs.



## TABLE DES MATIÈRES

<b>INTRODUCTION GÉNÉRALE</b>	<b>23</b>
<b>ETUDE BIBLIOGRAPHIQUE</b>	<b>29</b>
<b>I. Les 2,5-dicétopipérazines</b>	<b>30</b>
1. <i>Définition et propriétés du cycle 2,5-dicétopipérazine</i>	<i>30</i>
1.1 Définition du cycle 2,5-dicétopipérazine	30
1.2 Conformation du cycle 2,5-dicétopipérazine	31
1.3 Propriétés chimiques des 2,5-dicétopipérazines	31
2. <i>Diversité des sources naturelles productrices de 2,5-dicétopipérazines</i>	<i>32</i>
3. <i>Diversité structurale des 2,5-dicétopipérazines d'origine biologique</i>	<i>33</i>
4. <i>L'identification des 2,5-dicétopipérazines</i>	<i>33</i>
<b>II. Les dicétopipérazines, rôles physiologiques et activités biologiques</b>	<b>36</b>
1. <i>Des rôles physiologiques peu connus</i>	<i>36</i>
1.1 Dicétopipérazines, des molécules de communication	36
1.1.1 Dicétopipérazines interagissant avec des biosenseurs du <i>quorum sensing</i>	38
1.1.2 Dicétopipérazines et communication inter-espèce et inter-règne	39
1.1.3 Le cyclo(Phe-Pro), une molécule qui interfère avec la pathogénicité des <i>Vibrio</i>	40
1.1.4 Dicétopipérazines induisant la production de substances antimicrobiennes	40
1.1.5 Dicétopipérazines exacerbant l'altération des denrées alimentaires	41
1.2 D'autres rôles décrits pour les dicétopipérazines	41
1.2.1 Virulence de micro-organismes pathogènes	41
1.2.2 Molécules liant le fer	42
2. <i>Des activités biologiques variées</i>	<i>43</i>
2.1 Activités antibactériennes	43
2.2 Activités antitumorales/anticancéreuses	45
2.3 Activité neuroprotectrice	49
<b>III. Synthèse chimique des dicétopipérazines</b>	<b>50</b>
1. <i>Formation du cycle dicétopipérazine</i>	<i>50</i>
1.1 <i>Via</i> la formation de liaison amide	51
1.1.1 Cyclisation d'ester dipeptidique	51
1.1.2 Réaction de Ugi	52
1.1.3 Autres voies utilisant la formation de liaison amide	52
1.2 <i>Via</i> N-alkylation	53
1.3 Autres voies de synthèse	53
2. <i>Fonctionnalisation du cycle dicétopipérazine</i>	<i>54</i>
2.1 Fonctionnalisation sur les carbones	54

2.2	Fonctionnalisation sur les azotes	55
2.3	Réaction sur les carbones des carbonyles (C-2) (C-5)	55
<b>IV. Les voies de biosynthèse de dicétopipérazines utilisant les synthétases de peptides non ribosomiques (NRPS)</b>		<b>57</b>
1.	<i>Les synthétases de peptides non ribosomiques, des méga-complexes enzymatiques organisés en modules</i>	57
1.1	Le domaine d'adénylation reconnaît le substrat et catalyse son activation	58
1.2	Le domaine de thiolation permet le transfert des substrats activés d'un centre catalytique à l'autre	59
1.3	Le domaine de condensation catalyse la formation de la liaison peptidique	60
1.4	Le domaine de terminaison catalyse la libération du peptide	60
1.5	Les peptides non ribosomiques peuvent être modifiés au cours ou après leur synthèse	61
2.	<i>Voies de biosynthèse de dicétopipérazines dépendant des NRPS</i>	62
2.1	Voies de biosynthèse de la gliotoxine et des épithiodioxopipérazines : les gènes de la NRPS et des enzymes de modifications sont regroupées en un même cluster	64
2.2	La voie de biosynthèse des dicétopipérazines de type fumitremorgine utilise une enzyme de modification « extérieure » au cluster de la NRPS	65
2.3	La voie de biosynthèse de la thaxtomine : la NRPS prend en charge un acide aminé modifié au préalable	66
2.4	La biosynthèse de l'érythrocheline : l'acide aminé utilisé est synthétisé par un autre cluster distant du cluster contenant la NRPS	67
2.5	La biosynthèse de l'ergopeptine résulte de l'action conjointe de deux NRPS	68
2.6	La biosynthèse de dicétopipérazines comme produits secondaires	69
<b>V. Les voies de synthèse de dicétopipérazines dépendant des synthèses de cyclodipeptides ou CDPS</b>		<b>70</b>
1.	<i>Les synthèses de cyclodipeptides : mise en évidence et caractérisation biochimique et structurale</i>	70
1.1	Les CDPS, une nouvelle famille d'enzymes qui synthétisent principalement des cyclodipeptides hydrophobes	70
1.2	Les CDPS utilisent les aminoacyl-ARNt comme substrats	72
1.3	Les CDPS possèdent une structure réminiscente des aminoacyl-ARNt synthétases	72
1.4	Les CDPS utilisent un mécanisme catalytique de type ping-pong	75
1.5	Les bases moléculaires de l'interaction entre les CDPS et leurs substrats	78
2.	<i>Voies de biosynthèse de dicétopipérazines dépendant des CDPS</i>	81
2.1	Voies de biosynthèse de l'albonoursine	81
2.2	Voies de biosynthèse de l'acide pulcherriminique	83
2.3	Voie de biosynthèse de la mycocyclosine	84
<b>MATÉRIELS ET MÉTHODES</b>		<b>85</b>
<b>I. Analyse bio-informatique</b>		<b>86</b>
1.	<i>Identification des CDPS potentielles</i>	86
2.	<i>Calcul d'arbres phylogénétiques</i>	86
2.1	Arbre de mai 2013	86

2.2	Arbre de septembre 2014	86
3.	<i>Outils pour l'analyse des séquences des CDPS</i>	86
3.1	Prédiction avec HHPRED	86
3.2	Diagramme topologique	87
3.3	Logos de séquences	87
4.	<i>Analyse des environnements génomiques des CDPS</i>	87
<b>II.</b>	<b>Matériel commercial et techniques classiques de biologie moléculaire</b>	<b>87</b>
1.	<i>Techniques classiques de biologie moléculaire</i>	87
2.	<i>Souches bactériennes utilisées</i>	88
2.1	Souches pour la biologie moléculaire	88
2.2	Souches pour l'expression protéique	89
3.	<i>Plasmides commerciaux</i>	89
4.	<i>Milieux de culture</i>	90
<b>III.</b>	<b>Préparation des vecteurs pIJ194 et pIJ196</b>	<b>91</b>
1.	<i>Préparation du pIJ194</i>	91
1.1	Suppression du site <i>XbaI</i>	91
1.2	Ajout des sites <i>SpeI</i> et <i>XbaI</i> en préparant une cassette par PCR SOEing	92
1.3	Ajout d'une étiquette 6-Histidine	94
2.	<i>Préparation du vecteur pIJ196</i>	94
<b>IV.</b>	<b>Obtention des gènes de CDPS et d'enzymes de modification</b>	<b>95</b>
1.	<i>Amplification d'ADN génomique</i>	95
2.	<i>Gènes de synthèse</i>	97
<b>V.</b>	<b>Culture pour l'expression des CDPS et des enzymes de modification</b>	<b>97</b>
1.	<i>Culture en erlenmeyer</i>	97
2.	<i>Culture en plaque 24 puits</i>	98
<b>VI.</b>	<b>Identification des dicétopipérazines</b>	<b>99</b>
1.	<i>Standards chimiques</i>	99
2.	<i>Analyses des surnageants de culture par LC-MS/MS</i>	99
3.	<i>Purification de cyclodipeptides</i>	100
3.1	A partir de surnageant de culture	100
3.2	A partir de milieu réactionnel	101
4.	<i>Analyse en composition d'acides aminés</i>	101
5.	<i>Analyse par spectrométrie de masse haute résolution (HRMS)</i>	101

6.	<i>Analyses par résonance magnétique nucléaire (RMN)</i>	102
<b>VII.</b>	<b>Analyse protéique des culots bactériens</b>	<b>103</b>
1.	<i>Culots issus de la culture en erlenmeyer</i>	103
2.	<i>Culots issus de la culture en plaque 24 puits</i>	103
3.	<i>Analyse par SDS-PAGE</i>	103
<b>VIII.</b>	<b>Purification des protéines (réalisée au sein de l'équipe par Emmanuel Favry et Jérôme Seguin)</b>	<b>104</b>
1.	<i>Purification des CDPS (réalisée par Emmanuel Favry et Jérôme Seguin)</i>	104
2.	<i>Purification des dioxygénases dépendant du Fe(II) et du 2-oxoglutarate (réalisée par Emmanuel Favry)</i>	104
<b>IX.</b>	<b>Test d'activités <i>in vitro</i></b>	<b>105</b>
1.	<i>CDPS</i>	105
1.1	Test avec lysat soluble	105
1.2	Test couplé avec ARNt et aaRS purifiés	105
2.	<i>Dioxygénases dépendant du Fe(II) et du 2-oxoglutarate</i>	105
<b>CHAPITRE I: MÉTHODE DE CARACTÉRISATION MOYEN-DÉBIT DES ACTIVITÉS DES ENZYMES DES VOIES DÉPENDANT DES CDPS</b>		<b>107</b>
<b>I.</b>	<b>Mise au point du système de clonage</b>	<b>109</b>
1.	<i>Choix et mise en place du système de clonage</i>	109
1.1	Système à deux vecteurs	109
1.2	Système permettant de rassembler plusieurs enzymes dans un vecteur	109
2.	<i>Optimisation du vecteur d'expression pour les CDPS</i>	111
2.1	Le clonage de gènes de CDPS dans pQE60 induit une toxicité chez <i>E. coli</i>	111
2.2	L'introduction du gène <i>lacI</i> dans le plasmide pQE60 permet un clonage stable	112
3.	<i>Choix de l'origine des gènes</i>	113
<b>II.</b>	<b>Mise au point du test d'activité <i>in vivo</i> en moyen débit</b>	<b>114</b>
1.	<i>Choix des souches d'<i>E. coli</i> pour l'expression des CDPS et des enzymes de modification</i>	114
2.	<i>Expression avec un milieu auto-induit</i>	115
2.1	Pourquoi un milieu auto-induit ?	115
2.2	Choix de la température de culture	116
2.3	Milieu auto-induit et M15-pREP4 : un écueil à éviter	117
3.	<i>Protocole de culture en plaque 24 puits</i>	118
<b>III.</b>	<b>Introduction d'un nouveau mode de séparation chromatographique</b>	<b>119</b>
1.	<i>Analyse réalisée en routine au laboratoire</i>	119
2.	<i>La colonne dC18 n'est pas adaptée à l'analyse des cyclodipeptides polaires ou de petite taille</i>	120

3.	<i>La colonne Hypercarb permet d'analyser une nouvelle gamme de cyclodipeptides</i>	120
<b>IV.</b>	<b>Discussion, conclusion et perspectives</b>	<b>121</b>
<b>CHAPITRE II : MISE EN ÉVIDENCE ET CARACTERISATION DE NOUVELLES ACTIVITÉS CDPS</b>		<b>125</b>
<b>I.</b>	<b>Identification par bio-informatique de 80 CDPS potentielles et sélection de 49 CDPS à tester</b>	<b>127</b>
1.	<i>Identification des CDPS potentielles par PSI-BLAST</i>	127
2.	<i>Correction des séquences obtenues et création d'un arbre phylogénétique</i>	128
3.	<i>Sélection des CDPS à étudier</i>	129
<b>II.</b>	<b>41 nouvelles CDPS sur 49 sont actives : redéfinition des caractéristiques générales de ces enzymes</b>	<b>133</b>
1.	<i>41 des 49 CDPS sélectionnées actives in vivo</i>	133
2.	<i>Caractéristiques générales des CDPS actives</i>	135
3.	<i>Signature fonctionnelle</i>	136
<b>III.</b>	<b>Des sous-familles de CDPS aux résidus catalytiques différents</b>	<b>136</b>
1.	<i>Définition des sous-familles</i>	136
2.	<i>Caractéristiques fonctionnelle et structurale des sous-familles</i>	137
<b>IV.</b>	<b>Une plus grande diversité de cyclodipeptides</b>	<b>141</b>
1.	<i>Méthode pour l'identification des cyclodipeptides</i>	141
2.	<i>Activités de synthèse de cyclodipeptides des CDPS</i>	142
3.	<i>Diversité de cyclodipeptides apportée par les 41 nouvelles CDPS</i>	145
<b>V.</b>	<b>Vers la prédiction de la spécificité des CDPS</b>	<b>146</b>
1.	<i>Prédiction des résidus constituant les poches de liaison des substrats chez les CDPS</i>	147
2.	<i>Les CDPS de même activité montrent des séquences consensus pour les résidus constituant les poches P1 et P2</i>	148
3.	<i>Validation expérimentale du modèle prédictif</i>	151
<b>VI.</b>	<b>Vers une caractérisation biochimique plus approfondie des CDPS : l'étude <i>in vitro</i> des CDPS nouvellement identifiées</b>	<b>153</b>
1.	<i>Sélection et purification de 16 CDPS</i>	153
2.	<i>Analyse <i>in vitro</i> des activités des CDPS</i>	160
2.1	<i>Mise au point du test avec un lysat soluble de <i>E. coli</i>.</i>	161
2.2	<i>Résultats des tests <i>in vitro</i> des activités</i>	162
<b>VII.</b>	<b>Discussion, conclusion et perspectives</b>	<b>164</b>

<b>CHAPITRE III: MISE EN ÉVIDENCE DE NOUVELLES ACTIVITÉS DE MODIFICATION DANS LES VOIES DEPENDANT DES CDPS</b>	<b>167</b>
<b>I. Environnement génomique des CDPS : diversité et sélection d'enzymes de modification à étudier</b>	<b>169</b>
1. <i>Analyse des environnements génomiques des CDPS</i>	169
2. <i>Sélection des voies de biosynthèse à étudier</i>	180
<b>II. Mise en évidence d'activités de modification</b>	<b>181</b>
1. <i>Enzymes de modification sélectionnées et modifications potentielles opérées</i>	182
1.1 Déshydrogénase/réductase à chaîne moyenne	182
1.2 Aldéhyde déshydrogénases	182
1.3 Dioxygénases dépendant du Fe(II) et du 2-oxoglutarate	183
1.4 Monooxygénase à flavine	184
1.5 Méthyltransférase	186
2. <i>Identification de nouvelles activités de modification</i>	187
3. <i>Le cas de la méthyltransférase R</i>	198
<b>III. Enzymes de modification du cyclo(Ala-Glu) : activité <i>in vitro</i> et caractérisation structurale des produits</b>	<b>201</b>
1. <i>Reproduction <i>in vitro</i> de l'activité enzymatique d'une OG</i>	201
1.1 Purification de l'enzyme K	201
1.2 Caractérisation de l'activité <i>in vitro</i>	203
1.2.1 Produits obtenus <i>in vitro</i>	203
1.2.2 Etude cinétique préliminaire	205
1.2.3 Dépendance aux cofacteurs	205
2. <i>Elucidation structurales des dérivés du cyclo(Ala-Glu)</i>	207
2.1 Données HRMS	207
2.2 Analyses RMN	209
<b>IV. Discussion, conclusion et perspectives</b>	<b>213</b>
<b>CONCLUSION GÉNÉRALE</b>	<b>217</b>
<b>ANNEXES</b>	<b>221</b>
<i>ANNEXE 1 : EPREUVES DE L'ARTICLE ACCEPTE DANS NATURE CHEMICAL BIOLOGY</i>	222
<i>ANNEXE 2 : DÉTAILS DES RÉSULTATS HRMS OBTENUS SUR LE CYCLO(ALA-GLU) ET LES COMPOSÉS IJ18, IJ19, IJ20, PB2, PB3, PB4 ET PB6.</i>	235
<i>ANNEXE 3 : SÉLECTION DE SPECTRES RMN AYANT SERVI A L'ÉLUCIDATION STRUCTURALE DE PB2, PB3, PB4 ET PB6.</i>	238

## **TABLE DES FIGURES**

Figure 1 : Stratégie pour la découverte de produits naturels	25
Figure 2 : Structure générale des dicétopipérazines	30
Figure 3 : Exemples de dicétopipérazines : organismes d'origine et structures	32
Figure 4 : Mécanismes de fragmentation des cyclodipeptides protonés, reproduit de <sup>62</sup>	34
Figure 5 : Annotation du spectre de masse MS <sup>2</sup> du cyclo(Pro-Val) obtenu par ESI-MS en mode positif <sup>58</sup>	36
Figure 6 : Structure moléculaire générique des acyl-homosérine lactones	37
Figure 7 : Dicétopipérazines impliquées dans la virulence de micro-organismes pathogènes : structures et organismes d'origine	42
Figure 8 : Dicétopipérazines liant le fer : structures et organismes d'origine	43
Figure 9 : Structure de la bicyclomycine	45
Figure 10 : Structures de la tryprostatine B et tryprostatine A	46
Figure 11 : Positions variables de la tryprostatine A, d'après <sup>191</sup>	47
Figure 12 : Structures de la fumitremorgine C et de son analogue optimisé Kol143	47
Figure 13 : Structures de la phénylahistine, de la déshydrophénylahistine et de la plinabuline	48
Figure 14 : Structure générale des spirodicétopipérazines	48
Figure 15 : Voies de synthèse pour formation du cycle dicétopipérazine (d'après <sup>46</sup> )	51
Figure 16 : Synthèse de dicétopipérazine par la formation de liaison amide (inspiré de <sup>31</sup> )	51
Figure 17 : Isonitriles utilisés pour la réaction de Ugi : 1-isocyanocyclohexene et 1-isocyano-2(2,2-dimethoxyethyl)benzene	52
Figure 18 : Condensation d'acides aminés (d'après <sup>46</sup> )	53
Figure 19 : Cyclisation via $\alpha$ -haloacétamides (chloroacétamides) (d'après <sup>46</sup> )	53
Figure 20 : Cyclisation via acylation d'un énolate (d'après <sup>46</sup> )	54
Figure 21 : Schéma de réaction de la fonctionnalisation en C-1 et C-3 par une condensation aldolique	55
Figure 22 : Organisation de la tyrocidine synthétase	57
Figure 23 : Réaction au domaine d'adénylation	58
Figure 24 : Réaction au domaine de thiolation	59
Figure 25 : Modèle montrant les étapes permettant l'élongation du peptide sur un module, reproduit de <sup>275</sup>	60
Figure 26 : Réaction au domaine de condensation	60
Figure 27 : Réaction au domaine de thiolation	61
Figure 28 : Dicétopipérazines synthétisées par des voies de biosynthèse à NRPS	63
Figure 29 : Biosynthèse de la gliotoxine	64
Figure 30 : Biosynthèse des dicétopipérazines de type fumitremorgine	66
Figure 31 : Biosynthèse de la thaxtomine A	67
Figure 32 : Biosynthèse de l'érythrocheline	68
Figure 33 : Biosynthèse de l'ergopeptine	69

Figure 34 : Comparaison des trois structures cristallographiques de CDPS	73
Figure 35 : Superposition de la structure d'AlbC et de la tyrosyl-ARNt synthétase de <i>Methanococcus jannaschii</i> (TyrRS <sub>Mj</sub> ) en complexe avec son substrat L-tyrosine. Reproduit de <sup>361</sup>	74
Figure 36 : Etapes clés du mécanisme catalytique menant à la synthèse d'un cyclodipeptide	75
Figure 37 : Réaction du ZPCK sur le variant S37C d'AlbC et structure cristallographique du complexe mimant l'intermédiaire dipeptidyl-enzyme	76
Figure 38 : Mécanisme réactionnel proposé pour la formation d'un cyclodipeptide	78
Figure 39 : Régions d'AlbC impliquées dans la liaison à l'ARNt	79
Figure 40 : Voie de biosynthèse de l'albonoursine	82
Figure 41 : Voie de biosynthèse de la pulcherrimine	83
Figure 42 : Voie de biosynthèse de la mycocyclosine	84
Figure 43 : Cartes des plasmides pQE60, pRSF-Duet et pREP4	90
Figure 44 : Suppression du site <i>XbaI</i> par digestion, polymérisation d'ADN et ligature sur lui-même du vecteur.	92
Figure 45 : Création d'une cassette mutée contenant les sites <i>SpeI</i> et <i>XbaI</i> par PCR SOEing	93
Figure 46 : Carte du plasmide pIJ194 dérivé du pRSF-Duet commercial	94
Figure 47 : Carte du plasmide pIJ196 dérivé du pQE60 commercial	95
Figure 48 : Stratégie d'insertion d'un gène d'enzyme de modification dans pQE60	110
Figure 49 : Etapes pour la modification du pRSF-Duet	110
Figure 50 : Alignement du début de la séquence d'AlbC, de pc1814 et de pc1814long	111
Figure 51 : Toxicité liée au clonage des CDPS	112
Figure 52 : Stratégie expérimentale mise en place pour le criblage des activités de synthèse et l'expression des enzymes des voies à CDPS	122
Figure 53 : Alignement multiple partiel d'AlbC et des CDPS potentielles dont la partie N-terminale est manquante	128
Figure 54 : Alignement partiel avec AlbC de la CDPS potentielle encodée par SCAT_0901	129
Figure 55 : Arbre phylogénétique des 82 CDPS connues et CDPS potentielles identifiées dans les bases de données par bio-informatique (Mai 2013) : analyse préliminaire et sélection.	130
Figure 56 : Arbre phylogénétique des 82 CDPS connues et CDPS potentielles identifiées dans les bases de données par bio-informatique (Mai 2013) : résultats des tests d'activité et sous-familles	134
Figure 57 : Analyses SDS-PAGE de l'expression des CDPS potentielles inactives dans nos conditions expérimentales	135
Figure 58 : Signature fonctionnelle des CDPS représentée par un logo correspondant à la fréquence des acides aminés catalytiques aux positions 37, 40, 178, 182, 202 et 203 (en numérotation d'AlbC)	136
Figure 59 : Signature fonctionnelle pour les sous-familles de CDPS (NYH et XYP) et résidus catalytiques du membre de la famille potentielle SYQ	137
Figure 60 : Résidus catalytiques et résidus conservés dans les sous-familles NYH et XYP	139

Figure 61 : Insertions et délétions observées chez les CDPS 21, 23, 25, 26, 28 et 29 localisées sur une représentation topologique des éléments de structures secondaires communes aux trois CDPS de structures connues	140
Figure 62 : Comparaison des analyses LC-MS/MS des surnageants de culture de bactéries <i>E. coli</i> n'exprimant pas de CDPS (en noir) et exprimant la CDPS 37 (en rouge)	141
Figure 63 : Cyclodipeptides produits par la famille des CDPS	143
Figure 64 : Diversité de cyclodipeptides produits par les CDPS	145
Figure 65 : Résidus constituant les poches P1 et P2	147
Figure 66 : Motif de séquence des acides aminés constituant les deux poches P1 et P2 pour les groupes synthétisant le même cyclodipeptide	149
Figure 67 : Analyse SDS-PAGE de l'expression des CDPS et CDPS potentielles 1 à 46	156
Figure 68 : Analyse SDS-PAGE des CDPS purifiées	159
Figure 69 : Environnements génomiques des CDPS d'origine procaryote testées pour leur activité de synthèse de cyclodipeptides dans le chapitre II	173
Figure 70 : Environnement génomique de la CDPS 19 et synthénie	177
Figure 71 : Synthénie dans les environnements génomiques des CDPS produisant du cyclo(Xaa-Glu)	178
Figure 72 : Structure tridimensionnelle des dioxygénases dépendant du Fe(II) et du 2-oxoglutarate, adaptée de AlkB ( <i>Escherichia coli</i> alkylation repair protein B (PDB : 3BIE))	183
Figure 73 : Mécanisme du cycle catalytique des dioxygénases dépendant du Fe(II) et du 2-oxoglutarate (bleu) menant à la monoxydation du substrat (vert) et la formation de succinate et CO <sub>2</sub>	184
Figure 74 : Structures de la riboflavine, FMN et FAD	185
Figure 75 : Analyse SDS-PAGE de la production des enzymes de modification, hors méthyltransférase R	190
Figure 76 : Analyse SDS-PAGE de la production de méthyltransférase R	190
Figure 77 : Comparaison de l'analyse LC-MS/MS du surnageant de culture de bactéries <i>E. coli</i> exprimant la CDPS 37 en vert et co-exprimant la CDPS 37 et l'enzyme de modification potentielle J (en rouge)	191
Figure 78 : Fragmentation MS <sup>2</sup> (a.), MS <sup>3</sup> (b.) et MS <sup>4</sup> (c.) du cyclo(Ala-Glu) ( $m/z$ 201, $t_R$ =23,5 min (Hybercarb))	192
Figure 79 : Fragmentation MS <sup>2</sup> (a.), MS <sup>3</sup> (b.) et MS <sup>4</sup> (c.) du composé inconnu de $m/z$ 199 et $t_R$ =23,5 min (Hybercarb)	193
Figure 80 : Données MS <sup>n</sup> sous le pic 3 à $t_R$ =30,3 min (Hybercarb). Données MS (a.). Fragmentation MS <sup>2</sup> (506) (b.), MS <sup>3</sup> (506 → 308) (b.) et MS <sup>2</sup> (272,5) (c.)	194
Figure 81 : Fragmentation MS <sup>2</sup> du cyclo(Trp-Trp) (a.) et cyclo(Trp-Trp) méthylé (b.)	198
Figure 82 : Fragmentation MS <sup>3</sup> (373→242) du cyclo(Trp-Trp) (a.) et MS <sup>3</sup> (387→256) du cyclo(Trp-Trp) méthylé (b.)	199
Figure 83 : Purification de la protéine K sur une IMAC chargée en nickel	202
Figure 84 : Analyse SDS-PAGE des différentes étapes de la purification	203

Figure 85 : Analyse LC-MS/MS du milieu réactionnel en présence (en bleu) et en l'absence (en vert) de cyclo(Ala-Glu)	204
Figure 86 : Suivi pendant 60 min de la consommation du cyclo(Ala-Glu) et de la formation des composés PB2, PB3, PB4, PB5 et PB6 catalysées par la dioxygénase K	205
Figure 87 : Effets de la présence de cofacteurs, d'additifs et d'inhibiteurs sur la consommation de cyclo(Ala-Glu) par la dioxygénase K	206
Figure 88 : Structures des composés PB2a, PB2b, PB3a, PB3b, PB4 et PB6 (Z ou E) déterminées par RMN	211
Figure 89 : Activités de modification de cyclodipeptides connues dans les voies de biosynthèse dépendant des CDPS	216

## **TABLES DES TABLEAUX**

Tableau 1 : Ions ammonium et dérivés des 20 acides aminés protéinogéniques .....	35
Tableau 2 : Exemples de dicétopipérazines présentant une activité antibactérienne ou antibiotique .....	44
Tableau 3 : Cyclodipeptides synthétisés par <i>E. coli</i> recombinant exprimant les CDPS.....	71
Tableau 4 : Acides aminés accommodés par chacune des deux poches des CDPS caractérisées et les résidus constituant potentiellement ces deux poches. D'après <sup>361</sup> .....	80
Tableau 5 : Oligonucléotides utilisés pour la PCR SOEing <sup>1</sup> .....	92
Tableau 6 : Oligonucléotides utilisés pour l'amplification des gènes de CDPS .....	96
Tableau 7 : Oligonucléotides utilisés pour l'amplification des gènes de CDPS n° 47 à 49 .....	96
Tableau 8 : Oligonucléotides utilisés pour l'amplification du gène d'enzyme de modification R.....	97
Tableau 9 : Bilan des avantages et inconvénients des souches <i>E. coli</i> pour l'étude des activités de synthèse en moyen débit.....	115
Tableau 10 : Suivi de la croissance bactérienne et de la production de cyclo(Phe-Leu) en fonction du milieu et de la température de culture à 24 et 48 h .....	117
Tableau 11 : Suivi à 24, 30 et 48 h de la production de cyclo(Phe-Leu) en milieu auto-induit en fonction de la souche (BL21AI ou M15-pREP4) utilisée.....	117
Tableau 12 : Temps de rétention des cyclodipeptides polaires analysés par HPLC sur une colonne Atlantis dC18.....	120
Tableau 13 : Temps de rétention des cyclodipeptides polaires analysés par HPLC sur une colonne HyperCarb .....	121
Tableau 14 : CDPS sélectionnées pour l'étude, désignées par le locus de leur gène, le numéro d'identifiant au NCBI, l'organisme d'origine et le numéro attribué dans l'étude. ....	131
Tableau 15 : Gradients optimisés pour la purification des cyclodipeptides.....	142
Tableau 16 : Cyclodipeptides majoritaires produits par les CDPS nouvellement identifiées.....	144
Tableau 17 : Comparaison des résidus constituant les poches P1 et P2 chez les CDPS synthétisant du cyclo(Leu-Leu).....	148
Tableau 18 : Comparaison des résidus constituant les poches P1 et P2 chez les CDPS synthétisant du cyclo(Leu-Glu) ou du cyclo(Ala-Glu).....	151
Tableau 19 : Prédiction et validation de l'activité de synthèse de cyclodipeptides des CDPS.....	152
Tableau 20 : CDPS détectées dans les fractions soluble (S) et insoluble (I) préparées depuis des culots bactériens obtenus à l'issue du criblage d'activité de synthèse de cyclodipeptide.....	157
Tableau 21 : Résumé des observations réalisées au cours des purifications et quantité finale des CDPS évaluée par absorbance à 280 nm et par Bradford.....	160
Tableau 22 : Production de cyclo(Phe-leu) dans les différentes conditions testées.....	162
Tableau 23 : Test <i>in vitro</i> de l'activité de synthèse de cyclodipeptides <sup>1</sup> .....	163
Tableau 24 : Pourcentage d'identité des séquences entre les séquences des protéines conservées chez <i>G. effusa</i> , <i>N. brasiliensis</i> , <i>S. albulus</i> , <i>P. sp. M47T1</i> , <i>P. fluorescens</i> , <i>V. brasiliensis</i> et <i>B. lata</i> (sp. 383).....	179

Tableau 25 : Enzymes de modifications sélectionnées pour l'étude moyen-débit <i>in vivo</i> .....	181
Tableau 26 : Classification des FMO adapté de <sup>410</sup> .....	186
Tableau 27 : Composés apparaissant lorsque les enzymes de modification sélectionnées sont co-exprimées avec la CDPS correspondant (d'après l'annexe électronique 2).....	195
Tableau 28 : Produits de la réaction catalysée par l'enzyme K sur le cyclo(Ala-Glu).....	204
Tableau 29 : Conditions testées pour mesurer l'effet de la présence de cofacteurs, d'additifs et d'inhibiteurs .....	206
Tableau 30 : Masse haute résolution et formule brute probable pour les dérivés du cyclo(Ala-Glu).....	208
Tableau 31 : Attribution des signaux RMN des composés en solutions (DMSO-d <sub>6</sub> , 30 °C).....	212

## **TABLE DES ANNEXES ÉLECTRONIQUES**

Annexe électronique 1 : Informations disponibles dans les bases de données et séquences relatives aux enzymes étudiées (format .xlsx)

Annexe électronique 2 : Cyclodipeptides et dicétopipérazines retrouvés dans les surnageants de culture (format .xlsx)

Annexe électronique 3 : Identité de séquence entre les 60 séquences d'intérêt (format .xlsx)

Annexe électronique 4 : Résultats HHPRED pour les CDPS sélectionnées (format .xlsx)

Annexe électronique 5 : Alignement des 52 CDPS actives et des 8 CDPS inactives (formats .txt ou .pdf)

Annexe électronique 6 : Prédiction des résidus constituant les poches P1 et P2 (format .xlsx)

Annexe électronique 7 : Arbre phylogénétique des CDPS Sept 2014 (format .pdf)

Annexe électronique 8 : Prédiction des groupes de CDPS ayant la même activité de synthèse de cyclodipeptides (format .xlsx)

Annexe électronique 9 : Bases N1-N72 des ARNt disponibles dans les bases de données pour les organismes contenant une CDPS étudiée (format .xlsx)

## **ABRÉVIATIONS**

2OG : 2-oxoglutarate  
6×His : Etiquette polyhistidine  
aa-ARNt : Aminoacyl-ARNt  
aaRS : Aminoacyl-ARNt-synthétase  
ADH : Alcool-déshydrogénase  
ADN : Acide désoxyribonucléique  
AMPc : adénosine monophosphate cyclique  
ATP : adénosine triphosphate  
BCRP : Protéine de résistance du cancer du sein  
DFT : Théorie de la fonctionnelle de la densité  
DTT : Dithiothréitol  
ESI-MS<sup>n</sup> : Spectrométrie de masse à source d'ionisation par électronebulisation  
FAS : Synthèses d'acides gras  
FMO : Monooxygénase à flavine  
HPLC : Chromatographie liquide à haute performance  
IPNS : Isopénicilline-N-synthases  
IPTG : Isopropyl β-D-1-thiogalactopyranoside  
LC-MS/MS : Chromatographie liquide couplée à la spectrométrie de masse en tandem  
MAP2 : Protéine associée aux microtubules-2  
MDR : Déshydrogénase/réductase à chaîne moyenne  
MFS : Superfamille de facilitateurs majeurs  
MS<sup>n</sup> : Spectrométrie de masse par fragmentation multiple  
NRPS : Synthétase de peptides non ribosomiques  
OG : Dioxygénases dépendant du Fe(II) et du 2-oxoglutarate  
PCR : Réaction en chaîne de polymérisation  
PKS : Synthase de polycétides  
PMSF : Phénylméthylsulfonylfluoride  
PSP : Phosphosérine phosphatase  
RMN : Résonance magnétique nucléaire  
SDR : Déshydrogénase/réductase à chaîne courte  
SDS-PAGE : Électrophorèse sur gel de polyacrylamide en présence de dodécylsulfate de sodium  
TONB : Récepteur dépendant de TonB

# INTRODUCTION GÉNÉRALE

Les produits naturels, provenant en premier lieu des plantes, mais aussi des micro-organismes et des animaux, ont constitué depuis au moins cinq millénaires, la base des systèmes de médecines traditionnelles. Des traces de l'utilisation de pharmacopée à base de produits naturels ont été retrouvées dans toutes les grandes civilisations : Mésopotamie, Egypte pharaonique, Chine impériale, Antiquité gréco-romaine et monde Arabe durant le Moyen-Âge<sup>1</sup>. Avec la découverte, en 1929 par Alexander Fleming, de la pénicilline produite par le champignon filamenteux *Penicillium notatum*, nous sommes entrés dans l'« Âge d'Or des antibiotiques ». Cette découverte a également été à l'origine de la recherche de nouveaux agents bioactifs dans la nature et en particulier dans le monde microbien. Les micro-organismes sont une source riche de métabolites bioactifs aux structures diverses. Ils sont à l'origine de certaines des molécules les plus importantes de l'industrie pharmaceutique, à savoir de nombreux antibiotiques (pénicillines, céphalosporine, aminoglycoside, tétracycline), des agents immunosuppresseurs (cyclosporine et rapamycine), des agents hypolypémiants (mévastatine, lovastatine) et des agents antihelminthiques et antiparasitaires (ivermectines)<sup>1</sup>. Ainsi depuis 50 ans, les produits naturels et leurs dérivés semi-synthétiques sont une source importante de nouveaux principes actifs<sup>2</sup>.

Cependant, dans les années 1980 et 1990, un déclin d'intérêt pour les produits naturels s'est amorcé. L'un des principaux obstacles à l'utilisation des produits naturels est la difficulté que représente leur isolement depuis les sources où les principes actifs sont souvent présents en faible quantité, rendant leur utilisation dans des criblages à haut débit ardue. De plus, pendant quelques années, le nombre de nouveaux produits naturels identifiés a chuté, principalement à cause de l'isolement récurrent de produits déjà connus. Une partie de la communauté scientifique en était arrivée à penser que le potentiel thérapeutique des produits naturels avait été épuisé. Dans le même temps, le développement de techniques de chimie combinatoire a ouvert une nouvelle voie de diversité chimique dans laquelle les industries pharmaceutiques se sont « engouffrées »<sup>3,4</sup>. Mais la chimie combinatoire s'est révélée une source pauvre de molécules bioactives. De 1981 à 2010, une seule entité issue *de novo* d'une telle stratégie a été approuvée, alors même que cette dernière concentrait la majeure partie des efforts de l'industrie pharmaceutique.<sup>2</sup> De plus, suite à ce désintérêt pour les produits naturels au profit de la chimie combinatoire, une diminution du nombre de médicaments approuvés a été observée<sup>2</sup>. La chimie combinatoire, en réalité, se révèle beaucoup plus adaptée dans l'optimisation structurale de molécules « tête de série », c'est-à-dire ayant déjà montré une activité thérapeutique. Un regain d'intérêt pour les produits naturels s'est ensuite fait sentir, encouragé par l'idée qu'une petite chimiothèque de molécules bien choisies montrant une large diversité chimique et couvrant donc un espace chimique important<sup>5</sup> apportait plus de résultats qu'une grande chimiothèque constituée de molécules assemblées au hasard. La diversité structurale n'est pas le seul atout des molécules d'origine naturelle. Les produits naturels se distinguent aussi pour leurs activités hautement sélectives et spécifiques<sup>1</sup>. Selon Cragg et Newman, la bioactivité des produits naturels peut s'expliquer en faisant l'hypothèse que tous les produits naturels ont une activité de liaison à un récepteur, sélectionnée par l'évolution<sup>6</sup> ; la difficulté se résume alors à trouver quel est ce récepteur<sup>1</sup>.

Néanmoins, la richesse des produits naturels reste encore sous-exploitée. En ce qui concerne les micro-organismes, il est estimé que seulement 1 % des micro-organismes connus sont cultivables en laboratoire<sup>1</sup> et que seulement une infime partie des micro-organismes est connue<sup>7</sup>. L'étude des génomes bactériens révèle que chaque souche serait capable de produire des produits naturels, au moins entre 25 et 30 molécules différentes, et que 90% des molécules d'origine naturelle seraient « cryptiques »<sup>8</sup>. La mise au point de technologies de séquençages à haut débit marque le passage de l'ère pré-génomique à l'ère post-génomique avec une croissance exponentielle du nombre de génomes séquencés. L'entrée dans l'ère post-génomique et la découverte de voies menant à la biosynthèse de métabolites secondaires, notamment celles impliquant des mégacomplexes enzymatiques responsables de la synthèse de peptides non-ribosomiques et de polycétides, a changé la manière de rechercher de nouveaux produits naturels (Figure 1). Désormais, l'isolement de nouvelles molécules, *via* l'activité biologique, n'est plus la seule voie d'accès aux produits naturels et l'identification de clusters de gènes responsables de la biosynthèse de métabolites secondaires, selon une approche dite de « genome-mining », c'est-à-dire d'exploitation des données du génome, prend une place de plus en plus importante. En outre, avec l'explosion du séquençage génomique, les banques de données génomiques s'enrichissent de séquences provenant d'organismes extrémophiles ou marins et également de fragments de génomes provenant d'organismes non-cultivables (métagénome)<sup>1</sup>.

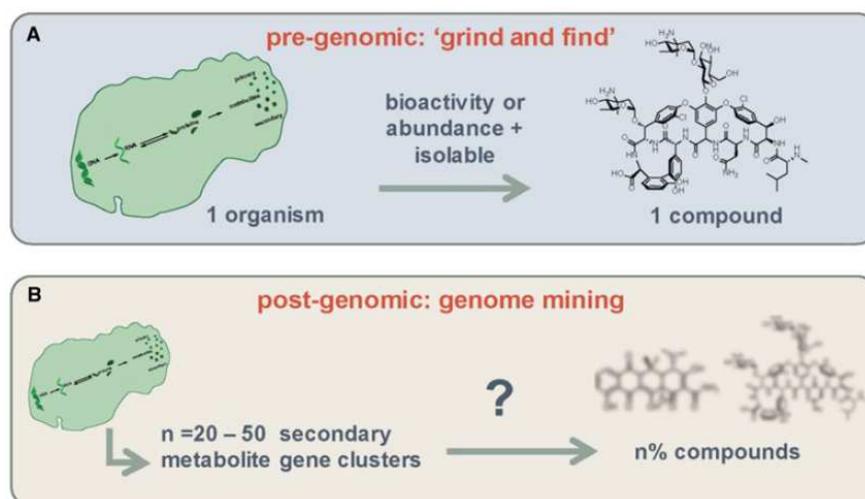


Figure 1 : Stratégie pour la découverte de produits naturels

a. Stratégie historique pour la découverte de nouveaux produits en « grind and find » (« mouliner et identifier »). b. Stratégie de l'ère post-génomique basée sur le « genome-mining » (« exploitation des données du génome »). Reproduite de<sup>9</sup>

Dans ce contexte, trois défis majeurs sont à relever. Le premier consiste à être capable d'activer l'expression de clusters de gènes de biosynthèse. En effet, l'activation de l'expression de ces clusters, souvent cryptiques dans l'hôte d'origine, est une condition essentielle à la synthèse des produits naturels. Le second vise à produire les molécules naturelles d'intérêt en quantité suffisante pour envisager le

développement de médicament, en proposant une alternative à une synthèse par voie chimique qui peut s'avérer ardue. Le troisième entend générer des analogues « non-naturels » de produits naturels dont l'activité biologique ou les propriétés physicochimiques ont été améliorées<sup>10</sup>.

Les stratégies utilisées pour les deux premiers défis se rejoignent, car elles consistent toutes deux à augmenter la production de produits naturels, soit au moment de leur découverte soit dans des stades plus avancés du développement d'un médicament. Des méthodes d'ingénierie utilisant l'hôte naturel ou des approches visant à modifier son environnement ont été développées pour permettre la production de produits naturels. Une reconstitution *in vitro* de la voie peut également être envisagée, bien que celle-ci nécessite l'expression recombinante et la purification de chaque enzyme de la voie, ce qui peut être laborieux. Mais c'est la surexpression hétérologue des gènes de la voie dans un hôte facilement manipulable qui reste l'option la plus employée<sup>10</sup>. A ces stratégies s'ajoutent des méthodes pour faciliter l'identification de métabolites des voies faiblement exprimées. La prédiction par des méthodes bioinformatiques de la structure des produits synthétisés permet notamment de rechercher spécifiquement les molécules attendues<sup>11</sup>.

En ce qui concerne le troisième défi, la connaissance des enzymes responsables de la biosynthèse des produits naturels permet d'envisager, par des approches de bio-ingénierie, la production des molécules « naturelles non naturelles » aux propriétés biologiques et pharmacologiques améliorées. Il est ainsi possible d'inactiver des gènes dans des voies mais aussi d'en introduire de nouveaux, de faire prendre en charge des substrats ou des précurseurs non naturels pour faire de la mutasynthèse, ou de recourir à l'évolution dirigée d'enzymes pour modifier ou étendre leur spécificité<sup>12,13</sup>. Les premières expériences dans le domaine ont été effectuées sur deux grandes classes de voies : l'une menant à la biosynthèse des polycétides et l'autre menant à la biosynthèse des peptides non ribosomiques<sup>14,6,15</sup>. Ces composés sont synthétisés par des mégasynthases multifonctionnelles appelées PKS (Polyketide Synthase) et NRPS (Non-ribosomal Peptide Synthetase). Ces mégacomplexes enzymatiques sont organisés en modules, chacun responsable de l'incorporation d'un précurseur dans le produit final. La nature modulaire de ces mégacomplexes enzymatiques se prête particulièrement bien à la manipulation génétique. Le squelette des peptides ou polycétides peut être principalement modifié en changeant la nature des monomères incorporés ou leur nombre<sup>16</sup>. Par ailleurs, des enzymes de modification ou « tailoring enzymes » codées dans les clusters de gènes de produits naturels sont responsables de la « maturation » de ces composés : elles catalysent une grande variété de réactions, incluant des glycosylations, des halogénations, des hydroxylations et diversifient ainsi grandement la structure des molécules. Les enzymes de modification sont donc également des candidates idéales pour des approches de bio-ingénierie<sup>10,16</sup>.

Les cyclodipeptides et leurs dérivés, les 2,5-dicétopipérazines, sont une classe de molécules naturelles particulièrement représentée chez les micro-organismes. Ces molécules possèdent une importante diversité structurale autour d'un noyau 2,5-dicétopipérazine commun. Elles présentent des activités biologiques et pharmacologiques remarquables et extrêmement variées. Certaines 2,5-dicétopipérazines présentent ainsi des activités anticancéreuses comme les phénylahistines<sup>17</sup> ou les tryprostatines<sup>18</sup>, d'autres possèdent des activités antibiotiques comme la bicyclomycine<sup>19</sup>, antifongiques

comme le cyclo(L-Phe-L-Pro)<sup>20</sup>, antivirales et immunosuppressives comme la gliotoxine<sup>21,22</sup>, ou neuroprotectrices comme le cyclo(L-His-L-Pro)<sup>23</sup>.

Jusque dans les années 2000, les seules voies décrites pour la synthèse de 2,5-dicétopipérazines faisaient intervenir des NRPS. Les travaux entrepris par l'équipe de Muriel Gondry, en collaboration avec l'équipe de Jean-Luc Pernodet, a mis en évidence l'existence de voies de biosynthèse des 2,5-dicétopipérazines indépendant des NRPS. Il s'agit de voies faisant intervenir une nouvelle famille d'enzymes : les synthases de cyclodipeptides (CDPS). Ces travaux ont débuté avec le déchiffrement de la voie de biosynthèse de l'albonoursine chez *Streptomyces noursei*. Cette voie fait intervenir deux enzymes : une CDPS, AlbC, responsable de la formation du cycle 2,5-dicétopipérazine du cyclo(L-Phe-L-Leu) et une enzyme de modification, la cyclodipeptide oxydase (CDO), responsable de la  $\alpha,\beta$ -déshydrogénation des chaînes latérales du cyclo(L-Leu-L-Phe) pour donner l'albonoursine<sup>24,25</sup>. D'autres enzymes apparentées à AlbC ont ensuite été identifiées et ont montré également des activités de synthèse de cyclodipeptide<sup>26,27</sup>. Au début de mon projet de thèse, neuf CDPS faisaient partie de la famille des CDPS. En plus de la voie de biosynthèse de l'albonoursine, deux autres voies de biosynthèse de 2,5-dicétopipérazines dépendant des CDPS avaient été caractérisées. Cependant, des études bio-informatiques avaient montré que les bases de données génomiques contiennent des gènes codant des CDPS potentielles, parfois associés à des gènes codant des enzymes de modification potentielles<sup>28,29</sup>. En 2012, une cinquantaine de gènes de CDPS avait été recensés par l'équipe et ce nombre n'a cessé d'augmenter à chaque interrogation des bases de données. Etant donné le potentiel thérapeutique des dicétopipérazines, l'élucidation de nouvelles voies de biosynthèse et l'identification de nouvelles dicétopipérazines apparaissent alors extrêmement pertinentes et attractives.

L'objectif de mon doctorat a donc été d'explorer la diversité moléculaire encodée par les voies de biosynthèse de 2,5-dicétopipérazines dépendant des CDPS. Je me suis intéressée aussi bien aux CDPS qu'aux enzymes de modification potentielles, la diversité de structure des 2,5-dicétopipérazines étant apportée aussi bien par les unes que par les autres. Devant le petit nombre de CDPS et de voies caractérisées comparé au nombre grandissant de gènes de CDPS potentielles identifiés dans les bases de données, un changement de stratégie s'imposait pour la découverte et le déchiffrement de nouvelles voies de biosynthèse dépendant de CDPS. La stratégie que nous avons adoptée au cours de mon doctorat à consister à étudier en parallèle plusieurs dizaines de CDPS et une vingtaine d'enzymes de modification.

Après une étude bibliographique, présentant les 2,5 dicétopipérazines, leurs activités biologiques et leurs voies de synthèse chimique et biologique, j'exposerai les résultats de mon travail. Ces derniers s'articulent en trois chapitres. Le premier chapitre présentera la mise au point d'une méthode de caractérisation moyen-débit des activités des enzymes des voies à CDPS. La stratégie que nous avons retenue nous a en effet imposé une première phase dans laquelle notre protocole d'étude des CDPS et des enzymes de modification a été adapté à nos objectifs. Le second chapitre a été consacré à la caractérisation de nouvelles activités CDPS, ce qui non seulement a permis d'élargir la gamme des cyclodipeptides produits par les CDPS mais a également apporté des informations sur la répartition phylogénétique des

CDPS et posé les bases de la spécificité des CDPS. Enfin, le dernier chapitre est consacré à l'étude des enzymes de modification. Les travaux présentés dans ce dernier chapitre ont mené à la caractérisation d'une nouvelle famille d'enzymes de modification dans les voies à CDPS et à la caractérisation de nouvelles dicétopipérazines.

# ETUDE BIBLIOGRAPHIQUE

## I. Les 2,5-dicétopipérazines

Les 2,5-dicétopipérazines (ou 2,5-dioxopipérazines ou pipérazines-2,5-diones) sont une famille de composés chimiques organiques cycliques. Ces molécules sont particulièrement abondantes dans la nature et sont connues en tant que produits naturels depuis le début du XX<sup>ème</sup> siècle. Elles ont été, pendant de nombreuses années, considérées comme des produits secondaires de réactions de dégradation d'oligo- ou de polypeptides, ne présentant que peu d'intérêt<sup>30,31</sup>. Depuis les années 1970, avec la découverte de l'éventail d'activités biologiques pouvant leur être associées, les 2,5-dicétopipérazines<sup>30,32-35</sup> suscitent un intérêt de plus en plus marqué et elles ont pu être isolées en grand nombre à partir de diverses sources biologiques.

Dans cette première partie d'étude bibliographique, je vais tout d'abord définir et caractériser ces molécules en tant qu'entités chimiques. Puis, je m'intéresserai aux 2,5-dicétopipérazines d'origine naturelle et présenterai leur diversité de sources et de structures. Je finirai cette partie en documentant la technique principale que j'ai utilisée pour leur identification au cours de mes travaux.

### 1. Définition et propriétés du cycle 2,5-dicétopipérazine

Les 2,5-dicétopipérazines se définissent par la présence du cycle 2,5 dicétopipérazine.

#### 1.1 Définition du cycle 2,5-dicétopipérazine

Le noyau 2,5-dicétopipérazine est un hétérocycle à six atomes possédant deux fonctions lactames. Il peut également être défini comme un noyau pipérazine possédant deux fonctions cétones, placées en opposition aux positions 2 et 5. La structure générale des 2,5-dicétopipérazines est représentée (Figure 2). Lorsque la formation de ce cycle provient de la condensation de deux acides  $\alpha$ -aminés, les 2,5-dicétopipérazines résultantes sont appelées cyclodipeptides ou dipeptides cycliques.

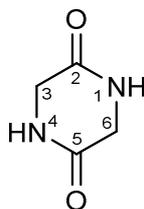


Figure 2 : Structure générale des dicétopipérazines

La numérotation des atomes du cycle 2,5-dicétopipérazines est indiquée.

Dans ce manuscrit, un cyclodipeptide formé à partir d'acides aminés protéinogéniques (de configuration L) est désigné par une nomenclature utilisant le code à trois lettres : cyclo(Xaa-Xaa). La stéréochimie des acides aminés de configuration D sera précisée en les faisant précéder de D et l' $\alpha$ - $\beta$ -

déshydrogénation de la chaîne latérale sera indiqué par  $\Delta$ . Par ailleurs, les termes dicétopipérazines et 2,5-dicétopipérazines seront employés indifféremment dans la suite du manuscrit.

## 1.2 Conformation du cycle 2,5-dicétopipérazine

La conformation adoptée par le cycle dicétopipérazine dépend des forces extérieures s'appliquant sur ce cycle, de la présence de substituants et de la nature de ceux-ci.

Le cycle non substitué (cyclo(Gly-Gly)) présente dans la plupart des cas observés (cristallographie aux rayons X<sup>36</sup>, RMN en solution<sup>37,38</sup>, analyse vibrationnelle et calcul par DFT<sup>39</sup>) une conformation plane, alors que des calculs *ab initio* et des spectres micro-onde suggèrent une conformation bateau préférentielle<sup>40</sup>. L'écart énergétique calculé entre les deux conformations étant faible (1.3-1.7 kcal/mol), il est probable que les forces s'appliquant sur le cycle, provenant du cristal ou de la solution, puissent imposer une conformation par rapport à l'autre<sup>41</sup>. Les dicétopipérazines symétriques *cis*-disubstituées adoptent préférentiellement, selon des expériences de chimie quantique, des conformations bateaux<sup>42</sup> qui peuvent être observées aplaties ou twistées voire planes pour le cyclo(Ser-Ser) à l'état solide. Les dicétopipérazines *trans*-disubstituées présentent des conformations chaises aplaties ou planes<sup>43</sup>. Les dicétopipérazines tétrasubstituées possèdent généralement une conformation plane<sup>43</sup>. Dans le cas de substituants particulièrement encombrés stériquement, la conformation est telle qu'elle minimise les interactions stériques. Par ailleurs, la conformation adoptée peut être grandement influencée par la tendance des substituants aromatiques à se superposer au cycle dicétopipérazine<sup>37</sup>. Enfin, dans les cyclodipeptides contenant une proline, la configuration bateau est stable car le cycle à cinq de la proline est fusionné avec le cycle dicétopipérazine, ce qui impose une configuration rigide<sup>44,45</sup>.

## 1.3 Propriétés chimiques des 2,5-dicétopipérazines

Au regard de ses propriétés chimiques, le cycle 2,5-dicétopipérazine constitue une « architecture » privilégiée en chimie médicinale. Les molécules possédant cet hétérocycle rigide et de conformation contrainte peuvent mimer une conformation peptidique préférentielle, tout en étant stables à la protéolyse contrairement aux peptides linéaires. Elles possèdent deux accepteurs et deux donneurs de liaison hydrogène qui leur permettent d'interagir avec des cibles protéiques (enzyme ou récepteur). Par ailleurs, le cycle possède six positions substituables dont quatre de manière stéréochimiquement contrôlée, ce qui permet d'introduire une importante diversité chimique<sup>46</sup>.

La solubilité des dicétopipérazines dépend de la nature des chaînes latérales qui les constituent. Le cycle dicétopipérazine, avec ses deux fonctions lactames qui établissent des liaisons hydrogène dans les solvants protiques, suffit à solubiliser les dicétopipérazines simples. Lorsque les chaînes sont plus complexes et hydrophobes, la solubilité des dicétopipérazines dans les solvants protiques mais également aprotiques diminue, probablement du fait d'interactions hydrogène intermoléculaires et d'interactions  $\pi$ - $\pi$ <sup>47,48</sup>.

## 2. Diversité des sources naturelles productrices de 2,5-dicétopipérazines

Les dicétopipérazines sont largement répandues dans le monde vivant.

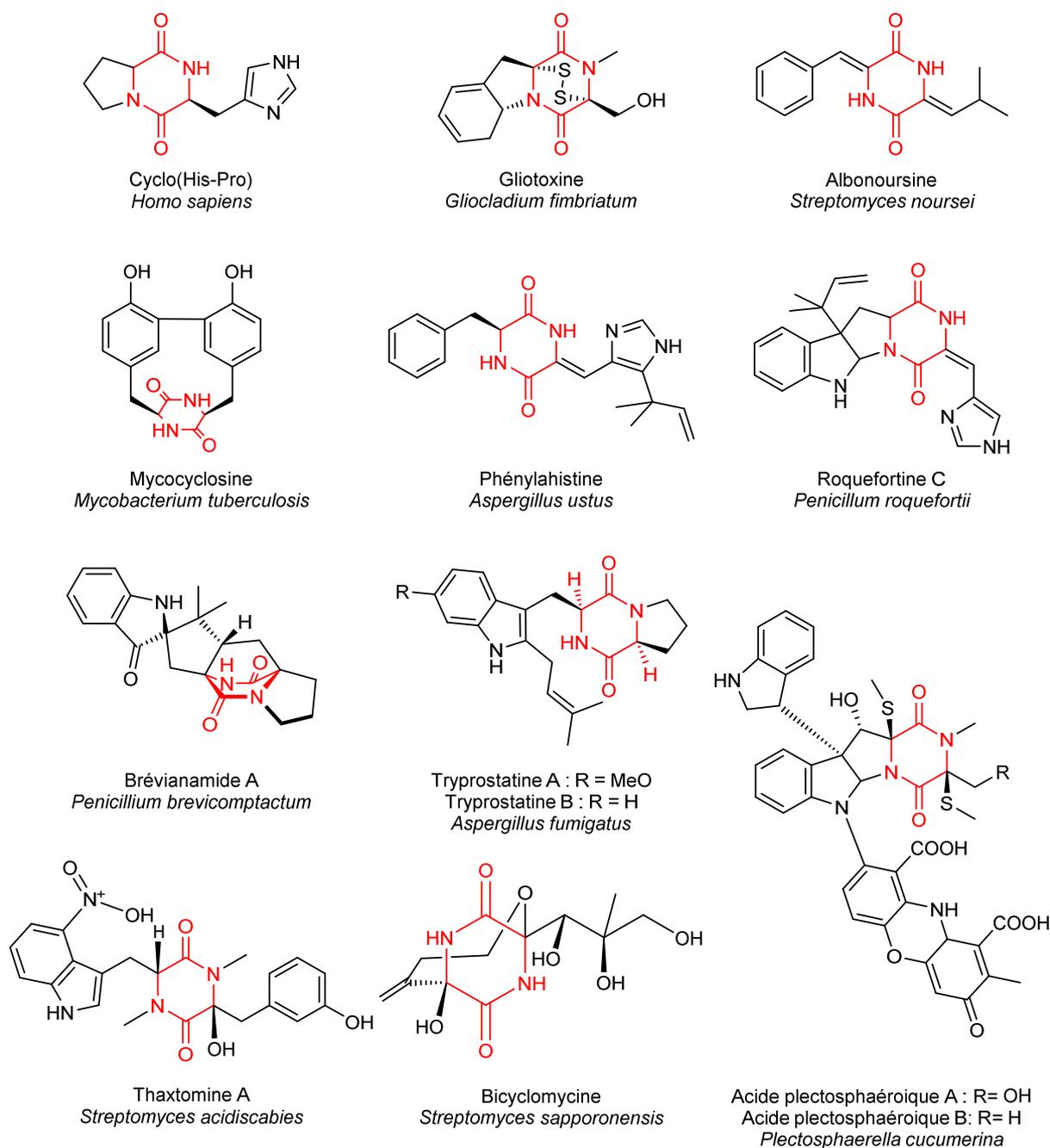


Figure 3 : Exemples de dicétopipérazines : organismes d'origine et structures

Le cycle dicétopipérazine est coloré en rouge.

Elles sont majoritairement produites comme métabolites secondaires par des microorganismes aussi bien d'origine terrestre que marine. Ainsi, elles sont retrouvées dans des bactéries, en particulier les actinobactéries (*Streptomyces noursei*<sup>49</sup>, *Streptomyces acidiscabies*<sup>50</sup>, *Streptomyces sapporonensis*<sup>34</sup>, *Mycobacterium*

*tuberculosis*<sup>29</sup>), mais aussi des champignons filamenteux (*Penicillium roquefortii*<sup>51</sup>, *Aspergillus ustus*<sup>52</sup>, *Aspergillus fumigatus*<sup>18</sup>, *Gliocladium fimbriatum*<sup>53</sup>, *Plectosphaerella cucumerina*<sup>54</sup>) (Figure 3).

Des animaux marins tels que des éponges, des étoiles de mer, des gorgones<sup>55</sup> ou des anémones<sup>27</sup> sont également recensés comme producteurs de dicétopipérazines.

Enfin, le cyclo(His-Pro) et le cyclo(Gly-Pro) sont les deux seules dicétopipérazines endogènes connues chez les mammifères<sup>56</sup>.

### 3. Diversité structurale des 2,5-dicétopipérazines d'origine biologique

Le cycle dicétopipérazine est particulièrement abondant dans la nature et il est retrouvé dans de nombreux produits naturels des plus simples aux plus complexes. Quelques exemples illustrant la diversité structurale des dicétopipérazines sont représentés (Figure 3).

Les dicétopipérazines les plus simples sont constituées par les cyclodipeptides, comme le cyclo(His-Pro)<sup>32</sup>. Viennent ensuite les dérivés plus ou moins complexes de cyclodipeptides, dans lesquels les chaînes latérales et/ou le noyau dicétopipérazine présentent des modifications. Les dérivés du cyclo(Trp-Xaa) en général et, plus précisément, ceux du cyclo(Trp-Pro) (tryprostatine A et B<sup>18</sup>, brevianamide A<sup>57</sup>) sont particulièrement abondants et parmi les plus diversifiés. Des dérivés du cyclo(Phe-Leu) (albonoursine<sup>49</sup>), du cyclo(Phe-His) (phénylahistine<sup>52</sup>), du cyclo(Tyr-Tyr) (mycocyclosine<sup>29</sup>), du cyclo(Ile-Leu) (bicyclomycine<sup>34</sup>) sont également retrouvés.

Des modifications de plusieurs natures peuvent être introduites sur les chaînes latérales : des isoprénylations (roquefortine C<sup>51</sup>, phénylahistine), des hydroxylations (bicyclomycine), des nitrations (thaxtomine A<sup>50</sup>), des additions de groupements méthoxy, des déshydrogénations (albonoursine, phénylahistine) ou des cyclisations de chaînes latérales (mycocyclosine). Les chaînes latérales substituées peuvent être retrouvées cyclisées sur l'un des atomes du cycle dicétopipérazine, comme pour la bicyclomycine (cyclisation probable de la chaîne latérale de l'isoleucine), la brevianamide A (cyclisation oxydante sur le carbone  $\alpha$  du noyau) ou la roquefortine C (cyclisation de la chaîne latérale du tryptophane sur l'azote N1 ou N4).

Le cycle dicétopipérazine peut lui aussi subir des substitutions variées, comme des *N*-méthylations (thaxtomine A) ou des *C*-hydroxylations (bicyclomycine). Il peut également présenter des thiométhylations (acides plectosphaéroïques) ou des thiolations (gliotoxine) qui mènent à la formation d'un pont disulfure sur le cycle dicétopipérazine.

Enfin, le cycle dicétopipérazine peut intervenir dans des architectures moléculaires importantes éloignées du simple cyclodipeptide, comme illustré avec l'acide plectosphaéroïque.

### 4. L'identification des 2,5-dicétopipérazines

Mon travail de thèse a requis l'identification d'un grand nombre de dicétopipérazines (cf. chapitres II et III). J'ai réalisé cette identification en utilisant deux techniques classiquement utilisées pour la

caractérisation structurale de produits naturels, à savoir des analyses de résonance magnétique nucléaire (RMN) classiques et de spectrométrie de masse d'ionisation par électronébulisation MS<sup>n</sup> (ESI-MS<sup>n</sup>). Cette dernière technique ayant été systématiquement utilisée en première intention dans mon étude, j'ai choisi de présenter dans ce paragraphe les données bibliographiques disponibles sur son utilisation pour l'identification des dicétopipérazines.

Il est à noter que les dicétopipérazines sont souvent retrouvées en faible quantité dans des milieux parfois complexes. La spectrométrie de masse en tandem couplée à la chromatographie liquide à haute performance (HPLC) s'est ainsi imposée comme la technique de référence pour la détection de ces molécules<sup>58-63</sup>. Lorsque des standards chimiques sont disponibles, la comparaison des temps de rétention et des profils de fragmentation permet l'identification des composés, sans nécessiter leur purification.

Dans le cas des cyclodipeptides, de nombreuses données de fragmentation tant sur les cyclodipeptides eux-mêmes que sur des peptides ou des acides aminés sont disponibles, permettant l'identification structurale de la plupart des cyclodipeptides, même en l'absence de standard chimique. En effet, les cyclodipeptides ont des spectres de masse caractéristiques, avec une signature de fragmentation spécifique<sup>64</sup>. Les mécanismes de fragmentation en ESI-MS<sup>n</sup> des cyclodipeptides ont été décrits par plusieurs études en mode positif<sup>59,62</sup> (Figure 4) et en mode négatif<sup>65</sup>.

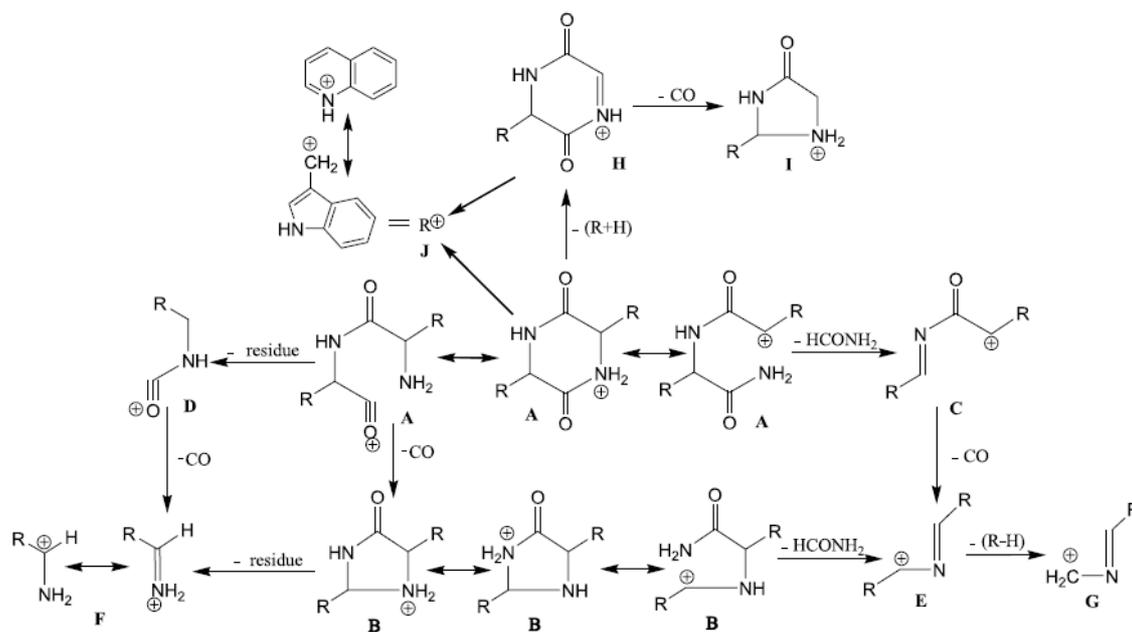


Figure 4 : Mécanismes de fragmentation des cyclodipeptides protonés, reproduit de <sup>62</sup>

En mode positif, la fragmentation d'une dicétopipérazine protonée fait apparaître une perte de  $m/z$  28 qui correspond au carbonyle (CO) et d'une perte de  $m/z$  45 qui correspond au HCONH<sub>2</sub><sup>59,62</sup>. Dans la majorité des cas, ces deux pertes ont lieu successivement dans un ordre ou l'autre (-28/-45 et -45/-28). La fragmentation d'un cyclodipeptide peut également faire apparaître des ions immonium et

d'autres ions dérivés, les uns et les autres étant caractéristiques des chaînes latérales des acides aminés le constituant<sup>58</sup>. L'ensemble des ions caractéristiques des chaînes latérales d'acides aminés protéinogéniques a été recensé et est présenté dans le (Tableau 1). Ainsi, la fragmentation du cycle dicétopipérazine et des chaînes latérales donnent des ions spécifiques qui permettent d'identifier les cyclodipeptides. A titre d'exemple, l'annotation du spectre de masse du cyclo(Pro-Val) est donnée en Figure 5. Néanmoins, la fragmentation classique d'un cyclodipeptide peut être modifiée lorsqu'il est constitué de certains acides aminés. Tel est le cas pour des acides aminés comme Met<sup>65</sup>, Glu ou Gln<sup>66</sup> dont les chaînes latérales possèdent des groupements chimiques sensibles à la fragmentation ou pour la proline qui est fusionnée avec le cycle dicétopipérazine<sup>59,67</sup>.

**Tableau 1 : Ions immonium et dérivés des 20 acides aminés protéinogéniques**

Acide aminé	Code à 3 lettres	Code à 1 lettre	Ion immonium <sup>1</sup>	Ion(s) dérivé(s) <sup>1</sup>
Acide aspartique	Asp	D	<i>88</i>	<i>70</i>
Acide glutamique	Glu	E	<i>102</i>	-
Alanine	Ala	A	44	-
Arginine	Arg	R	<i>129</i>	59, 70, 73, 87, 100, 112
Asparagine	Asn	N	<i>87</i>	<i>70</i>
Cystéine	Cys	C	<i>76</i>	-
Glutamine	Gln	Q	101	56, 84, <i>129</i>
Glycine	Gly	G	30	-
Histidine	His	H	<b>110</b>	<i>82, 121, 123, 138, 166</i>
Isoleucine	Iso	I	<b>86</b>	44, 72
Leucine	Leu	L	<b>86</b>	44, 72
Lysine	Lys	K	<i>101</i>	70, 84, 112, 129
Méthionine	Met	M	<i>104</i>	61
Phénylalanine	Phe	F	<b>120</b>	<i>91</i>
Proline	Pro	P	<b>70</b>	-
Sérine	Ser	S	60	-
Thréonine	Thr	T	74	-
Tryptophane	Trp	W	<b>159</b>	<i>77, 117, 130, 132, 170, 171</i>
Tyrosine	Tyr	Y	<b>136</b>	<i>91, 107</i>
Valine	Val	V	72	41, 55, 69

Source : Banque de données Mascot [http://www.matrixscience.com/help/fragmentation\\_help.html](http://www.matrixscience.com/help/fragmentation_help.html) à partir de <sup>68,69</sup>.

<sup>1</sup>Les signaux intenses sont en gras, tandis que les faibles sont en italique.

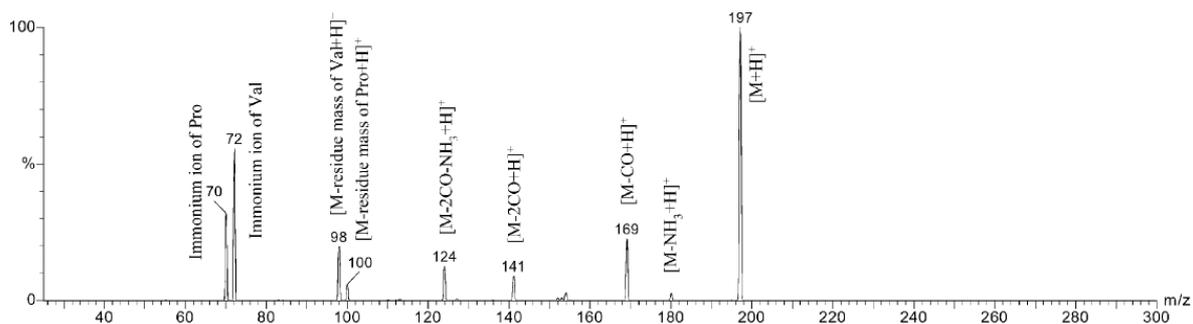


Figure 5 : Annotation du spectre de masse MS<sup>2</sup> du cyclo(Pro-Val) obtenu par ESI-MS en mode positif<sup>58</sup>

## II. Les dicétopipérazines, rôles physiologiques et activités biologiques

L'intérêt pour les dicétopipérazines est motivé, d'une part, par la recherche du rôle physiologique qu'elles peuvent jouer pour l'organisme sécréteur et, d'autre part, par l'utilisation biologique et/ou pharmacologique qui peut en être faite. Toutefois, il est parfois difficile de séparer activité biologique et rôle physiologique. En effet, une molécule possédant une activité antibiotique peut être considérée du point de vue de son potentiel thérapeutique ou du rôle de défense qu'elle remplit pour la bactérie sécrétrice.

Je présenterai dans un premier temps les données bibliographiques relatives aux rôles physiologiques des dicétopipérazines, en insistant particulièrement sur leur rôle en tant que molécules de communication, rôle le plus abondamment décrit dans la littérature. Dans un second temps, je décrirai les activités biologiques et pharmacologiques de ces molécules qui peuvent parfois également être reliées à un rôle physiologique.

### 1. Des rôles physiologiques peu connus

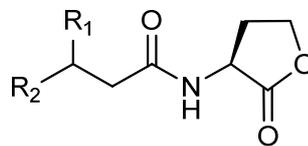
Le rôle physiologique des dicétopipérazines reste encore, dans la majorité des cas, à être élucidé. Cependant, ces molécules sont le plus souvent présentées comme des molécules de communication de cellule à cellule.

#### 1.1 Dicétopipérazines, des molécules de communication

La communication cellulaire, en particulier chez les bactéries, emprunte bien souvent un mécanisme connu sous le nom de détection du *quorum* (*quorum sensing* en anglais). Le *quorum sensing* est basé sur la sécrétion et la détection de petites molécules capables de diffuser librement dans le milieu et à travers les membranes des cellules. Le terme de *quorum sensing* est principalement associé à la capacité des bactéries à percevoir et répondre à leur densité de population. Ce système de communication leur permet

d'acquérir un comportement communautaire et est impliqué dans la régulation de l'expression des gènes. Si, initialement, le *quorum sensing* a été décrit comme la communication entre les bactéries d'une même communauté, il apparaît aujourd'hui que les molécules du *quorum sensing* peuvent permettre de communiquer entre différentes espèces de bactéries<sup>70</sup>, voire entre des bactéries et des eucaryotes<sup>71</sup>, notamment dans le cas de la communication pathogène-hôte<sup>72-75</sup>. Le *quorum sensing* joue un rôle dans de nombreux mécanismes tels que l'arrêt de la croissance bactérienne<sup>76,77</sup>, la survie en phase stationnaire<sup>78</sup>, la sécrétion de protéines<sup>79</sup>, la régulation du transfert de plasmide de conjugaison<sup>80,81</sup>, la synthèse d'antibiotiques<sup>82,83</sup> ou l'expression de certains facteurs de virulence<sup>84-95</sup>. Le *quorum sensing* peut également être responsable du passage de la symbiose à la virulence ou d'une phase planctonique à un biofilm<sup>96,97</sup>.

Le mécanisme de *quorum sensing* a été découvert chez une bactérie marine bioluminescente, *Vibrio fischeri*. Cette bactérie colonise l'organe lumineux de poissons marins ou de calmars. Elle est retrouvée au sein de cet organe à de très fortes concentrations (10<sup>10</sup> bactéries/ml) et produit de la lumière<sup>98,99</sup>. L'auto-inducteur régulant la production de l'enzyme luciférase, responsable de la production de lumière, a été identifié comme étant une acyl-homosérine lactone (acyl-HSL ou AHL) (Figure 6). Lorsque la concentration en bactérie augmente, la concentration en AHL augmente également jusqu'à atteindre un seuil qui déclenche la synthèse de luciférase et la bactérie devient bioluminescente. Les gènes responsables de ce phénomène sont *luxI* qui code une synthase qui produit l'AHL et *luxR* qui code le récepteur lequel interagit avec la molécule « signal ».



**Figure 6 : Structure moléculaire générique des acyl-homosérine lactones**

Ce type de régulation reposant sur le même type d'auto-inducteur n'est pas isolé et a été retrouvé chez de nombreuses autres bactéries à Gram négatif<sup>100,101</sup>.

Chez les bactéries à Gram positif, le mécanisme de *quorum sensing* utilise des oligopeptides comme auto-inducteurs. Ceux-ci interagissent avec un système à deux composants<sup>102</sup>, constitué d'une histidine kinase transmembranaire qui détecte le stimulus environnemental spécifique et d'un régulateur de réponse capable de déclencher l'expression de gènes cibles<sup>102</sup>.

Un autre type d'auto-inducteur, AI-2 (pour autoinducer-2), a été découvert chez *Vibrio harveyi* et a été proposé comme étant un auto-inducteur universel permettant la communication entre différentes espèces<sup>103</sup>. Si cette idée est assez répandue dans la littérature<sup>104,105</sup>, l'universalité de cet auto-inducteur est mise en doute par plusieurs auteurs. En effet, s'il existe, chez la plupart des bactéries, des gènes homologues à ceux permettant la synthèse de AI-2, les gènes homologues aux récepteurs de AI-2 ne

seraient présents que chez les espèces de *Vibrio*<sup>106</sup>. Dans une revue parue en 2007, Turovsky *et al.*<sup>107</sup> pointent également plusieurs biais, expérimentaux notamment, qui ont pu mener à des surinterprétations.

Un dernier type de système intervenant dans le *quorum sensing* a été proposé pour les interactions entre les bactéries et les eucaryotes. Il s'agit du système AI-3/épinéphrine/norépinéphrine, mis en évidence dans le cas des bactéries *Escherichia coli* responsable de diarrhées sanglantes<sup>72</sup> (EHEC pour Enterohemorrhagic *E. coli*), mais qui existerait chez d'autres organismes<sup>104</sup>.

Afin d'étudier le *quorum sensing*, certains micro-organismes ont été sélectionnés ou développés pour en faire des modèles rapporteurs ou des biosenseurs de ce phénomène. En plus de *Vibrio fischeri* et *Vibrio harveyi* avec lesquels le phénomène a été découvert, on peut citer *Chromobacterium violaceum*, pour lequel l'activation du *quorum sensing* se traduit par la synthèse d'un pigment violet, *E. coli* - pSB401 ou pHV2001- qui a été développée par ingénierie génétique pour contenir un système bioluminescent activable par des auto-inducteurs de type AHL et *Agrobacterium tumefaciens* NT1 pZLR4 qui possède un système rapporteur basé sur la synthèse de la  $\beta$ -galactosidase pouvant hydrolyser le substrat chromogène X-gal en composé devenant bleu en présence d'oxygène<sup>108</sup>.

### 1.1.1 Dicétopipérazines interagissant avec des biosenseurs du *quorum sensing*

Des dicétopipérazines provenant d'organismes variés ont montré des capacités à interférer avec le *quorum sensing*. Certaines de ces molécules ont été isolées à partir de bactéries à Gram négatif. En 1999, Holden *et al.* ont identifié trois dicétopipérazines différentes dans les surnageants de culture de plusieurs bactéries : le cyclo( $\Delta$ Ala-Val) chez *Pseudomonas aeruginosa*, *Proteus Mirabilis*, *Citrobacter freundii* et *Enterobacter agglomerans*, le cyclo(Tyr-Pro) chez *P. aeruginosa*, *P. mirabilis* et *C. freundii*, et le cyclo(Phe-Pro) chez *Pseudomonas fluorescens* et *Pseudomonas alcaligenes*. Ces trois dicétopipérazines sont capables d'activer des biosenseurs considérés comme spécifiques des AHL, bien que les concentrations nécessaires soient bien supérieures à celle de l'auto-inducteur naturel. Ces molécules agissent également comme antagoniste de l'auto-inducteur naturel, ce qui suggère qu'il existe une compétition entre les dicétopipérazines testées et l'auto-inducteur pour la liaison au récepteur de type LuxR<sup>109</sup>.

Quatre cyclodipeptides produits par *Pseudomonas putida*, (le cyclo(Pro-Tyr), le cyclo(Pro-Leu), le cyclo(Phe-Pro), et le cyclo(Val-Leu)) ont été isolés grâce à une purification guidée par l'activation de biosenseurs à AHL. Chacun de ces cyclodipeptides synthétisés chimiquement active au moins un des biosenseurs testés mais pas nécessairement celui qui a permis son identification, laissant supposer que la co-purification d'une AHL pourrait être à l'origine des activations observées lors de la purification<sup>110</sup>.

Une étude réalisée avec 332 bactéries marines à Gram positif a révélé la capacité du cyclo(Pro-Tyr) à inhiber la bioluminescence de *Vibrio harveyi*<sup>111</sup>, un pathogène marin intéressant à utiliser comme rapporteur du *quorum sensing* car ce dernier est particulièrement bien caractérisé et répond à trois types d'auto-inducteurs (AHL, CAI-1 et AI-2)<sup>112</sup>.

Enfin, un exemple de cyclodipeptide interférant dans le *quorum sensing* a été recensé. Il s'agit du cyclo(Pro-Val) produit par *Haloterrigena hispanica*, une archée extrêmement halophile<sup>113</sup>.

Il est à noter que des études remettent en cause le rôle de molécules « signal » du *quorum sensing* attribué aux dicétopipérazines. Des dicétopipérazines issues de synthèse chimique ne présentent pas les propriétés d'activation du *quorum sensing* qui avaient pourtant permis de les isoler à partir de surnageant de culture. Comme suggéré par les auteurs eux-mêmes<sup>110</sup> et par d'autres<sup>114</sup>, il semblerait que l'utilisation de méthodes séparatives non suffisamment résolutive pourraient mener à la collecte de fractions dans lesquelles une dicétopipérazine en quantité majoritaire et une AHL en quantité minoritaire seraient présentes. Etant donné que les AHL sont actives à des concentrations bien inférieures à celles nécessaires pour les dicétopipérazines<sup>109</sup>, les AHL co-purifiées pourraient être à l'origine de l'activation du système de *quorum sensing* qui est observée dans ces cas<sup>110,114</sup>. Par ailleurs, Campbell *et al.* ont synthétisé et testé vis-à-vis de plusieurs biosenseurs de *quorum sensing* une bibliothèque de dicétopipérazines de type cyclo(Pro-Xaa), Xaa étant un acide aminé protéinogénique ou non-naturel<sup>115</sup>. Contrairement aux résultats obtenus lors des études précédentes, aucun des cyclodipeptides contenant des acides aminés naturels n'a pu activer ou inhiber les rapporteurs du *quorum sensing* utilisés. Seuls deux cyclodipeptides non naturels (cyclo(Pro-L-4-Cl-Phe) et cyclo(Pro-L-4-I-Phe)) auxquels s'ajoutent six dicétopipérazines non naturelles d'une étude ultérieure<sup>116</sup>, se sont révélés capables de moduler la réponse des biosenseurs. D'après les auteurs, ces résultats contradictoires pourraient être liés à l'utilisation, dans ce travail, d'organismes rapporteurs du *quorum sensing* qui ne surexpriment par les récepteurs de type LuxR, contrairement aux autres études<sup>115</sup>.

Les concentrations élevées nécessaires à l'apparition d'interférences avec le *quorum sensing*, le questionnement sur la signification physiologique de celles-ci, l'absence de découverte de système utilisant uniquement les dicétopipérazines comme auto-inducteurs ont contribué à freiner les recherches dans ce domaine<sup>117,118</sup>. Néanmoins, d'autres études soulignent le rôle des dicétopipérazines comme molécules de communication aux implications physiologiques variées – avec ou non une interférence identifiée dans le mécanisme de *quorum sensing* -, comme l'illustrent les exemples décrits ci-après.

### 1.1.2 Dicétopipérazines et communication inter-espèce et inter-règne

Le cyclo(Phe-Pro) et le cyclo(Tyr-Pro) sont produits par *Lactobacillus reuteri*, une souche endogène du microbiote vaginal<sup>119</sup>. Les deux cyclodipeptides ont été identifiés comme des molécules « signal » capables d'interférer avec le système de détection du quorum *agr* de *S. aureus*, pathogène opportuniste de la flore vaginale. Le système *agr* est un régulateur clé des gènes de virulence et réprime l'expression de la toxine responsable du syndrome du choc toxique. Ces deux cyclodipeptides sont non seulement un exemple de molécules servant à la communication inter-espèce mais également à la régulation de la virulence d'un pathogène par une souche commensale.

Un exemple de communication inter-règne a également été montré par Ortiz-Castro *et al.*<sup>1</sup>. Ces derniers ont mis en évidence que, chez *P. aeruginosa*, la biosynthèse du cyclo(Pro-Val), du cyclo(Pro-Phe) et

du cyclo(Pro-Tyr) est régulée par un signal de type AHL et que ces cyclodipeptides peuvent agir comme des molécules « signal » et promouvoir la croissance d'*Arabidopsis thaliana*. Leur activité proviendrait de l'homologie structurale, notamment la structure planaire, qui existe entre ces molécules et l'auxine, une phytohormone jouant un rôle dans le développement et la croissance des végétaux.

### 1.1.3 Le cyclo(Phe-Pro), une molécule qui interfère avec la pathogénicité des *Vibrio*

Le cyclo(Phe-Pro) a été identifié dans les surnageants de culture de *Vibrio vulnificus* via l'activation du *quorum sensing* d'un organisme biosenseur (*Vibrio fischeri*). Il a également été retrouvé dans les surnageants d'autres espèces de *Vibrio*, à savoir *V. harveyi*, *V. cholerae* et *V. parahaemolyticus*. Il induit l'expression du gène *ompU* chez tous ces organismes. La protéine OmpU est impliquée dans de nombreuses activités liées à la pathogénicité (résistance aux peptides antimicrobiens, tolérance aux acides biliaires et organiques, formation de biofilm, attachement des cellules et probablement adhésion).

Chez *V. cholerae*, le cyclo(Phe-Pro) induit également l'expression du gène *ctxA* qui code une des sous-unités de la toxine cholérique, le facteur majeur de virulence de l'agent pathogène du choléra<sup>120</sup>. Les travaux de Bina *et al.* montrent que si le cyclo(Phe-Pro) affecte bien la virulence de *V. cholerae*, il induirait plutôt une diminution de l'expression des deux principaux facteurs de virulence, la toxine cholérique et la protéine d'adhésion TCP (toxin-correlated pilus)<sup>121</sup>. Le mécanisme de régulation utilisé par le cyclo(Phe-Pro) est associé à la protéine ToxR<sup>122</sup> qui régule l'expression de nombreux facteurs de virulence (toxine cholérique, protéine TCP, protéine OmpU). L'élucidation de la cascade de signalisation utilisée par le cyclo(Phe-Pro) a ouvert la voie à l'utilisation de cyclodipeptides dans le traitement du choléra. Ainsi, parmi les cyclodipeptides hétérologues testés, le cyclo(Val-Val) a démontré une activité de diminution de virulence plus importante que le cyclo(Phe-Pro)<sup>123</sup>, confirmant la possibilité d'utiliser les cyclodipeptides pour diminuer la gravité du choléra.

### 1.1.4 Dicétopipérazines induisant la production de substances antimicrobiennes

Cinq cyclodipeptides (le cyclo(Pro-Phe), le cyclo(Pro-Tyr), le cyclo(Ala-Val), le cyclo(Pro-Leu), et le cyclo(Pro-Val)) ont été identifiés dans les surnageants de culture de *Burkholderia cepacia* CF-66. Ils pourraient jouer un rôle dans la production d'une substance antifongique par la bactérie<sup>105</sup>.

*Pseudoalteromonas* sp. NJ6-3-1, une bactérie retrouvée chez une éponge de mer, synthétise trois cyclodipeptides (le cyclo( $\Delta$ Val-Val), le cyclo(Phe-Val) et le cyclo(Pro-Leu)). Les auteurs ont montré que l'un deux, le cyclo(Phe-Val) est capable d'induire la production de substances antibactériennes par NJ6-3-1<sup>124</sup>.

### 1.1.5 Dicétopipérazines exacerbant l'altération des denrées alimentaires

En recherchant les molécules « signal » impliquées dans l'altération de denrées alimentaires par *Shewanella baltica*, deux équipes ont identifié quatre cyclodipeptides (le cyclo-(Pro-Gly), le cyclo-(Pro-Leu), le cyclo-(Leu-Leu) et le cyclo-(Pro-Phe)). Le microorganisme saprophyte est responsable de la détérioration du poisson *Pseudosciaena crocea* (courbine jaune) et de la crevette à patte blanche (*Litopenaeus vannamei*). Ces quatre cyclodipeptides ont la propriété d'augmenter les capacités de détérioration de *S. baltica*<sup>125</sup> tandis que le cyclo(Pro-Leu) s'est révélé capable d'augmenter le niveau de transcription du gène du récepteur LuxR, de favoriser la formation de biofilm et la croissance bactérienne chez *S. baltica*<sup>126</sup>.

Ces différents exemples montrent que les dicétopipérazines peuvent jouer le rôle de molécules de communication mais d'autres rôles leur ont aussi été attribués comme présenté ci-après.

## 1.2 D'autres rôles décrits pour les dicétopipérazines

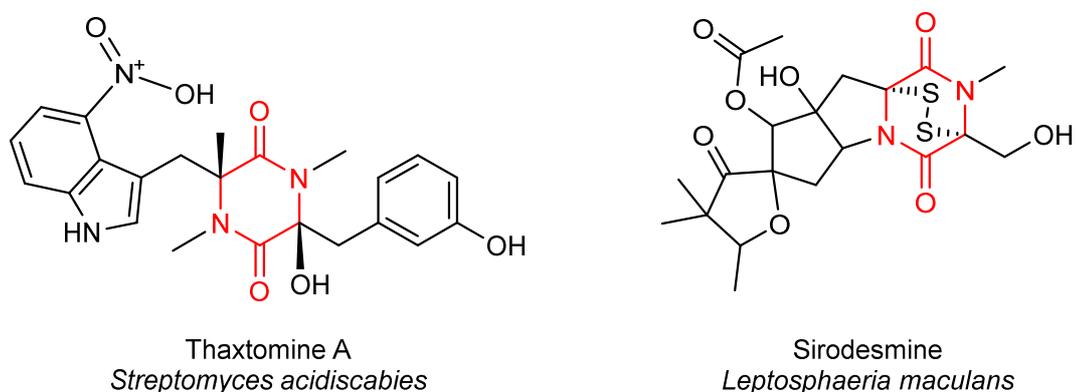
### 1.2.1 Virulence de micro-organismes pathogènes

Plusieurs dicétopipérazines sont décrites comme étant impliquées dans la virulence de micro-organismes, comme par exemple les molécules de la famille des thaxtomines<sup>127</sup> ou encore la sirodesmine PL<sup>128</sup>.

La gale commune des plantes est une pathologie qui touche notamment des légumes économiquement importants telles les pommes de terre. La gale commune se traduit par l'apparition de lésions au niveau des organes souterrains des plantes (racines et tubercules) et touche particulièrement les zones de croissance active. Cette maladie est causée par des bactéries du genre des *Streptomyces*. Plus d'une douzaine d'espèces de *Streptomyces* associées à la gale commune ont été identifiées à ce jour, les trois les plus étudiées étant *S. scabies*, *S. acidiscabies* et *S. turgidiscabies*<sup>127</sup>. Toutes ces espèces produisent de la thaxtomine A (Figure 7) ou une autre molécule de la famille des thaxtomines. De nombreux résultats ont montré que ces dicétopipérazines interviennent directement dans la virulence de ces espèces. Tout d'abord, la thaxtomine est présente dans les tissus des plantes infectées et, une fois extraite et ajoutée sur des tissus sains, elle induit les symptômes de la gale commune<sup>129</sup>. Enfin, Healy *et al.* ont observé que des souches mutantes pour la voie de biosynthèse de la thaxtomine A ne sont plus virulentes<sup>130</sup>. La thaxtomine A affecterait les plantes en inhibant la biosynthèse de cellulose<sup>131</sup>.

Quant à la sirodesmine PL (Figure 7), elle est le principal métabolite secondaire sécrété par l'espèce *Leptosphaeria maculans* et elle serait impliquée dans la maladie de "la jambe noire" chez l'espèce végétale *Brassica napus* (utilisée dans l'alimentation animale, la production d'huile végétale et de biocarburants)<sup>132</sup>. Cette phytotoxine est une épipolythiodioxopipérazine (ETP) caractérisée par la présence d'un pont disulfure très réactif (cf. IV.2.1 Voies de biosynthèse de la gliotoxine et des épithiodioxopipérazines : les gènes de la NRPS et des enzymes de modifications sont

regroupées en un même cluster). Sa production chez *L. maculans* est régulée par la disponibilité des acides aminés qui la constituent<sup>133</sup>.



**Figure 7 : Dicétopipérazines impliquées dans la virulence de micro-organismes pathogènes : structures et organismes d'origine**

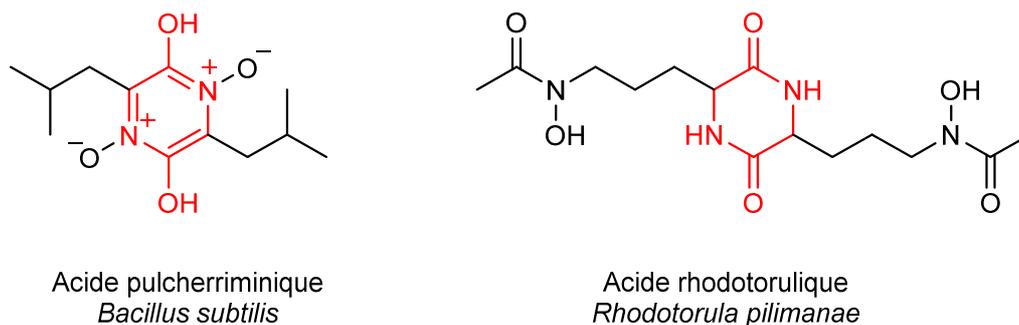
Après ces deux exemples de dicétopipérazines cytotoxiques qui sont impliquées dans des phénomènes de virulence, des molécules ayant acquis des propriétés leur permettant de fixer le fer vont être présentées.

### 1.2.2 Molécules liant le fer

Le fer est un élément essentiel pour la majorité des organismes. Il est présent dans de nombreux cofacteurs, en particulier le cofacteur hémique, et dans des complexes Fer-Soufre. Sa présence est indispensable à plusieurs processus métaboliques, dont la synthèse de désoxyribonucléotides, la phosphorylation oxydative et le transport d'électrons. En dépit de sa relative abondance dans la nature, la concentration en fer disponible pour les cellules est très limitée (de l'ordre de  $10^{-18}$  M).

Lors d'une carence en fer, de nombreux champignons microscopiques synthétisent et sécrètent des sidérophores qui sont des molécules de faible masse moléculaire, de 500 à 1000 Da, ayant une grande affinité pour le fer. Le complexe fer-sidérophore est ensuite capté par le microorganisme grâce à des protéines de transport. Parmi les sidérophores, on trouve une dicétopipérazine, l'acide rhodotorulique (Figure 8), constituée du dimère de la  $\delta$ -N-acétyl-L- $\delta$ -N-hydroxyornithine. L'acide rhodotorulique a été isolé à partir de cultures déficientes en fer de *Rhodotorula pilimanae* pour sa capacité à lier fortement le fer<sup>134</sup>.

L'acide pulcherrimique, produit notamment chez *Bacillus subtilis*, est une molécule de cyclo(Leu-Leu) avec deux N-hydroxylations sur le noyau dicétopipérazine qui confèrent à cette molécule une capacité à fixer fortement le fer, formant ainsi la pulcherrimine (Figure 8, cf. V.2.2 Voies de biosynthèse de l'acide pulcherriminique)<sup>135</sup>. Les complexes entre les molécules d'acide pulcherrimique et les atomes de fer forment des pigments rouges chez plusieurs espèces de *Bacilles* et chez une levure, *Metschnikowia pulcherrima*, dont le rôle physiologique est encore inconnu<sup>136</sup>.



**Figure 8 : Dicétopipérazines liant le fer : structures et organismes d'origine**

Ces exemples de rôles physiologiques de dicétopipérazines concernent les micro-organismes. Néanmoins, deux dicétopipérazines sont décrites comme ayant un rôle physiologique chez les mammifères. Ainsi, le cyclo(His-Pro), isolé chez des mammifères dont l'homme, atténue les effets sédatifs de l'éthanol et joue un rôle dans la perception de la douleur, dans la modulation de l'appétit et dans la thermorégulation, ceci suivant un mécanisme dopaminergique<sup>56</sup>. Le cyclo(Pro-Gly), identifié dans le cerveau du rat, serait une substance facilitant la mémorisation et ayant une activité anxiolytique<sup>137</sup>. Ces deux molécules et leurs activités seront présentées avec plus de détails dans le paragraphe suivant concernant les activités biologiques des dicétopipérazines car leurs dérivés peuvent être utilisés comme agents neuroprotecteurs (cf. II.2.3 Activité neuroprotectrice)

## 2. Des activités biologiques variées

Les activités biologiques des dicétopipérazines ont été récemment répertoriées par A.D Borthwick dans une revue très complète sur le sujet<sup>46</sup>. Les dicétopipérazines présentent principalement des activités antibactérienne, antifongique, antivirale, antitumorale, neuroprotectrice, anxiolytique, anti-inflammatoire, bioherbicide<sup>46</sup>, mais également, bien que moins documentées, des activités immunosuppressive<sup>138,139</sup>, insecticide<sup>140-143</sup>, antiparasitaire<sup>144-146</sup>, cardiomodulatrice<sup>147,148</sup>, vasodilatatrice<sup>149</sup> ou des effets toxiques<sup>150</sup> (neurotoxique<sup>33,151-153</sup> ou génotoxique<sup>154</sup>).

L'objectif de cette présentation bibliographique ne sera donc pas de présenter les activités biologiques des dicétopipérazines de manière exhaustive mais d'illustrer l'intérêt de ces molécules en recherche thérapeutique en donnant des exemples de molécules aux activités antibactérienne, antitumorale et neuroprotectrice.

### 2.1 Activités antibactériennes

Nombreuses sont les dicétopipérazines d'origine naturelle ou chimique, simples cyclodipeptides ou architectures plus complexes, qui présentent des activités antibactériennes (Tableau 2). Ces activités peuvent être faibles<sup>155</sup> voire très faibles<sup>156</sup>, modérées<sup>34</sup> ou particulièrement remarquables<sup>157</sup>.

**Tableau 2 : Exemples de dicétopipérazines présentant une activité antibactérienne ou antibiotique**

Dicétopipérazine	Microorganisme d'origine <sup>1</sup>	Bactérie(s) sensible(s)	Références
Albonoursine ou cyclo( $\Delta$ Phe- $\Delta$ Leu)	<i>Streptomyces noursei</i>	<i>Bacillus spp.</i>	35
Alternarosine A	<i>Alternaria raphani</i>	<i>Escherichia coli</i> , <i>Bacillus subtilis</i>	156
Avrainvillamide	<i>Aspergillus ochraceus</i>	Bactéries multi-résistantes	158
Bis-N-norgliovictine (dérivé 3,6-dimethylthio)	<i>Verticillium hemipterigenum</i>	<i>Mycobacterium tuberculosis</i>	159
Cairomycline B	<i>Streptomyces sp. (AS-C-19)</i>	Bactérie Gram positif	160
Chaetocine	<i>Chaetomium minutum</i>	<i>Staphylococcus aureus</i> , <i>Sarcina lutea</i> , <i>Streptococcus faecalis</i> , <i>Proteus mirabilis</i> , <i>Shigella flexneri</i> , <i>E. coli</i>	161
Cyclo(Phe-Pro)	<i>Pseudomonas rhizosphaerae</i>	Marnie fouling bacteria	162
Cyclo(Phe-Pro)	-	Bactéries Gram positif et Gram négatif	155
	<i>Actinomyces AMLK-335</i>	Entérocoques résistants à la vancomycine (VRE)	163
Cyclo(D-Pro-D-Phe)	<i>Marine bacteria CF-20 c-148</i>	<i>Vibrio anguillarum</i>	164
Cyclo-trans-4-OH-(D-Pro-D-Phe)	<i>Marine bacteria CF-20 c-148</i>	<i>V. anguillarum</i>	164
Brevianamide F ou cyclo(Trp-Pro)	-	Bactéries Gram positif et Gram négatif	155
	<i>Streptomyces sp. TN58</i>	<i>S. aureus</i> , <i>Micrococcus luteus</i>	165
Bicyclomycline	<i>Streptomyces sapporonensis</i> , <i>Streptomyces aizunensis</i>	Bactéries Gram négatif et une bactérie Gram positif, <i>Micrococcus luteus</i> .	34,166
Cyclo(Tyr-Pro)	<i>Pseudomonas rhizosphaerae</i>	Bactéries marines responsables de bioencrassement (biofouling)	162
Cyclo(Tyr-Pro)	-	Bactéries Gram négatif	155
Cyclo(Leu-Pro)	<i>Streptomyces sp. KH-614</i>	Entérocoques résistants à la vancomycine (VRE)	167
cyclo(D-Pro-D-Leu)	<i>Marine bacteria CF-20 c-148</i>	<i>V. anguillarum</i>	164
Cyclo(D-Pro-D-Ile)	<i>Marine bacteria CF-20 c-148</i>	<i>V. anguillarum</i>	164
Cyclo(Val-Pro)	<i>Pseudomonas rhizosphaerae</i>	Marnie fouling bacteria	162
Cyclo(D-Pro-D-Val)	<i>Marine bacteria CF-20 c-148</i>	<i>V. anguillarum</i>	164
Cyclo(Tyr-Ser)	-	Bactéries Gram positif et Gram négatif	168
Cyclo(Tyr-Ile),	<i>Pseudomonas rhizosphaerae</i>	Marnie fouling bacteria	162
Cyclo(Trp-Trp)	-	Bactéries Gram positif et Gram négatif	155
Cyclo(His-Tyr)	-	Bactéries Gram positif	169
Cyclo(Pro-Pro)	<i>Lucilia sericata</i>	<i>M. luteus</i> et <i>Pseudomonas aeruginosa</i>	170
Dimère cycloserine	-	<i>M. tuberculosis</i>	171
Dehydrogliotoxine	<i>Gliocladium flavofuscum</i> <i>Gliocladium virens</i>	<i>M. tuberculosis</i>	172
Dehydroxybisdethiobis(methylthio)gliotoxine	<i>Pseudallescheria sp.</i>	<i>S. aureus</i> multirésistant et <i>S. aureus</i> résitant à la méthicilline	173
Epicoccine A	<i>Epicoccum nigrum</i>	<i>Bacillus subtilis</i>	174
Gliotoxine	<i>Gliocladium flavofuscum</i> <i>Gliocladium virens</i>	<i>M. tuberculosis</i>	157
Macrolide MPC1001	<i>Cladorrhinum sp. KY4922</i>	Bactéries Gram positif ( <i>S. aureus</i> )	175
Rodriguesine A et B	<i>Didemnum sp.</i>	<i>Streptococcus mutans</i> , <i>S. aureus</i> , <i>S. P. aeruginosa</i>	176
Vertihemiptellides A et B	<i>Verticillium hemipterigenum</i>	<i>M. tuberculosis</i>	159
Roquefortine C	<i>Penicillium sp.</i>	Bactéries Gram positif	177

<sup>1</sup>La présence d'un tiret « - » indique que la dicétopipérazine est d'origine chimique.

La bicyclomycline et ses dérivés sont un exemple notable de l'activité antibactérienne des dicétopipérazines. La bicyclomycline (Figure 9) est une dicétopipérazine pontée, isolée de *Streptomyces sapporonensis*<sup>34,178</sup> et *Streptomyces aizunensis*<sup>166,179</sup>, qui montre des activités d'inhibition vis-à-vis d'un large

spectre de bactéries à Gram négatif et d'une bactérie à Gram positif, *Micrococcus luteus*. Elle présente une faible toxicité chez l'humain et est utilisée pour traiter des diarrhées chez l'homme et chez certains animaux (veau et porc).

Sa structure ainsi que son mécanisme d'action sont uniques parmi les antibiotiques<sup>19</sup>. La bicyclomycine cible le facteur de terminaison de la transcription, la protéine Rho<sup>180</sup>. Elle agit comme un inhibiteur réversible et non compétitif de l'adénosine triphosphate (ATP)<sup>181</sup> et mène à la synthèse de transcrits anormalement longs<sup>182</sup>. Des structures cristallographiques de complexe bicyclomycine:Rho ont été obtenues et ont permis de définir le site de liaison de la dicétopipérazine sur la protéine et de donner les bases de son mécanisme d'action<sup>19,183</sup>. Afin d'améliorer l'efficacité et le spectre d'action de la bicyclomycine, plusieurs analogues synthétiques ou semi-synthétiques ont été testés. Cette étude relation structure-activité a révélé que le triol en C1 et le bicyclic [4.2.2] sont essentiels à l'activité<sup>184</sup>, contrairement à l'exométhylène en C5 et C5a<sup>185</sup>. La substitution de ce méthylène par des groupements 3-formylphenylsulfanyl<sup>186,187</sup>, formyl<sup>188</sup> ou 3-fluoro-benzylsulfanyl<sup>189</sup> a permis d'améliorer l'activité du composé.

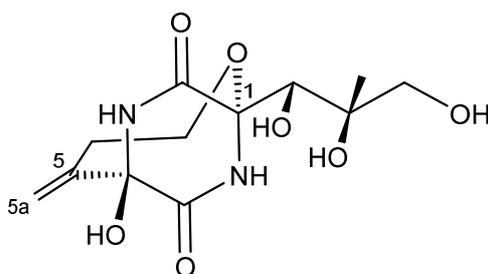


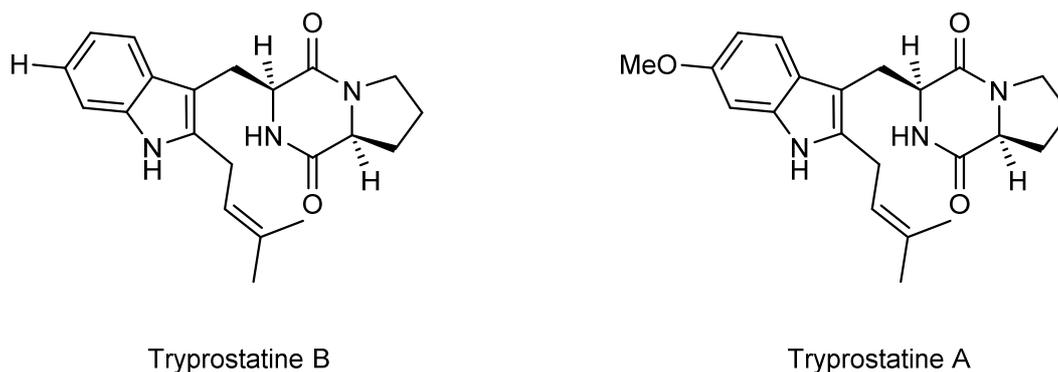
Figure 9 : Structure de la bicyclomycine

Les dicétopipérazines naturelles antibactériennes sont donc nombreuses et constituent une réserve de molécules d'intérêt thérapeutique. Il est également probable que ces molécules aient un rôle physiologique important chez l'hôte en tant que molécules de défense.

## 2.2 Activités antitumorales/anticancéreuses

Les exemples développés dans ce paragraphe traitent de dicétopipérazines naturelles qui ont, pour certaines, été soumises à des modifications chimiques pour optimiser leur activité. Le dernier exemple concerne des molécules obtenues directement par voie chimique.

Des dérivés de cyclo(Trp-Pro) sont répertoriés comme des molécules aux activités anticancéreuses notables. C'est le cas par exemple des tryprostatines A et B isolées de *Aspergillus fumigatus*<sup>18</sup>.



**Figure 10 : Structures de la tryprostatine B et tryprostatine A**

La tryprostatine B (Figure 10) a été décrite comme étant un inhibiteur du cycle cellulaire chez les mammifères<sup>18</sup>, même si d'autres études ont montré que son action ne semblait pas spécifique du cycle cellulaire<sup>190</sup>. Il a été montré que la tryprostatine B est un inhibiteur faible de la topoisomérase II<sup>191</sup>. Certaines disparités ont été observées dans l'évaluation de l'activité anticancéreuse de la tryprostatine B. Elle pourrait s'expliquer par la cytotoxicité de certains de ses produits de dégradation<sup>191,192</sup>. L'étude de l'activité et du mécanisme d'action des différents stéréoisomères de la tryprostatine B a montré que seul le diastéréoisomère DL possède une activité cytotoxique remarquable sur les trois lignées cellulaires de cancer testées (sein, poumons, prostate)<sup>191,193</sup>. De plus, ces études montrent que le L-Trp est indispensable pour l'activité cytotoxique, ce qui permet de commencer à définir les éléments nécessaires à l'activité anticancéreuse des tryprostatines. Le mécanisme d'action de ce composé ne repose pas sur une activité d'inhibition de la topoisomérase II ou de la polymérisation de la tubuline, mais sur un autre mécanisme encore inconnu<sup>193</sup>. Par ailleurs, des essais préliminaires ont montré qu'il serait cytotoxique sur des cellules humaines normales<sup>191</sup>.

La tryprostatine A (Figure 10), quant à elle, est un inhibiteur de la topoisomérase II et de la polymérisation de la tubuline grâce à son interaction avec MAP2 (microtubule-associated-protein)<sup>190</sup>. La tryprostatine A est également décrite comme un inhibiteur de la protéine BCRP (breast cancer resistance protein). La protéine BCRP est un membre de la super famille des transporteurs de type ABC qui donnent aux cellules tumorales un phénotype de résistance pléiotropique atypique<sup>194</sup>. Ainsi, en comparaison avec la tryprostatine B qui ne diffère que par l'absence du substituant méthoxy sur le noyau indole, la tryprostatine A est moins cytotoxique et présente un mécanisme d'action identifié et original qui inhibe à la fois la polymérisation de la tubuline et la protéine BCRP. Ces éléments ont encouragé des études de relations structure-activité pour comprendre et améliorer l'activité anticancéreuse de ce composé naturel<sup>191</sup>. L'énantiomère, les 2 diastéréoisomères et 20 autres analogues de la tryprostatine A modifiés en quatre positions (Figure 11) ont été synthétisés et testés.

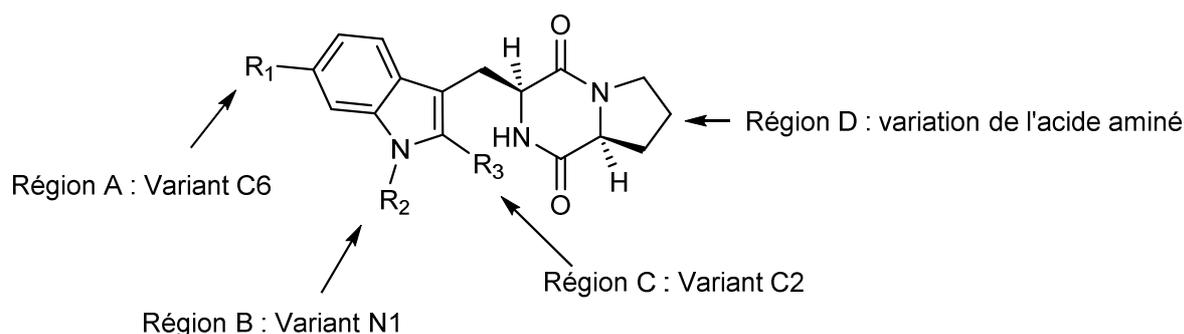


Figure 11 : Positions variables de la tryprostatine A, d'après<sup>191</sup>

Il ressort de cette étude que la configuration LL de la tryprostatine A et le groupement méthoxy sont essentiels pour l'inhibition de la topoisomérase II et de la polymérisation de la tubuline. La présence du 2-isoprenyl en R3 (région C) est essentielle à l'inhibition de la prolifération cellulaire, sauf si son absence est compensée par la présence d'un groupement benzylique ou allylique en R2 (région B). Le groupement méthoxy en R1 peut être substitué par un groupement thiocyanate (NCS) ou azide (N3). La L-proline peut être remplacée par d'autres L-acides aminés (région D). La nature des substituants du noyau indole en R2 et R3 influence le mécanisme d'action de l'analogue et met en lumière la versatilité du squelette tryprostatine comme modèle pour la recherche d'agents chimiothérapeutiques<sup>191</sup>.

Le dérivé cyclisé de la tryprostatine A, la fumitremorgine C (Figure 12), isolée également d'*A. fumigatus*, présente aussi une activité d'inhibition de la protéine BCRP. Etant donné que la fumitremorgine C est neurotoxique et provoque des convulsions chez la souris et d'autres animaux, plus d'une quarantaine d'analogues de ce composé ont été synthétisés et testés pour identifier de nouveaux inhibiteurs de BCRP non toxiques<sup>195-197</sup>. Ce travail a mené à la synthèse d'un analogue, le Kol143 (Figure 12), inhibiteur sélectif et puissant de la BCRP, utilisable *in vivo*.

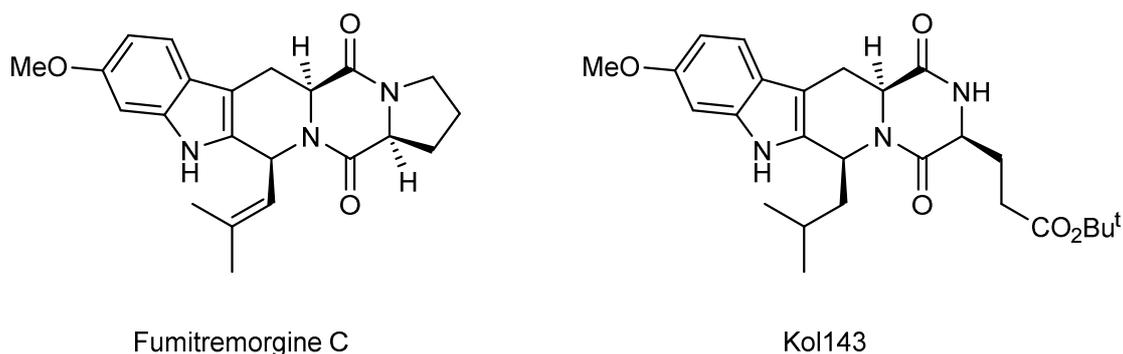
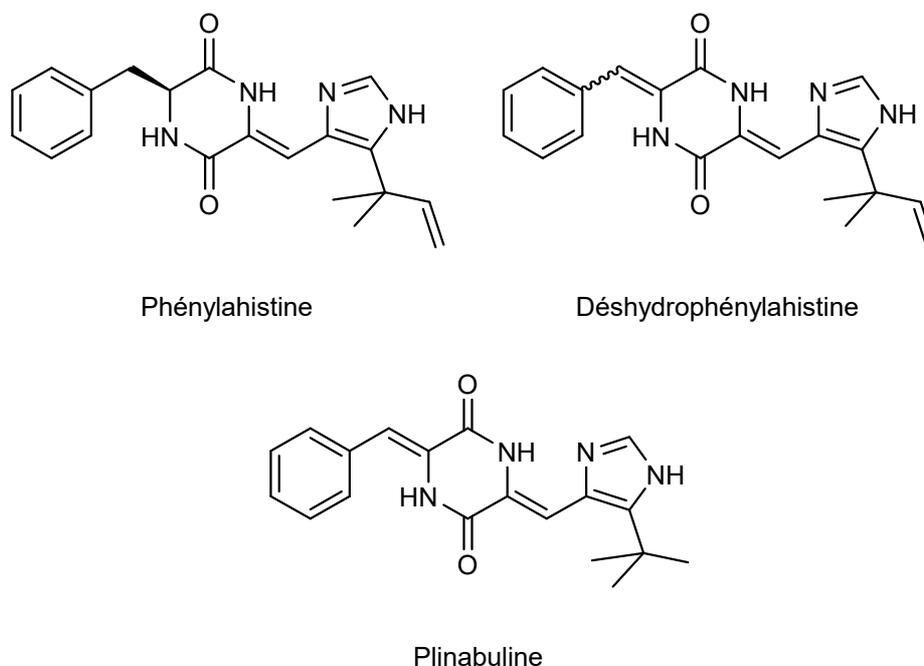


Figure 12 : Structures de la fumitremorgine C et de son analogue optimisé Kol143

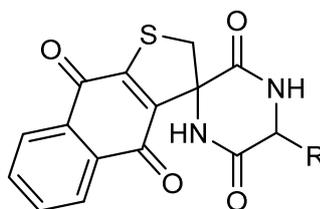
La (-)-phénylhistine produite par *Aspergillus ustus* est également une molécule aux propriétés anticancéreuses. Elle agit en inhibant la polymérisation des microtubules par la fixation sur la tubuline. A nouveau, une série d'analogues, notamment des cyclodipeptides mono ou di-déshydrogénés, a été étudiée

pour optimiser l'activité biologique de ce composé. L'analogue doublement insaturé de la phénylahistine, la déshydrophénylahistine, s'est révélé particulièrement prometteur en démontrant des activités supérieures au taxol, la vinblastine ou la vincristine. Une amélioration supplémentaire a été apportée en substituant l'isoprényle par un tertio-butyle pour donner la plinabuline qui est actuellement en phase II de développement clinique (Figure 13).



**Figure 13 : Structures de la phénylahistine, de la déshydrophénylahistine et de la plinabuline**

Des spirodicétopipérazines (Figure 14), des molécules synthétiques, ont montré, sur plusieurs lignées de cellules tumorales, des activités cytotoxiques équivalentes voire supérieures à la doxorubicine, un agent chimiothérapeutique de type quinone. Elles ont également montré des activités contre des lignées tumorales ayant développé une résistance pléiotropique. Elles agiraient par liaison à l'ADN en intercalant le motif tricyclique planaire dans l'hélice de l'ADN tandis que le noyau dicétopipérazine substitué interagirait avec la partie externe du double brin<sup>198,199</sup>.



**Figure 14 : Structure générale des spirodicétopipérazines**

Ces quelques exemples montrent notamment comment, à partir d'un composé naturel identifié pour son action anticancéreuse, il est possible d'optimiser la structure moléculaire afin d'obtenir un composé autant voire plus actif et utilisable dans un traitement chimiothérapeutique. En effet, certains analogues de dicétopipérazines naturelles permettent de conserver l'activité de cette dernière tout en éliminant ses effets délétères (cytotoxicité pour les cellules normales, neurotoxicité). Par ailleurs, la variété des mécanismes d'action décrits pour des molécules comportant le noyau dicétopipérazine ou pour des molécules ayant le même squelette tryprostatine est à noter et illustre la versatilité de ces motifs.

### 2.3 Activité neuroprotectrice

Le cyclo(His-Pro) est un métabolite de l'hormone thyroïdienne (TRH pour thyrotropin-releasing hormone) chez les mammifères. L'hormone thyroïdienne, produite par l'hypothalamus, est de nature tripeptidique ((pyro)-Glu-His-Pro-NH<sub>2</sub>) et agit dans le contrôle du système neuroendocrinien. Le principal mécanisme responsable du catabolisme de cette hormone au sein du système nerveux central est l'hydrolyse de l'acide pyroglutamique terminal par la pyroglutamyl aminopeptidase. Le dipeptide linéaire obtenu se cyclise alors pour donner le cyclo(His-Pro)<sup>200</sup>. Ce dernier est donc synthétisé de manière endogène *in vivo*. Il est ubiquitaire dans le système nerveux central et a également été retrouvé dans l'appareil digestif et la prostate ainsi que dans plusieurs fluides corporels<sup>30</sup>.

Il a été montré que le cyclo(His-Pro) agit comme anti-inflammatoire au niveau des cellules gliales en contrant le cycle d'inflammation, délétère pour les neurones, induit par un agent pro-inflammatoire (les lipopolysaccharides)<sup>201</sup>. Son mécanisme d'action serait double : 1) inactivation de la voie à NF-κB qui mène à la mort neuronale en augmentant la réponse pro-inflammatoire par la production de cytokines pro-inflammatoires et de dérivés réactifs de l'oxygène et 2) activation de Nrf2 qui augmente la protection antioxydante par la synthèse de glutathion notamment<sup>202</sup>. Le cyclo(His-Pro) internalisé par les neurones a également une activité neuroprotectrice<sup>203</sup>. Il a en effet démontré une activité neuroprotectrice sur les cellules PC12 (lignées utilisées comme modèle de cellules neuronales) en diminuant le stress oxydatif induit par les agents pro-inflammatoires<sup>23,202,204</sup>.

Puisque le cyclo(His-pro) diminue l'inflammation des cellules gliales et induit une réponse neuroprotectrice chez les neurones, il apparaît comme un potentiel agent thérapeutique. De plus le cyclo(His-Pro) exogène est très bien absorbé oralement, est stable vis-à-vis de l'activité des peptidases et traverse la barrière hémato-encéphalique<sup>200,205</sup>. Cependant, le cyclo(His-Pro) a également, en plus de son action sur le système nerveux central, des effets électrophysiologiques, endocriniens et cardiovasculaires<sup>23</sup>. Des analogues du cyclo(His-Pro) ont été évalués afin d'obtenir un composé qui posséderait uniquement l'activité neuroprotectrice désirée. Une première génération de dérivés du cyclo(His-Pro) a été développée en remplaçant l'imidazole du cyclo(His-Pro) par un groupement di-tert-butylphénol car le phénol agirait comme capteur de dérivés réactifs de l'oxygène<sup>206</sup>. Une seconde génération de dérivés du cyclo(His-Pro), dans laquelle la chaîne histidine est remplacée par un cycle à 3, 4 ou 5 a été développée et a également

montré de bonnes activités neuroprotectrices *in vitro* et *in vivo*<sup>207–209</sup>. En s'appuyant sur ces études, d'autres dicétopipérazines ont été identifiées comme agents neuroprotecteurs<sup>210</sup>.

Le seul autre cyclodipeptide endogène connu chez les mammifères est le cyclo(Pro-Gly), identifié dans le cerveau du rat qui serait une substance facilitant la mémorisation et ayant une activité anxiolytique<sup>137</sup>.

Ces deux cyclodipeptides naturels ont ouvert la voie à l'utilisation des dicétopipérazines en tant qu'agents neuroprotecteurs. Plusieurs dicétopipérazines analogues de la TRH ont ainsi été développées et la proline a été identifiée comme un élément clé de l'activité neuroprotectrice. Il est à noter qu'une autre classe de dicétopipérazines, contenant des insaturations dans les chaînes latérales, a montré des activités d'augmentation des fonctions sensorielles et motrices et de piégeage de radicaux libres. Ces deux classes de dicétopipérazines (analogues de la TRH et insaturés) pourraient trouver des applications dans le traitement de maladies neurodégénératives comme celles d'Alzheimer ou de Parkinson<sup>211</sup>.

Alors qu'ils étaient, à l'origine, considérés comme de simples produits de dégradation, les dicétopipérazines révèlent peu à peu leur extraordinaire diversité de structures et d'activités biologiques. Même si leur rôle physiologique n'est pas encore élucidé dans la grande majorité des cas, les dicétopipérazines naturelles représentent un réservoir formidable de molécules à tester et à optimiser pour des applications thérapeutiques. Cette richesse naturelle a stimulé les recherches sur la synthèse totale de dicétopipérazines naturelles et de leurs analogues. Dans la partie suivante, je vais présenter les voies de synthèse chimique les plus courantes pour l'obtention de dérivés dicétopipérazines. Je m'efforcerai au cours de cette présentation de mettre en avant les atouts et les faiblesses de ces stratégies. Les méthodes existant pour la formation du cycle dicétopipérazine puis celles permettant la fonctionnalisation de cycle vont être détaillées.

### III. Synthèse chimique des dicétopipérazines

Le cycle dicétopipérazine étant une architecture de choix pour la chimie médicinale, une panoplie de stratégies pour la synthèse chimique de composés comportant ce cycle a été développée.

#### 1. Formation du cycle dicétopipérazine

Plusieurs méthodes de synthèse chimique en solution ou en phase solide permettent la formation du noyau dicétopipérazine. Les trois principales sont la formation d'une liaison amide (A), la *N*-alkylation (B) et la *C*-acylation (C). A ceci s'ajoutent les cyclisations tandem *via* N1-C2/C3-N4 (D) et *via* C2-N1-C6 (E)<sup>46</sup> (Figure 15).

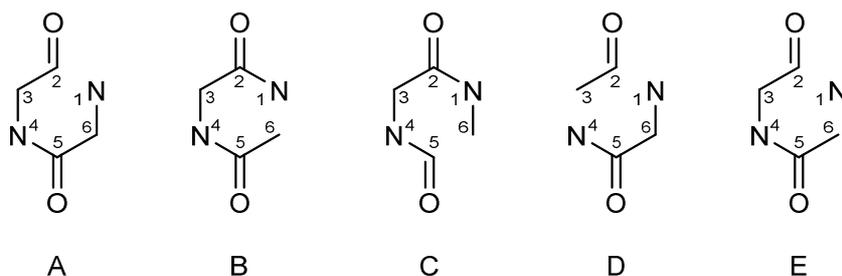


Figure 15 : Voies de synthèse pour formation du cycle dicétopipérazine (d'après<sup>46</sup>)

### 1.1 Via la formation de liaison amide

Il existe principalement deux voies de synthèse utilisant la formation d'une liaison amide. La première utilise un dipeptide de type 1 présentant à une extrémité une fonction amine et à l'autre extrémité une fonction ester qui réagissent ensemble pour former le cycle. La deuxième utilise la réaction de Ugi entre un isonitrile, un acide (le cas échéant un acide aminé), un aldéhyde et une amine pour former le dipeptide 2 qui se cyclise (Figure 16).

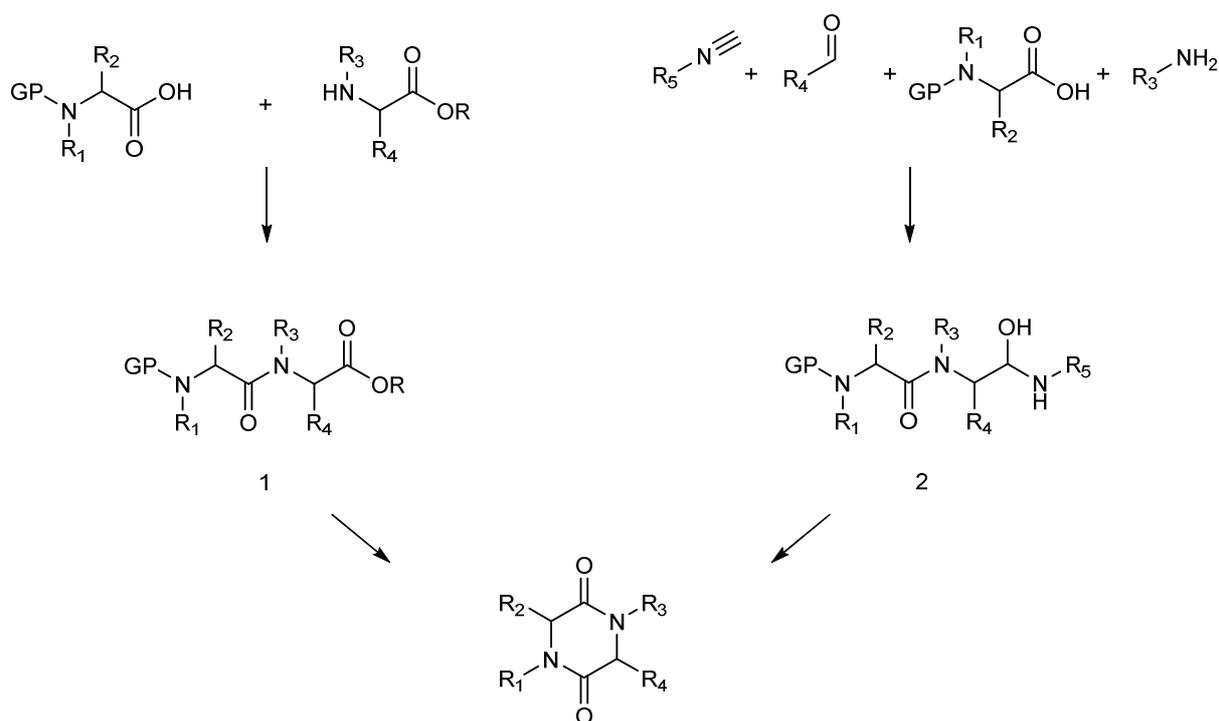


Figure 16 : Synthèse de dicétopipérazine par la formation de liaison amide (inspiré de<sup>31</sup>)

#### 1.1.1 Cyclisation d'ester dipeptidique

L'ester dipeptidique 1 est obtenu en couplant un acide aminé dont l'amine est protégée et un acide aminé portant une fonction ester. Une fois l'ester dipeptidique déprotégé, il peut réagir sur lui-même pour former le cycle dicétopipérazine. L'efficacité de la cyclisation peut dépendre des chaînes latérales des

acides aminés. Pour favoriser la cyclisation, une réaction à chaud en milieu acide<sup>212</sup> ou basique<sup>213</sup> peut être opérée bien que ces conditions puissent favoriser une racémisation. Une nouvelle méthode efficace quels que soient les acides aminés utilisés, écologique et limitant l'épimérisation a été récemment décrite. Elle consiste à réaliser la réaction en solvant aqueux en chauffant par micro-onde<sup>214</sup>. Cette méthode a été optimisée en déprotégeant l'amine par irradiation micro-onde permettant ainsi de réaliser les deux dernières étapes (déprotection et cyclisation) de manière simultanée<sup>215,216</sup>.

Des synthèses en phase solide utilisant le même schéma réactionnel sont aussi connues. Le clivage depuis la résine peut alors se faire après<sup>217</sup>, avant la cyclisation<sup>218</sup> ou simultanément<sup>219</sup>.

### 1.1.2 Réaction de Ugi

La réaction de Ugi permet d'obtenir à partir d'un isonitrile, d'un acide (aminé), d'un aldéhyde et d'une amine un précurseur acyclique (2) avec des rendements et une pureté optique équivalents à ceux obtenus par la précédente méthode<sup>220</sup>. Cette voie de synthèse donne accès à une plus grande diversité chimique étant donné qu'un grand choix de précurseurs (acide carboxylique, aldéhyde, amine) sont disponibles commercialement, contrairement aux acides aminés nécessaires au couplage peptidique<sup>46</sup>. L'isonitrile, quant à lui, est choisi de manière à donner une amide terminale suffisamment labile pour permettre la cyclisation, tel que le 1-isocyanocyclohexène<sup>221</sup> ou le 1-isociano-2(2,2-diméthoxyéthyl)benzène<sup>222</sup> (Figure 17). Une fois l'intermédiaire dipeptidique linéaire obtenu par la réaction de Ugi, un traitement acide permet de déprotéger l'amine et d'activer l'amide en C-terminal pour permettre la cyclisation, sans affecter d'autres fonctionnalités chimiques et la stéréochimie de la molécule.

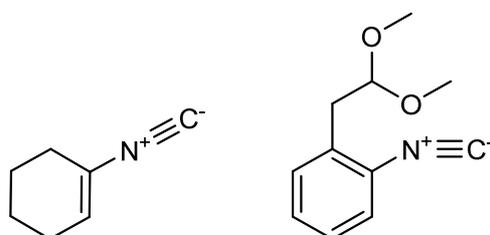


Figure 17 : Isonitriles utilisés pour la réaction de Ugi : 1-isocyanocyclohexène et 1-isociano-2(2,2-diméthoxyéthyl)benzène

Le dipeptide 2 peut également être synthétisé en phase solide à partir d'isonitrile sur résine. La cyclisation peut avoir lieu après clivage du dipeptide<sup>223</sup> ou simultanément en réalisant un clivage cyclisant<sup>224</sup>. Dans ce dernier cas, une cyclisation à 90°C sans acide acétique doit être préférée, afin d'éviter l'épimérisation du produit final. D'une manière générale, la synthèse en phase solide permet d'obtenir des produits plus purs que ceux obtenus par la synthèse en solution.

### 1.1.3 Autres voies utilisant la formation de liaison amide

Il est possible d'obtenir une 2,5-dicétopipérazine par condensation directe de deux acides aminés, mais avec généralement un rendement faible<sup>225</sup> (Figure 18). L'utilisation de phosgène pour activer l'un des

acides aminés<sup>226</sup> ou d'agent de couplage peptidique<sup>227</sup> permet de favoriser la réaction. Mais, c'est l'activation par micro-onde associée à des agents de couplage peptidique qui permet d'obtenir les meilleurs rendements (85-95 %) en un temps court (5min) sans opérer d'épimérisation<sup>228</sup>. Il a été montré que l'introduction d'une faible proportion de liquide ionique dans le milieu réactionnel exposé aux micro-ondes permettait de préparer des dicétopipérazines non symétriques et chirales, ou symétriques à partir d'acides aminés non protégés<sup>229</sup>. L'utilisation de la réaction d'Aza-Wittig a été également décrite comme voie possible pour la synthèse de 2,5-dicétopipérazines<sup>230</sup>.

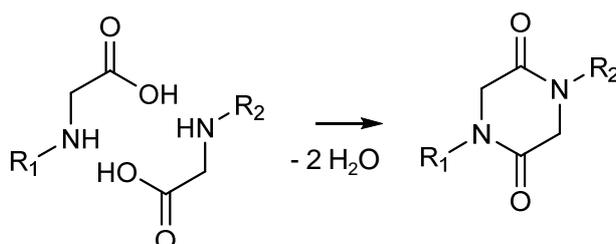


Figure 18 : Condensation d'acides aminés (d'après <sup>46</sup>)

### 1.2 Via N-alkylation

La N-alkylation est la seconde voie la plus utilisée pour la synthèse de 2,5-dicétopipérazines. En présence de base, le précurseur linéaire  $\alpha$ -haloacétamide est capable de se cycliser pour former des 2,5-dicétopipérazines 1,4 disubstituées symétriques (Figure 19). Le substituant de l'azote a une influence sur le rendement de cyclisation. Ainsi les substituants aryles ou alkyles sont bien tolérés<sup>231</sup>, alors que les substituants électroattracteurs empêchent la cyclisation<sup>232</sup>. La réaction de Ugi peut également être mise à profit pour produire le précurseur  $\alpha$ -haloacétamide même si cette méthode a tendance à provoquer l'épimérisation des centres stéréogènes<sup>233</sup>. Il est à noter que d'autres voies de synthèse basées sur la réaction de Ugi combinée à une réaction de Aza-Mickaël<sup>234</sup> ou une réaction de Diels-Alder<sup>235</sup> peuvent être utilisées pour la formation des cycles dicétopipérazines.

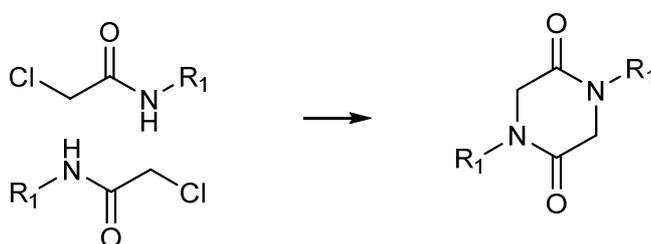


Figure 19 : Cyclisation via  $\alpha$ -haloacétamides (chloroacétamides) (d'après <sup>46</sup>)

### 1.3 Autres voies de synthèse

Comme l'ont montré Peng et Clive, la formation du cycle dicétopipérazine peut également être obtenue avec un très bon rendement, *via* N-alkylation par une cyclisation intramoléculaire résultant de la réaction d'un énolate sur le carbonyle d'un carbamate phénylique<sup>236</sup> (Figure 20). Une cyclisation radicalaire

a également été décrite pour la synthèse de l'avrainvillamide<sup>237</sup>. La transformation par liaison N1-C2/C3-N4, appelée cyclisation tandem, est obtenue par la réaction d'un acide  $\alpha$ -aminé avec un électrophile bivalent. Un exemple de ce type de voie est décrit pour la synthèse de tétrahydro- $\beta$ -carboline et tétrahydroisoquinoline<sup>238</sup>. La formation simultanée de la liaison N1-C2 et N1-C6 permet une introduction aisée de substituants en position N1. Un exemple de ce type de réaction est donné pour la synthèse des analogues de  $\beta$ -carboline. L'amine primaire réagit avec l'haloester précédemment formé puis avec l'ester pour former le cycle<sup>239</sup>.

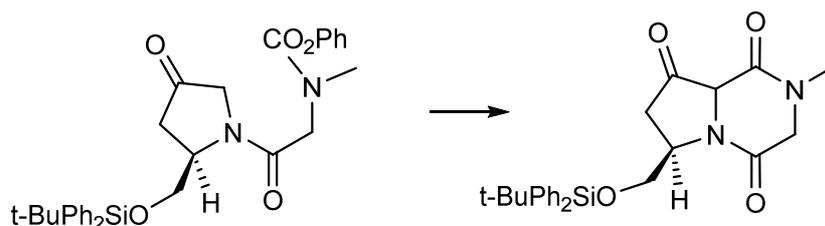


Figure 20 : Cyclisation via acylation d'un énolate (d'après<sup>46</sup>)

## 2. Fonctionnalisation du cycle dicétopipérazine

Si le choix des précurseurs peut permettre d'introduire au moment de la formation du cycle dicétopipérazine une variété de groupements chimiques, la fonctionnalisation du cycle déjà formé est également possible. Le cycle peut ainsi être modifié en 6 positions.

### 2.1 Fonctionnalisation sur les carbones

Plusieurs méthodes ont été développées pour fonctionnaliser en C-3 et C-6 de manière régio-contrôlée et stéréo-contrôlée.

Le passage par des intermédiaires énolates permet une C-alkylation qui nécessite néanmoins la protection des deux amides. Un contrôle diastéréoisomérique peut être obtenu en choisissant un groupement protecteur approprié<sup>240</sup>, par exemple un inducteur chiral<sup>241</sup>. Le type d'agent alkylant utilisé influence également le diastéréoisomère obtenu en fonction notamment de son encombrement stérique<sup>242</sup>. Les dicétopipérazines peuvent également être fonctionnalisées en utilisant un précurseur portant un groupe alcoxycarbonyle qui peut être éliminé par hydrolyse suivie d'une décarboxylation<sup>243</sup>. Cette méthode permet d'introduire des groupements aryle<sup>244</sup> et thioether<sup>245</sup>.

Une halogénéation des dicétopipérazines peut être obtenue par voie radicalaire ou électrophile. La régio-sélectivité de la réaction peut être contrôlée par le choix des N-substituants qui activent ou désactivent les positions carbones adjacentes<sup>246</sup>. A partir de ces composés halogénés notamment bromés, une large gamme de substitution nucléophile peut avoir lieu : SR<sup>247</sup>, OR<sup>248</sup>, NR<sub>2</sub><sup>249</sup>, alkyl<sup>250</sup> et aryl<sup>251</sup>.

L'introduction d'un atome d'oxygène en C-3 ou en C-6 peut être obtenue par l'oxydation d'un énolate, mais des méthodes plus douces sont parfois nécessaires, notamment lors de la synthèse totale de

produits naturels. Une voie radicalaire a ainsi été développée et utilise la formation d'un peroxyde, obtenu par irradiation sous oxygène, et sa réduction pour obtenir l'hydroxyde attendu<sup>252,253</sup>.

Le cycle dicétopipérazine peut également être fonctionnalisé en condition basique par condensation aldolique entre un aldéhyde et l'intermédiaire énolate du cycle<sup>254</sup> (Figure 21).

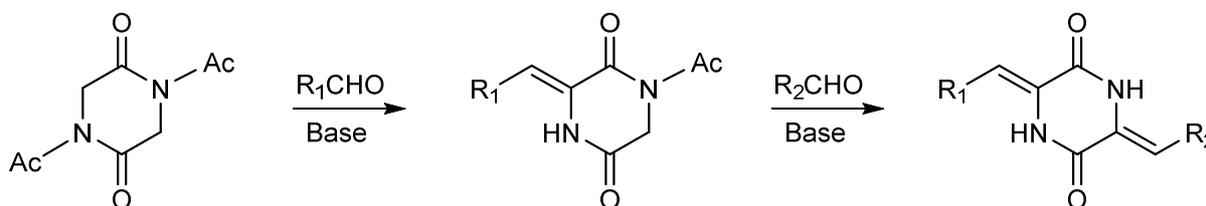


Figure 21 : Schéma de réaction de la fonctionnalisation en C-1 et C-3 par une condensation aldolique

## 2.2 Fonctionnalisation sur les azotes

L'azote de la lactame du cycle peut être alkylé en utilisant de l'hydride de sodium<sup>255,256</sup> comme base mais, dans ces conditions, une compétition avec l'alkylation du carbone est possible<sup>257</sup>. Le choix de l'agent alkylant peut permettre de favoriser l'une ou l'autre des alkylations. Dans le cas d'une réaction sur le cyclo(Pro-Xaa), les conditions d'alkylation peuvent également provoquer une épimérisation<sup>258</sup>.

L'azote du cycle peut également opérer en condition acide une cyclisation intramoléculaire en réagissant sur un noyau indole<sup>259</sup>. Cette cyclisation peut être accompagnée d'une bromation (en présence de Br<sub>2</sub>)<sup>260,261</sup> ou d'une oxydation<sup>262,263</sup>. L'azote du cycle peut également réagir avec un aldéhyde pour donner des dérivés *N*- $\alpha$ -hydroxyalkyles générant des espèces *N*-acyliminium qui, en condition acide, se cyclisent en un cycle aromatique *via* une réaction de Friedel-Craft intramoléculaire<sup>264</sup>. Enfin, des arylations intermoléculaires ou intramoléculaires sont possibles sur l'azote.

## 2.3 Réaction sur les carbones des carbonyles (C-2) (C-5)

La réactivité du carbonyle ou de la lactame du cycle dicétopipérazine peut être utilisée pour obtenir une grande variété de dérivés : pipérazines<sup>265,266</sup>, cycle thiolactame<sup>267</sup>, 3,6-dihydro-2,5-diéthoxy-pyrazine<sup>268</sup>, 1,4-dihydropyrazine et 1,4-pyrazine<sup>269</sup>, pyrazino[2,1-b]quinazoline-3,6-dione<sup>270</sup>, pipérazine-3-one pontée<sup>264,271</sup>, cyclols et cyclodepsipeptides<sup>272</sup>. Néanmoins, les modifications apportées sur le cycle dicétopipérazine sont telles que les molécules obtenues ne sont plus des dicétopipérazines. Les voies chimiques menant à ces molécules ne seront donc pas discutées.

La chimie des dicétopipérazines a donc été énormément développée ces 30 dernières années du fait du potentiel pharmacologique de ces molécules. Elle présente néanmoins plusieurs difficultés liées

notamment à la réactivité des six positions fonctionnalisables. En effet, une compétition existe entre les positions N-1/N-4, les positions C-3/C-6, et les positions C/N. Le contrôle de la régio-sélectivité des réactions est donc primordial et nécessite généralement des opérations de protection et de déprotection qui sont incompatibles avec une approche de chimie verte. Le contrôle stéréochimique est l'autre défi à relever lors de la synthèse de dicétopipérazines naturelles ou d'inspiration naturelle. Les méthodes de formation ou de fonctionnalisation du cycle doivent permettre d'éviter la racémisation de la molécule, l'épimérisation de centres stéréogènes et de contrôler le diastéréoisomère obtenu. Enfin, lorsque le motif dicétopipérazine est inséré dans des architectures plus importantes, dans la synthèse totale de molécules naturelles complexes par exemple, il convient de trouver des conditions de réaction douces qui n'altèrent pas d'autres fonctions chimiques de la molécule. Dans une approche de chimie verte, l'utilisation d'enzymes pourrait donc venir compléter ou suppléer la synthèse par voie chimique, notamment dans des étapes délicates où la régio-sélectivité et le contrôle stéréochimique sont requis. Pour ce faire, une approche de choix est l'identification et la caractérisation des enzymes constituant les voies de biosynthèse des dicétopipérazines. De plus, la découverte de nouvelles voies menant à la synthèse de nouvelles dicétopipérazines naturelles pourrait permettre d'enrichir le réservoir de dicétopipérazines au potentiel thérapeutique.

Peu de voies de biosynthèse de dicétopipérazines sont caractérisées. Jusque dans les années 2000, il était considéré que les dicétopipérazines étaient synthétisées uniquement dans des voies impliquant les synthétases de peptides non ribosomiques (NRPS pour Non Ribosomal Peptide Synthetases). En 2002, la première voie utilisant une enzyme n'appartenant pas à la famille des NRPS a été identifiée pour la synthèse de l'albonoursine. En effet, cette voie utilise une enzyme qui a été le premier représentant identifié d'une nouvelle famille d'enzymes : les synthases de cyclodipeptides ou CDPS.

Après une présentation générale des NRPS, quelques exemples de voies de biosynthèse de dicétopipérazines utilisant les NRPS seront présentés. Une description plus exhaustive des CDPS sera ensuite proposée avant de présenter les voies à CDPS connues dans la littérature au début de mon travail de thèse.

## IV. Les voies de biosynthèse de dicétopipérazines utilisant les synthétases de peptides non ribosomiques (NRPS)

### 1. Les synthétases de peptides non ribosomiques, des méga-complexes enzymatiques organisés en modules

Les NRPS appartiennent à la grande famille des mégasynthèses qui regroupent également les synthèses de polycétides (polyketide synthases, PKS) et les synthèses d'acides gras (fatty acid synthases, FAS). Les NRPS sont des complexes multienzymatiques responsables de la synthèse de petits peptides (2 à 50 acides aminés). Leur caractérisation et leur purification ont débuté à la fin des années 1960. Mais leur fonctionnement n'a commencé à être compris qu'au cours des années 1970, avec les travaux des équipes de Lipmann puis de Marahiel<sup>273</sup>. Les NRPS sont de taille importante et composées d'une ou plusieurs chaînes polypeptidiques (ou sous-unités). Elles sont organisées de manière modulaire. Chaque module est une « section » d'une chaîne polypeptidique et est généralement responsable de l'incorporation d'un acide aminé dans le peptide final. A titre d'exemple, la NRPS responsable de la biosynthèse de la tyrocidine est représentée (Figure 22). Cette NRPS de 1,3 MDa est composée de trois chaînes polypeptidiques, TycA, TycB, TycC, et contient au total 10 modules. Dans cet exemple, TycB est composée de trois modules et est responsable de l'incorporation de trois acides aminés<sup>274</sup>.

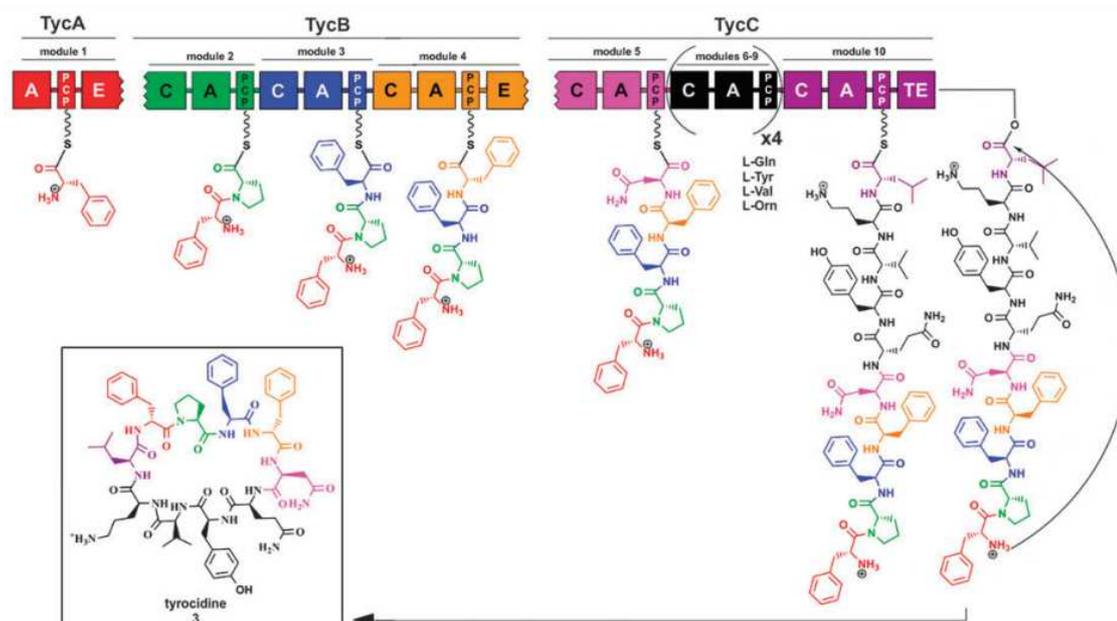


Figure 22 : Organisation de la tyrocidine synthétase

La NRPS responsable de la synthèse de la tyrocidine est codée par trois gènes *tycA*, *tycB* et *tycC*. Chacun de ces gènes code une des sous-unités de la NRPS : TycA, TycB et TycC, respectivement constituée de un, trois et six modules. Chaque module est responsable de l'incorporation d'un acide aminé (de la même couleur que le module) et est divisé en domaine. (A : domaine d'adénylation, PCP : domaine de thiolation, C : domaine de condensation, TE : domaine de thioestérification, E : domaine d'épimérisation ; reproduit de<sup>273</sup>

Chaque module est divisé en domaine, chacun représentant une « unité » enzymatique qui catalyse une étape de la réaction. Les domaines peuvent être reconnus par des motifs de séquence hautement conservés. Trois domaines responsables de la reconnaissance et de l'activation du substrat, de la propagation de la chaîne polypeptidique naissante aux différents sites catalytiques et de la formation de la liaison peptidique, sont indispensables à la synthèse du peptide. Il s'agit respectivement des domaines d'adénylation (A), de thiolation (T ou PCP pour Peptidyl Carrier Protein) et de condensation (C) <sup>275</sup>. A l'exception du premier module qui ne comporte pas de domaine de condensation et est appelé module d'initiation, chaque module du complexe enzymatique contient ces trois domaines dans l'ordre suivant C-A-T et est appelé module d'élongation<sup>276,277</sup>. Le dernier module du complexe enzymatique est appelé module de terminaison et contient souvent un domaine thioestérase chargé notamment de la libération du peptide synthétisé.

### 1.1 Le domaine d'adénylation reconnaît le substrat et catalyse son activation

Le domaine d'adénylation sélectionne l'acide aminé et l'active en présence d'ATP en formant un acide aminé adénylé (Figure 23). S'il l'on souhaite réaliser un parallèle avec la synthèse ribosomique, cette étape correspondrait au chargement de l'acide aminé sur l'ARNt par une aminoacyl-ARNt synthétase. Ce premier domaine comprend environ 550 acides aminés<sup>275</sup>. Il est à noter que, en assurant le choix des substrats, les domaines d'adénylation déterminent la structure finale du peptide synthétisé.

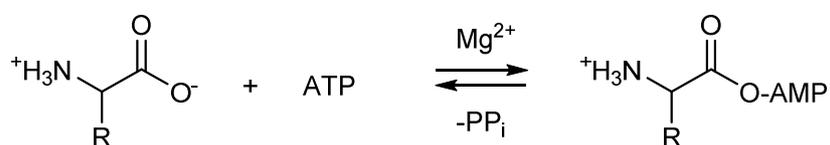


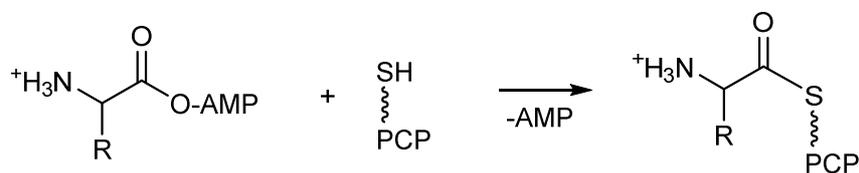
Figure 23 : Réaction au domaine d'adénylation

Les travaux de Stachelhaus et Marahiel ont énormément contribué à la compréhension du mode de reconnaissance de l'acide aminé par le domaine d'adénylation. La résolution d'une structure du domaine d'adénylation de la phénylalanine (PheA) de la gramicidine S synthétase en complexe avec son substrat, la phénylalanine, et l'AMP<sup>278</sup> a permis d'identifier la poche de liaison de la phénylalanine qui est constituée de dix résidus. En s'appuyant sur des données structurales et des données d'identité de séquences, Marahiel *et al.* ont émis l'hypothèse que les chaînes principales des domaines d'adénylation adoptaient la même conformation quel que soit l'acide aminé reconnu. Ils ont alors supposé que les dix résidus identifiés dans la poche de liaison constituaient une signature de spécificité<sup>278</sup>. En alignant les séquences de 160 autres domaines d'adénylation, pour lesquels la spécificité était connue, avec celle de la PheA, les dix résidus de ces domaines supposés impliqués dans la liaison du substrat ont pu être déterminés. Un arbre phylogénétique des domaines d'adénylation calculé en utilisant uniquement des

séquences composées de ces dix résidus montre un « rassemblement » des domaine d'adénylation en fonction de leur spécificité<sup>279,280</sup>. Ces résidus peuvent être donc considérés comme une signature de séquence à l'instar des « codons » de la synthèse non ribosomique et peuvent être utilisés pour prédire la spécificité des domaines d'adénylation non testés. De même que le code génétique, ce code non ribosomique est dégénéré car il existe des signatures différentes pour l'incorporation du même acide aminé<sup>279</sup>. Ce code a, par la suite, été mis à jour et enrichi<sup>275</sup>, et permet aujourd'hui de prédire la spécificité du substrat avec une justesse relativement élevée. De nos jours, des outils bio-informatique basés sur ce code de reconnaissance sont disponibles en ligne pour prédire la spécificité de chaque domaine d'adénylation d'une NRPS comme NRPS predictor<sup>281,282</sup> (<http://nrps.informatik.uni-tuebingen.de>) ou PKS/NRPS Analysis Web-site<sup>283</sup> (<http://nrps.igs.umaryland.edu>).

### 1.2 Le domaine de thiolation permet le transfert des substrats activés d'un centre catalytique à l'autre

L'acide aminé activé par le domaine d'adénylation est transféré au domaine de thiolation (T ou PCP) qui contient environ 100 résidus. Son activité repose sur la présence d'un cofacteur 4'-phosphopantéthéinyle (4'PP) qui a été lié au préalable à une sérine conservée par la phosphopantéthéinyle-transférase. Ce bras phosphopantéthéinyle a un rôle de stabilisation, en liant l'acide aminé activé par formation d'un thioester <sup>284,285</sup> (Figure 24) ainsi qu'un rôle de transfert du substrat vers les centres de catalyse de la liaison peptidique.



**Figure 24 : Réaction au domaine de thiolation**

L'affinité du cofacteur 4'PP pour les différents domaines évolue en fonction de son chargement et permet la mobilité dans le module de l'acide aminé et de la chaîne peptidique en élongation (Figure 25). Elle est ainsi maximum pour le domaine d'adénylation quand le cofacteur est libre, ce qui permet le chargement de l'acide aminé sur le cofacteur (a). Elle se modifie alors pour être maximum envers le domaine de condensation (n) qui précède le domaine T, afin de permettre la formation de la liaison peptidique entre l'acide aminé activé et la chaîne peptidique en formation (b). Enfin, elle se modifie une dernière fois pour être maximale envers le site de condensation qui le suit<sup>286</sup>(c) (n+1). Le domaine de thiolation est donc responsable du transfert de l'acide aminé activé et des intermédiaires peptidiques d'un centre catalytique à l'autre.

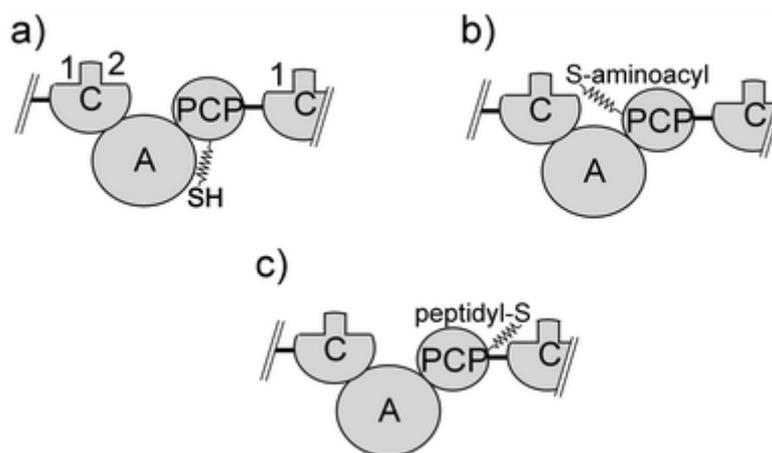


Figure 25 : Modèle montrant les étapes permettant l'élargissement du peptide sur un module, reproduit de<sup>275</sup>

### 1.3 Le domaine de condensation catalyse la formation de la liaison peptidique

Le domaine de condensation (C) (environ 450 acides aminés) assure la formation de la liaison peptidique entre l'acide aminé activé porté par le domaine de thiolation du module n, jouant le rôle de nucléophile, et le peptide ou l'acide aminé porté par le domaine de thiolation du module n-1 jouant le rôle d'électrophile.<sup>287,288</sup> (Figure 26). Le domaine de condensation possède donc deux sites : l'un pour accueillir la chaîne en élargissement provenant du module n-1 (site donneur (1), (Figure 25) et un autre pour accueillir l'acide aminé à ajouter provenant du module n (site accepteur (2)). Par ailleurs, en plus de la spécificité du domaine A, il a été montré que le domaine de condensation joue également un rôle dans la spécificité.

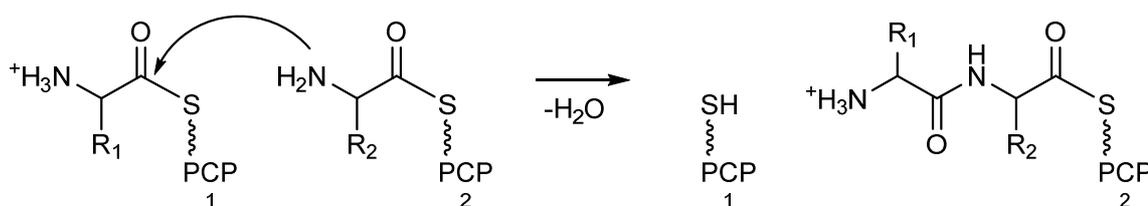


Figure 26 : Réaction au domaine de condensation

### 1.4 Le domaine de terminaison catalyse la libération du peptide

Le dernier module du complexe enzymatique est appelé terminaison et possède souvent un domaine thioestérase (Te) d'environ 250 résidus qui permet la libération du peptide synthétisé. Le peptide formé à la dernière étape de condensation est transféré au résidu sérine du site actif du domaine Te. Après cette acylation, le domaine Te est désacylé par une attaque nucléophile du NH<sub>2</sub> de l'extrémité N-terminale (par exemple dans le cas de la tyrocidine, Figure 22) ou par un groupe fonctionnel d'une des chaînes latérales du peptide (amine, ester ou hydroxyle). Cela provoque une cyclisation intramoléculaire ou

macro-cyclisation du peptide par la formation de fonction amide ou ester. (Figure 27). Le domaine Te peut également catalyser la di- ou trimérisation du peptide. Enfin, si l'attaque nucléophile est opérée par une molécule H<sub>2</sub>O (par hydrolyse), le peptide obtenu est alors linéaire<sup>289</sup>. Il est à noter que si le domaine Te est délété, la libération du produit n'a pas lieu. La libération du peptide permet également la régénération de la NRPS qui peut recommencer une nouvelle synthèse.

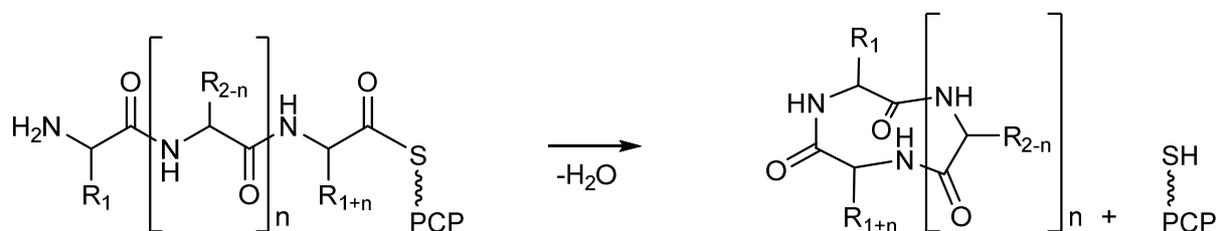


Figure 27 : Réaction au domaine de thiolation

### 1.5 Les peptides non ribosomaux peuvent être modifiés au cours ou après leur synthèse

En plus des domaines déjà décrits, d'autres domaines peuvent être présents dans les modules afin de réaliser des modifications sur la chaîne polypeptidique pendant sa synthèse.

Ainsi les chaînes latérales des cystéines, des sérines et des thréonines peuvent subir des hétérocyclisations réalisées par des domaines Cy (450 acides aminés) qui remplacent le domaine C classique en assurant également la formation de la liaison peptidique<sup>290,291</sup>. Pour transformer des thiazolines ou oxazolines en thiazoles ou oxazoles, la présence d'un domaine d'oxydation appelé Ox est nécessaire<sup>292-294</sup>. Les atomes d'azote du squelette peptidique peuvent subir des méthylation. Cette modification, fréquemment retrouvée chez les peptides d'origine fongique<sup>295</sup>, est introduite par un domaine de N-méthylation (N-Mt) de 420 acides aminés et nécessite la S-adénosylméthionine (SAM) comme cofacteur<sup>296-299</sup>. Des C-méthylations sont également possibles et sont assurées par des domaines C-Mt<sup>290</sup>. En plus du contrôle stéréochimique assuré par le domaine A, certaines NRPS sont capables de contrôler la stéréochimie des acides aminés incorporés par le biais de domaines d'épimérisation (E) (450 acides aminés) qui réalisent en réalité une racémisation de l'acide aminé<sup>300</sup>. Le domaine C suivant a alors la charge de sélectionner uniquement l'énantiomère/diastéréoisomère D du mélange L/D produit<sup>301</sup>. On recense également la présence de domaines permettant la formation de liaison avec des acides gras<sup>302,303</sup> ou de domaines réductase (R) qui transforment l'acide carboxylique terminal en aldéhyde ou alcool<sup>304</sup>. Enfin, des enzymes hybrides NRPS-PKS permettent d'introduire des unités acétates ou propionates dans des structures peptidiques<sup>305-307</sup> (bléomycine A2<sup>308</sup>) ou des unités d'acides aminés dans des polycétides (épothilone<sup>309</sup>).

Des modifications peuvent avoir lieu une fois que le peptide formé est libéré par la NRPS. Ces modifications sont catalysées par des enzymes appelées « tailoring enzymes » qui peuvent être des glycosyl

transférases, des halogénases ou des hydroxylases. Les gènes codant ces enzymes et les NRPS sont souvent associés au niveau génomique et forment des clusters de gènes de biosynthèse<sup>310</sup>. Des exemples de ce type de modifications vont être donnés dans la partie suivante qui illustre la biosynthèse de dicétopipérazines.

## **2. Voies de biosynthèse de dicétopipérazines dépendant des NRPS**

Douze voies de biosynthèse dépendant des NRPS sont à ce jour identifiées. Dans la majorité des cas, la voie est dédiée à la biosynthèse de la dicétopipérazine comme pour la gliotoxine<sup>311</sup>, la brévianamide F et ses dérivés appelés fumitremorgines<sup>312</sup>, l'érythrochéline<sup>313</sup>, l'ergotamine<sup>314</sup>, la roquefortine C<sup>315</sup>, l'acétylazonalénine<sup>316</sup>, la thaxtomine A<sup>130</sup>, la sirodesmine PL<sup>128</sup>, le coprogène<sup>317</sup> ou l'acide dimérumique<sup>318</sup>. Dans les autres cas, les dicétopipérazines sont des produits secondaires de voies de biosynthèse de peptides de taille plus importante, comme le cyclo(D-Phe-Pro) et la cyclomarazine A<sup>319,320</sup> (Figure 28). Même si elles font toutes intervenir une NRPS, ces voies ne sont pas organisées de manière similaire. Afin d'illustrer ces différentes organisations, je vais présenter plusieurs exemples de voies.

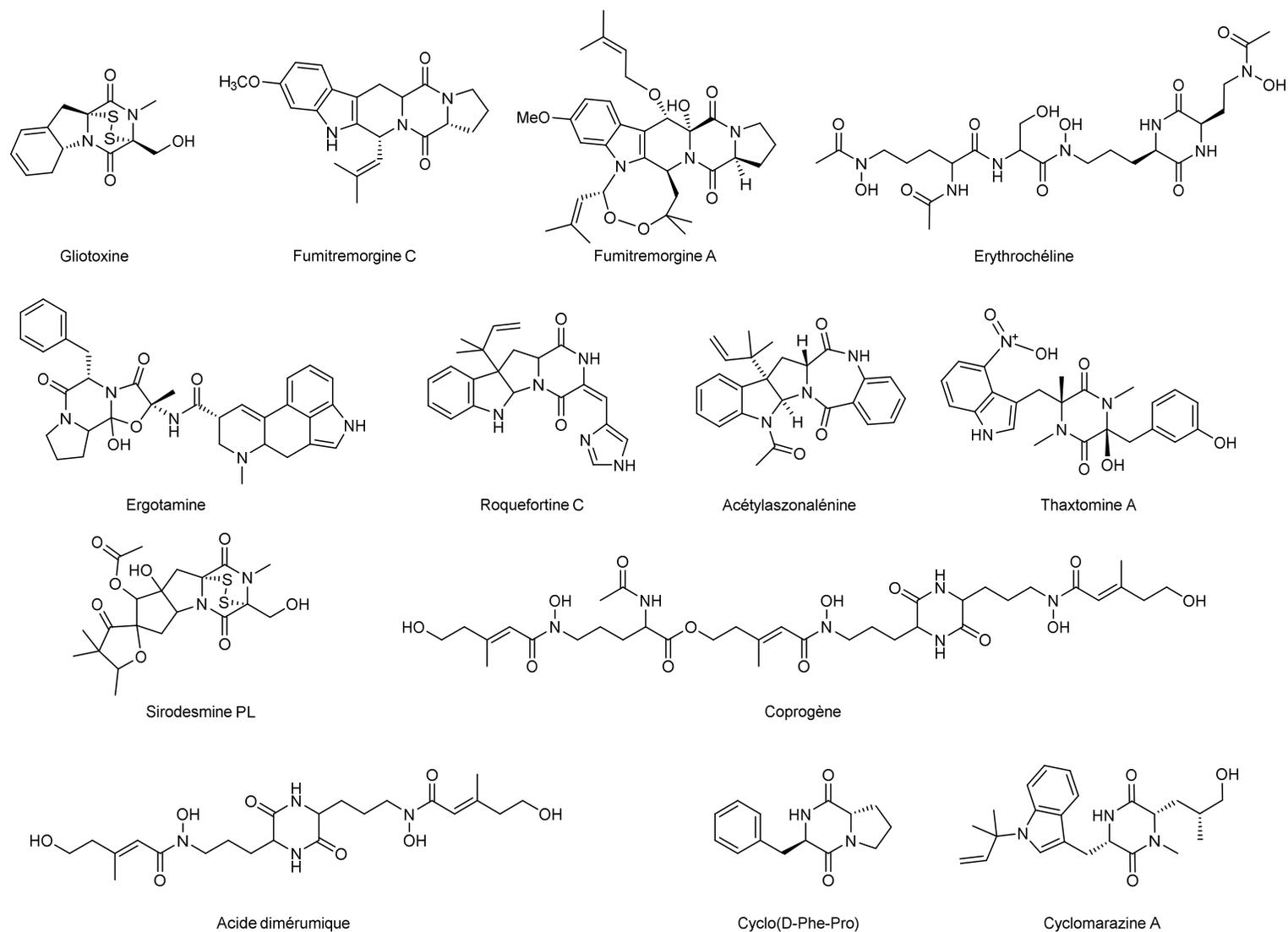


Figure 28 : Dicotopiprazines synthétisées par des voies de biosynthèse à NRPS

## 2.1 Voies de biosynthèse de la gliotoxine et des épidthiodioxopipérazines : les gènes de la NRPS et des enzymes de modifications sont regroupées en un même cluster

La gliotoxine est une mycotoxine isolée à l'origine de *G. fimbriatum* et produite par plusieurs souches de champignons, parmi lesquels *A. fumigatus*, agent causal d'aspergilloses invasives. Elle joue un rôle important dans la pathobiologie de ce dernier et a démontré des activités de modulation de la réponse immunitaire et une capacité à induire l'apoptose chez plusieurs lignées de cellules<sup>150,321</sup>. Elle possède une structure particulière dans laquelle un pont disulfure, indispensable à son activité biologique, relie les carbones 3 et 6 du noyau dicétopipérazine et est considérée comme le prototype de la classe des épidthiodioxopipérazines. Etant donné sa structure particulière et son activité remarquable, la voie de biosynthèse de ce métabolite a été particulièrement étudiée. Les gènes responsables de la synthèse de la gliotoxine sont regroupés en un seul cluster de 13 gènes dans le génome de *A. fumigatus*<sup>128,322,323</sup> (Figure 29). Globalement, une NRPS (GliP)<sup>311</sup> est responsable de la synthèse du motif dicétopipérazine. Un cytochrome P450 (GliC) et une glutathion S-transférase (GliG) participent à la formation de la liaison C-S entre les carbones du cycle dicétopipérazine et le glutathion (GSH). Les étapes menant à l'obtention des deux fonctions thiols libres ne sont pas encore élucidées mais semblent impliquer une dipeptidase (GliJ) et une enzyme de type 1-aminocyclopropane-1-carboxylate synthase (GliI). Enfin une oxydase (GliT) est responsable de la formation du pont disulfure<sup>321</sup>. Ce cluster sert de modèle aux autres clusters responsables de la synthèse des épidthiodioxopipérazines, parmi lesquels celui de la sirodesmine PL<sup>128,322</sup>.

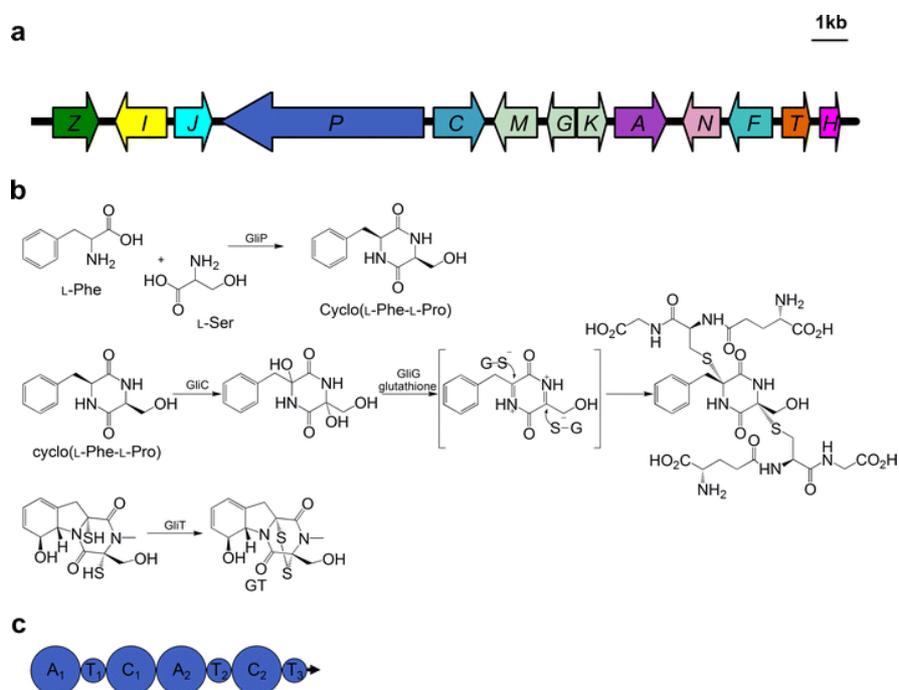


Figure 29 : Biosynthèse de la gliotoxine

a) Cluster de gènes pour la voie de biosynthèse de la gliotoxine chez *A. fumigatus*. B) Etapes clés de la voie biosynthèse de la gliotoxine c) Organisation modulaire de la NRPS GliP. Reproduit de<sup>324</sup>

## 2.2 La voie de biosynthèse des dicétopipérazines de type fumitremorgine utilise une enzyme de modification « extérieure » au cluster de la NRPS

Les alcaloïdes de type fumitremorgine sont des dérivés de la brévianamide F (cyclo(Pro-Trp)) tels que les tryprostatines et le verruculogène<sup>324</sup>. Certains de ces composés ont montré des activités biologiques intéressantes. Par exemple, les tryprostatines A et B et la fumitremorgine C ont des activités anticancéreuses, (cf. II.2.2 Activités antitumorales/anticancéreuses), le verruculogène et la fumitremorgine B sont des agents génotoxiques et la fumitremorgine A a des effets provoquant des tremblements de la tête et du corps chez la souris<sup>325</sup>. Durant les vingt dernières années, le verruculogène et la fumitremorgine A ont été isolés de nombreux champignons et en particulier de *A. fumigatus* et *Neosartorya fischeri*<sup>326</sup>. Un cluster de gènes pour la synthèse de ces composés a été identifié chez *A. fumigatus* Af293, *A. fumigatus* A1163 et *N. fischeri* NRRL181 (Figure 30). Ce cluster comprend huit gènes et les identités entre les gènes orthologues des trois espèces sont comprises entre 68-100%<sup>326-328</sup>. Les fonctions des enzymes codées par ces gènes ont été élucidées par des approches d'inactivation de gènes et/ou d'expression d'enzymes recombinantes chez *E. coli*. La synthèse de ce type d'alcaloïdes commence par la condensation des deux acides aminés L-tryptophane et L-proline en brévianamide F qui est catalysée par une NRPS (FtmPS ou FtmA)<sup>312</sup>. Puis la prényltransférase FtmPT1 permet d'obtenir la tryprostatine B<sup>329</sup>. A partir de cette dernière, la tryprostatine A est obtenue par action du cytochrome P450 FtmP450-1 et de la méthyltransférase FtmMT<sup>328</sup>. Les deux autres cytochromes P450 (FtmP450-2 et FtmP450-3) codés par le cluster sont responsables de la cyclisation de la tryprostatine A en fumitremorgine C et de l'hydroxylation de cette dernière pour donner la 12,13-hydroxyfumitremorgine C<sup>328</sup>. Ce dernier composé est ensuite prénylé par la deuxième prényltransférase FtmPT2, ce qui conduit à la fumitremorgine B<sup>330</sup> qui peut être convertie en verruculogène par FtmOx1<sup>331</sup>. Enfin un dernier gène, n'appartenant pas au cluster précédent, a été identifié chez *N. fischeri* NRRL181 et code une prényltransférase FtmPT3 qui réalise une O-prénylation d'un groupement hydroxyle sur le verruculogène pour donner la fumitremorgine A<sup>326</sup>. Aucun homologue de ce gène n'a été trouvé dans le génome d'*A. fumigatus* Af293 et *A. fumigatus* A1163. Ces souches ne possèdent donc pas le potentiel génétique pour synthétiser la fumitremorgine A.

Ainsi, la voie de biosynthèse de la fumitremorgine A a été complètement élucidée chez *N. fischeri* et est constituée d'une NRPS et de plusieurs enzymes de modification dont l'une n'est pas « clustérisée » au niveau génomique.

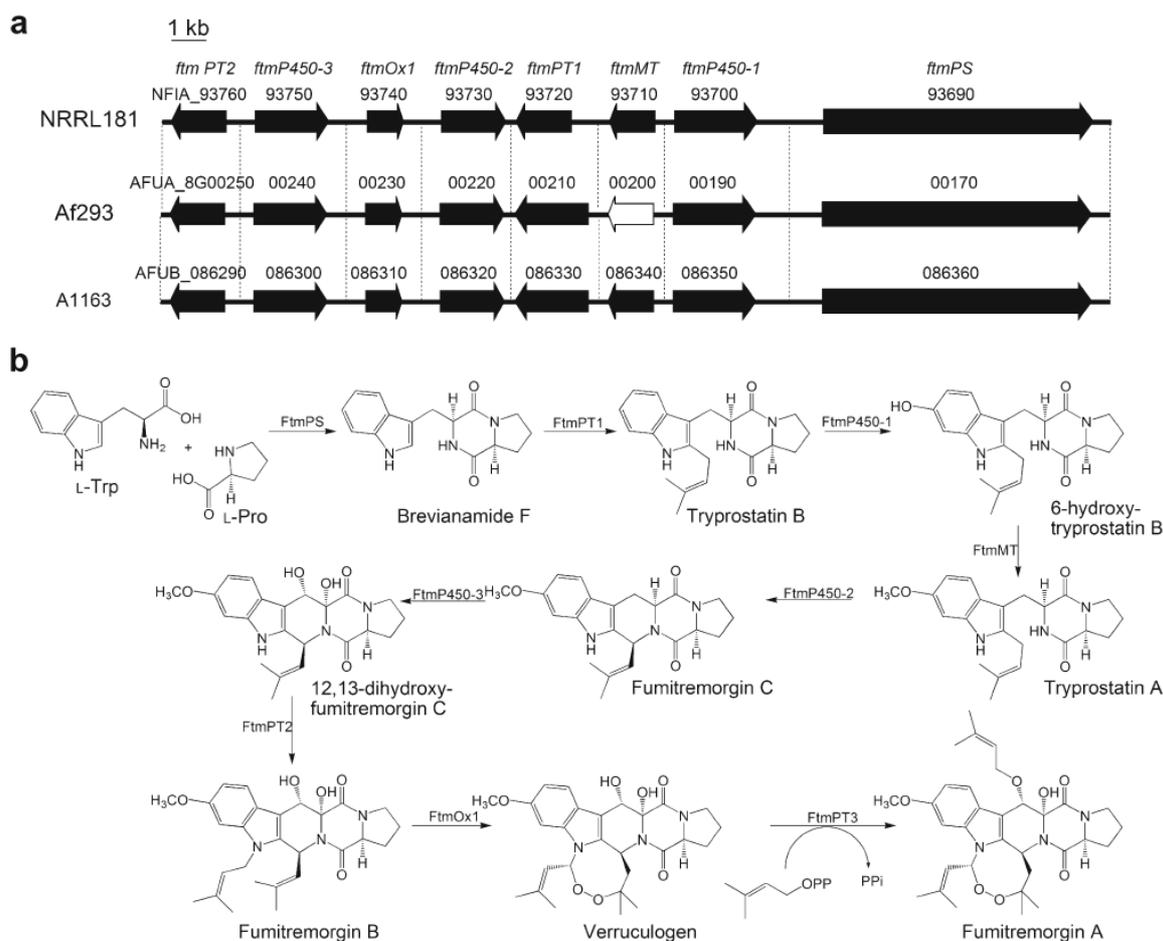


Figure 30 : Biosynthèse des dicétopipérazines de type fumitremorgine

a) Cluster de gènes pour la voie de biosynthèse des dicétopipérazines de type fumitremorgine chez *A. fumigatus* *Af293* and *A1163* and *N. fischeri* *NRRL18*. Les gènes en noir ont été testés expérimentalement b) Voie de biosynthèse proposée pour des dicétopipérazines de type fumitremorgine chez *N. fischeri* *NRRL18*. Reproduit de<sup>324</sup>

### 2.3 La voie de biosynthèse de la thaxtomine : la NRPS prend en charge un acide aminé modifié au préalable

Les phytotoxines de type thaxtomine, produites par des *Streptomyces* phytopathogènes, sont à l'origine de gale de racines et de tubercules et provoquent une nécrose des tissus<sup>129,332</sup>(cf. II.1.2.1 Virulence de micro-organismes pathogènes). Ce sont des cyclodipeptides formés à partir de L-phénylalanine et de L-4-nitrotryptophane. Les gènes responsables de la synthèse des thaxtomines sont regroupés en un cluster retrouvés sur les chromosomes de plusieurs *Streptomyces*<sup>333</sup> (Figure 31). Une synthase d'oxyde nitrique TxtD génère de l'oxyde nitrique (NO) à partir de L-arginine<sup>334</sup>. Le cytochrome P450 TxtE catalyse la réaction de nitration du L-tryptophane pour donner le L-4-nitrotryptophane, en utilisant NO (précédemment synthétisé) et O<sub>2</sub> comme co-substrats<sup>335</sup>. La NRPS codée par *txtA* et *txtB* permet la synthèse du motif dicétopipérazine N<sup>7</sup>-N méthylé à partir de cet acide aminé modifié, de la L-phénylalanine et de la S-adenosyl-L-méthionine<sup>130,336</sup>. Enfin le cytochrome P450 TxtC réalise

l'hydroxylation du noyau aromatique de la phénylalanine, ce qui conduit à la thaxtomine A<sup>337</sup>. Les autres thaxtomines peuvent se différencier par l'absence d'un *N*-méthyl ou de l'hydroxylation<sup>338</sup>.

Ainsi l'introduction de modification peut se faire avant ou après la synthèse du cycle dicétopipérazine.

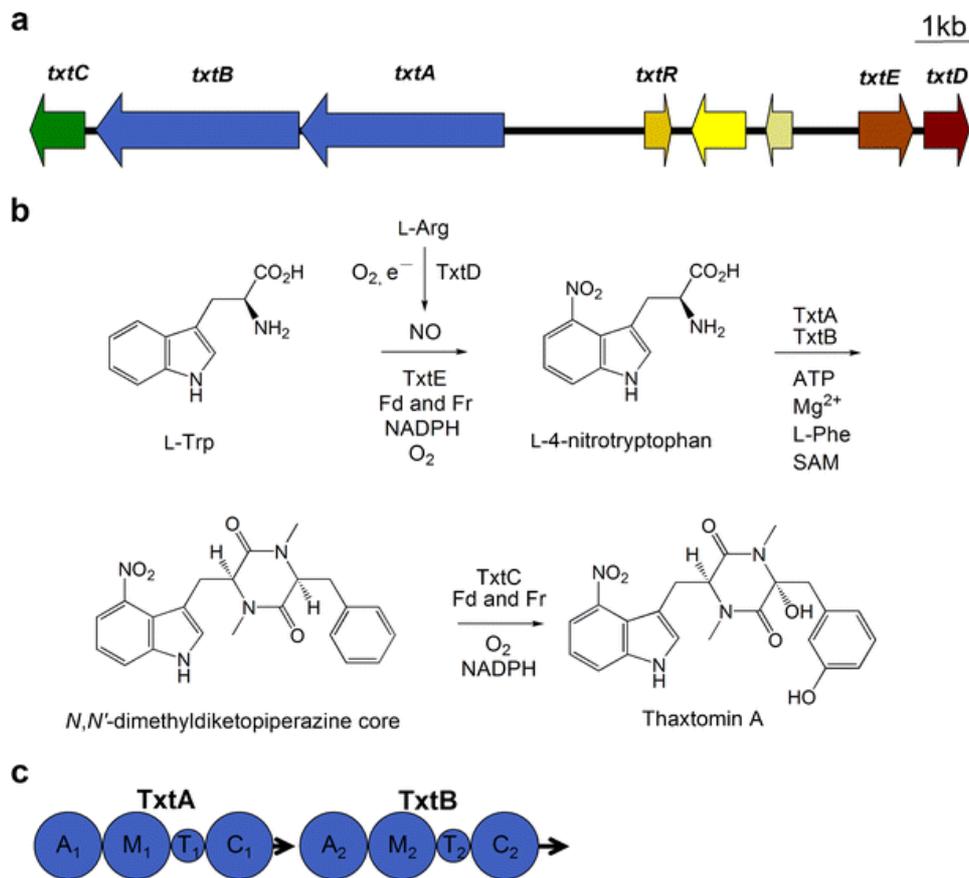


Figure 31 : Biosynthèse de la thaxtomine A

a) Cluster de gènes pour la voie de biosynthèse de la thaxtomine A chez *S. turgidiscabies*. b) Etapes clés de la voie biosynthèse de la thaxtomine A. c) Organisation modulaire de la NRPS TxtA et TxtB. Reproduit de<sup>324</sup>

## 2.4 La biosynthèse de l'érythrocheline : l'acide aminé utilisé est synthétisé par un autre cluster distant du cluster contenant la NRPS

L'érythrocheline est une dicétopipérazine potentiellement sidérophore produite par *Saccharopolyspora erythraea*. Contrairement au mécanisme classique de synthèse non ribosomique par les NRPS, la synthèse de l'érythrocheline implique deux clusters de gènes (*nrps5* et *nrps1*) codant chacun une NRPS<sup>313</sup> (Figure 32). La NRPS de *nrps1* contient une mutation qui la rend inactive. Le substrat de la NRPS EcrD de *nrps5* est synthétisé par deux enzymes, l'une codée dans *nrps5* et l'autre dans *nrps1*. En effet, la monooxygénase dépendante de la flavine ErcB, située dans le cluster *nrps5*, réalise la  $\delta$ -*N*-hydroxylation de la L-ornithine. La Mcd, située dans le cluster *nrps1*, catalyse la décarboxylation du malonyl-CoA pour donner l'acétyl-CoA. Puis elle transfère cet acétyle à la  $\delta$ -*N*-hydroxyl-L-ornithine pour

donner le L- $\delta$ -acétyl-  $\delta$ -N-hydroxyornithine, substrat de la NRPS ErcD. EcrD synthétise alors un térapeptide et libère l'érythrocheline après cyclisation intramoléculaire entre les résidus 3 et 4 pour donner le cycle dicétopipérazine<sup>313</sup>.

Ainsi cette synthèse fait intervenir de manière imbriquée des enzymes appartenant à des clusters éloignés sur le génome.

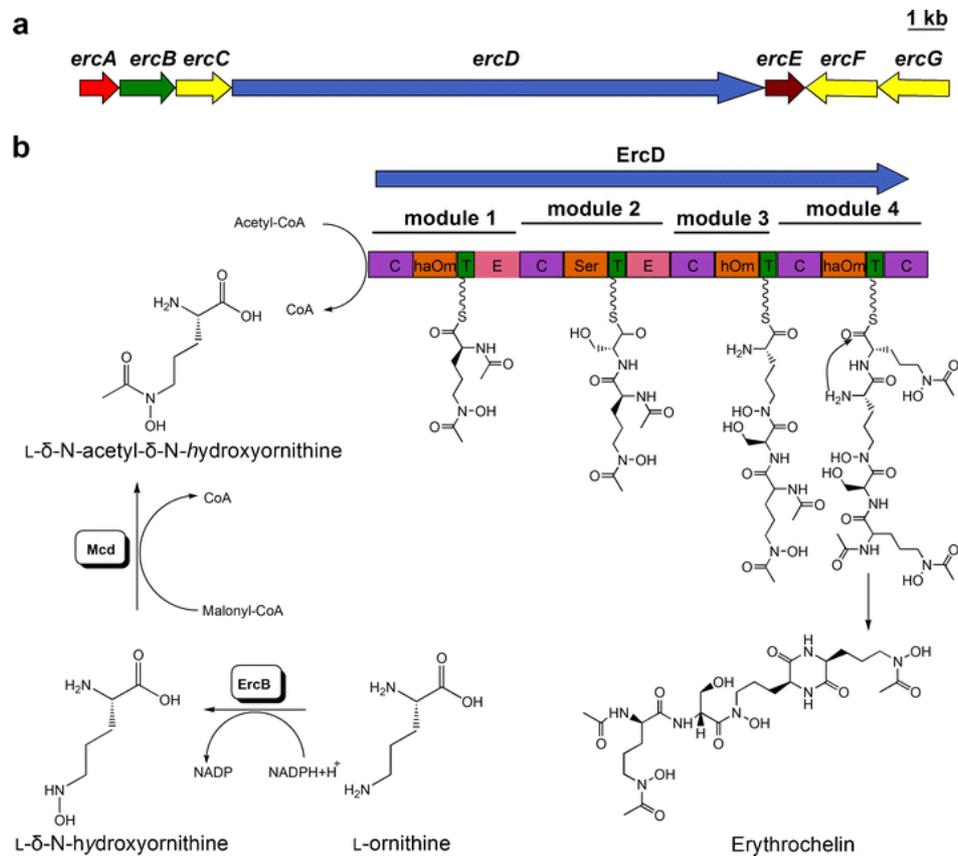


Figure 32 : Biosynthèse de l'érythrocheline

a) Cluster de gènes pour la voie de biosynthèse de la érythrocheline chez *S. erythraea*. b) Mécanisme proposé pour la voie biosynthèse de l'érythrocheline. Reproduit de<sup>324</sup>

## 2.5 La biosynthèse de l'ergopeptine résulte de l'action conjointe de deux NRPS

Les ergopeptines sont des alcaloïdes, isolés de *Claviceps purpurea*, qui présentent une structure tripeptidique reliée par une liaison amide au carbonyle de l'acide D-lysergique. (Figure 33). Certaines ergopeptines comme l'ergotamine et la dihydroergotamine sont utilisées comme traitement de la migraine aiguë et de l'algie vasculaire de la face. Leur synthèse nécessite l'utilisation de deux NRPS situées sur le même cluster : LPS1 et LPS2. La LPS2 serait responsable de l'activation de l'acide D-lysergique et de sa présentation par l'intermédiaire de son domaine T, afin qu'il puisse être pris en charge par la LPS1. La LPS1, ayant activé chacun des trois acides aminés nécessaires (Ala, Phe et Pro) grâce à ses trois modules, synthétise le tripeptide relié à l'acide D-lysergique<sup>314,339-341</sup>. Le composé contenant le cycle dicétopipérazine

est libéré sous forme de lactame puis une hétérocyclisation, vraisemblablement catalysée par un P450, est alors nécessaire pour obtenir l'ergopeptine finale<sup>342,343</sup>.

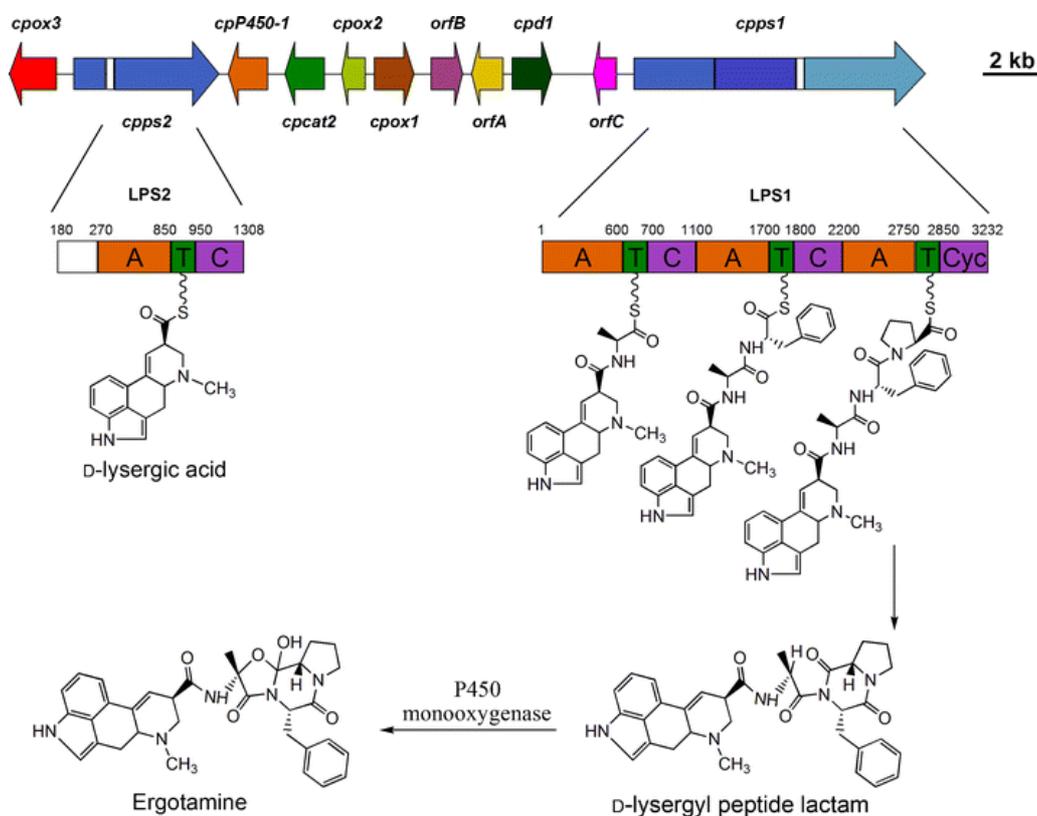


Figure 33 : Biosynthèse de l'ergopeptine

Cluster de gènes pour la voie de biosynthèse de l'ergopeptine chez *C. purpurea* P1 ; Mécanisme proposé pour la voie de biosynthèse de l'ergopeptine. Reproduit de<sup>324</sup>

## 2.6 La biosynthèse de dicétopipérazines comme produits secondaires

Certaines dicétopipérazines sont formées par des NRPS dédiées à la synthèse de peptides plus longs. La formation du cyclodipeptide résulterait dans ces cas d'une cyclisation intramoléculaire spontanée d'un peptide linéaire en cours de formation. Par exemple, le cyclo(D-Phe-Pro) et la cyclomarazine A sont produits respectivement par des NRPS dédiées à la synthèse du décapeptide tyrocidine A<sup>319</sup> et de la cyclomarazine A<sup>320</sup>. Dans le cas du cyclo(D-Phe-Pro), l'intermédiaire D-Phe-Pro lié au second module est intrinsèquement instable et se cyclise naturellement. Dans le cas de la cyclomarazine A, le clivage depuis la NRPS serait facilité par une thioestérase de type II<sup>344</sup>.

Ainsi, les dicétopipérazines obtenues par les NRPS peuvent résulter d'une synthèse dédiée ou représenter un produit secondaire obtenu spontanément ou non lors de la synthèse de peptides non ribosomiques plus longs. Une ou plusieurs NRPS sont responsables de l'obtention du cycle

dicétopipérazine tandis que des enzymes de modifications, regroupées ou non dans le même cluster, apportent des modifications (nitration, oxydation, méthylation,...) avant ou après la formation du cycle. Les voies à NRPS produisent des dicétopipérazines de taille et de complexité variées et incorporent des acides aminés protéinogéniques ou non protéinogéniques. La synthèse d'un cycle dicétopipérazine nécessite au moins deux modules de NRPS, soit une enzyme d'environ 200 kDa.

Or, une autre voie du métabolisme secondaire, utilisant des enzymes de taille bien inférieure, permet également de synthétiser des dicétopipérazines. La première occurrence de ce type de voie a été mise en évidence au laboratoire en 2002 pour la synthèse de l'albonoursine. Ces voies utilisent des enzymes appartenant à une nouvelle famille : les synthases de cyclodipeptides ou CDPS. Je vais dans un premier temps présenter les connaissances acquises sur les CDPS depuis leur découverte avant de présenter les voies à CDPS connues dans la littérature au début de mon travail.

## **V. Les voies de synthèse de dicétopipérazines dépendant des synthases de cyclodipeptides ou CDPS**

### **1. Les synthases de cyclodipeptides : mise en évidence et caractérisation biochimique et structurale**

L'albonoursine est une dicétopipérazine produite par *Streptomyces noursei* et *Streptomyces albus*. Isolée au début des années 1960, sa voie de biosynthèse est restée pendant longtemps inconnue et a été finalement élucidée en 2002, au sein du laboratoire (cf. V.2.1 Voies de biosynthèse de l'albonoursine). Elle contient le premier membre de la famille qui deviendra la famille des CDPS. Il s'agit de AlbC, une enzyme de petite taille (239 acides aminés, 26,8 kDa), qui est à elle seule responsable de la synthèse de deux cyclodipeptides majoritaires : le cyclo (Phe-Leu), intermédiaire de l'albonoursine, et le cyclo(Phe-Phe). La séquence de cette enzyme présente une similitude modérée avec uniquement deux séquences présentes dans les bases de données au moment de l'étude (Rv2275 de *M. tuberculosis* (34 % d'identité) et YvmC de *Bacillus subtilis* (29% d'identité). Elle n'est pas apparentée au NRPS et ne présente pas de motif connu de liaison de l'ATP<sup>25</sup>. Des études ultérieures ont permis d'identifier les substrats, de déterminer la structure, de proposer un mécanisme catalytique et de postuler les bases de la spécificité pour cette famille d'enzyme.

#### **1.1 Les CDPS, une nouvelle famille d'enzymes qui synthétisent principalement des cyclodipeptides hydrophobes**

Au début de mon projet de thèse, l'activité de synthèse de cyclodipeptides de neuf CDPS dont AlbC était connue. En effet, l'équipe de Muriel Gondry, en collaboration avec celle de Jean-Luc Pernodet (I2BC, Orsay), avait identifié dans les bases de données génomiques huit protéines présentant une similitude modérée avec AlbC ainsi qu'une taille similaire et avait caractérisé leurs activités<sup>26,27</sup>.

Huit proviennent de différents phyla bactériens : Firmicutes (*Bacillus*, *Staphylococcus*), Actinobacteria (*Corynebacterium*, *Mycobacterium*) et  $\gamma$ -Proteobacteria (*Photobabudus*)<sup>26</sup>. La neuvième est retrouvée chez un animal primitif, l'anémone de mer *Nematostella vectensis*<sup>27</sup>. Un alignement des séquences de ces protéines et de AlbC révèle que seulement sept résidus sont strictement conservés : Gly35, Ser37, Gly79, Tyr128, Tyr178, Glu182, Tyr202 (numérotation selon la séquence de AlbC). De plus, aucune séquence consensus n'émerge<sup>345</sup>.

Lorsque AlbC est exprimée de manière recombinante chez *E. coli*, elle synthétise des cyclodipeptides qui sont retrouvés dans les surnageants de culture<sup>25</sup>. Chacune des autres CDPS a été exprimée chez *E. coli* et les surnageants de culture ont été analysés par chromatographie liquide couplée à la spectrométrie de masse (LC-MS/MS). Chaque enzyme produit majoritairement un cyclodipeptide et plusieurs cyclodipeptides minoritaires révélant la spécificité relâchée de ces enzymes. Dans chaque cas, un acide aminé semble être incorporé de manière préférentielle et permet de définir un profil de cyclodipeptides produits (Tableau 3). Il est important de noter que les cyclodipeptides formés sont constitués d'acides aminés hydrophobes tels que Phe, Tyr, Leu et Trp.

Au cours de mon projet de thèse, trois autres CDPS provenant d'actinomyces (*Actinosynema* et *Nocardiopsis*) ont été identifiées par deux autres équipes<sup>346–348</sup>. Elles synthétisent également des cyclodipeptides constitués d'acides aminés hydrophobes et l'une d'entre elles (Amir\_4627) produit un unique cyclodipeptide (le cyclo(Trp-Trp)), ce qui en faisait la seule CDPS à ne pas présenter de spécificité relâchée (Tableau 3).

**Tableau 3 : Cyclodipeptides synthétisés par *E. coli* recombinant exprimant les CDPS**

CDPS	Organisme	Cyclodipeptides produits <sup>1</sup>	Produits préférentiels
AlbC	<i>Sreptomycetes noursei</i>	cFL (54), cFY (20), cFF (19), cFM (10), cYM (6), cYL (5), cLM (4), cLL (2), cMM (1), cYA (<1), cVL (<1), cYY (<1)	cFX
Rv2275	<i>Mycobacterium tuberculosis</i>	cYY (29), cYF (2), cYL (<0.4), cYA (<0.3), cYM (0.1), cYW (0.1)	cYX
YvmC-Blic	<i>Bacillus licheniformis</i>	cLL (71), cLM (6), cLF (4), cXX (1)	cLX
YvmC-Bsub	<i>Bacillus subtilis</i>	cLL (97), cLF (7), cLM (5), cXX (0.9)	cLX
YvmC-Bthu	<i>Bacillus thuringiensis</i>	cLL (60), cLF (8), cLM (3), cFF (1), cXX (0.2)	cLX
pSHaeC06	<i>Staphylococcus haemolyticus</i>	cLL (118), cLM (10), cLF (8), cXX (0.5)	cLX
Plu0297	<i>Photobabudus luminescens</i>	cLL (17), cLF (<1), cLM (<1)	cLX
Jk0923	<i>Corynebacterium jeikeium</i>	cLL (0.5), cLM (0.2)	cLX
Nvec-CDPS2	<i>Nematostella vectensis</i>	cWF (1), cWL (1), cWM (1), cFF (0.1–0.5), cFM (0.1–0.5), cFA (0.1–0.5), cYL (0.1–0.5), cYF (0.1–0.5), cYM (0.1–0.5), cLA (0.1–0.5), cWG (<0.1), cFL (<0.1)	cWX
Ndas_1148 <sup>2</sup>	<i>Nocardiopsis dassonvillei</i>	cFY(34), cFF(30), cFL(6), cYL, cYY, cFA, cFM	cFX
Amir_4627	<i>Actinosynema mirum</i>	cWW(29)	cWW

<sup>1</sup>Cyclodipeptides indiqués dans le code à une lettre, X représentant un acide aminé variable. Lorsque la donnée était disponible, la quantité de cyclodipeptide produit en mg par litre de culture est indiquée entre parenthèse. <sup>2</sup>La CDPS de *N. alba*, homologue à Ndas\_1148 (62% d'identité de séquence) n'apparaît pas dans ce tableau car son activité chez *E. coli* recombinante n'a pas été étudiée<sup>347</sup>.

## 1.2 Les CDPS utilisent les aminoacyl-ARNt comme substrats

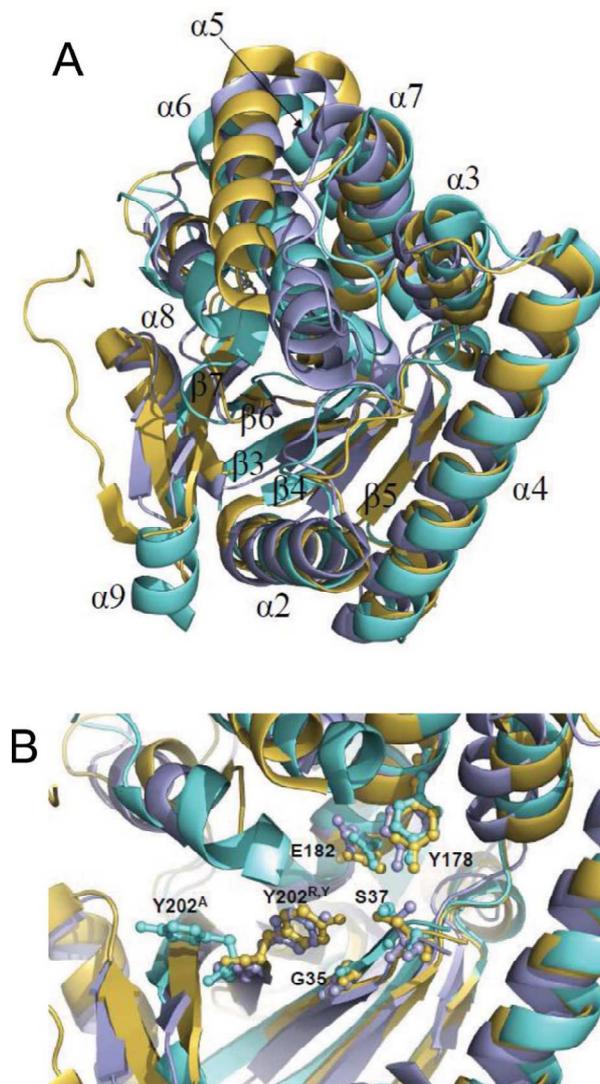
Au sein du laboratoire, la CDPS de AlbC de *S. noursei* a été utilisée comme modèle pour étudier la famille des CDPS et déterminer, dans un premier temps, ses caractéristiques biochimiques. La question qui se posait alors était la suivante : comment AlbC, une enzyme d'une aussi petite taille, peut-elle catalyser la formation de deux liaisons peptidiques alors que cela nécessiterait chez les NRPS l'utilisation d'une enzyme constituée d'au moins deux modules et donc de taille bien plus importante (pratiquement dix fois plus importante) ? Par ailleurs, tous les processus biologiques connus de formation de la liaison peptidique passent par l'activation des acides aminés comme cela se voit avec les NRPS et la formation d'acides aminés adénylés. Dans le cas des CDPS, aucune séquence consensus de liaison d'un éventuel cofacteur pour activer un acide aminé *via* un mécanisme connu ne peut être identifiée. Il se pose donc également la question de l'activation de ces acides aminés par les CDPS.

Il a été montré par des approches *in vitro* que les substrats utilisés par AlbC sont des acides aminés chargés sur des ARNt (aa-ARNt) et que l'ATP n'est pas nécessaire à la réaction. Les aa-ARNt étant des formes déjà activées d'acides aminés, l'enzyme n'a pas besoin de réaliser l'activation du substrat elle-même. Ceci explique également pourquoi aucune zone de liaison à l'ATP n'avait été identifiée.<sup>26</sup> L'utilisation de substrats aa-ARNt a été montrée pour tous les autres membres de la famille identifiés au laboratoire<sup>26,27</sup>. Cette caractéristique de la famille la place à l'intersection entre le métabolisme primaire et le métabolisme secondaire<sup>349</sup> et vient documenter les rares familles d'enzymes connues qui utilisent des aa-ARNt comme substrats pour divers processus biochimiques<sup>291,350,351</sup>. Brièvement, il s'agit des glutamyl réductases pour la biosynthèse de porphyrine<sup>352</sup>, des aminoacyl-phosphatidylglycérol transférases MprF pour l'amination de phospholipides dans la membrane cellulaire<sup>353</sup> et de VlmA, un homologue de MprF, pour la biosynthèse de l'antibiotique valanimycine<sup>354</sup>. On retrouve également des enzymes catalysant la formation de liaisons peptidiques : les aminoacyl-ARNt-protéine transférases qui modifient la partie N-terminale des protéines pour leur dégradation subséquente, les aminoacyl-transférases de type Fem qui interviennent dans la biosynthèse du peptidoglycane<sup>355,356</sup> et PacB qui est impliquée dans la formation de l'antibiotique pacidamycine<sup>357</sup>. Toutes ces enzymes qui forment des liaisons peptidiques appartiennent à la même famille structurale des *N*-acétyl-transférases caractérisées par un repliement de type GNAT (GCN5-related *N*-acetyltransferase)<sup>356-360</sup> mais ce n'est pas le cas pour les CDPS, comme indiqué ci-dessous.

## 1.3 Les CDPS possèdent une structure réminiscente des aminoacyl-ARNt synthétases

Les structures cristallographiques de trois CDPS ont été publiées : celles de AlbC (*Streptomyces noursei*) à une résolution de 1,9 Å<sup>361</sup>, de Rv2275 (*Mycobacterium tuberculosis*) à 2,0 Å<sup>362</sup>, et de YvmC-Blic (*Bacillus subtilis*) à 1,7-2,8 Å<sup>363</sup>. Malgré une faible identité de séquence (environ 27 %), ces trois protéines possèdent une forte homologie structurale, avec un repliement commun de type  $\alpha/\beta$  compact. Le feuillet  $\beta$  central est constitué de cinq brins  $\beta$  parallèles ( $\beta$ 3- $\beta$ 7) reliés par des hélices  $\alpha$  ( $\alpha$ 2- $\alpha$ 9). Les brins  $\beta$ 3- $\beta$ 5, l'hélice  $\alpha$ 2 et la partie C-terminale de l'hélice  $\alpha$ 4 forment un repliement de type Rossmann qui, de même

que les deux brins supplémentaires ( $\beta 6$ - $\beta 7$ ), se superpose très bien dans les trois structures<sup>345</sup> (Figure 34A.). Les trois protéines possèdent une poche similaire, nommée P1, qui est accessible depuis la surface et est délimitée par les hélices  $\alpha 6$ - $\alpha 8$  et les brins  $\beta 3$ - $\beta 4$  et  $\beta 6$ <sup>361</sup>. Cinq des sept acides aminés strictement conservés chez les CDPS (Gly35, Ser37, Tyr178, Glu182 et Tyr202, numérotation de AlbC) sont positionnés à l'entrée de cette poche<sup>345</sup>, tandis que le fond de la poche est constitué de résidus hydrophobes non conservés<sup>362,363</sup>(Figure 34B). Ces protéines présentent également une zone riche en résidus chargés positivement, majoritairement positionnée sur l'hélice  $\alpha 4$ , que l'on appellera « patch basique ». Plusieurs données suggèrent qu'elle pourrait être une des zones d'interaction avec l'ARNt<sup>361,363</sup>.



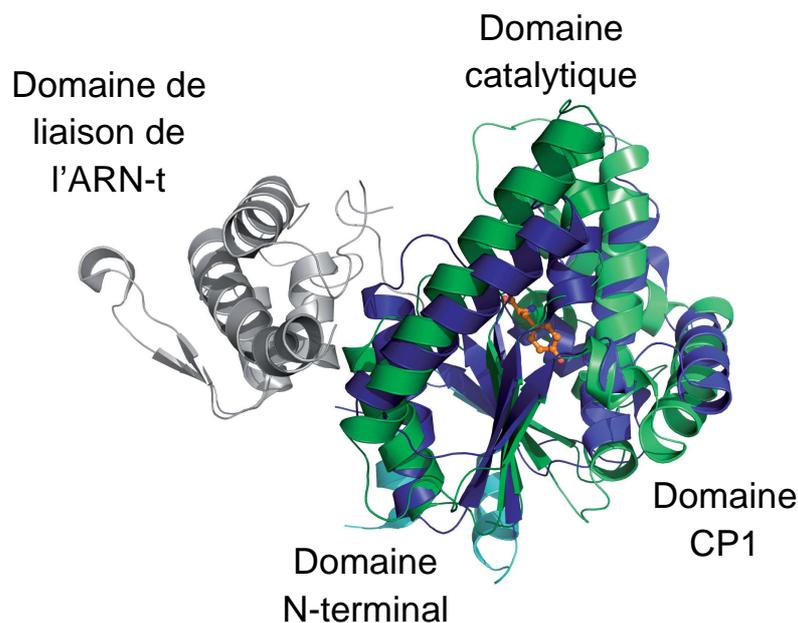
**Figure 34 : Comparaison des trois structures cristallographiques de CDPS**

**A. Superposition de la structure d'AlbC (bleu), Rv2275 (jaune) et YvmC (violet). La numérotation des hélices  $\alpha$  et des feuillets  $\beta$  a été reportée. B. Superposition des résidus conservés dans la poche et numérotés selon la numérotation d'AlbC. Reproduit de<sup>345</sup>**

AlbC, Rv2275 et YvmC-Blic présentent une forte homologie structurale avec le domaine catalytique des aminoacyl-ARNt synthétases (aaRS) de classe I, en particulier celui des tyrosyl-ARNt synthétases (TyrRS) et tryptophanyl-ARNt synthétases (TrpRS) (classe Ic) malgré une faible identité de

séquence avec ces dernières (moins de 16%)<sup>345,361–363</sup>. Les aaRS catalysent la formation des aa-ARNt suivant un processus en deux étapes : tout d’abord, elles activent les acides aminés *via* l’utilisation d’ATP pour former des intermédiaires aminoacyl-adenylates, puis elles transfèrent ces intermédiaires sur les ARNt qui leur sont dédiés. Ces enzymes possèdent une architecture modulaire, avec un domaine catalytique qui est le siège de la réaction d’aminocyclation et un domaine de reconnaissance de l’anticodon (ou domaine de liaison à l’ARNt).

Les aaRS de classe Ic possèdent un domaine catalytique présentant un repliement de type Rossmann et un domaine CP1 (connective polypeptide domaine 1) qui sont tous les deux retrouvés chez les CDPS. Par ailleurs, la poche de liaison de l’acide aminé substrat des TyrRS et TrpRS est très similaire à la poche P1 retrouvée chez les CDPS. Cependant, des différences notables existent entre les deux familles d’enzymes. Ainsi, les CDPS ne possèdent ni domaine de reconnaissance de l’anticodon ni domaine de liaison à l’ATP (Figure 29). De plus, les aminoacyl-ARNt synthétases de classe Ic sont organisées en homodimères alors qu’AlbC et YvmC-Blic sont actives en tant que monomères<sup>361,363</sup>. Quant à Rv2275, il pourrait s’organiser en homodimère mais son mode de dimérisation serait différent de celui des synthétases précitées et cette dimérisation ne serait pas requise pour la catalyse<sup>362</sup>. Il est à noter également que les aaRS synthétases de classe I ne possèdent pas le « patch basique » retrouvés chez les CDPS.



**Figure 35 : Superposition de la structure d’AlbC et de la tyrosyl-ARNt synthétase de *Methanococcus jannaschii* (TyrRS<sub>Mj</sub>) en complexe avec son substrat L-tyrosine. Reproduit de<sup>361</sup>**

Sur le plan phylogénétique, les CDPS sont des membres de la famille des protéines à domaines « rossmanoïdes » de superclasse HUP, qui ont probablement évolué des TyrRS ou des TrpRS<sup>28</sup>. Au cours de l’évolution, les CDPS auraient perdu le domaine de liaison à l’ARNt, ainsi que la capacité à activer les

acides aminés. En revanche, elles ont acquis de nouveaux résidus catalytiques leur permettant d'utiliser les produits de leurs ancêtres comme substrats pour la synthèse de cyclodipeptides<sup>345</sup>.

#### 1.4 Les CDPS utilisent un mécanisme catalytique de type ping-pong

Plusieurs équipes ont travaillé à l'élucidation du mécanisme réactionnel intervenant dans la synthèse des cyclodipeptides par les CDPS, ce qui a permis finalement de proposer le mécanisme global présenté (Figure 36)

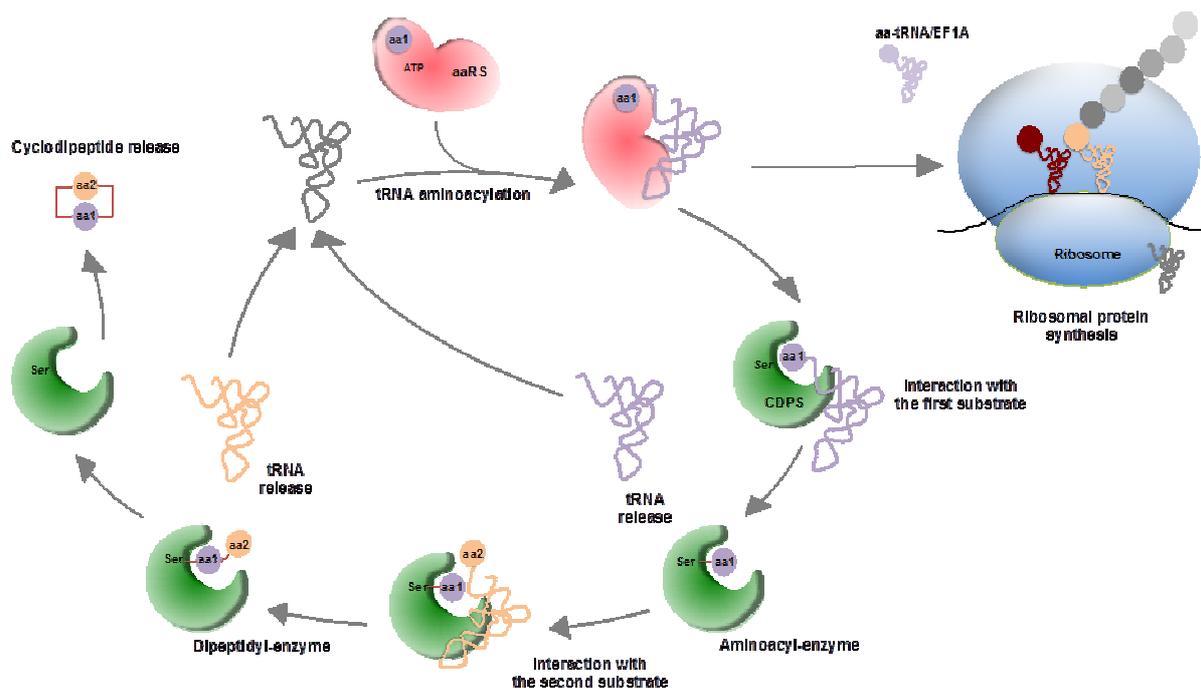


Figure 36 : Etapes clés du mécanisme catalytique menant à la synthèse d'un cyclodipeptide

Les CDPS détournent les aa-ARNt synthétisés par les aminoacyl-ARNt synthétases de leur rôle canonique dans la synthèse des protéines. En effet, elles vont utiliser deux aminoacyl-ARNt comme substrats pour former un cyclodipeptide. Les CDPS vont tout d'abord interagir avec un premier aa-tRNA. Pour AlbC, Rv2275, YvmC-Blic et Nvec-CDPS2, il a été supposé, grâce à des expériences de mutagenèse dirigée et des analyses utilisant des substrats radioactifs, que l'entité aminoacyle de ce substrat est transférée sur une sérine conservée (en position 37 chez AlbC), ce qui conduit à la formation d'une aminoacyl-enzyme<sup>27,361-363</sup>. L'existence de cette aminoacyl-enzyme a été confirmée pour AlbC et Nvec-CDPS2 grâce à une analyse par spectrométrie de masse de fragments de digestion trypsique des enzymes ayant été au préalable incubées pendant des temps courts avec leurs substrats. Puis l' aminoacyl-enzyme interagit avec le second aa-ARNt et l'entité aminoacyle de ce second substrat est transférée sur l' aminoacyl-enzyme pour donner une dipeptidyl-enzyme<sup>364</sup>. Finalement, une cyclisation intramolécule se produisant sur l'intermédiaire dipeptidyle conduit à la formation du cyclodipeptide.

Récemment, une étude menée au laboratoire a permis d'identifier cet intermédiaire dipeptidyl-enzyme mais aussi de préciser le mécanisme catalytique. Un tel intermédiaire étant particulièrement instable, il semblait raisonnable de tenter de bloquer la réaction avant la cyclisation du dipeptide. Afin d'identifier les résidus impliqués dans la cyclisation, la structure cristallographique d'un complexe mimant cet intermédiaire a tout d'abord été déterminée (Figure 37).

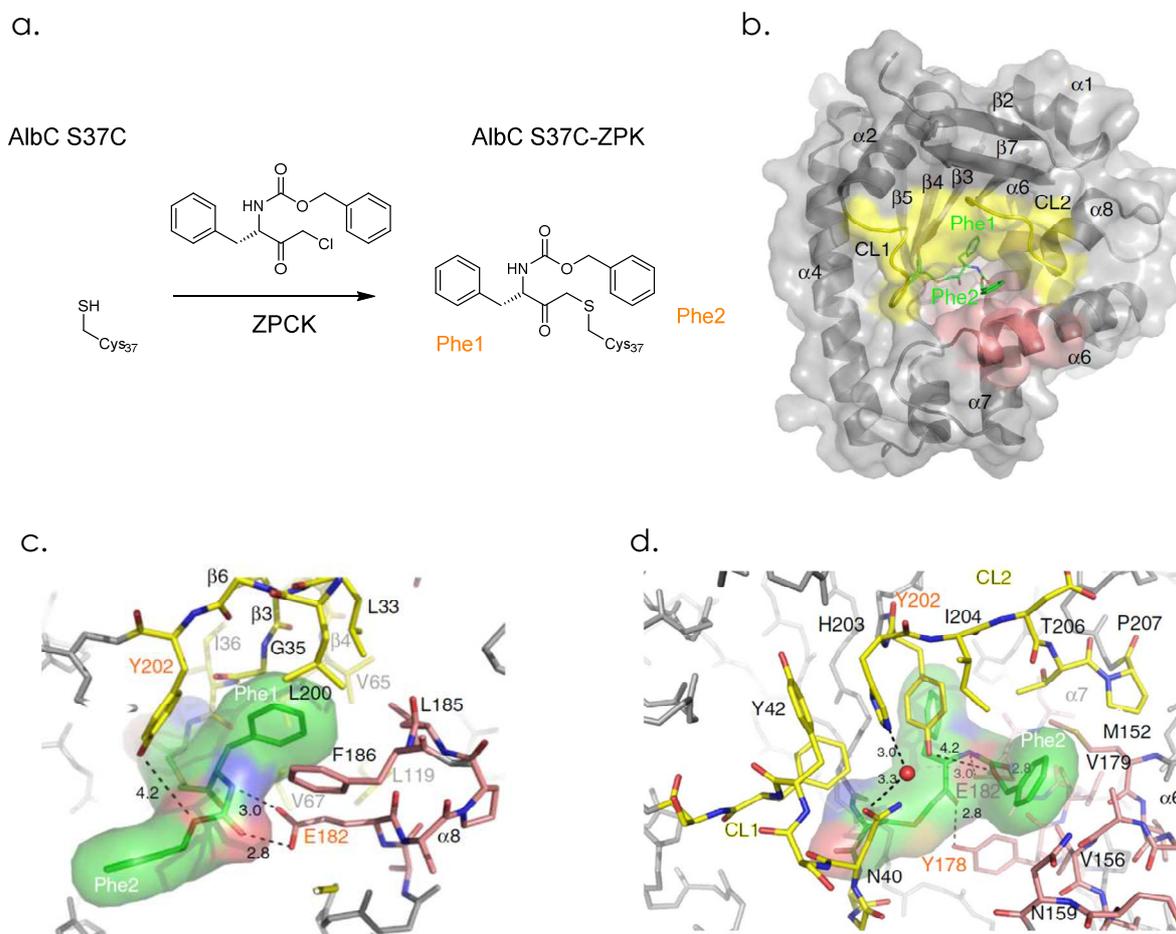


Figure 37 : Réaction du ZPCK sur le variant S37C d'AlbC et structure cristallographique du complexe mimant l'intermédiaire dipeptidyl-enzyme

a. Le ZPCK (*N*-carbobenzyloxy-*L*-Phe chlorométhyl cétone) réagit avec la cystéine 37 du variant S37C d'AlbC pour donner l'enzyme marquée appelée AlbC S37C-ZPK. b. Structure complète du mime de l'intermédiaire dipeptidyl-enzyme. c. Zoom de la poche P1 contenant Phe du premier substrat (Phe1) en représentation bâton. d. Zoom de la poche plus large P2 contenant Phe du deuxième substrat (Phe2) en représentation bâton. Les éléments de structures secondaires sont indiqués. La protéine est en gris, les régions autour des poches sont colorées CL1 et CL2 (CL1: 33–43, CL2: 202–207), l'extrémité N-terminale de  $\beta$ 3,  $\beta$ 4,  $\beta$ 6 sont en jaune. Les résidus 151 à 159 de  $\alpha$ 6 et 172-182 de  $\alpha$ 7 sont en rose. Le ZPK (ZPCK lié) est représenté en vert, les résidus catalytiques sont indiqués en orange. Reproduit de<sup>365</sup>

Le complexe a été obtenu en faisant réagir le ZPCK (*N*-carbobenzyloxy-*L*-Phe chlorométhyl cétone), un analogue du dipeptide Phe-Phe, avec un variant d'AlbC dans lequel la sérine catalytique est remplacée par une cystéine (S37C). Ce variant conserve l'aptitude à former une acyl-enzyme mais celle-ci est beaucoup moins réactive. La structure d'AlbC S37C-ZPK (*N*-carbobenzyloxy-*L*-Phe méthyl cétone), mime du dipeptidyl-enzyme AlbC-*L*-Phe-*L*-Phe, a été résolue à 2,9 Å. Elle montre un groupement benzyle

du ZPK, mime de la chaîne latérale de Phe1 du dipeptide, ancré dans la poche catalytique P1 d'AlbC. Le second groupement benzyle du ZPK, mime de la chaîne latérale de Phe2 du dipeptidyle, occupe une seconde poche plus large, nommée P2, qui n'est pas visible dans la structure tridimensionnelle d'AlbC seule. Cette poche P2 est délimitée par deux boucles contenant les résidus catalytiques (résidus 36-43 et 202-207), l'extrémité C-terminale de  $\alpha 6$  et la partie centrale de  $\alpha 7$ . A partir de cette structure, un modèle du dipeptidyl-AlbC a pu être obtenu et utilisé pour identifier les résidus impliqués dans la cyclisation du dipeptide. Ainsi, le résidu conservé Tyr202 semble pouvoir jouer le rôle de base pour déprotoner l'amine primaire de Phe2 afin qu'elle réalise une attaque nucléophile sur l'ester activé<sup>366</sup> pour cycliser le dipeptide (Figure 37). Pour ralentir cette étape, Tyr202 a été muté en Phe. Le variant AlbC Y202F a alors été incubé en présence de Phe-ARNt<sup>Phe</sup>, soumis à une trypsinolyse et analysé par spectrométrie de masse. Des fragments du dipeptidyl-enzyme dans lesquels Ser37 apparaît avec un adduit Phe-Phe ont été identifiés, montrant que le mécanisme passe par un dipeptidyl-enzyme<sup>364</sup>. De plus, ce résultat confirmerait le rôle du résidu Tyr202 dans la cyclisation car l'adduit Phe-Phe n'est pas observé pour AlbC sauvage dans les mêmes conditions expérimentales.

La structure d'AlbC S37C-ZPK a également permis de confirmer et/ou de revisiter le rôle des autres résidus catalytiques et de proposer le mécanisme présenté (Figure 38). Quand le premier substrat se lie, l'entité aminoacyl Phe1 est accommodée dans la poche P1 avec ses groupements amino et carbonyle correctement positionnés par des liaisons hydrogène avec les chaînes latérales des résidus conservés Asp182 et Tyr178 (étape 1). L'hydroxyle de la Tyr178 polarise le carbonyle de l'ester du substrat et facilite ainsi l'attaque nucléophile par le résidu conservé Ser37. L'activation de la Ser37 résulte probablement d'un mécanisme concerté de transfert de protons impliquant les deux hydroxyles adjacents du nucléotide A76 de l'entité ARNt, comme cela a été décrit pour l' aminoacyl-transférase FemX<sup>360</sup>. L' aminoacyl-enzyme est ainsi formée et l'hydroxyle de Tyr178 pourrait aider à sa stabilisation (étape 2). Puis, le second substrat se lie ; le groupe ammonium de Phe1 déprotoné par le carboxyle de Asp182 peut alors attaquer le groupe carbonyle de l'entité Phe2 du deuxième substrat (étape 3), ce qui conduit à la formation de l'intermédiaire dipeptidyl-enzyme (étape 4). Le résidu conservé Tyr202 est alors bien positionné - grâce à un réseau de liaisons hydrogène impliquant les résidus essentiels Asn40 et His203 - pour déprotoner le groupe ammonium de Phe2 qui peut alors attaquer la liaison ester de l'enzyme, ce qui conduit à la formation du cyclodipeptide (étape 5).

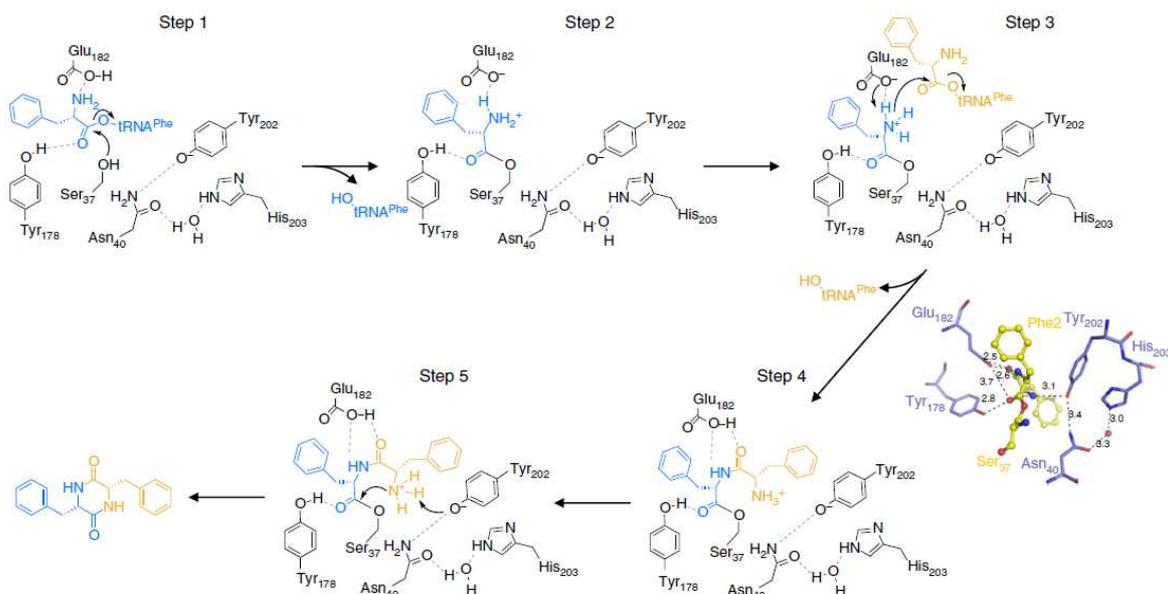


Figure 38 : Mécanisme réactionnel proposé pour la formation d'un cyclodipeptide

Le mécanisme menant à la formation du cyclo(Phe-Phe) catalysé par AlbC est proposé. L'acide aminé du premier substrat est en bleu, celui du deuxième en jaune. Les interactions possibles sont indiquées en pointillés.

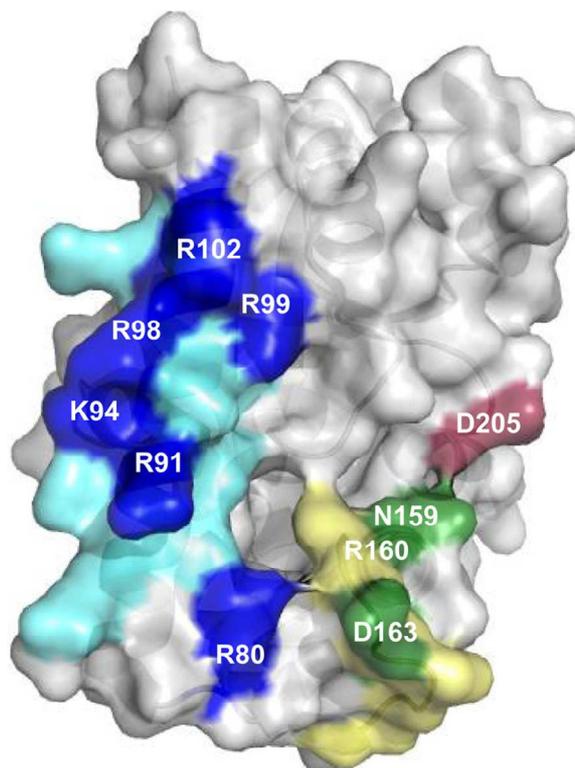
En ce qui concerne le mécanisme chez les autres CDPS, seule la formation de l'acyl-enzyme a été identifiée expérimentalement, mais la conservation de la structure tridimensionnelle chez les trois CDPS de structure connue et des résidus catalytiques permet d'envisager un mécanisme similaire.

### 1.5 Les bases moléculaires de l'interaction entre les CDPS et leurs substrats

Si le mécanisme des CDPS est presque entièrement élucidé, certaines questions restent encore en suspens notamment concernant les sites de liaison des substrats aux CDPS et les déterminants de la spécificité des CDPS.

Les substrats des CDPS sont complexes : ils possèdent une entité aminoacyle et une entité ARNt. Beaucoup de questions se posent quant au rôle de chaque entité dans la liaison à la CDPS, et ce pour le premier et le second substrat. Il apparaît tout d'abord que les CDPS reconnaissent différemment chacun des deux substrats et possèdent un site de liaison spécifique pour chacun d'entre eux, comme le montrent deux études récentes menées au laboratoire sur la CDPS AlbC<sup>364,365</sup>. AlbC synthétise majoritairement un hétérocyclodipeptide, le cyclo(Phe-Leu), ce qui offre la possibilité de distinguer la liaison de chacun des deux substrats. Il a tout d'abord été montré que le Phe-ARNt<sup>Phe</sup> est utilisé comme premier substrat et le Leu-ARNt<sup>Leu</sup> comme deuxième substrat. En effet, la formation de l'intermédiaire aminoacyl-enzyme qui résulte du transfert de l'entité aminoacyle du premier substrat n'a pu être mise en évidence qu'avec le Phe-ARNt<sup>Phe</sup>. Par ailleurs, l'utilisation de variants de AlbC lors d'expériences de mesure d'activité *in vitro* de

synthèse de cyclodipeptides et de détection de la formation d'aminoacyl-enzymes a permis de montrer que l'entité ARNt du premier substrat interagit avec le « patch basique » situé essentiellement sur l'hélice  $\alpha 4$ , alors que celle du second substrat implique des résidus situés sur les boucles  $\alpha 6$ - $\alpha 7$  et  $\beta 6$ - $\alpha 8$  (Figure 39). Quant aux entités aminoacyles des deux substrats, elles interagissent chacune avec une poche différente, à savoir la poche P1 pour le premier substrat et la poche P2 pour le deuxième substrat, poches qui sont toutes deux bien définies dans la structure AlbC S37C-ZPK (Figure 37). La combinaison de ces deux études permet ainsi d'avoir une vision globale de l'interaction de chacun des substrats.



**Figure 39 : Régions d'AlbC impliquées dans la liaison à l'ARNt**

La surface de la structure d'AlbC est colorée en gris. Le patch basique localisé sur l'hélice  $\alpha 4$  est coloré en bleu, la boucle  $\alpha 6$ - $\alpha 7$  est en vert, le résidu D205 de la boucle  $\beta 6$ - $\alpha 8$  est en rouge foncé. Les résidus dont l'implication dans l'interaction avec l'ARNt a été montrée par mutagenèse sont en foncé. Reproduit de<sup>365</sup>

L'utilisation de mutants de Phe-ARNt<sup>Phe</sup> et Leu-ARNt<sup>Leu</sup> comme substrats de AlbC a également permis de montrer que la spécificité de reconnaissance du premier substrat est essentiellement due à l'entité aminoacyle de celui-ci et non à son entité ARNt<sup>365</sup>. Cela avait déjà été montré avec la substitution par Asn d'un résidu du fond de poche, Leu200, qui avait fait passer le cyclodipeptide majoritairement produit de cyclo(Phe-Leu) à cyclo(Tyr-Leu)<sup>361</sup>, Asn étant le résidu présent à cette position chez Rv2275 qui produit du cyclo(Tyr-Tyr)<sup>362</sup>. Concernant le deuxième substrat, la spécificité de reconnaissance dépend de l'entité ARNt avec notamment la paire de base N1-N72 qui est un déterminant clé mais aussi de l'entité aminoacyle<sup>365</sup>.

Tableau 4 : Acides aminés accommodés par chacune des deux poches des CDPS caractérisées et les résidus constituant potentiellement ces deux poches. D'après<sup>364</sup>

CDPSs	Poche 1 (P1)									Poche 2 (P2)																		
	Acides aminés accommodés			Résidus constituant P1						Acides aminés accommodés			Résidus constituant P2															
AlbC	F	L	M	Y	L <sub>33</sub>	G <sub>35</sub>	V <sub>65</sub>	V <sub>67</sub>	L <sub>119</sub>	L <sub>185</sub>	F <sub>186</sub>	L <sub>200</sub>	F	Y	M	A	V	M <sub>152</sub>	V <sub>153</sub>	R <sub>154</sub>	A <sub>155</sub>	V <sub>156</sub>	V <sub>157</sub>	M <sub>158</sub>	N <sub>159</sub>	T <sub>204</sub>	T <sub>206</sub>	P <sub>207</sub>
Rv2275	Y				V <sub>84</sub>	G <sub>86</sub>	V <sub>116</sub>	T <sub>118</sub>	M <sub>170</sub>	L <sub>236</sub>	F <sub>237</sub>	N <sub>251</sub>	Y	F	L	M	A	L <sub>203</sub>	V <sub>204</sub>	R <sub>205</sub>	R <sub>206</sub>	F <sub>207</sub>	L <sub>208</sub>	S <sub>209</sub>	T <sub>210</sub>	Q <sub>255</sub>	L <sub>257</sub>	P <sub>258</sub>
YvmC-Blic	L				L <sub>33</sub>	G <sub>35</sub>	L <sub>65</sub>	A <sub>67</sub>	F <sub>117</sub>	F <sub>187</sub>	F <sub>188</sub>	L <sub>202</sub>	L	F	M			M <sub>150</sub>	S <sub>151</sub>	H <sub>152</sub>	A <sub>153</sub>	A <sub>154</sub>	I <sub>155</sub>	-	L <sub>156</sub>	R <sub>206</sub>	W <sub>208</sub>	K <sub>209</sub>
YvmC-Bsub	L				L <sub>33</sub>	G <sub>35</sub>	L <sub>65</sub>	A <sub>67</sub>	F <sub>117</sub>	F <sub>187</sub>	F <sub>188</sub>	L <sub>202</sub>	L	F	M			M <sub>150</sub>	S <sub>151</sub>	R <sub>152</sub>	E <sub>153</sub>	A <sub>154</sub>	I <sub>155</sub>	-	I <sub>156</sub>	R <sub>206</sub>	W <sub>208</sub>	K <sub>209</sub>
YvmC-Bthu	L				I <sub>33</sub>	G <sub>35</sub>	L <sub>65</sub>	A <sub>67</sub>	F <sub>117</sub>	F <sub>187</sub>	F <sub>188</sub>	L <sub>202</sub>	L	F	M			M <sub>150</sub>	S <sub>151</sub>	H <sub>152</sub>	A <sub>153</sub>	A <sub>154</sub>	I <sub>155</sub>	-	L <sub>156</sub>	K <sub>206</sub>	W <sub>208</sub>	E <sub>209</sub>
PSHaeC06	L				I <sub>24</sub>	G <sub>26</sub>	L <sub>56</sub>	A <sub>58</sub>	F <sub>109</sub>	F <sub>178</sub>	F <sub>179</sub>	M <sub>193</sub>	L	F	M			M <sub>142</sub>	S <sub>143</sub>	L <sub>144</sub>	Q <sub>145</sub>	A <sub>146</sub>	L <sub>147</sub>	Q <sub>148</sub>	S <sub>149</sub>	A <sub>197</sub>	W <sub>199</sub>	E <sub>200</sub>
Plu0297	L				L <sub>29</sub>	G <sub>31</sub>	L <sub>61</sub>	P <sub>63</sub>	F <sub>116</sub>	F <sub>186</sub>	Y <sub>187</sub>	L <sub>201</sub>	L	F	M			M <sub>149</sub>	S <sub>150</sub>	F <sub>151</sub>	Q <sub>152</sub>	A <sub>153</sub>	I <sub>154</sub>	-	-	R <sub>205</sub>	W <sub>207</sub>	S <sub>208</sub>
Jk0923	L				I <sub>10</sub>	G <sub>12</sub>	L <sub>42</sub>	P <sub>44</sub>	I <sub>96</sub>	A <sub>162</sub>	Y <sub>163</sub>	L <sub>177</sub>	L	F	M			M <sub>129</sub>	S <sub>130</sub>	R <sub>131</sub>	A <sub>132</sub>	A <sub>133</sub>	C <sub>134</sub>	Q <sub>135</sub>	-	T <sub>181</sub>	W <sub>183</sub>	P <sub>184</sub>
Amir_4627	W				V <sub>33</sub>	G <sub>35</sub>	V <sub>65</sub>	P <sub>67</sub>	M <sub>119</sub>	F <sub>184</sub>	F <sub>185</sub>	N <sub>199</sub>	W					M <sub>152</sub>	S <sub>153</sub>	R <sub>154</sub>	E <sub>155</sub>	V <sub>156</sub>	L <sub>157</sub>	A <sub>158</sub>	L <sub>159</sub>	R <sub>203</sub>	L <sub>205</sub>	P <sub>206</sub>
Ndas-1148	F	Y			L <sub>37</sub>	G <sub>39</sub>	L <sub>69</sub>	A <sub>71</sub>	L <sub>123</sub>	L <sub>185</sub>	L <sub>186</sub>	C <sub>200</sub>	Y	F	L	A	M	H <sub>156</sub>	V <sub>157</sub>	R <sub>158</sub>	H <sub>159</sub>	V <sub>160</sub>	I <sub>161</sub>	G <sub>162</sub>	A <sub>163</sub>	A <sub>204</sub>	M <sub>206</sub>	P <sub>207</sub>
Nvec2	W	F			L <sub>29</sub>	G <sub>31</sub>	F <sub>61</sub>	P <sub>63</sub>	P <sub>117</sub>	Y <sub>194</sub>	I <sub>195</sub>	F <sub>209</sub>	F	L	M	G	A	E <sub>150</sub>	V <sub>151</sub>	A <sub>152</sub>	R <sub>153</sub>	V <sub>154</sub>	L <sub>155</sub>	G <sub>156</sub>	K <sub>157</sub>	R <sub>213</sub>	W <sub>215</sub>	P <sub>216</sub>

Les cyclodipeptides synthétisés par AlbC, Rv2275, YvmC-Blic, YvmC-Bsub, YvmC-Bthu, PSHaeC06, Plu0297 et Jk0923 sont identifiés dans<sup>26</sup>. Ceux synthétisés par Amir\_4627, Ndas\_1148, et Nvec-CDPS2 sont identifiés dans <sup>27,346,348</sup>. L'acide aminé accommodé par la poche P1 correspond à l'acide aminé préférentiellement incorporé dans les cyclodipeptides ; la poche P2 peut accommoder plusieurs acides aminés, sauf pour Amir\_4627. Les acides aminés faiblement incorporés sont en gris. Les résidus constituant P1 et P2 sont déduits des structures cristallographiques d'AlbC S37C-ZPK (PDB ID, 4Q24), d'AlbC (PDB ID, 3OQV), de Rv2275(PDB ID, 2X9Q) et de YvmC(PDB ID, 3OQJ) et sont respectivement notés en rose et surligné en jaune.

La structure de AlbC S37C-ZPK (Figure 31) ayant permis d'identifier les poches P1 et P2 de liaison des entités aminoacyles qui sont donc essentielles pour la spécificité, il a été possible de comparer les résidus correspondants qui définissent les poches des CDPS caractérisées biochimiquement tout en tenant compte des cyclodipeptides produits (Tableau 4). Cette comparaison a permis de confirmer que la poche P1 a un caractère hydrophobe compatible avec l'accommodation de résidus hydrophobes et de montrer que la poche P2 contient des résidus peu conservés, ce qui pourrait expliquer la spécificité de substrat relâchée pour le deuxième résidu incorporé dans les cyclodipeptides. Cependant, cette comparaison n'a pas permis de trouver des règles pouvant régir la spécificité des CDPS.

Les CDPS sont donc des enzymes de petite taille qui catalysent la formation de deux liaisons peptidiques à partir de deux aminoacyl-ARNt selon un mécanisme ping-pong. Si les premières études sur la spécificité ont montré que les entités aminoacyle et ARNt jouent un rôle dans la reconnaissance des substrats, elles n'ont cependant pas permis de déchiffrer le code de spécificité des CDPS, qui, à l'instar du code non ribosomique, permettrait de prédire les cyclodipeptides synthétisés par une CDPS.

## 2. Voies de biosynthèse de dicétopipérazines dépendant des CDPS

Au début de mon étude, seulement trois voies dépendant de CDPS avaient été déchiffrées : les voies de biosynthèse de l'alboursine<sup>25</sup>, de l'acide pulcherriminique<sup>367</sup> et de la mycocyclosine<sup>29</sup>. Ces voies de biosynthèse impliquent une CDPS et une seule enzyme de modification du cyclodipeptide. Les gènes correspondants sont regroupés sur le génome, ce qui a facilité leur identification et la caractérisation de la voie. Malgré cette simplicité apparente, les produits finaux peuvent être assez inattendus, comme dans le cas de la mycocyclosine ou de la pulcherrimine du fait des activités des enzymes de modification constituant ces voies de biosynthèse.

### 2.1 Voies de biosynthèse de l'albonoursine

L'albonoursine (cyclo(déhydro-Leu-déhydro-Phe)) est une molécule aux propriétés antibiotiques qui est synthétisée par plusieurs souches de *Streptomyces*. L'albonoursine résulte d'une double déshydrogénation entre les carbones C $\alpha$  et C $\beta$  du cyclo(Phe-Leu)(Figure 40). Cette activité de déshydrogénation du cyclo(Phe-Leu) avait été identifiée et étudiée à partir d'extraits bactériens<sup>368-370</sup>. En 2001 et 2002, le groupe de gènes responsable de sa biosynthèse chez *Streptomyces noursei* a été identifié au laboratoire et l'activité des enzymes correspondantes a été caractérisée<sup>24,25</sup>. Ce travail a permis de découvrir le premier membre de la famille des CDPS, AlbC.

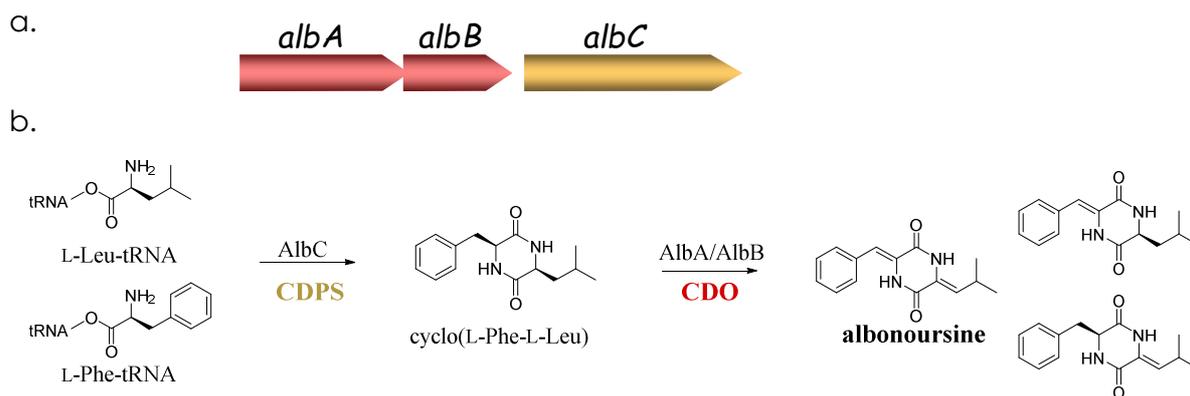


Figure 40 : Voie de biosynthèse de l'albonoursine

a. Cluster de gènes pour la voie de biosynthèse de l'albonoursine. b. Etapes de la synthèse de l'albonoursine

AlbC synthétise le cyclo(Phe-Leu), précurseur de l'albonoursine. Une seconde enzyme, codée par *albA* et *albB* et nommée oxydase de cyclodipeptide (CDO), est responsable de la déshydrogénation entre les carbones C $\alpha$  et C $\beta$  du cyclodipeptide pour donner l'albonoursine. La CDO est une enzyme multimérique de haute masse moléculaire, composée de 2 types de sous-unités AlbA (23kDa), une flavoprotéine, et AlbB (11 kDa) ; les deux sous-unités sont essentielles à l'activité CDO. AlbB est co-purifiée avec AlbA mais est retrouvée en quantité inférieure (un dixième). Les séquences de *albA* et *albB* se recouvrent sur 41 nucléotides, ce qui suggère qu'il existe une régulation durant la synthèse des deux sous-unités (Figure 40).

CDO catalyse tout d'abord la formation du cyclodipeptide en mono-déshydrodicétopipérazine, puis en bis-déshydrodicétopipérazine. *In vitro*, les deux configurations Z et E sont observées pour les doubles liaisons tandis que l'albonoursine isolée de culture de *S. noursei* ne contient que la configuration Z<sup>371</sup>. Un mécanisme de déshydrogénation, compatible avec l'apparition de forme Z et E, impliquerait le passage par une imine qui se réarrangerait pour donner le composé  $\alpha,\beta$ -déshydrogéné.

La spécificité de substrat de CDO a été testée *in vitro* en utilisant comme substrats des dipeptides linéaires et cycliques. Aucun des peptides linéaires testés n'est substrat de CDO alors que tous les dipeptides cycliques le sont, suggérant que la présence du cycle dicétopipérazine est essentielle pour l'activité de l'enzyme. Les cyclodipeptides contenant des chaînes aromatiques et hydrophobes sont les meilleurs substrats et la nature du deuxième acide aminé semble affecter le taux de déshydrogénation du premier acide aminé<sup>24</sup>.

Il est intéressant de noter que, du fait que CDO agisse uniquement sur les cyclodipeptides, son activité de conversion peut être utilisée pour révéler la présence de cyclodipeptides dans un milieu<sup>372</sup>. CDO peut également être utilisée pour la bioconversion des cyclodipeptides<sup>373</sup> et peut ainsi constituer un outil permettant d'augmenter la diversité des dicétopipérazines synthétisées par ingénierie combinatoire.

## 2.2 Voies de biosynthèse de l'acide pulcherriminique

La pulcherrimine est un pigment extracellulaire produit par de nombreuses espèces de bactéries et quelques levures après une culture dans un milieu enrichi en Fe(III). La pulcherrimine est un complexe ferrique, dans lequel les quatre oxygènes de l'acide pulcherriminique coordonnent deux atomes de Fe. L'acide pulcherriminique est donc un précurseur de la pulcherrimine. Le cluster responsable de la synthèse de l'acide pulcherriminique a été identifié chez *Bacillus subtilis*<sup>374</sup> ; une CDPS, YvmC est responsable de la synthèse du cyclo(Leu-Leu)<sup>26</sup> dont le noyau dicétopipérazine est ensuite oxydé par un cytochrome P450, Cyp134A1(=CypX), pour donner l'acide pulcherriminique (Figure 41).

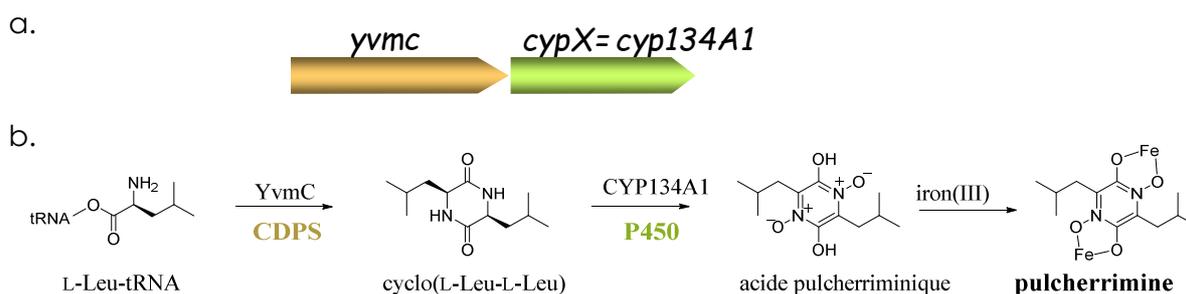


Figure 41 : Voie de biosynthèse de la pulcherrimine

a. Cluster de gènes pour la voie de biosynthèse de la pulcherrimine. b. Étapes de la synthèse de la pulcherrimine

De manière générale, les cytochromes P450 sont des hémoprotéines de type mono-oxygénases qui utilisent l'oxygène moléculaire pour catalyser en général des réactions d'oxydation sur des molécules diverses<sup>375,376</sup>. Les cytochromes P450 peuvent être impliqués dans des processus de dégradation de composés et présentent dans ce cas une spécificité particulièrement relâchée. Ils peuvent également intervenir dans des voies de biosynthèse et leur action est alors particulièrement spécifique. Les cytochromes P450 ont besoin, pour être actifs, de partenaires qui leur fournissent les électrons nécessaires au mécanisme catalytique. En l'absence de chaîne de transfert d'électrons, un substrat de cytochrome P450 va se lier au site actif mais ne sera pas transformé<sup>377</sup>. La liaison du substrat à un cytochrome P450 entraîne une modification des caractéristiques des propriétés structurales de l'hème, qui est utilisée pour déterminer l'affinité de l'enzyme pour son substrat.

Schlichting *et al.* ont caractérisé la structure tridimensionnelle des formes libres et liées à cyclo(Leu-Leu) de Cyp134A1 de *B. subtilis* et ont déterminé l'efficacité de l'enzyme pour le cyclo(Leu-Leu) mais également pour différents analogues<sup>367</sup>. De même que CDO, Cyp134A1, peut transformer d'autres cyclodipeptides que son substrat naturel, parfois même avec une affinité plus importante. Il semble avoir une préférence pour les chaînes latérales de type leucine ou de taille importante et hydrophobe par rapport aux chaînes plus légères. Cyp134A1 réalise sur le cyclo(Phe-Leu) une oxydation similaire à celle faite sur le cyclo(Leu-Leu), tandis que le cyclo(Met-Met) semble subir une réaction d'oxydation d'un autre type menant potentiellement à une aromatisation du cycle. Cyp134A1 apparaît donc comme une enzyme capable de réaliser des oxydations variées sur des substrats différents.

### 2.3 Voie de biosynthèse de la mycocyclosine

Le gène de la CDPS de *Mycobacterium tuberculosis*, Rv2275, synthétisant du cyclo(Tyr-Tyr) est associé au niveau génomique avec un gène codant un cytochrome P450 (CYP121A1). Ces deux enzymes constituent la voie de biosynthèse de la mycocyclosine<sup>26,29</sup> (Figure 42).

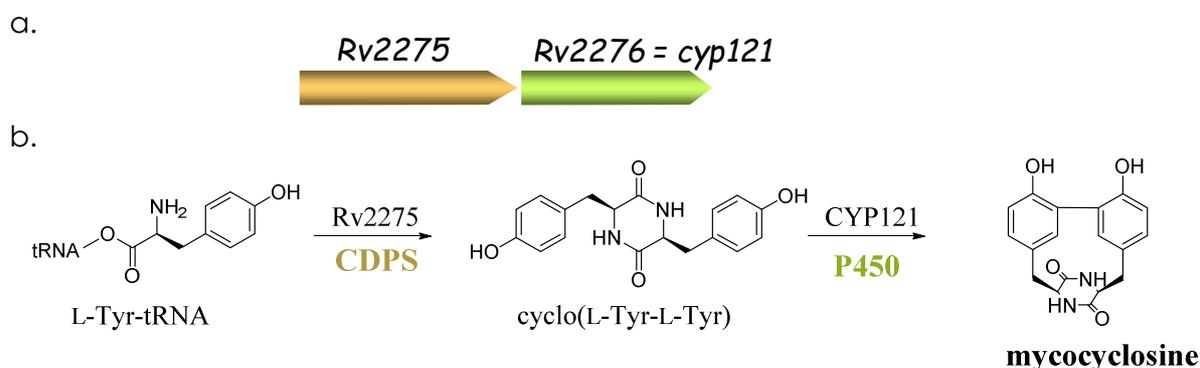


Figure 42 : Voie de biosynthèse de la mycocyclosine

a. Cluster de gènes pour la voie de biosynthèse de la mycocyclosine. b. Etapes de la synthèse de la mycocyclosine

Sur le plan du rôle physiologique, un certain nombre d'indices laissent à penser que cette voie revêt une importance particulière. Tout d'abord, les gènes *rv2275* et *cyp121* sont retrouvés exprimés chez un certain nombre de souches du complexe *M. tuberculosis*<sup>378,379</sup>. Cette association au niveau génétique a conduit l'équipe de Muriel Gondry à étudier la transformation du cyclo(Tyr-Tyr) par CYP121A1, montrant ainsi que CYP121A1 catalyse la formation d'une liaison C-C entre les noyaux tyrosyles du cyclo(Tyr-Tyr) pour donner la mycocyclosine<sup>29</sup>, une molécule qui n'a jamais été isolée de *M. tuberculosis* ou d'une autre source naturelle. Le mécanisme catalytique utilisé par CYP121 impliquerait la formation de radicaux tyrosyles et des mouvements de chaînes latérales du cyclo(Tyr-Tyr). Sur la base de structures tridimensionnelles et d'études biochimiques avec des analogues de substrats, la spécificité de réaction et de substrat de cette enzyme a été étudiée ; il s'avère que CYP121 est une enzyme très spécifique, contrairement à CDO et CYP134A1<sup>29,380,381</sup>.

Au début de mon étude, seules ces trois voies dépendant des CDPS étaient décrites. Avec l'augmentation du nombre de séquences de génomes disponibles, de plus en plus de nouvelles séquences de CDPS potentielles sont retrouvées. Ces CDPS potentielles sont associées au niveau génomique avec des gènes codant des enzymes non décrites dans les voies à CDPS, suggérant une plus grande variété dans les enzymes de modification<sup>345</sup>. Il est à noter qu'au cours de mes travaux, deux autres voies dépendant des CDPS ont été décrites par une autre équipe. Il s'agit de la voie de biosynthèse de nocazines<sup>346</sup> et de la voie de biosynthèse des dérivés du c(Trp-Trp)<sup>348</sup>.

# MATÉRIELS ET MÉTHODES

# I. Analyse bio-informatique

## 1. Identification des CDPS potentielles

Les bases de données de séquences du NCBI sont interrogées pour rechercher des gènes codant pour ces CDPS potentielles. Des recherches par PSI-BLAST à partir de séquences de plusieurs CDPS caractérisées, sont conduites avec un seuil de « E-value » de 0,005. A chaque itération, les séquences non pertinentes sont retirées pour l'itération suivante. Six à sept itérations ont été nécessaires pour obtenir un set de séquences stable.

## 2. Calcul d'arbres phylogénétiques

### 2.1 Arbre de mai 2013

Un alignement multiple des 82 séquences de CDPS potentielles (cf. Chapitre II. I.2 Correction des séquences obtenues et création d'un arbre phylogénétique) est calculé par Muscle<sup>382</sup> dans SeaView<sup>383</sup>. Cet alignement n'est pas ajusté manuellement mais l'alignement correct des résidus catalytiques et des résidus conservés est vérifié. L'arbre est calculé en utilisant le programme PhyML (v 3.1)<sup>384</sup> dans SeaView avec la méthode du maximum de vraisemblance (modèle de substitution : LG, amélioration de la typologie de l'arbre : NNI, évaluation de la fiabilité des branches : test aLRT (SH-like)). Le logiciel iTOL<sup>385,386</sup> est utilisé pour améliorer et éditer les représentations graphiques de l'arbre obtenu.

### 2.2 Arbre de septembre 2014

Après des recherches par PSI-BLAST réalisées comme détaillées ci-dessus, environ 300 séquences de CDPS potentielles ont été obtenues (cf. I.1 Identification des CDPS potentielles). Nous avons corrigé les séquences tronquées en N-terminal quand cela était possible et nous avons retiré du set les séquences tronquées en C-terminal, les séquences redondantes et les séquences dont la sérine catalytique manquait (notamment les CDPS potentielles 12 et 15). Les séquences des CDPS 25 et 26 ont été ajoutées au set. Un alignement multiple de ces séquences a été réalisé en utilisant Muscle<sup>382</sup> dans SeaView<sup>383</sup> et a été ajusté manuellement pour aligner les résidus conservés et les résidus catalytiques. Les insertions ou les domaines C-terminaux qui ne s'alignaient pas ont été supprimés de l'alignement multiple. L'arbre a été calculé comme décrit ci-dessus.

## 3. Outils pour l'analyse des séquences des CDPS

### 3.1 Prédiction avec HHPRED

L'outil HHPred du site du Max Planck Institute a été utilisé pour prédire les structures secondaires et le mode de repliement pour chaque CDPS potentielle en interrogeant la banque de données de protéine RCSB (RCSB PDB) avec les paramètres par défaut (méthode de génération des alignements de séquences multiples,

HHblits ; seuil de « E-value » pour la génération d'alignements de séquences multiples,  $10^{-3}$ ; mode d'alignement, local)<sup>387</sup>.

### 3.2 Diagramme topologique

Les diagrammes topologiques des trois CDPS de structures connues ont été obtenus avec PDBSum sur le site EMBL-EBI et un diagramme topologique consensus a été déduit en conservant les structures secondaires retrouvées dans les trois structures cristallographiques de CDPS.

### 3.3 Logos de séquences

Les logos de séquences des CDPS sont des empilements de symboles représentant les acides aminés dans le code à une lettre<sup>388</sup>. Ils sont obtenus sur le site WebLogo<sup>389</sup>. Chaque empilement représente une position dans la séquence. La taille des symboles dans l'empilement indique la fréquence relative de chaque acide aminé à cette position.

## 4. Analyse des environnements génomiques des CDPS

Une zone de 5 à 10 kpb autour de chaque séquence codant une CDPS procaryote est analysée. La fonction probable des protéines codées par les gènes présents dans cette zone est déterminée en utilisant le serveur du NCBI et l'outil MicroScope d'annotation des génomes, développé par le LabGem (<http://www.genoscope.cns.fr/agc/microscope/home/index.php>). Pour utiliser l'outil MicroScope, chaque génome codant une CDPS étudiée a fait l'objet d'une demande d'ajout dans la base de données. L'identification de la fonction probable des protéines codées s'appuie sur l'annotation préexistante dans les bases de données, l'identification de domaines conservés (Conserved Domain Database du NCBI) et l'annotation fonctionnelle de protéines suffisamment homologues (60% d'identité de séquence) repérées par BLAST (NCBI) ou par l'outil de visualisation des synthénies ou conservation de l'environnement génomique (MicroScope).

## II. Matériel commercial et techniques classiques de biologie moléculaire

### 1. Techniques classiques de biologie moléculaire

Les techniques classiques de biologie moléculaire employées sont faites selon Sambrook et Russel<sup>390</sup>.

**Digestion.** Les enzymes de restriction sont fournies par Thermo Scientific (Life Technologies) sous leur formulation FastDigest lorsqu'elles étaient disponibles et utilisées selon les recommandations du fournisseur.

**Purification d'ADN.** Les fragments d'ADN issus de digestions ou de PCR sont séparés sur gel d'agarose 1% et extraits en utilisant le kit QIAquick Gel Extraction Kit (Qiagen) selon le protocole fourni.

**Ligature.** La ligase d'ADN du phage T4 (T4 DNA Ligase, Life technologies) est utilisée, selon les recommandations du fournisseur, pour catalyser la jonction entre deux molécules d'ADN.

**Transformation.** Des bactéries chimio-compétentes fraîchement décongelées (100µl) (cf. II.2 Souches bactériennes utilisées) sont mises en présence de 5 à 100 ng de plasmide dans un tube préalablement incubé dans la glace. Après une incubation de 30 minutes (min) dans la glace, les bactéries subissent un choc thermique à 42°C pendant 45 secondes (s). Elles sont ensuite placées dans de la glace pendant 2 min, additionnées de 900 µl de SOC et incubées 30 min à 37°C en l'absence d'antibiotique, puis étalées sur LB-agar contenant l'antibiotique approprié (cf. II.4 Milieux de culture).

**PCR sur colonies.** Pour tester la présence d'un insert introduit par ligature dans le vecteur de clonage, nous avons criblé plusieurs clones par PCR sur colonies en utilisant le DreamTaq Green PCR Master Mix (2×) (Thermo Scientific, Life Technologies). La matrice d'ADN est apportée par le prélèvement d'une fraction de la colonie à tester. Dans la plupart des cas, des amorces s'hybridant en amont et en aval du site de clonage sont utilisées. 3 min de dénaturation initiale à 98°C permettent de lyser les bactéries et de dénaturer l'ADN. Puis 30 cycles composés d'une phase de dénaturation à 98°C de 30 s, suivie d'une phase d'hybridation à 57°C de 30 s et d'une phase d'élongation à 72°C de 1 min sont réalisés avant une élongation finale à 72°C de 15 min. La température de la phase d'hybridation et le temps de la phase d'élongation peuvent varier respectivement en fonction de la température de fusion des amorces sur la matrice et de la longueur de la séquence amplifiée.

**Purification de plasmide.** Les purifications de plasmides d'ADN sont réalisées à partir des cultures liquides en utilisant le kit GenElute™ Plasmid Miniprep (Sigma) ou le kit Wizard® Plus Midipreps DNA Purification System (Promega) et en suivant le protocole fourni. Les plasmides sont ensuite dialysés contre 100 volumes d'eau PPI (eau pour préparation injectable) sur des membranes 0.025 µm VSWP (Merck Millipore Lifescience) pour les miniprep, ou précipités à l'éthanol en présence d'acétate d'ammonium et repris dans 30 µl d'eau PPI pour les midiprep.

**Séquençage.** Les séquences des régions codant pour chaque enzyme étudiée dans ce travail ont été vérifiées par séquençage chez un prestataire (Eurofins Genomics).

## 2. Souches bactériennes utilisées

### 2.1 Souches pour la biologie moléculaire

La souche d'*E. coli* DH5α (*fhuA2 ΔlacU169 phoA glnV44 gyrA96 recA1 relA1 endA1 thi-1 hsdR17 φ80' lacZΔM15*) (Invitrogen) a été utilisée pour les expériences de clonage et d'amplification de plasmide. Cette souche ainsi que la souche XL1-blue (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI<sup>q</sup> lacZΔM15 Tn10 (Tet<sup>r</sup>)]*) ont été utilisées pour la congélation des bactéries transformées par les plasmides

obtenus au cours de l'étude. Des cellules compétentes de ces souches ont été préparées en suivant la méthode d'Inoue<sup>391</sup> avec une compétence mesurée d'environ  $10^8$  ufc/ $\mu$ g de pUC18.

## 2.2 Souches pour l'expression protéique

Les souches d'*E. coli* BL21AI (F<sup>-</sup> *ompT hsdS<sub>B</sub> gal dcm araB::T7RNAP-tetA*) (Invitrogen) et M15 (F<sup>-</sup>, *thi lac mtl  $\Phi$ 80 lac $\Delta$ M15*) (Qiagen) sont utilisées pour les cultures permettant l'expression des CDPS et des enzymes de modification en vue de l'analyse des dicétopipérazines synthétisées ou de la purification des protéines. Ces bactéries sont rendues compétentes à  $10^5$  ufc/ $\mu$ g de pUC18 selon un protocole décrit par Hanahan<sup>390,392</sup>.

La souche BL21AI contient une insertion chromosomique (*araB::T7RNAP*) codant pour l'ARN polymérase du phage T7 dont l'expression est inductible par l'arabinose. Ce type de souche est utilisable avec des vecteurs qui placent les gènes des protéines clonées sous le contrôle d'un promoteur T7. Ceci permet de n'induire l'expression des gènes clonés qu'en présence d'arabinose.

La souche M15 pour sa part permet une répression de l'expression des gènes par la présence du plasmide pREP4 (cf. II.3 Plasmides commerciaux).

## 3. Plasmides commerciaux

Trois types de plasmides commerciaux ont été utilisés au cours du projet. (Figure 43).

Le plasmide **pQE60** est destiné au clonage des gènes de CDPS. Il possède une origine de réplication ColE1, une cassette de résistance à l'ampicilline, une séquence codant une étiquette poly-histidine (His-Tag ou 6xHis) en 3' du gène cloné. Le gène cloné est sous contrôle d'un promoteur fort constitutif T5. Ce dernier est suivi d'opérateurs *lacO*, sur lesquels la protéine LacI se fixe pour inhiber la transcription chez *E. coli*. Cette inhibition peut être levée et la transcription induite par la présence d'IPTG (isopropyl  $\beta$ -D-1-thiogalactopyranoside), un analogue non métabolisable du lactose.

Le plasmide **pRSF-Duet** est destiné au clonage des enzymes de modification. Il possède une origine de réplication RSF1030 et une cassette de résistance à la kanamycine, compatibles avec l'utilisation concomitante du vecteur pQE60. Il possède deux unités de transcription contiguës, chacune ayant son propre site de clonage multiple afin de pouvoir exprimer, depuis le même vecteur, deux protéines différentes, l'une pouvant être suivie par une étiquette 6xHis, l'autre par une étiquette S (S-Tag). Cette propriété du vecteur ne sera pas exploitée dans notre étude. Les gènes clonés sont sous contrôle du promoteur T7, qui est un promoteur fort non constitutif, c'est-à-dire qu'il nécessite l'utilisation d'une souche capable de produire l'ARN polymérase du phage T7, comme la souche BL21AI (cf. II.2.2 Souches pour l'expression protéique)

Le plasmide **pREP4** est un plasmide portant une origine de réplication p15A et un gène de résistance à la kanamycine. Il permet l'expression constitutive de la protéine LacI codée par le gène *lacI*. LacI se fixe sur les

opérateurs *lacO* localisés en amont des gènes, et empêche ainsi la transcription par l'ARN polymérase. Cette inhibition peut être levée par l'ajout d'IPTG qui, en se liant à LacI, diminue son affinité pour les sites *lacO*.

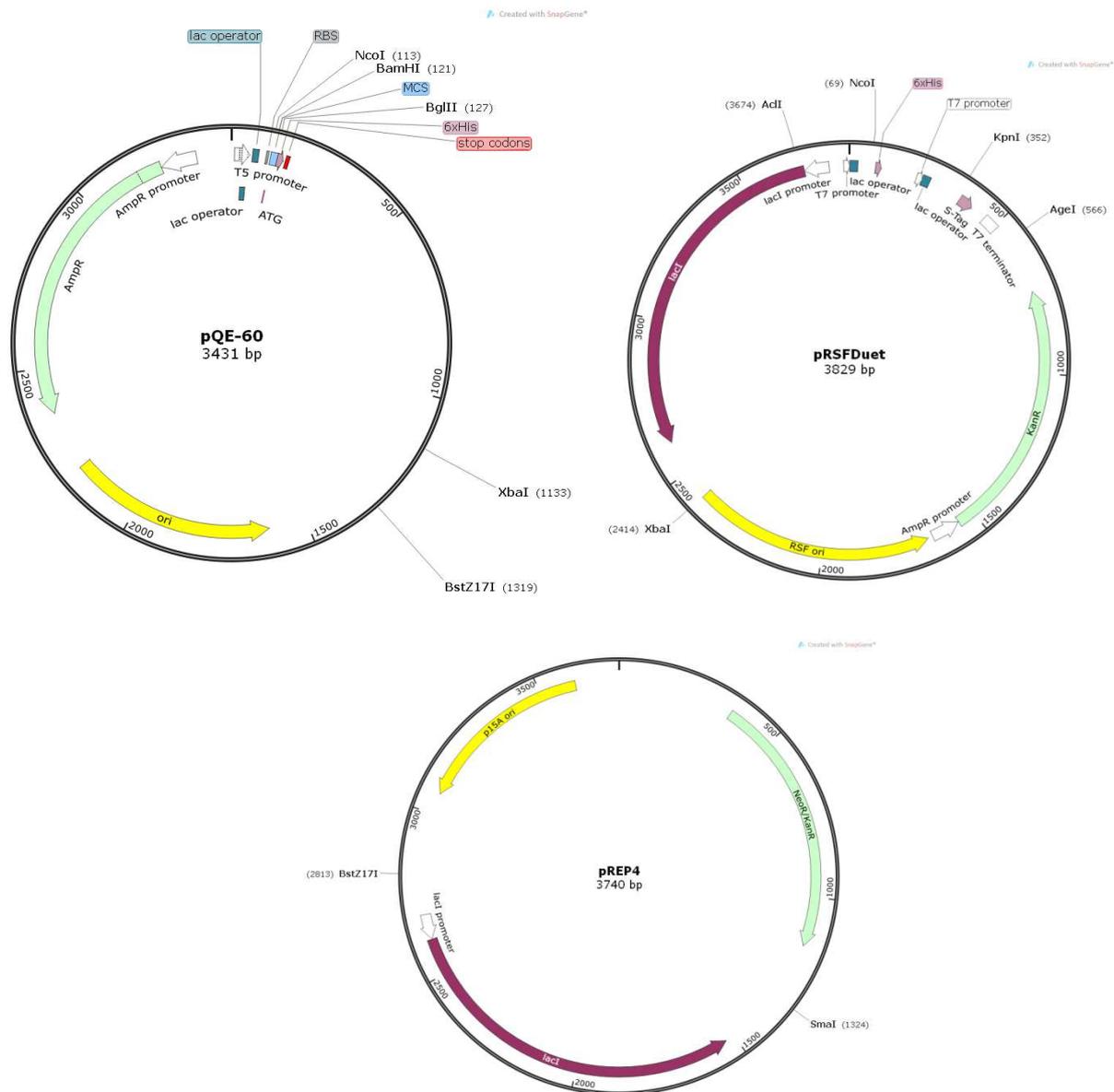


Figure 43 : Cartes des plasmides pQE60, pRSF-Duet et pREP4

Les caractéristiques du plasmide sont indiquées par des flèches annotées. Les origines de répliation sont indiquées en jaunes, les gènes de résistance en vert, les promoteurs en blanc, les étiquettes en rose, le gène *lacI* est en violet foncé. La position des sites de restriction impliqués dans la suite du travail est également repérée. Les cartes de plasmide ont été réalisées avec le logiciel SnapGene.

#### 4. Milieux de culture

Pour les cultures en boîte de Pétri, nous avons utilisé un milieu LB-agar comme décrit par Sambrook et Russel<sup>390</sup>. En fonction de la (des) résistance(s) présentée(s) par les plasmides à sélectionner ou à maintenir, une solution d'ampicilline (concentration finale 200 µg/ml) ou de kanamycine (concentration finale 25 µg/ml) est ajoutée. Les cultures permettant l'expression des CDPS et des enzymes de modification en vue de l'analyse,

dans le surnageant de culture, des dicétopipérazines produites sont réalisées en milieu minimum M9<sup>390</sup> supplémenté avec une solution de vitamines et une solution d'oligoéléments comme décrit précédemment<sup>26</sup>. La source de carbone est apportée par une solution de glucose (0,5%), de glycérol (0,5%) ou une solution de sucre 50x5052 (0,5% de glycérol, 0,05 % de glucose, 0,2 % de lactose) ou 50x5052<sup>2</sup> (0,5% de glycérol, 0,05 % de glucose, 0,05 % d'arabinose, 0,2 % de lactose). Les concentrations indiquées sont les concentrations finales dans le milieu. Les solutions 50x5052 et 50x5052<sup>2</sup> permettent de réaliser une auto-induction de l'expression des gènes sous le contrôle des opérons lactose et arabinose respectivement et sont extraites ou inspirées du travail de Studier sur les milieux autoinduits<sup>393</sup>. Le milieu LB liquide est utilisé pour les cultures permettant les amplifications de plasmides ou les expressions de CDPS en vue de leur purification. Le milieu ZYP-5052<sup>2</sup> correspond au milieu ZYP-5052 décrit par Studier<sup>393</sup> dans lequel la solution de 50x5052 est remplacée par la solution 50x5052<sup>2</sup>. Ce milieu est utilisé pour les cultures permettant l'expression des enzymes de modification en vue de leur purification. La seule modification apportée à la composition du milieu est que la solution 1000× « metals mix » a été remplacée par la solution d'oligoélément utilisée classiquement avec le M9. Les cultures sont réalisées en présence d'un antibiotique correspondant à la résistance apportée par le plasmide à maintenir, à savoir l'ampicilline à 100 µg/ml pour le pQE60 et la kanamycine à 25µg/ml pour le pRSF-Duet et le pREP4.

### III. Préparation des vecteurs pIJ194 et pIJ196

#### 1. Préparation du pIJ194

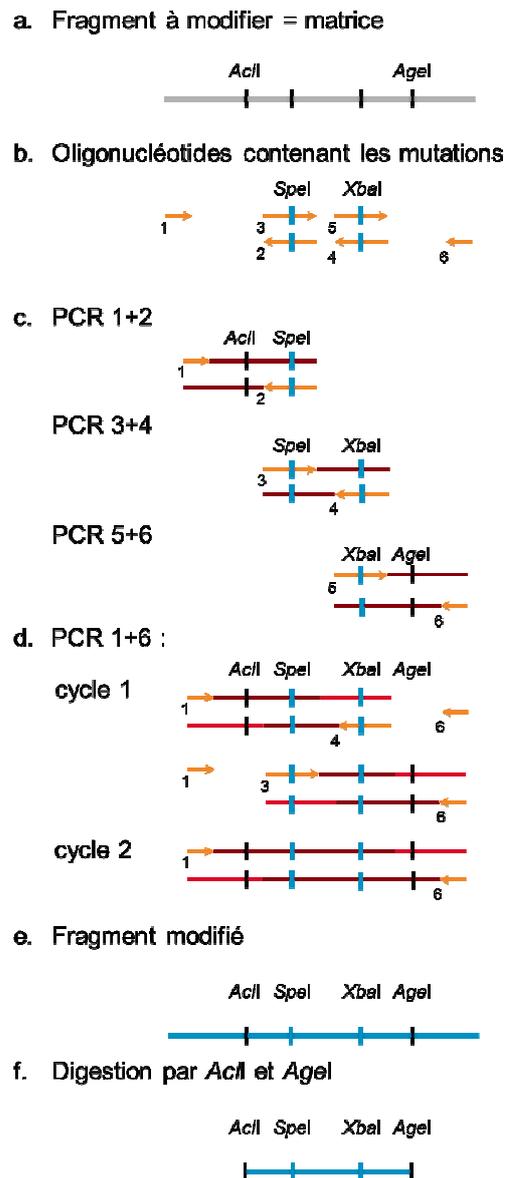
Le plasmide pIJ194 dérive du pRSF-Duet. Il contient un site unique SpeI en amont des unités de transcription, un site XbaI unique en aval des unités de transcription. Pour les clonages des gènes d'enzymes de modification, un site NcoI localisé dans le site de clonage multiple de la première unité de transcription et un site KpnI localisé dans le site de clonage multiple de la seconde unité de transcription seront utilisés, laissant une seule unité de transcription permettant l'expression d'une protéine sans étiquette. Par la suite, une étiquette 6×His sera ajoutée en modifiant directement le vecteur pRSFDuet contenant le gène de l'enzyme de modification. La construction du pIJ194 est décrite ci-dessous.

##### 1.1 Suppression du site *XbaI*

Etant donné qu'un site *XbaI* unique doit être introduit dans le plasmide pRSF-Duet à proximité du site de clonage, celui présent en position 2414 doit être supprimé. Le vecteur a été digéré par l'enzyme de restriction *XbaI* en donnant des extrémités cohésives. L'ADN polymérase de phage T4 (New England BioLabs), en présence de dNTP (New England BioLabs), est ensuite utilisée pour synthétiser des « double-brins » à la place des « simple-brins » des extrémités cohésives. Une ligation sur lui-même du vecteur permet, à partir des extrémités franches, d'obtenir un vecteur pRSF-Duet sans site *XbaI* (pRSF-Duet(*XbaI*-)) (Figure 44).



45 c.). Les amplicons obtenus sont purifiés sur gel d'agarose et servent de matrice à la PCR suivante qui fait intervenir uniquement les oligonucléotides 1 et 6. Lors de cette PCR, les brins d'ADN se chevauchant s'hybrident et servent de matrice à la polymérase. On obtient ainsi, dès le premier cycle de PCR, des fragments combinant deux des amplicons précédents. Selon le même principe d'hybridation des brins se chevauchant, l'amplicon de la cassette mutée complète est obtenu au cycle suivant (Figure 45 d.). À la fin de la PCR, cet amplicon contenant les deux sites *SpeI* et *XbaI* à introduire et les sites *AcI* et *AgeI* est purifié sur gel d'agarose (Figure 45 f.).



**Figure 45 : Création d'une cassette mutée contenant les sites *SpeI* et *XbaI* par PCR SOEing**

L'amplicon final et le vecteur pRSF-Duet(*XbaI*-) sont digérés par *AcI* et *AgeI* et purifiés sur gel d'agarose. Une ligature de la cassette mutée dans le vecteur pRSF-Duet(*XbaI*-) est ensuite réalisée pour obtenir le vecteur désiré, nommé pIJ194 selon la nomenclature mise en place dans le laboratoire (Figure 46).



et *SmaI* par ligature. La présence du gène *lacI* dans le sens désiré (identique au sens de réplication du plasmide) est vérifiée par PCR sur colonies puis par séquençage. Le vecteur ainsi obtenu est nommé pIJ196 (Figure 47).

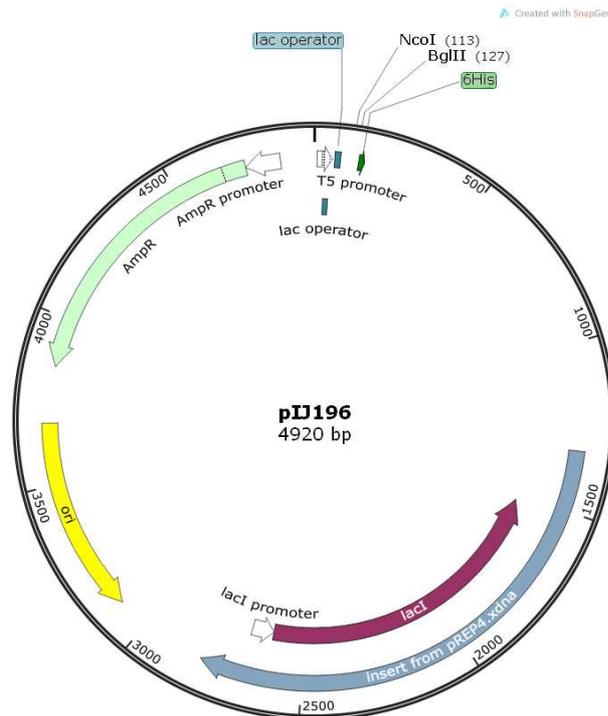


Figure 47 : Carte du plasmide pIJ196 dérivé du pQE60 commercial

Les caractéristiques du plasmide sont indiquées par des flèches annotées. L'origine de réplication est indiquée en jaune, le gène de résistance en vert, les promoteurs en blanc, les étiquettes en vert, le gène *lacI* est en violet foncé. La position des sites de restriction impliqués dans la suite du travail est également repérée. La carte du plasmide a été réalisée avec le logiciel SnapGene.

#### IV. Obtention des gènes de CDPS et d'enzymes de modification

Les gènes de CDPS et d'enzymes de modification potentielles ont été obtenus soit par amplification par PCR à partir d'ADN génomique, soit par synthèse de gène auprès de GeneArt et Lifetechnologies pour les gènes de CDPS ou d'Eurofins pour les gènes d'enzymes de modification.

##### 1. Amplification d'ADN génomique

Les premiers essais de clonage et d'analyse de dicétopipérazines produites ont été réalisés à partir de gènes amplifiés par PCR à partir d'ADN génomique extrait des souches bactériennes suivantes : *Sphingobium japonicum* UT26S (ADN fourni par Y. Nagata, Tohoku University, Sendai, Japan), *Candidatus Protochlamydia amoebophila* UWE25 (ADN fourni par G. Schwammel, University of Vienna, Vienna, Austria) et *Streptomyces* sp. AA4 (ADN obtenu par l'intermédiaire de J.-L. Pernodet, I2BC, Orsay).

Les gènes de CDPS (*SJA\_c1-32620*, *pc1814* et *pc1814long*) sont clonés dans le vecteur pQE60 ou le pIJ196 entre les sites *NcoI* et *BamHI*. Les séquences codantes sont amplifiées par PCR (30 s de dénaturation initiale à 98°C puis 30 cycles composés d'une phase de dénaturation à 98°C de 10 s, suivie d'une phase d'hybridation à 58,5°C de 30 s et d'une phase d'élongation à 72°C de 15 s, et enfin 10 min d'élongation finale à 72°C) en utilisant des amorces conçues pour ajouter un site *NcoI* ou *BspHI* (compatible avec *NcoI*) recouvrant le codon initiateur et un site *BamHI* immédiatement après le dernier codon codant (Tableau 6).

**Tableau 6 : Oligonucléotides utilisés pour l'amplification des gènes de CDPS**

Gènes amplifiés	Amorce (5'-3') dans le sens de lecture (FOR) et dans le sens inverse (REV) <sup>1</sup>	
SJA_c1-32620	FOR	taagactgaagaaactcattc <b>ATGA</b> AATATCGGCGACTATGGCCACGC
	REV	taagactgaagaaactcatt <b>TCAGGATCC</b> ATAAGAGTCGGCAGCCTCGATTTCAGAC
pc1814	FOR	taagactgaagaaactcattgg <b>ccATGG</b> AAACCGAAAAACCG
	REV	gattttttcttttgaataac <b>TTAGGATCC</b> ATTTTGCATGATATTATGAAGG
pc1814long	FOR	ttaattaatgaaaattgccag <b>ccATGG</b> CCAATAAATATACATGGAGAAAGAATTTTAC
	REV	gattttttcttttgaataac <b>TTAGGATCC</b> ATTTTGCATGATATTATGAAGG

<sup>1</sup>Les sites de restrictions ajoutés sont représentés en bleu. La séquence codante est en majuscule.

Trois des gènes de CDPS potentielles testées lors du criblage moyen débit ont été obtenus par amplification d'ADN génomique extrait des souches suivantes : *Streptomyces* sp. AA4 (*Amycolatopsis* sp. AA4) (ADN obtenu par l'intermédiaire de J.-L. Pernodet, I2BC, Orsay), *Streptomyces cattleya* NRRL 8057 (ADN fourni par V. Barbe, Génoscope, Evry) et *Streptomyces venezuelae* ATCC 10712 (ADN fourni par M Bibb, John Innes Centre, Norwich, UK) (Tableau 7). L'extraction et l'amplification par PCR ont été réalisées au laboratoire de Jean-Luc Pernodet (I2BC, Orsay). Pour les détails de la procédure, se référer à la section « matériels et méthodes » de l'article (Annexe 1).

**Tableau 7 : Oligonucléotides utilisés pour l'amplification des gènes de CDPS n° 47 à 49**

CDPS	Gènes amplifiés	Amorce (5'-3') dans le sens de lecture (FOR) et dans le sens inverse (REV) <sup>1</sup>	
47	SSMG_00951	FOR	CCATGGCGATCAGCAACGGCTTCCGC
		REV	AGATCTGGCTGGGCGGATGGTGGCGT
48	SCAT_0198	FOR	CCATGGCGCTGCATCGAACGTCCTTCTC
		REV	AGATCTCGGCTCGGCGGGCAGTTGCT
49	SVEN_1864	FOR	CCATGGCGACTCAGGCGGCTTTTCGCC
		REV	AGATCTTCGCGCGGCTCCTCGGTCGG

Le gène d'enzyme de modification (SSMG\_00950) de *Streptomyces* sp. AA4 (*Amycolatopsis* sp. AA4) est cloné dans le vecteur pIJ194 entre les sites *NcoI* et *KpnI*. La séquence codante est amplifiée par PCR (30 s de dénaturation initiale à 98°C puis 30 cycles composés d'une phase de dénaturation à 98°C de 10 s, suivie d'une phase d'hybridation-élongation à 72°C de 45 s et enfin 10 min d'élongation finale à 72°C) en utilisant des

amorces conçues pour ajouter un site *BspHI* compatible avec *NcoI* recouvrant le codon initiateur et un site *BamHI* précédés de 2 codons « stop » immédiatement après le dernier codon codant (Tableau 8).

**Tableau 8 : Oligonucléotides utilisés pour l'amplification du gène d'enzyme de modification R**

Enzyme de modification	Gène amplifié	Amorce (5'-3') dans le sens de lecture (FOR) et dans le sens inverse (REV) <sup>1</sup>	
R	SSMG_00950	FOR	taagactgaagaaactcattcattc <b>ATG</b> ACGAACCCCGCCGACCAG
		REV	taagactgaagaaactcattcat <b>gtaccctat</b> TACGGCTTTGTCGCGGTAAGAGTCAAAC

<sup>1</sup>Les sites de restrictions ajoutés sont représentés en bleu. Les codons stop sont en violet. La séquence codante est en majuscule.

## 2. Gènes de synthèse

Les gènes de synthèse codant les CDPS potentielles (1 à 46 et 50 à 63) sont obtenus auprès de GeneArt et Lifetechnologies. Les séquences sont issues de la sélection par bio-informatique (cf Chapitre II. I Identification par bio-informatique de 80 CDPS potentielles et sélection de 49 CDPS à tester). La séquence à l'extrémité N-terminale a été allongée ou raccourcie dans quelques cas (cf. Chapitre II. I.2 Correction des séquences obtenues et création d'un arbre phylogénétique). L'insertion des gènes de synthèse dans le vecteur pIJ196 est réalisée entre les sites *NcoI* et *BglII*. Un site *NcoI* (CCATGG) est conçu au début de la séquence codante de manière à ce que l'ATG constitue le codon initiateur. Dans le cas où le résidu suivant n'était pas compatible avec un codon de type Gxx, un codon alanine est ajouté. Un site *BglII* est ajouté après le dernier codon codant. Les séquences des gènes synthétisés sont regroupées dans l'annexe électronique 1.

Les gènes de synthèse codant les enzymes de modifications (A à Q) sont obtenus auprès d' Eurofins Genomics. Les séquences sont déterminées par bio-informatique. L'insertion des gènes se fait entre *NcoI* et *KpnI* sur le pIJ194. De même que pour les gènes de synthèse de CDPS, le site *NcoI* est ajouté pour recouvrir le codon initiateur et nécessite parfois d'introduire une alanine comme deuxième résidu. Le site *BglII* est introduit pour sa part après deux codons « stop ».

## V. Culture pour l'expression des CDPS et des enzymes de modification

### 1. Culture en erlenmeyer

Les premiers essais pour tester l'activité des CDPS et des enzymes de modification ont été effectués conformément à la méthode décrite précédemment<sup>26</sup>. Les CDPS et enzymes de modification testées sont exprimées dans *E. coli* M15-pREP4, BL21AI ou BL21AI-pREP4. 50 µl de bactéries chimio-compétentes sont transformées avec 50 à 100 ng de plasmide selon le protocole de transformation classique (cf. II.1 Techniques classiques de biologie moléculaire). Lors d'une co-transformation, chaque plasmide doit au moins être apporté

à 100 ng. Les bactéries transformantes obtenues sur LB-agar sont cultivées dans 4 ml de milieu M9 supplémenté (cf. II.4 Milieux de culture) contenant 0,5 % de glycérol, 100 µg/ml d'ampicilline (et 25 µg/ml de kanamycine dans le cas de l'utilisation de pREP4). Après une nuit sous agitation à 37°C, 1 ml (2 ml) de cette pré-culture est (sont) utilisé(s) pour ensemer 25 ml (60 ml) de M9. Lorsque la culture atteint une  $DO_{600} = 0,6$ , une solution d'IPTG est ajoutée pour atteindre 1mM de concentration finale et la croissance est continuée à 20°C pendant 20 heures. La culture est alors centrifugée à 4000 *g* et à 4°C pour récolter, d'une part, le surnageant de culture qui va être analysé par LC-MSMS et, d'autre part, le culot bactérien dont le contenu protéique peut être analysé par électrophorèse sur gel de polyacrylamide en présence de dodécylsulfate de sodium (SDS-PAGE). Des témoins négatifs sont réalisés en cultivant une souche transformée avec un plasmide ne portant aucun gène de CDPS ou d'enzyme de modification.

## 2. Culture en plaque 24 puits

Les enzymes testées sont exprimées dans *E. coli* M15-pREP4, BL21AI ou BL21AI-pREP4. 25 µl de cellules compétentes sont transformés avec 50 à 100 ng de plasmide dans un puits d'une plaque 24 puits « deepwell » (D. Dutcher) selon le protocole de transformation classique (cf II.1 Techniques classiques de biologie moléculaire). Une fois le choc thermique réalisé, les cellules sont diluées dans 475 µl de SOC. Après 30 min d'incubation à 37°C, 25 µl sont étalés sur du LB-agar contenant les antibiotiques appropriés, coulé dans des plaques 6 puits. Les colonies transformantes obtenues servent à ensemer 2 ml de M9 supplémentés (cf. II.4 Milieux de culture) contenant 0,5 % de glucose, 100 µg/ml d'ampicilline (et 25 µg/ml de kanamycine dans le cas de l'utilisation de pREP4). Ces pré-cultures, réalisées dans des plaques 24 puits également, sont recouvertes d'un film poreux stérile (VWR) et incubées sous agitation (200 tours/min) à 37°C pendant une nuit. Au matin, 100 µl de chaque pré-culture sont ajoutés dans un puits contenant 2 ml de M9 supplémentés auquel on ajoute les antibiotiques appropriés et une solution de glycérol (pour M15-pREP4), de 50x5052 (pour BL21AI et BL21AI-pREP4 transformées uniquement par un plasmide pIJ196) ou de 50x50522 (pour BL21AI transformées par un plasmide pIJ194) (cf. II.4 Milieux de culture). Dans le cas de BL21AI et BL21AI-pREP4, la plaque recouverte d'un film poreux est incubée sous agitation (200 tours/min) 3 h 30 à 37°C puis 24 h à 20°C. Dans le cas de M15-pREP4, un suivi de croissance bactérienne est nécessaire. Quatre dilutions au 1/20<sup>ème</sup> de la culture contenue dans chaque puits sont réalisées dans les puits d'une plaque 96 puits. L'absorbance des puits est alors mesurée et calculée par le lecteur de plaque (ELx800, BIO-TEK INSTRUMENT). Une fois que la culture de la majorité des puits a atteint une  $A_{600}$  de 0,6 (au bout de 4-5 h à 37°C, 200 tours/min), 4 µl d'IPTG 1M sont introduits dans chaque puits. L'incubation est prolongée à 20°C et 200 tours/min pendant 24 h. A la fin de la culture, la croissance bactérienne finale est mesurée (cf. ci-dessus) et les plaques sont centrifugées à 3000 *g* et 4°C pendant 30 min. 500 µl de surnageant sont prélevés puis les culots bactériens sont débarrassés du surnageant de culture restant, séchés et congelés à -80°C, avant d'être analysés par SDS-PAGE.

## VI. Identification des dicétopipérazines

### 1. Standards chimiques

Les standards chimiques utilisés sont obtenus auprès de Bachem pour le cyclo(Gly-Gly), le cyclo(Gly-Ala), le cyclo(Ala-Ala), le cyclo(Leu-Gly), le cyclo(Val-Val), le cyclo(Glu-Ala), le cyclo(Pro-Leu), le cyclo(Pro-Met), le cyclo(Phe-Pro), le cyclo(Tyr-Ser), le cyclo(Tyr-Pro), le cyclo(His-Phe) et de Sigma pour le cyclo(Leu-Ala) et le cyclo(Leu-Val).

Le cyclo(Phe-Ala), le cyclo(Phe-Val), le cyclo(Phe-Thr), le cyclo(Phe-Asn) et le cyclo(Phe-Glu) sont obtenus par synthèse chimique selon des procédures déjà publiées<sup>394</sup>. Pour plus de détails sur la purification et la caractérisation structurale de ces cyclodipeptides, se reporter au manuscrit de l'article disponible en annexe 1.

### 2. Analyses des surnageants de culture par LC-MS/MS

Les surnageants de culture, collectés après centrifugation des cultures bactériennes, sont acidifiés avec de l'acide trifluoroacétique (TFA) ou de l'acide formique (AF) (2% final). Ils sont analysés par HPLC en phase inverse couplée à la spectrométrie de masse par électro-ébulisaison en utilisant un système 1100 HPLC d'Agilent associé à un spectromètre de masse à trappe ionique (Esquire HCT, Bruker Daltonik GmbH). Deux colonnes chromatographiques sont utilisées pour l'analyse des surnageants de culture : Altantis dC18 (4.6 mm x 150 mm, 3  $\mu\text{m}$ , 100  $\text{\AA}$ , Waters) et Hypercarb (4.6 mm x 150 mm, 5  $\mu\text{m}$ , 250  $\text{\AA}$ , ThermoScientific). Les gradients utilisés sont les suivants (éluant A : 0.1% (v/v) acide formique dans  $\text{H}_2\text{O}$  ; éluant B: 0,1% (v/v) acide formique dans acétonitrile/ $\text{H}_2\text{O}$  (90/10) ; débit : 0,6 ml/min).

Gradient 1

Temps (min)	% Eluant A	% Eluant B
0	100	0
10	100	0
60	50	50
62	0	100
67	0	100
70	100	0
90	100	0

Gradient 2

Temps (min)	% Eluant A	% Eluant B
0	85	15
10	85	15
60	35	65
62	0	100
67	0	100
70	85	15
90	85	15

Gradient 3

Temps (min)	% Eluant A	% Eluant B
0	100	0
5	85	15
55	35	65
57	0	100
62	0	100
65	100	0
105	100	0

Le gradient 1 est le gradient qui a été utilisé dans la plupart des analyses. Les gradients 2 et 3 ont été utilisés dans de rares cas, lorsque les composés analysés étaient particulièrement hydrophobes.

La détection des analytes par spectrométrie de masse est réalisée à partir du temps  $t=8$  min en mode positif et les paramètres sont optimisés pour la détection d'une large gamme de cyclodipeptides comme décrit précédemment<sup>26,29</sup>.

### 3. Purification de cyclodipeptides

Les cyclodipeptides et produits dérivés sont purifiés par HPLC semi-préparative en utilisant LaPrep $\Sigma$  VWR. Les purifications ont été réalisées à un débit de 4,75 ml/min.

#### 3.1 A partir de surnageant de culture

Les cyclodipeptides cyclo(Tyr-Val), cyclo(Gly-Asn), cyclo(Gly-Val), cyclo(Ala-Pro), cyclo(Leu-Glu) et cyclo(L-cystine) ont été purifiés à partir de surnageants de cultures de *E. coli* recombinant exprimant respectivement les CDPS 22, 27, 8, 18, 36 et 9. Le cyclo(Tyr-Val) a été purifié en utilisant une colonne LiCART 250-10 Purospher®STAR RP-18 endcapped (10 mm  $\times$  250 mm, 5  $\mu$ m, 120 Å, Merck). Le cyclo(Gly-Asp), le cyclo(Gly-Val), le cyclo(Ala-Pro), le cyclo(Leu-Glu) et le cyclo(L-cystine) ont été purifiés avec une colonne Hypercarb (10 mm  $\times$  150 mm, 5  $\mu$ m, 250 Å, Thermo Scientific). Le gradient utilisé pour la purification a été optimisé pour chaque cyclodipeptide (cf. Chapitre II. IV.1 Méthode pour l'identification des cyclodipeptides). Les cyclodipeptides purifiés ont été obtenus à une pureté > 90%, d'après l'estimation réalisée à partir du chromatogramme UV de l'analyse HPLC enregistré à 220 nm. L'analyse a été réalisée sur la colonne Atlantis dC18 (4.6 mm  $\times$  150 mm, 3  $\mu$ m, 100 Å, Waters) pour le cyclo(Tyr-Val) (gradient 1 ; éluant A : 0,1% (v/v) TFA dans H<sub>2</sub>O ; éluant B : 0,1% (v/v) TFA dans acétonitrile/H<sub>2</sub>O (90/10) ; débit : 0,6 ml/min) et sur la colonne Hypercarb (4.6 mm  $\times$  150 mm, 5  $\mu$ m, 250

Å, ThermoScientific) pour le cyclo(Gly-Asn), le cyclo(Gly-Val), le cyclo(Ala-Pro), le cyclo(Leu-Glu) et le cyclo(L-cystine) en utilisant le même gradient que pour la purification à 0,6 ml/min. Les cyclodipeptides obtenus ont été analysés en déterminant leur composition en acide aminés (cf. VI.4 Analyse en composition d'acides aminés) et par analyse de masse haute résolution. (cf. VI.5 Analyse par spectrométrie de masse haute résolution (HRMS)).

### 3.2 A partir de milieu réactionnel

Les composés PB2, PB3, PB4 et PB6 sont purifiés à partir du milieu réactionnel de la transformation *in vitro* du cyclo(Ala-Glu) par l'enzyme K. Les conditions de la transformation sont celles décrites ci-dessous (cf. IX.2 Dioxygénases dépendant du Fe(II) et du 2-oxoglutarate) en portant la concentration en enzymes à 10 µM et la concentration en substrat à 20 mM. Après une incubation de 21 h, le milieu réaction est acidifié à 3% de TFA et purifié sur une colonne Hypercarb (10 mm × 150 mm, 5 µM, 250 Å, Thermo Scientific) en utilisant le gradient 1. Les fractions contenant les produits d'intérêt sont collectées puis lyophilisées. Les composés ont été obtenus à une pureté > 95% pour les composés PB4 et PB6. Le composé PB2 est obtenu à une pureté > 90% mais la fraction comporte 6 % de PB3. Le composé PB3 est obtenu à ~ 65% de pureté et la fraction contient ~ 30% de PB2. Ces valeurs ont été estimées à partir du chromatogramme UV de l'analyse HPLC enregistré à 220 nm. L'analyse a été réalisée sur la colonne Hypercarb (4.6 mm x150 mm, 5 µm, 250 Å, ThermoScientific) (gradient 1 ; débit : 0,6 ml/min). Les composés obtenus ont été analysés par RMN.

## 4. Analyse en composition d'acides aminés

Cette analyse a été réalisée par Steven Dubois (CEA/iBiTec-S/SIMOPRO). Les échantillons en « duplicate » ou « triplicate » (10 µl à environ 100 µM dans 1% DMSO/eau MilliQ ou 10 µl de fraction collectée en sortie d'HPLC préparative) sont séchés sous vide, dans un tube en verre scellé en utilisant le système PicoTag (Waters) et hydrolysés sous vapeur de HCl 6N avec un cristal de phénol pendant 17 h dans une étuve à 110°C. Les échantillons hydrolysés sont dissous dans 20 à 50 µl d'eau Milli-Q et 5 à 20 µl de cet hydrolysat HCl (contenant au moins 200 pmol de chaque acide aminé) sont analysés par HPLC d'échange ionique et quantifiés après dérivation à la ninhydrine en sortie de colonne sur l'analyseur automatique d'acides aminés aminoTac JLC-500/V (JEOL). Des solutions standards d'acides aminés sont utilisées pour la calibration au début de chaque série d'analyses. Aucune donnée probante n'a été obtenue pour le cyclo(L-cystine).

## 5. Analyse par spectrométrie de masse haute résolution (HRMS)

**MALDI-TOF/TOF** (Analyse réalisée par Robert Thai ; CEA/iBiTec-S/SIMOPRO) : 0,5 µl de l'échantillon provenant de la fraction collectée en sortie d'HPLC préparative (à environ 100 µM) est déposé sur une plaque de MALDI standard en utilisant la méthode de la goutte séchée avec 0,5 µl de solution de matrice d'acide α-cyano-4-hydroxycinnamique (4-HCCA) à 10 mg/ml dans 50%

CH<sub>3</sub>CN/H<sub>2</sub>O/0,1% TFA. Les spectres sont enregistrés en utilisant un spectromètre MALDI-TOF/TOF 4800 (AB-SCIEX) en mode réflectron dans la gamme de  $m/z$  100-600. Chaque spectre résulte de 1000 tirs (20 positions différentes dans chaque dépôt et 50 tirs par sous-spectre) et une calibration interne est appliquée en utilisant le  $m/z$  de la matrice 4-HCCA et de ses clusters (adduits multiples). Cet instrument a été utilisé pour la détermination des masses exactes des composés suivants : HRMS ( $m/z$ ): cyclo(Tyr-Val), [MH<sup>+</sup>] calculée pour C<sub>14</sub>H<sub>19</sub>N<sub>2</sub>O<sub>3</sub> 263,1396, mesurée à 263,1393; cyclo(Gly-Val), [MH<sup>+</sup>] calculée pour C<sub>7</sub>H<sub>13</sub>N<sub>2</sub>O<sub>2</sub> 157,0977, mesurée à 157,0974; cyclo(Ala-Pro), [MH<sup>+</sup>] calculée pour C<sub>8</sub>H<sub>13</sub>N<sub>2</sub>O<sub>2</sub> 169,0977, mesurée à 169,0966; cyclo(Leu-Glu), [MH<sup>+</sup>] calculée pour C<sub>11</sub>H<sub>19</sub>N<sub>2</sub>O<sub>4</sub> 243,1345, mesurée à 243,1324.

**ESI-Orbitrap (LTQ-Orbitrap XL+ ETD, ThermoFisher Scientific)**: Le spectromètre se caractérise par un couplage entre un piège linéaire (LTQ) et un analyseur Orbitrap. Les échantillons ont été analysés en mode d'ionisation positive à 3,8 kV. La gamme de  $m/z$  est de 50 à 500  $m/z$ . La résolution est fixée à 30 000. La température du capillaire de transfert est de 275°C. Le voltage du capillaire est fixé à 40 V, le gaz auxiliaire à 0. L'échantillon est appliqué à un débit de 3,5 µl/min. Pour les expériences MS<sup>n</sup>, deux modes de fragmentation induite par des collisions de basse énergie (CID) et de haute énergie (HCD) ont été utilisés. L'énergie de collision normalisée appliquée a varié de 35 à 80 %.

L'instrument disponible au CEA/SPI (Saclay) a été utilisé pour la détermination des masses exactes de cyclo(Gly-Asn), [MH<sup>+</sup>] calculée pour C<sub>6</sub>H<sub>10</sub>N<sub>3</sub>O<sub>3</sub> 172,0722, mesurée à 172,0713 ; cyclo(L-cystine), [MH<sup>+</sup>] calculée pour C<sub>6</sub>H<sub>9</sub>N<sub>2</sub>O<sub>2</sub>S<sub>2</sub> 205,0100, mesurée à 205,0096. Cette analyse été réalisée par François Fenaille ; CEA/iBiTec-S/SPI).

L'instrument disponible à la plateforme de spectrométrie de masse et de protéomique de l'Institut de Biologie Paris-Seine (Paris) a été utilisé pour les analyses des composés IJ18, IJ19, IJ20, PB2, PB3, PB4 et PB6. Les détails des résultats obtenus sont disponibles en annexe 2. Ces analyses ont été réalisées en collaboration avec Lucrèce Matheron (UPMC/ Plate-forme spectrométrie de masse et protéomique)

## 6. Analyses par résonance magnétique nucléaire (RMN)

Les expériences ont été réalisées en collaboration avec Isabelle Correia (UPMC/Laboratoire des Biomolécules). Les expériences de RMN ont été enregistrées sur un spectromètre Avance III (Bruker) équipé d'une cryosonde TCI et opérant à une fréquence de 500 MHz pour <sup>1</sup>H. Les échantillons lyophilisés (≈ 1 mg) sont dissous dans du DMSO-*d*<sub>6</sub> (Eurisotop) et les spectres sont enregistrés à 30°C. Les résonances <sup>1</sup>H et <sup>13</sup>C sont attribuées grâce aux expériences 1D <sup>1</sup>H, 1D <sup>13</sup>C, 2D <sup>1</sup>H-<sup>1</sup>H COSY, 2D <sup>1</sup>H-<sup>13</sup>C HSQC, 2D <sup>1</sup>H-<sup>15</sup>N HSQC et 2D <sup>1</sup>H-<sup>13</sup>C HMBC. Les déplacements chimiques <sup>1</sup>H et <sup>13</sup>C sont calibrés sur le signal du solvant (respectivement 2,50 et 39,5 ppm). Les expériences RMN sont traitées et analysées avec le programme TOPSPIN 2.0 (Bruker). Une expérience NOESY sur les composés PB2a et PB2b a permis de déterminer la stéréochimie du C $\alpha$  Glu.

## **VII. Analyse protéique des culots bactériens**

### **1. Culots issus de la culture en erlenmeyer**

Les cellules sont reprises dans 2,5 ml de tampon de lyse glacé (50 mM Tris-HCl pH 7,5, glycérol 5 %), congelées et lysées par une presse d'Eaton. Après décongélation, une solution de MgCl<sub>2</sub> (concentration finale 2,8 mM), 1 mM de phénylméthylsulfonylfluoride (PMSF) et 1 µl de benzonase (concentration finale 100 U/ml) sont ajoutés. Après une incubation de 1 h sur un lit de glace, le lysat est centrifugé à 22000 g pendant 15 min. 2,5 µl de tampon de charge 5× (250 mM Tris HCl pH 6,8, 500 mM dithiothréitol (DTT), 10% SDS, 50% glycérol) sont ajoutés à 10 µl du surnageant de lyse obtenus et sont dénaturés pendant 5 min à 98°C. Les culots de lyse sont rincés avec du tampon de lyse et sont repris dans 1 ml de tampon de charge 1× puis dénaturés pendant 30 min à 98°C. Les échantillons (12,5 µl pour le surnageant de lyse et 10 µl pour le culot) sont chacun analysés par SDS-PAGE.

### **2. Culots issus de la culture en plaque 24 puits**

Une lyse chimique inspirée de précédents travaux de l'équipe<sup>395</sup> est réalisée directement dans les puits des plaques 24 puits. Chaque culot est repris dans 400 µl de tampon de lyse glacé (100 mM Tris HCl, pH 8.0, 150 mM NaCl, 0.1% Triton X100, 0.5 mg/ml lysosyme, 10 µM phosphoramidon, 1 mM PMSF)<sup>395</sup>. Après 1,5 h sur un lit de glace, une solution de MgCl<sub>2</sub> (concentration finale 10 mM) et 5 unités de benzonase sont ajoutées et les échantillons sont incubés 1 h dans un bain glacé avant d'être centrifugés à 3000 g et à 4°C pendant 45 min. 50 µl de tampon de charge 5× sont ajoutés aux 200 µl de surnageant de lyse prélevés. Le culot de lyse est repris avec 200 µl de tampon de charge 1×. Ces échantillons sont dénaturés à 98°C pendant 20 minutes pour les fractions insolubles (culot) et 5 minutes pour les fractions solubles (surnageant). 10 µl de chaque fraction sont analysés par SDS-PAGE.

### **3. Analyse par SDS-PAGE**

Un gel dénaturant d'acrylamide 12 % de 10 puits est réalisé. Un marqueur de poids moléculaire est mis à migrer en même temps que les échantillons (Precision Plus Protein Standards Unstained, BioRad). La migration est réalisée à 200 V pendant 45 min à 1 h. Pour la migration de plusieurs gels en parallèle, le système Mini-Protean® 3 Dodeca™ Cell (BioRad) est utilisé. Les gels sont révélés au bleu de Coomassie et visualisés avec le Gel Doc EZ system (BioRad).

## **VIII. Purification des protéines (réalisée au sein de l'équipe par Emmanuel Favry et Jérôme Seguin)**

### **1. Purification des CDPS (réalisée par Emmanuel Favry et Jérôme Seguin)**

Les CDPS recombinantes ont été exprimées dans BL21AI-pREP4 transformées par le plasmide pIJ196-genXX (XX étant le n° de la CDPS). La production et la purification ont suivi le protocole déjà décrit par l'équipe<sup>26</sup> mais se sont arrêtées à la fin de l'étape de purification par affinité (colonne HisTrap). Les coefficients d'extinction molaire des protéines ont été estimés en utilisant ProtParam sur le site ExPASy.

### **2. Purification des dioxygénases dépendant du Fe(II) et du 2-oxoglutarate (réalisée par Emmanuel Favry)**

Les dioxygénases dépendantes du Fe(II) et de l'oxoglutarate sont exprimées dans BL21AI transformées par le pIJ194-genK. La souche est cultivée dans 25 ml de LB contenant 0,5 % de glucose et 100 µg/ml de kanamycine. Après une incubation sous agitation à 37°C pendant une nuit, cette pré-culture (10 ml) est utilisée pourensemencer 1 l de culture de ZYP-5052<sup>2</sup> (cf. II.4 Milieux de culture) contenant 100 µg/ml de kanamycine. Les bactéries sont cultivées à 37°C, puis lorsque l'A<sub>600</sub> atteint 0,6, la culture est poursuivie sous agitation à 20°C pendant 24 h. La culture est centrifugée à 2000 g pendant 40 min à 4°C, puis le culot bactérien est repris dans 50 ml de tampon de cassage (Tris HCl 50 mM pH 8, NaCl 0,5 M, imidazole 20 mM, PMSF 1 mM, benzamidine hydrochloride 1 mM, glycerol 5%, DTT 1 mM), congelé et lysé en utilisant une presse d'Eaton. Après décongélation sur lit de la glace, le lysat est traité par de la benzonase (20 unités/ml) en présence de 2 mM de MgCl<sub>2</sub>, est incubé 1 h sur un lit de glace puis est centrifugé à 2000 g pendant 40 min à 4°C. La fraction soluble est congelée dans de l'azote liquide et conservé à -80°C. Une fois décongelé, l'extrait protéique est purifié sur une colonne HisTrap HP de 5 ml (GE-Healthcare Bio-Sciences). Il est chargé sur la colonne équilibrée avec le tampon A (Tris HCl 50 mM pH 8, NaCl 0,5 M, imidazole 20 mM, DTT 1 mM, glycérol 5%). La protéine est éluée avec un gradient d'imidazole (20 mM à 1 M) de 100 ml avec un débit de 1ml/min. Les fractions contenant la protéine K sont rassemblées et sont dialysées contre 3 × 1,5 L de tampon de dialyse (Hepes 20 mM pH 7,5, NaCl 0,1 M, DTT 1 mM, benzamidine hydrochloride hydrate 0,5 mM, PMSF 1 mM) pendant 36 h à 4°C (Spectra/Por, Interchim ; MWCO : 12-14000 Da ; Vol/length : 2 mL/cm). La protéine purifiée obtenue est concentrée avec une unité de concentration de type Amicon (seuil 10000 Da) à environ 35 mg/ml puis congelée dans l'azote liquide pour être conservée à -80°C. Environ 67 mg de protéine ont été obtenus à partir d'un litre de culture.

## IX. Test d'activités *in vitro*

### 1. CDPS

#### 1.1 Test avec lysat soluble

Un lysat soluble d' *E. coli* ne produisant pas de CDPS est préparé comme précédemment décrit par l'équipe<sup>26</sup>. La réaction est réalisée dans 100 µl avec 40 µl de lysat soluble dans le tampon suivant : 5 mM ATP, 0,1 mM β-mercaptoéthanol, 15 mM MgCl<sub>2</sub>, 50 mM KCl, 0,1 mM EDTA et 50 mM HEPES-KOH pH 7,5. Le milieu est pré-incubé à 30°C pendant 15 min. Pour les premiers essais avec AlbC, 1, 5 ou 10 µM d'AlbC dans un tampon 0 ou 200 mM imidazole sont ajoutés. Pour les reconstitutions d'activités *in vitro* des CDPS purifiées (cf. Chapitre II. VI.1 Sélection et purification de 16 CDPS), 1 ou 10 µM d'enzyme purifiée est (sont) ajouté(s). Le milieu réactionnel est alors incubé pendant une nuit à 30°C. La réaction est arrêtée par 2% (v/v) de TFA. Après centrifugation à 20000 g, l'échantillon est analysé par LC-MS/MS selon la même méthode que les surnageants de culture (cf. VI.2 Analyses des surnageants de culture par LC-MS/MS).

#### 1.2 Test couplé avec ARNt et aaRS purifiés

Pour les CDPS n'ayant pas montré d'activité dans le test avec le lysat soluble, un test couplé est réalisé. Dans ce test, les aaRS génèrent *in situ* les substrats aa-ARNt des CDPS. Les tests sont réalisés dans 100µl dans le même tampon que les tests avec lysat soluble (5 mM ATP, 0,1 mM β-mercaptoéthanol, 15 mM MgCl<sub>2</sub>, 50 mM KCl, 0,1 mM EDTA et 50 mM HEPES-KOH pH 7,5) avec 50 µM de chaque acide aminé (Amino Acid Mixture, Complete, 1 mM de chacun des 20 acides aminés essentiels, Promega), 160 µM ARNt total (ARNt de *E. coli*, Sigma) et 0,18 mg/ml de aaRS total (Sigma) et 10 µl de CDPS purifiée. La réaction est réalisée à 30°C avec une pré-incubation de 15 min avant l'ajout de la CDPS. La réaction est continuée pendant une heure avant d'être traitée comme précédemment décrit (cf. IX.1.1 Test avec lysat soluble)

### 2. Dioxygénases dépendant du Fe(II) et du 2-oxoglutarate

**Test générique.** Le test est adapté de celui décrit par Mazmouz *et al.*<sup>396</sup>. Il est réalisé dans un volume final de 200 µl et contient 50 mM HEPES pH 7,5, 100 µM DTT, 50 µM FeSO<sub>4</sub>, 2 mM L-ascorbate, 500 µM 2-oxoglutarate, 100 µg/ml catalase (d'origine de foie de bœuf), 5 µM de dioxygénase purifiée et 1 mM cyclo(Ala-Glu). La réaction est débutée par l'ajout de cyclo(Ala-Glu) au milieu réactionnel, préalablement pré-incubé à 30°C, et est incubée à 30°C pendant 1 h. La réaction est arrêtée par l'ajout d'acide formique à une concentration finale de 2% (v/v). L'échantillon est ensuite centrifugé à 20000 g pendant 30 min avant d'être analysé par HPLC (Elite LaChrom VWR Hitachi) ou LC-MS/MS sur la colonne colonne Hypercarb

(4.6 mm x150 mm, 5  $\mu$ m, 250 Å, ThermoScientific) en utilisant le gradient 1 décrit ci-dessus (cf. VI.2 Analyses des surnageants de culture par LC-MS/MS).

**Etude cinétique préliminaire et dépendance aux cofacteurs.** Pour réaliser le suivi de la réaction dans le temps, la réaction est réalisée dans 1 ml final et 150  $\mu$ l sont prélevés et acidifiés pour l'analyse à des temps d'incubation variable (t=0 min, 1 min, 3 min, 10 min, 30 min, 60 min). Pour étudier la dépendance de la réaction aux constituants du milieu, plusieurs conditions de réaction sont testées comme décrit dans la partie résultats.

**Optimisation et changement d'échelle de la réaction de biotransformation.** Afin de pouvoir réaliser la biotransformation du cyclo(Ala-Glu) pour la purification des produits et leur analyse subséquente par RMN, une optimisation et un changement d'échelle de la réaction sont entrepris. Des essais dans 500  $\mu$ l de milieu réactionnel sont réalisés. Plusieurs concentrations en substrat (5, 10 et 20 mM) et en enzyme (10 et 20  $\mu$ M) sont testées. La concentration en 2-oxoglutarate est toujours en excès de 1,2 fois par rapport au substrat. À des temps d'incubations variables (1h30, 5h, 14h, 21h et 28h), 50 ou 150  $\mu$ l de milieu réactionnel sont prélevés, acidifiés et analysés par HPLC. Les conditions retenues pour la biotransformation sont : 20 mM cyclo(Ala-Glu), 24 mM 2-oxoglutarate, 10  $\mu$ M dioxygénase K purifiée, les autres concentrations restant identiques à ce qui a été décrit ci-dessus. La réaction est réalisée dans 5 ml répartis en 10 tubes. Le milieu réactionnel est alors incubé pendant 21 h à 30°C, puis acidifiée à 2% final (v/v) avec de l'acide formique.

CHAPITRE I:  
MÉTHODE DE CARACTÉRISATION MOYEN-  
DÉBIT DES ACTIVITÉS DES ENZYMES DES  
VOIES DEPENDANT DES CDPS

Au début de mon travail de thèse, seulement neuf CDPS et trois voies de biosynthèse associées à trois de ces CDPS avaient été décrites dans la littérature<sup>26,27,25,29,367</sup>. Ces CDPS synthétisent majoritairement des cyclodipeptides contenant des chaînes latérales hydrophobes, laissant penser que le potentiel de biosynthèse de ces voies se limite à ce type de composés. Cependant, plusieurs analyses bio-informatiques<sup>28,345</sup> suggéraient la présence de nombreuses voies dépendant des CDPS dans les bases de données génomiques.

Afin d'explorer au mieux la diversité structurale des dicétopipérazines biosynthétisées par les voies à CDPS identifiées dans les bases de données, nous avons mis au point un système pour caractériser en parallèle plusieurs activités des enzymes constituant ces voies (CDPS et enzymes de modification). L'objectif des travaux décrits dans ce chapitre a donc été de développer, en moyen débit, une méthode de clonage, d'expression et d'analyse de l'activité des CDPS et des enzymes de modifications potentielles.

Le protocole utilisé en routine au laboratoire pour caractériser les activités de synthèse de CDPS constitue le point de départ de ce développement. Dans ce protocole, les gènes des CDPS sont amplifiés à partir de l'ADN génomique du microorganisme, puis clonés dans le vecteur commercial pQE60 pour leur expression chez *E. coli*. Le plasmide ainsi obtenu sert à la transformation de bactéries M15-pREP4 ou BL21AI, qui expriment, après induction à l'IPTG, la CDPS de manière recombinante. Cette dernière synthétise alors les cyclodipeptides qui sont retrouvés dans le milieu de culture<sup>26</sup>. L'identification des cyclodipeptides présents dans les surnageants de culture obtenus après centrifugation des cultures est réalisée par HPLC en phase inverse couplée à la spectrométrie de masse par électronébulisation. La colonne chromatographique utilisée est une colonne dC18 avec *end-capping* permettant son utilisation en 100% H<sub>2</sub>O. Par ailleurs, les fractions soluble et insoluble des culots bactériens sont obtenues après cassage par presse d'Eaton et analysées par gel SDS-PAGE pour déterminer le contenu protéique et évaluer le taux d'expression des CDPS. Ce protocole très efficace pour l'étude d'une CDPS ne prévoit pas la co-expression de CDPS et d'enzymes de modification. De plus, il ne permet pas d'analyser plus de quelques CDPS simultanément, du fait des volumes de culture importants et de la technique de cassage utilisée.

Le développement du nouveau protocole a donc consisté en la conception et l'optimisation d'un système de clonage à plusieurs plasmides adapté aux études des CDPS et des enzymes de modification (partie I), la mise au point d'un test d'activité en moyen débit (partie II), et l'introduction d'un nouveau mode de séparation chromatographique complémentaire de celui existant, améliorant la détection des cyclodipeptides (partie III).

# I. Mise au point du système de clonage

## 1. Choix et mise en place du système de clonage

### 1.1 Système à deux vecteurs

Mon projet de thèse s'intéresse aux enzymes des voies impliquant les CDPS, et comprend donc l'étude des CDPS et des enzymes de modifications, seules ou associées. Le système de plasmides à mettre au point devait donc permettre de pouvoir exprimer concomitamment une ou plusieurs enzymes dans une seule bactérie *E. coli*.

Avant mon arrivée au laboratoire, un système de plasmides avait été mis au point par Pascal Belin. Il comprenait deux plasmides ayant des origines de répllication différentes et des cassettes de résistance aux antibiotiques différentes également : pQE60 (Qiagen ; origine de répllication ColE1 ; résistance à l'ampicilline; ~40 copies/cellule) et pRSF-Duet (Novagen ; origine de répllication RSF1030 ; résistance à la kanamycine; > 100 copies/cellule). Dans notre étude, le pQE60 est dédié au clonage des gènes de CDPS entre les sites *NcoI* et *BamHI* ou *BglII*. Le pRSF-Duet est dédié au clonage des enzymes de modification entre les sites *NcoI* et *KpnI*.

### 1.2 Système permettant de rassembler plusieurs enzymes dans un vecteur

Ce système à deux plasmides permet donc de co-exprimer une CDPS et une enzyme de modification au sein d'une même bactérie. Cependant, les voies à CDPS peuvent contenir plus d'une enzyme de modification. Le système d'expression doit donc donner la possibilité de co-exprimer plusieurs enzymes de modification en même temps.

Le système retenu permet de cloner, les uns derrière les autres, les différents gènes d'enzymes de modification, sous contrôle de leur propre promoteur et suivis de leur propre terminateur de transcription. Nous avons utilisé les sites de restriction *SpeI* et *XbaI* qui sont compatibles pour encadrer chaque unité de transcription en respectivement 5' et 3'. Ainsi l'unité *SpeI-XbaI* à ajouter sera clonée de manière directionnelle dans le site *XbaI* de l'unité à agrandir, donnant une double unité de transcription encadrée par *SpeI* et 5' et *XbaI* en 3' (Figure 48). Ce système s'inspire des travaux de Watanabe et Oikawa pour exprimer chez *E. coli* les différents gènes de voies de biosynthèse dépendant de NRPS et PKS<sup>397</sup>.

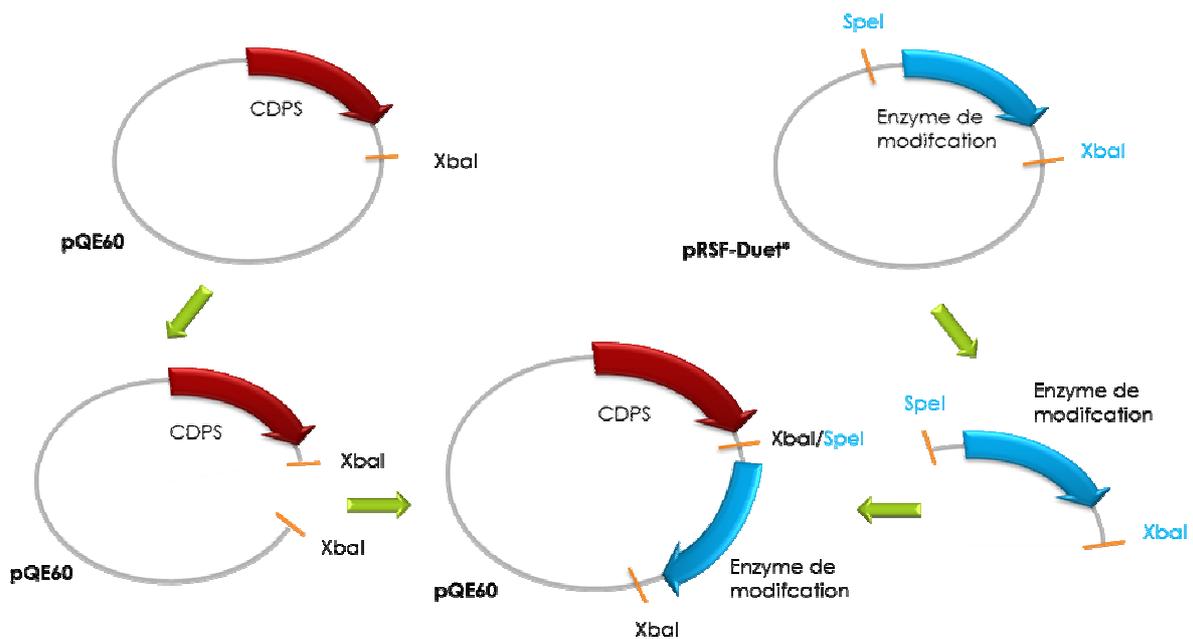


Figure 48 : Stratégie d'insertion d'un gène d'enzyme de modification dans pQE60

La première étape a consisté à supprimer du pRSF-Duet le site *XbaI* en position 2414 (cf. Matériels et méthodes III.1.1 Suppression du site *XbaI*) pour obtenir le vecteur pRSF-Duet(*XbaI*-). La seconde étape a consisté à ajouter, de part et d'autre des sites de clonages, un site *SpeI* et un site *XbaI*. Afin d'ajouter ces deux sites simultanément, une cassette mutée du site de clonage contenant ces deux sites de restriction a été préparée par la technique de PCR SOEing (Splicing by Overlap Extension). Cette dernière a ensuite été insérée dans le pRSF-Duet(*XbaI*-) grâce aux sites *AclI* situé en amont du site de clonage et *AgeI* situé en aval pour obtenir le vecteur désiré, baptisé pIJ194 selon la nomenclature mise en place dans le laboratoire (Figure 49) (cf. Matériels et méthodes III.1.2 Ajout des sites *SpeI* et *XbaI* en préparant une cassette par PCR SOEing).

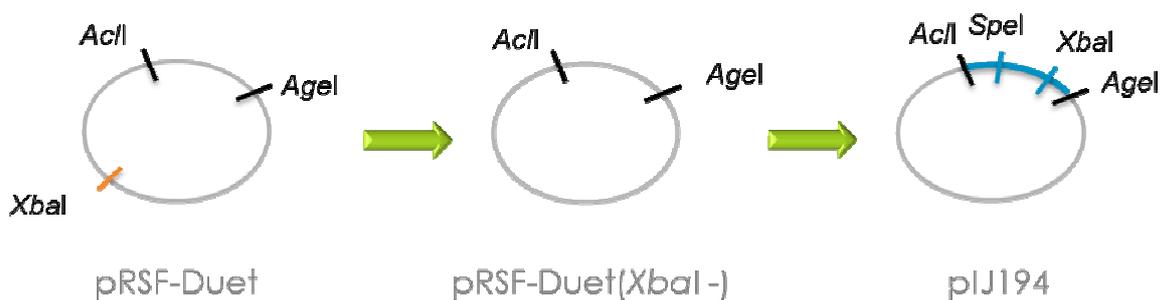


Figure 49 : Etapes pour la modification du pRSF-Duet

Suppression du site *XbaI* puis remplacement du fragment contenu entre *AclI* et *AgeI* par une cassette contenant les sites *SpeI* et *XbaI*

De cette manière, une CDPS et une enzyme de modification peuvent être rassemblées sur le même plasmide pQE60. En effet, une digestion du vecteur pIJ194 par les enzymes de restrictions *SpeI* et *XbaI* permet d'extraire un insert contenant le gène de l'enzyme de modification, ainsi que le promoteur

(T7) et les séquences opérateur *lacO*. Il peut alors être inséré dans le vecteur pQE60 digéré en *XbaI*, les sites *SpeI* et *XbaI* étant compatibles (Figure 48). Un seul site *XbaI* est alors reformé et permet si nécessaire d'incorporer un nouveau gène.

La solution retenue permet donc de rassembler sur le même vecteur pQE60 une CDPS et une ou plusieurs enzymes de modification. Ce vecteur peut alors être utilisé seul ou en combinaison avec un plasmide de type pIJ194 portant une autre enzyme de modification.

## 2. Optimisation du vecteur d'expression pour les CDPS

Les premiers essais de clonages de gènes de CDPS dans le vecteur pQE60 ont été réalisés avec des fragments d'ADN amplifiés à partir d'ADN génomique (I.3 Choix de l'origine des gènes). Deux CDPS ont été sélectionnées pour les premiers essais car elles faisaient partie de la sélection de CDPS potentielles à étudier (cf Chapitre II. I.3 Sélection des CDPS à étudier) et leur ADN génomique était disponible au laboratoire. Il s'agit de SJA\_c1-32620 de *Spingobium japonicum* UT26S et de pc1814 de *Candidatus* Protochlamydia amoebophila UWE25. Les CDPS potentielles sont désignées ici par le locus de leur gène. Dans le cas de pc1814, deux versions du gène ont été amplifiées. La première correspond à la séquence correspondant à pc1814 dans les bases de données. La seconde, nommée pc1814long, a été allongée de 34 résidus en N-terminal, car un codon initiateur plus probable, compte tenu des alignements de séquence protéique avec les CDPS connues, a été déterminé (Figure 50).

```

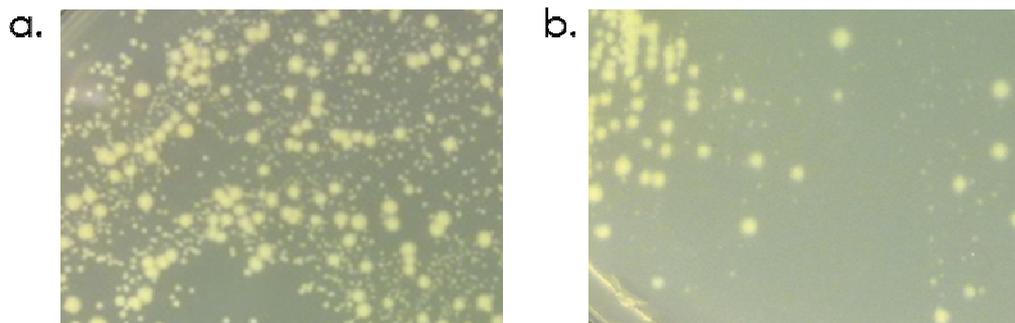
AlbC          -----M LAGLVPAPDH GMREEILGDR SRLIRQRGEH ALIGI---SA GNSYFSQKNT VMLLQWAGQR
pc1814        -----M LAGLVPAPDH GMREEILGDR SRLIRQRGEH ALIGI---SA GNSYFSQKNT VMLLQWAGQR
pc1814long    MNKYTWKKNF TAILSNQKEI IKGKLAFKTE ETHWMETEKP VIILVSTHSS FHEKISGDLK MNAFVSTIRN

```

Figure 50 : Alignement du début de la séquence d'AlbC, de pc1814 et de pc1814long

### 2.1 Le clonage de gènes de CDPS dans pQE60 induit une toxicité chez *E. coli*

En observant les colonies obtenues lors du clonage des gènes de CDPS SJA\_c1-32620, pc1814, et pc1814long, une hétérogénéité dans la taille des colonies est observable. Des colonies de taille normale « côtoient » des colonies de plus petite taille (Figure 51a.). Lorsque la présence d'un insert dans le site de clonage est testée par PCR sur colonies, seulement 13% des clones testés pour les colonies de taille normale sont positifs contre 100 % pour les colonies de petite taille. Les plasmides issus des clones positifs des colonies de taille normale sont séquencés et tous présentent dans le gène de CDPS, soit une mutation qui introduit un codon « stop » précoce, soit une délétion qui induit un décalage de phase puis un codon « stop ». Dans ces deux cas, seule une protéine abortive peut être synthétisée. Lorsqu'une petite colonie est repiquée sur une boîte de Pétri, elle donne des colonies de petite taille mais aussi des colonies de taille plus importante (Figure 51b.), ce qui suggère une instabilité dans la croissance qui ne permet pas de travailler avec la souche.



**Figure 51 : Toxicité liée au clonage des CDPS**

**a. Colonies obtenues après transformation par le vecteur pQE60-SJA\_c1-32620 ; des colonies de taille classique et de petite taille sont visibles, b. Colonies obtenues après le repiquage de l'une des petites colonies.**

L'absence d'insertion d'un gène ou l'insertion d'un gène non fonctionnel corrèle avec une croissance normale de la bactérie. A l'inverse, l'insertion d'un gène fonctionnel semble induire une toxicité qui provoque une diminution de la vitesse de croissance bactérienne. Cette toxicité associée au clonage d'un gène fonctionnel peut être due à la faible expression de la CDPS dans des conditions de non-induction, ce que l'on appelle une fuite au niveau d'un promoteur. Le système promoteur utilisé sur le pQE60 est composé d'un promoteur constitutif, PT5, suivi de sites opérateurs *lacO* sur lesquels peuvent se fixer le répresseur de l'opéron lactose, LacI. En se liant, LacI empêche l'ARN polymérase de synthétiser l'ARN correspondant. Une fuite souvent observée avec ce type de système est due à une trop faible quantité de LacI dans la cellule pour lier efficacement tous les sites *lacO* présents (le plasmide pQE60 est présent à environ 50 copies par cellule). Une alternative consiste à augmenter le nombre de copies du gène *lacI* dans la cellule afin d'augmenter la quantité de protéine LacI synthétisée.

## **2.2 L'introduction du gène *lacI* dans le plasmide pQE60 permet un clonage stable**

Une cassette contenant le gène *lacI* sous contrôle de son propre promoteur, a été obtenue depuis le vecteur pREP4 (Qiagen). Elle a été incorporée dans le plasmide pQE60 pour donner le plasmide pIJ196. Les clonages des trois gènes de CDPS dans ce nouveau plasmide ont été réalisés (entre les sites *NcoI* et *BglII* (compatible avec *BamHI*)). Les colonies obtenues présentaient une taille homogène et plus de 60% des clones testés par PCR sur colonies possédaient un insert de la taille attendue entre les sites *NcoI* et *BglII* de pQE60. Les séquençages des plasmides issus de clones positifs n'ont montré aucune mutation dans les séquences codant les gènes de CDPS.

L'insertion du gène *lacI* permet donc d'augmenter l'efficacité du clonage en diminuant vraisemblablement les fuites d'expression responsables de la toxicité observée précédemment.

Les modifications apportées au pRSF-Duet et au pQE60 pour donner respectivement le pIJ194 et le pIJ196 permettent de disposer d'un système à deux plasmides compatibles pour le clonage des gènes de

CDPS et d'enzymes de modification. Le pIJ196 a notamment été optimisé pour réduire les fuites d'expression afin d'offrir un système de clonage plus stable.

### 3. Choix de l'origine des gènes

Au commencement du projet, nous avons choisi d'amplifier les gènes des CDPS à étudier biochimiquement à partir de l'ADN génomique des organismes d'origine. Cette option présentait notamment les avantages d'être peu onéreuse et de s'affranchir des erreurs de séquençage qui peuvent être présentes dans les bases de données.

Néanmoins, nous avons été rapidement confrontés à la difficulté d'obtenir rapidement l'ADN ou les souches d'intérêt. En effet, ces derniers ne sont pas nécessairement disponibles dans les collections et les auteurs sollicités pour l'envoi d'un échantillon d'ADN ou de souche ne répondent pas systématiquement. Certaines souches sont pathogènes ou sont des bactéries intracellulaires obligatoires qu'il est difficile de manipuler dans des conditions de laboratoire classique. De plus, le pourcentage de GC chez certains organismes possédant une CDPS potentielle est supérieur à 70%, ce qui est souvent associé à une moins bonne amplification des gènes lors de PCR<sup>398</sup>. Par ailleurs, les CDPS peuvent être difficiles à produire chez *E. coli* car les gènes amplifiés par PCR ne sont pas optimisés pour l'expression chez cet organisme.

En revanche, l'utilisation de gènes de synthèse offre de nombreux avantages. Les codons peuvent être optimisés pour l'expression chez *E. coli*. Les structures secondaires sur les acides nucléiques peuvent être éliminées. Dans l'optique de faire une étude de plusieurs dizaines de CDPS en parallèle, l'utilisation des gènes de synthèse permet de standardiser les manipulations à réaliser sur chaque gène en choisissant un système unique de clonage pour tous les gènes, par exemple deux sites de restriction bien définis. De plus, les capacités de synthèse des fournisseurs sont telles que plusieurs dizaines de gènes de 500 pb à 1 kpb peuvent être obtenus en quelques semaines. Tous les gènes présents dans les bases de données peuvent être commandés dans un temps très raisonnable. Le seul bémol à apporter à l'utilisation des gènes de synthèse est l'investissement financier plus important qu'il constitue. Ce dernier est néanmoins compensé par l'économie de temps réalisée d'une part sur l'obtention des gènes et d'autre part sur les problèmes évités.

Il a donc été décidé d'opter pour des gènes synthétiques à la place de gènes naturels.

Les différents éléments permettant un clonage stable, standardisé et adapté aux clonages de plusieurs dizaines de gènes ont été mis en place. Cependant, à mon arrivée au laboratoire, l'étude des activités des CDPS était effectuée enzyme par enzyme, les cultures étant réalisées individuellement en erlenmeyer. Afin de permettre la caractérisation rapide de plusieurs dizaines de CDPS, nous avons établi

un protocole moyen débit en plaque 24 puits pour la culture et la production des CDPS et la préparation des différents extraits.

## II. Mise au point du test d'activité *in vivo* en moyen débit

Le point de départ de cette optimisation a été le protocole utilisé en routine au laboratoire (cf. Introduction de ce chapitre, Matériels et méthodes V.1 Culture en erlenmeyer et VII.1 Culots issus de la culture en erlenmeyer.) Plusieurs aspects du protocole devaient être adaptés pour réaliser un test d'activité moyen débit. Le choix de la souche d'expression s'est révélé particulièrement important et délicat et a constitué un fil rouge au cours de l'optimisation et des tests. En ce qui concerne les conditions de culture, nous avons essentiellement comparé la culture en milieu auto-induit à une culture en milieu classique et déterminé la température de culture la plus adaptée. Enfin, nous avons réalisé le passage à une culture en plaque 24 puits.

### 1. Choix des souches d'*E. coli* pour l'expression des CDPS et des enzymes de modification

Le protocole initial utilise pour la production la souche M15 possédant le plasmide pREP4 (M15-pREP4). Cette souche est recommandée par le fournisseur du plasmide pQE60 utilisé pour le clonage des gènes de CDPS, car elle permet notamment de limiter les fuites au niveau du promoteur sur le pQE60. Ce système M15-pREP4/pQE60 est utilisé depuis plusieurs années dans le laboratoire et a donné entière satisfaction.

Cependant, dans le cadre de mon projet, la possibilité d'exprimer une CDPS et une ou plusieurs enzymes de modification nous a amené à envisager l'utilisation d'une autre souche. En effet, les gènes d'enzymes de modification clonés dans pIJ194 seront sous contrôle d'un promoteur de l'ARN polymérase de T7 (absente dans la souche M15-pREP4) et véhiculés par un plasmide portant un gène de résistance à la kanamycine, tout comme le vecteur pREP4. Notre choix de souche s'est tourné vers BL21AI, qui dérive d'*E. coli* B, et porte sur un prophage le gène de l'ARN polymérase de T7 sous contrôle d'un promoteur inductible à l'arabinose. Il est donc possible d'induire l'expression de gènes sous contrôle d'un promoteur PT7 avec cette souche. De plus, il sera possible de différencier l'expression de CDPS dont le gène est sous contrôle d'un promoteur T5 suivi de site *lacO* (inductible par l'IPTG, quelle que soit la souche) de l'expression d'enzymes de décoration, dont les gènes sont sous contrôle d'un promoteur T7 suivi de site *lacO* (inductible dans BL21AI par la combinaison d'IPTG et d'arabinose).

La souche BL21AI apparaissait alors comme la souche idéale pour notre étude. Néanmoins, lorsque nous avons réalisé les premiers tests pour la mise au point du clonage avec AlbC (CDPS modèle utilisée au laboratoire, exprimée depuis pQE60) et pc1814long (CDPS exprimée depuis pIJ196), nous avons constaté une faible efficacité de transformation des bactéries BL21AI.

En effet, aucun des essais de transformation de BL21AI avec pIJ196-pc1814long n'a permis d'obtenir des bactéries transformées. L'intégrité du plasmide a été vérifiée sur gel et des bactéries transformées avec les souches DH5 $\alpha$  et M15-pREP4 ont été obtenues sans difficulté. Nous avons alors formulé l'hypothèse que la souche BL21AI pouvait être plus sensible que les autres souches à la toxicité due aux fuites d'expression. Les transformations étant réalisées sur un milieu riche (LB agar), pour lequel des cas d'induction par le milieu avaient été rapportés par Studier<sup>393</sup>. Afin d'augmenter la répression catabolique, les transformations ont été étalées sur un milieu LB supplémenté avec 0,2 % de glucose, mais sans succès.

Seule l'utilisation d'une souche de BL21AI-pREP4 a permis d'obtenir des transformants, mais elle empêche la co-transformation du pIJ196 et du pIJ194. Dans un premier temps, l'utilisation de cette souche n'a pas été envisagée de façon systématique car les problèmes de clonages rencontrés avec le gène pc1814long pouvaient ne pas se reproduire avec les autres gènes à tester. Cependant, si la souche BL21AI s'avère trop peu efficace pour la transformation par les plasmides pIJ196-CDPS, il nous faudra envisager l'utilisation de BL21AI-pREP4. Pour la co-transformation avec un plasmide pIJ194 portant un gène d'enzyme de modification, on pourra alors modifier le pREP4 de manière à remplacer son gène de résistance à la kanamycine par un gène de résistance au chloramphénicol ou un gène de résistance à streptomycine et spectinomycine.

Aucune des souches testées ne s'est révélée parfaitement adaptée à notre étude au cours de la mise au point du protocole (Tableau 9). Ceci nous a conduits à utiliser les deux souches (BL21AI et M15-pREP4) parallèlement pour les tests d'activités des nouvelles CDPS.

**Tableau 9 : Bilan des avantages et inconvénients des souches *E. coli* pour l'étude des activités de synthèse en moyen débit**

Souche	Exprime l'ARN polymérase de T7	Compatible avec l'utilisation de pIJ194	Obtention de transformants <sup>1</sup>
M15-pREP4	NON	NON	OUI
BL21AI	OUI	OUI	NON
BL21AI-pREP4	OUI	NON	OUI

<sup>1</sup>lors de la transformation avec le pIJ196-pc1814.

## 2. Expression avec un milieu auto-induit

### 2.1 Pourquoi un milieu auto-induit ?

Le milieu auto-induit est décrit comme une alternative intéressante pour les systèmes de culture et d'expression moyen et haut-débit. En effet, le nombre important de cultures à effectuer en parallèle rend fastidieux le suivi de croissance et l'ajout d'inducteur nécessaires pour une induction classique.

Le milieu auto-induit décrit par Studier<sup>393</sup> permet de déclencher une induction sans avoir à manipuler la culture ni à ajouter un quelconque composé au cours de la croissance. Cette auto-induction

se déclenche au niveau de promoteurs inductibles par une source de carbone (lactose, arabinose, etc ...) et est soumise à la répression catabolique, c'est à dire une répression de l'induction observée lorsque la croissance est faite en présence de glucose. Les milieux auto-induits décrits sont constitués de trois sources de carbone : du glycérol en quantité importante (0,5%), du glucose en faible quantité (0,05%) et du sucre inducteur en quantité importante également (0,2 %).

La consommation préférentielle du glucose s'accompagne d'une diminution dans la cellule du taux d'adénosine monophosphate cyclique (AMPc), ce qui a pour conséquence d'empêcher l'induction au niveau des promoteurs gouvernant la dégradation d'autres sources de carbone (répression catabolique). A l'épuisement du glucose, le glycérol est consommé et la concentration d'AMPc dans la cellule augmente, entraînant une levée de la répression catabolique et l'induction depuis les promoteurs inductibles par une source de carbone en présence de ladite source de carbone.

Les milieux auto-induits ont été décrits pour les gènes dont l'expression est sous le contrôle de l'opéron lactose, mais nous avons adapté les sources de carbone, en ajoutant une faible quantité d'arabinose (0,05 %), pour qu'une auto-induction se fasse avec les gènes dont l'expression est induite par l'arabinose.

De plus, l'auto-induction permet d'obtenir une meilleure reproductibilité, d'atteindre une plus forte concentration bactérienne et une plus forte expression des protéines.

## **2.2 Choix de la température de culture**

L'effet du passage au milieu auto-induit et l'effet de la température de culture ont été étudiés. Quatre cultures de BL21AI transformées par pQE60-AlbC sont réalisées en erlenmeyer. Trois d'entre elles utilisent du milieu auto-induit, tandis que la dernière utilise un milieu classique avec une induction à l'IPTG. Trois conditions de température sont testées avec le milieu auto-induit : 37°C, 37°C jusqu'à ce que la culture soit légèrement trouble puis 20°C et enfin 20°C. La densité bactérienne est évaluée par l'absorbance à 600 nm et la production de cyclodipeptide par l'aire sous le pic chromatographique correspondant au cyclodipeptide majoritaire (cyclo(Phe-Leu)) (Tableau 10).

Pour la culture en milieu auto-induit pour laquelle la température d'incubation passe de 37°C à 20°C, la production de cyclo(Phe-Leu) est du même ordre de grandeur que le témoin induit à l'IPTG. Cette condition apparaît être un bon compromis entre une culture à 20°C un peu lente qui ne produit pas suffisamment de cyclodipeptides en 24 h et une culture à 37°C, trop rapide qui ne semble pas favoriser la production de cyclodipeptides à 48 h. C'est donc cette condition qui est retenue pour la suite de l'optimisation.

**Tableau 10 : Suivi de la croissance bactérienne et de la production de cyclo(Phe-Leu) en fonction du milieu et de la température de culture à 24 et 48 h**

Induction	Température de la culture	24 h		48 h	
		A <sub>600</sub>	Cyclo(Phe-Leu) <sup>1</sup> (unité arbitraire)	A <sub>600</sub>	Cyclo(Phe-Leu) (unité arbitraire)
Auto-induction	37°C	2,98	3140	4,19	4809
Auto-induction	37°C puis 20°C	1,81	2955	4,69	7216
Auto-induction	20°C	1,31	1517	4,42	6660
IPTG	37°C puis 20°C	1,63	3553	4,45	7058

<sup>1</sup>Aire sous le pic d'un chromatogramme HPLC enregistré à 214 nm.

### 2.3 Milieu auto-induit et M15-pREP4 : un écueil à éviter

Avant que l'utilisation du pREP4 ne vienne à bout des difficultés de transformation avec la souche BL21AI, il avait été envisagé d'utiliser uniquement la souche M15-pREP4, étant donné qu'aucun problème de clonage n'avait été recensé. Nous avons donc voulu utiliser la souche M15-PREP4 en milieu auto-induit. De très faibles quantités de cyclodipeptides ont été observées dans les surnageants après 24 h de culture de bactéries M15-pREP4 transformées par pQE60-AlbC. Cette souche poussant moins vite que les BL21AI, nous avons alors émis l'hypothèse que l'induction de l'expression de la CDPS avait lieu beaucoup plus tardivement et que les cyclodipeptides produits seraient mieux détectables à l'issue d'une culture plus longue. Des surnageants de culture de M15-pREP4 et de BL21AI après 24 h, 30 h et 48 h de culture ont alors été analysés et comparés (Tableau 11). La production de cyclo(Phe-Leu) est pratiquement nulle, quel que soit le temps de culture en M15-pREP4 alors que la souche BL21AI produit d'importantes quantités de cyclo(Phe-Leu). L'hypothèse d'une auto-induction plus tardive est donc invalidée. Néanmoins, la production, certes faible mais non nulle, de cyclodipeptide laisse à penser que l'induction de l'expression de la protéine n'a pas lieu comme elle le devrait.

**Tableau 11 : Suivi à 24, 30 et 48 h de la production de cyclo(Phe-Leu) en milieu auto-induit en fonction de la souche (BL21AI ou M15-pREP4) utilisée**

Souche	Plasmide testé	Milieu	Cyclo(Phe-Leu) <sup>1</sup> (unité arbitraire)		
			24 h	30 h	48 h
BL21AI	pQE60-AlbC	Auto-induction	10006	16310	13744
M15-pREP4	pQE60-AlbC	Auto-induction	215	278	429

<sup>1</sup>Aire sous le pic d'un chromatogramme HPLC enregistré à 214 nm.

En s'intéressant de plus près au mécanisme d'induction par le lactose de l'expression des protéines sous contrôle de l'opéron lactose, une explication satisfaisante a été trouvée. En l'absence de glucose, le lactose présent dans le milieu pénètre dans les bactéries et est métabolisé par la  $\beta$ -galactosidase pour

donner l'allolactose, un isomère du lactose. C'est l'allolactose qui interagit avec la protéine LacI et l'empêche de se lier à l'opérateur *lacO*. Cela lève l'inhibition et induit l'expression du gène cloné. Or, en étudiant le génotype de la souche M15-pREP4, on s'aperçoit que la souche est *lac-*, c'est-à-dire n'exprime pas de  $\beta$ -galactosidase fonctionnelle et n'est donc pas capable de métaboliser le lactose en allolactose. Le lactose seul n'est pas capable d'induire l'expression des gènes clonés. Par ailleurs, il est à noter que l'IPTG est un analogue de l'allolactose et non du lactose.

L'auto-induction de l'expression des gènes de CDPS n'a donc pas lieu avec la souche M15-pREP4, ce qui explique les faibles quantités de cyclodipeptides détectées.

Le milieu auto-induit permet donc d'obtenir une croissance bactérienne satisfaisante et une production de cyclodipeptides au moins aussi bonne qu'avec le protocole classique, sans nécessiter les manipulations exigées par l'induction classique à l'IPTG. Néanmoins, seule la souche BL21AI permet de bénéficier des avantages du milieu auto-induit. L'utilisation de la souche M15-pREP4 nécessitera donc une induction manuelle avec ajout d'IPTG.

### 3. Protocole de culture en plaque 24 puits

Afin de pouvoir mener en parallèle l'étude de l'activité de dizaines de CDPS ou d'enzymes de modification potentielles, le protocole de culture initial a été modifié.

Dans le protocole initial, les cultures étaient réalisées individuellement dans des erlenmeyers. Le volume de surnageant obtenu était de 20 ml alors que seulement 50 à 100  $\mu$ l de surnageant sont nécessaires pour une analyse LC-MS/MS. Par ailleurs, des travaux menés au sein du laboratoire avaient montré que l'analyse du contenu protéique d'un culot bactérien pouvait se faire à partir de quelques centaines de  $\mu$ l de culture<sup>395</sup>. Le volume de culture pouvait ainsi être réduit pour permettre une culture en plaque.

Nous avons choisi de faire les cultures dans des plaques de 24 puits « deepwell » de 10 ml de capacité. Avec un volume de 2 ml par puits, le surnageant de culture et le culot bactérien obtenus sont suffisants pour plusieurs analyses LC-MS/MS et pour l'analyse du contenu protéique par gel SDS-PAGE. Aucun phénomène de décantation de la culture de 2 ml n'est observé lorsqu'elle est agitée à 200 tours/min dans l'incubateur, ce qui indique que l'agitation est suffisante pour homogénéiser la culture.

Les transformations bactériennes sont réalisées simultanément dans une plaque 24 puits, puis ces dernières sont étalées sur du LB-agar disposé dans des puits de plaques 6 puits. Les transformants obtenus servent à l'ensemencement des pré-cultures de 2 ml dans les puits de plaques 24 puits. Les pré-cultures sont réalisées dans du milieu minimum. Après une nuit d'incubation à 37°C sous agitation, 100  $\mu$ l de pré-culture sont prélevés pour ensemercer 2 ml de milieu auto-induit de type M9. Au bout de 3 h 30, temps

après lequel la culture est considérée comme légèrement turbide, la culture est passée de 37 à 20°C, puis l'incubation est continuée jusqu'à 24 h. Cette incubation de quelques heures à 37°C permet à la culture de s'amorcer et de démarrer plus rapidement. Les culots bactériens sont séparés du surnageant de culture par centrifugation directe de la plaque à 2800 g. 500 µl de surnageant sont prélevés et acidifiés, tandis que les culots bactériens sont séchés et congelés à -80°C. Le contenu protéique soluble et insoluble des culots obtenus est analysé par SDS-PAGE. Pour réaliser la lyse nécessaire à l'analyse du contenu protéique des culots bactériens en format 24 puits, une lyse chimique est choisie (cf. Matériels et méthodes VII.2 Culots issus de la culture en plaque 24 puits). La solution de lyse utilisée est inspirée des travaux de l'équipe<sup>395</sup> et permet d'obtenir des profils de séparation soluble/insoluble satisfaisants (cf. Figure 67 et Figure 75).

La culture bactérienne en plaque 24 puits permet la miniaturisation des échantillons et leur traitement en parallèle et elle est donc particulièrement adaptée à l'étude que nous souhaitons mener.

### **III. Introduction d'un nouveau mode de séparation chromatographique**

Les cyclodipeptides produits par les CDPS recombinantes sont retrouvés dans les surnageants de culture après 24 et 48 h de culture<sup>26</sup>. La chromatographie liquide couplée à la spectrométrie de masse en électro-ébulisaison est une méthode d'analyse particulièrement adaptée pour l'identification des dicétopipérazines (cf. Étude bibliographique. I.4 L'identification des 2,5-dicétopipérazines) et est la méthode de référence utilisée au laboratoire pour l'analyse des surnageants de culture.

#### **1. Analyse réalisée en routine au laboratoire**

Un milieu minimum contient uniquement des sels, des sources de carbone, des oligoéléments et des vitamines, contrairement à un milieu riche constitué de tryptones et d'extraits de levures. Ainsi, les surnageants issus de culture en milieu minimum peuvent être analysés par HPLC sans nécessiter de prétraitement lourd de l'échantillon, le bruit de fond faible permettant de détecter les molécules d'intérêt. Les surnageants de culture sont uniquement acidifiés à l'aide du TFA et ensuite centrifugés pour faire tomber les précipités avant d'être injectés sur la colonne. Les analyses de surnageants de culture sont réalisées par HPLC en phase inverse sur une colonne dC18 Atlantis (Waters). Cette colonne est résistante à une gamme de pH étendue et supporte l'utilisation d'éluant 100 % aqueux. Les surnageants de culture contiennent de nombreux sels (nécessaires à la croissance bactérienne) qui ne doivent pas être injectés dans la trappe ionique, car ils pollueraient la source dès le début de l'analyse et engendreraient un bruit de fond important au cours de l'analyse. Ces sels sont élués pendant la phase de lavage du gradient (10 premières minutes du gradient) (cf. Matériels et méthodes. VI.2 Analyses des surnageants de culture par LC-MS/MS). L'analyse par spectrométrie de masse ne commence qu'à partir de 8 min une fois que les sels ont été éliminés de la colonne.

## 2. La colonne dC18 n'est pas adaptée à l'analyse des cyclodipeptides polaires ou de petite taille

L'utilisation de cette colonne nous donnait satisfaction pour l'analyse de cyclodipeptides avec des chaînes latérales hydrophobes. Cependant, l'un des objectifs des tests d'activité de CDPS était de rechercher si certaines nouvelles CDPS étaient capables de synthétiser des cyclodipeptides incorporant des acides aminés polaires. La capacité de rétention de la colonne dC18 pour des cyclodipeptides polaires de petite taille a été testée. Cinq cyclodipeptides disponibles commercialement (cyclo(Gly-His), cyclo(Gly-Ser), cyclo(Gly-Ala), cyclo(Gly-Gln) et cyclo(Gly-Glu)) ont été analysés sur dC18 en utilisant le gradient d'éluion 1 (cf. Matériels et méthodes VI.2 Analyses des surnageants de culture par LC-MS/MS). Les temps de rétention de ces composés sont tous inférieurs au temps à partir duquel l'acquisition des paramètres de masse est lancée ( $t=8$  min) (Tableau 12). Dans ces conditions d'analyse, ces composés ne peuvent donc pas être identifiés, s'ils sont présents dans les surnageants de culture. Une autre colonne d'analyse doit être utilisée afin de pouvoir analyser les cyclodipeptides polaires.

**Tableau 12 : Temps de rétention des cyclodipeptides polaires analysés par HPLC sur une colonne Atlantis dC18**

Cyclodipeptides	Temps de rétention (min)
Cyclo(Gly-His)	4,5
Cyclo(Gly-Ser)	5,1
Cyclo(Gly-Ala)	7,0
Cyclo(Gly-Gln)	6,1
Cyclo(Gly-Glu)	7,1

## 3. La colonne Hypercarb permet d'analyser une nouvelle gamme de cyclodipeptides

Notre choix s'est porté sur la colonne Hypercarb de ThermoScientific. Cette colonne est composée à 100% de carbone graphite poreux. Le mode de séparation est décrit comme orthogonal à celui des C18, c'est-à-dire que les mécanismes de séparation sont complètement différents des mécanismes de séparation sur une colonne de type C18. Cette colonne est présentée par le fournisseur comme stable sur une large gamme de pH (1-14) et est connue pour avoir une forte rétention des analytes polaires et pour permettre de séparer des composés de structures proches.

Les cinq cyclodipeptides précédents sont analysés sur cette nouvelle colonne. Les temps de rétention obtenus sont supérieurs à ceux obtenus avec la colonne dC18 et supérieurs à 8 min, ce qui permet une acquisition de leur spectre de masse. Ce type de composés peut désormais être détecté lorsqu'il est présent dans un surnageant de culture.

**Tableau 13 : Temps de rétention des cyclodipeptides polaires analysés par HPLC sur une colonne HyperCarb**

Cyclodipeptides	Temps de rétention (min)
Cyclo(Gly-His)	19,2
Cyclo(Gly-Ser)	19,1
Cyclo(Gly-Ala)	19,4
Cyclo(Gly-Gln)	20,4
Cyclo(Gly-Glu)	21,8

Les cyclodipeptides hydrophobes sont en revanche trop retenus sur la colonne Hypercarb et sont élués sous forme de pics élargis ou lors du lavage de la colonne. Cette colonne ne peut donc pas être utilisée seule pour l'analyse des surnageants de culture.

C'est pourquoi, nous avons choisi d'analyser chaque surnageant de culture sur la colonne dC18 pour détecter et identifier les cyclodipeptides moyennement hydrophobes à très hydrophobes et la colonne Hypercarb pour les cyclodipeptides polaires et de petite taille.

#### **IV. Discussion, conclusion et perspectives**

Cette phase de mise au point et de test a permis de développer un système de criblage *in vivo* moyen-débit des activités de synthèse de dicétopipérazines des enzymes des voies à CDPS. La stratégie mise en place s'organise en plusieurs étapes à savoir : i) la commande de gènes de synthèse, ii) le sous-clonage dans un système de vecteur adapté limitant les fuites d'expression et permettant la co-expression de plusieurs enzymes, iii) la transformation de M15-pREP4 (avec une efficacité importante), de BL21AI (avec un risque de toxicité) ou de BL21AI-pREP4, iv) la culture dans un système moyen débit (étalement sur LB-agar, pré-culture et culture dans un milieu minimum à induire (M15-pREP4) ou auto-induit (BL21AI)), v) l'analyse du contenu protéique des culots bactériens avec une lyse chimique *in situ* et vi) l'analyse LC-MS/MS des surnageants de culture sur deux colonnes aux modes de séparation orthogonaux (Figure 52).

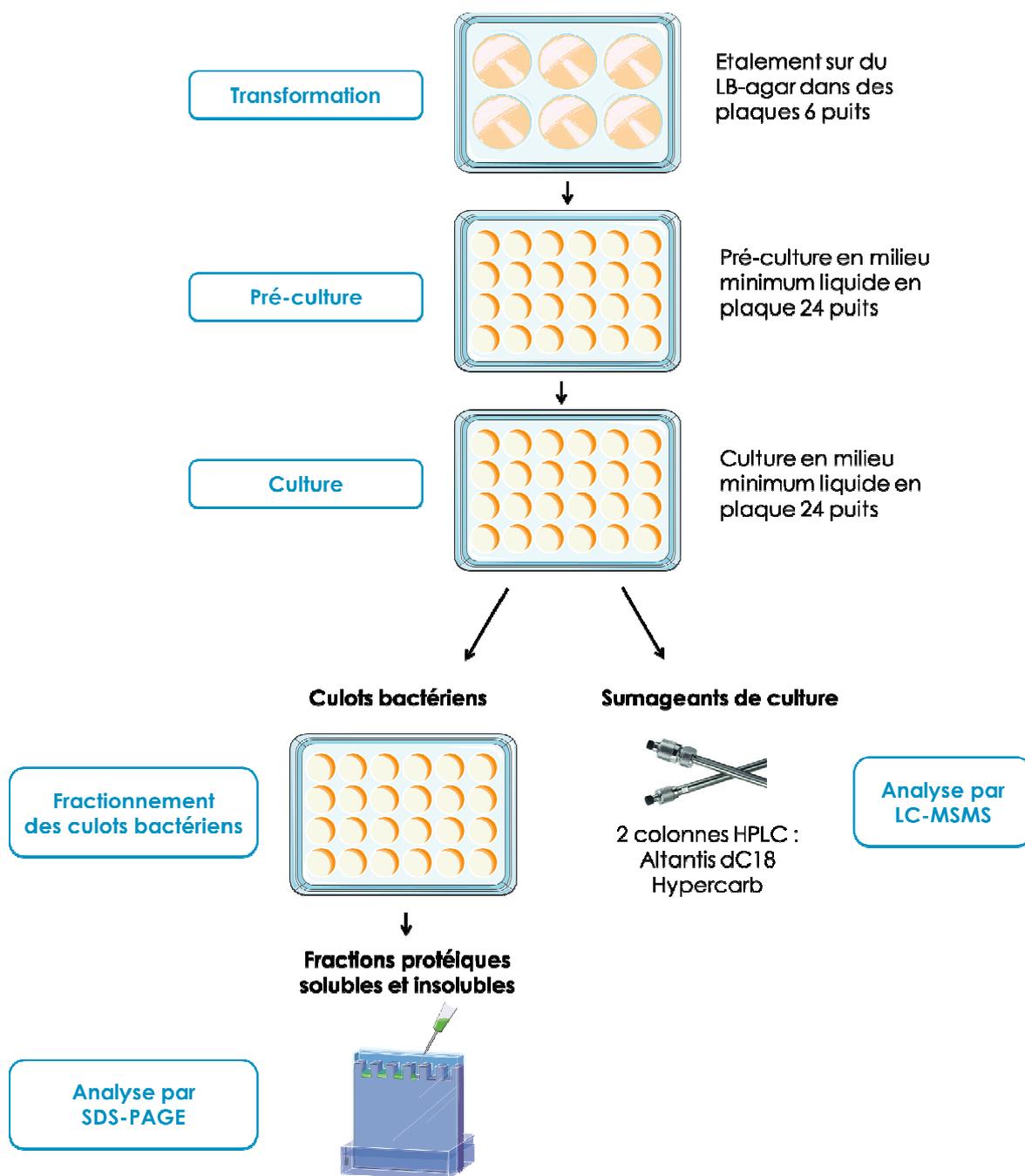


Figure 52 : Stratégie expérimentale mise en place pour le criblage des activités de synthèse et l'expression des enzymes des voies à CDPS

Au cours de cette optimisation, la répression des fuites d'expression s'est révélée un point particulièrement important. A deux reprises, lors du clonage des gènes des CDPS dans le pQE60 puis lors de la transformation des cellules BL21AI avec le pIJ196-pc1814long, nous avons rencontré des problèmes pour obtenir des transformants. Certaines séquences de CDPS semblent induire une toxicité due à des fuites d'expression. Ces fuites sont d'ailleurs visibles lorsque la souche M15-pREP4 est cultivée en milieu minimum auto-induit. En effet, les niveaux de production des cyclodipeptides ne sont pas nuls alors que

l'expression des protéines n'était pas induite du fait de l'absence d'activité de la  $\beta$ -galactosidase. Dans les deux cas, l'efficacité du clonage ou de la transformation a été rétablie en augmentant le nombre de copie du gène *lacI*, et donc la concentration intracellulaire de la protéine LacI, responsable de la répression de l'expression des gènes clonés. Une autre solution décrite par Studier<sup>393</sup> aurait pu mener à l'obtention de colonies pour les transformations difficiles. Elle consiste à étaler les bactéries transformées sur milieu solide défini non inducteur plutôt que sur un milieu riche non défini. En effet, l'activité inductrice du LB (liée à la présence d'une faible quantité de lactose) peut suffire à provoquer l'expression de la CDPS, et dans le cas où celle-ci est toxique, empêcher la formation de colonies.

L'origine de la toxicité de la CDPS pc1814long reste néanmoins en suspens. En effet, la toxicité provient-elle de la protéine elle-même, de son activité qui détourne les aa-ARNt destinés à la synthèse protéique ou des cyclodipeptides produits ? Etant donné que la toxicité observée lors des transformations semble due à des fuites d'expression, la protéine devrait être produite en faible quantité et le détournement des aa-ARNt qu'elle opère ne devrait pas être suffisant pour induire une toxicité. Etant donné que les dicétopipérazines et les cyclodipeptides sont réputés pour présenter des activités biologiques, on peut supposer que le cyclodipeptide produit par la CDPS pc1814long possède une activité antibiotique.

Par ailleurs, les essais réalisés ont montré les difficultés à trouver une souche ayant les caractéristiques désirées quant à l'efficacité de clonage et la praticité d'utilisation. En effet, la solution retenue pour diminuer les fuites d'expression a consisté à transformer uniquement des souches portant le pREP4, ce qui empêche la combinaison avec le pIJ194. L'étude de l'activité des CDPS n'est pas touchée par cette adaptation. En revanche, l'étude des enzymes de modification en association avec des CDPS présentant une toxicité pour BL21AI risque d'être compliquée et nécessitera des essais au cas par cas. Si l'utilisation de BL21AI-pREP4 se révèle indispensable, la résistance du plasmide pREP4 pourra être modifiée en remplaçant la cassette de résistance à la kanamycine par une cassette de résistance au chloramphénicol ou à la streptomycine/spectinomycine. Les deux souches testées ont été retenues pour les études ultérieures car la comparaison des résultats des deux souches permettra également de conforter la reproductibilité des résultats.

Enfin, l'utilisation d'une nouvelle colonne chromatographique dont le mode séparation est orthogonal à celle précédemment utilisée permet d'accéder à l'identification d'une nouvelle gamme de cyclodipeptides, polaires et de plus petite taille. Cependant, le manque de standards chimiques disponibles commercialement nous empêche de tester cette colonne pour la rétention de composés chargés positivement (cyclodipeptides incorporant Lys ou Arg). Il n'est donc pas à exclure que certains cyclodipeptides produits ne soient pas détectés en utilisant l'une ou l'autre des colonnes. Il est à noter également que l'analyse LC-MS/MS des surnageants sur deux phases stationnaires double le temps d'analyse sur la machine et le temps nécessaire à l'exploitation des données. Ce type de méthode d'analyse reste envisageable dans le cas d'une analyse moyen débit mais ne pourrait pas être utilisée à une échelle plus importante en l'état.

Nous avons désormais à disposition une stratégie de criblage des activités des enzymes des voies à CDPS. Cette dernière va être utilisée pour cribler les activités de synthèse, dans un premier temps, des CDPS potentielles disponibles dans les bases de données et dans un deuxième temps, des enzymes de modification associés à ces CDPS.

CHAPITRE II :  
MISE EN ÉVIDENCE ET  
CARACTERISATION DE NOUVELLES  
ACTIVITÉS CDPS

Comme je l'ai décrit dans le chapitre précédent, nous avons développé une stratégie expérimentale pour la mise en évidence de nouvelles activités enzymatiques constituant les voies de biosynthèse de dicétopipérazines. Cette stratégie comprend les différentes étapes nécessaires allant du clonage des gènes jusqu'à l'analyse des dicétopipérazines produites par les enzymes correspondantes. La mise en place de cet outil répond au besoin d'étudier l'activité d'un plus grand nombre d'enzymes en parallèle et en un temps réduit.

La famille des CDPS, au moment de son identification en 2009, ne comptait que huit membres : AlbC de *S. noursei*, Rv2275 de *M. tuberculosis*, Plu0297 de *P. luminescens*, Jk0923 de *C. jeikeium*, pSHaeC06 de *S. haemolyticus*, YvmC\_Bthu de *B. thuringiensis*, YvmC\_Blic de *B. licheniformis* et YvmC\_Bsub de *B. subtilis*<sup>26</sup>. Aucun autre membre potentiel n'avait été identifié dans les bases de données. Avec l'explosion des séquençages de génomes, les bases de données s'enrichissent chaque année de nouvelles séquences, dont certaines comportent des gènes de CDPS potentielles. Entre 2009 et 2012, plusieurs études<sup>28,345</sup> ont mis en évidence un nombre croissant des séquences de CDPS potentielles tandis que, dans le même temps, seule une CDPS (Nvec-CDPS2 de *Nematostella vectensis*)<sup>27</sup>, a été caractérisée biochimiquement. Plusieurs questions relatives à ces nouvelles séquences de CDPS potentielles se posaient alors. Ces enzymes, malgré les particularités de séquence que certaines présentent, sont-elles des CDPS actives ? Si oui, quelle est la nature des cyclodipeptides biosynthétisés ? Et, devant le nombre croissant de CDPS potentielles, serait-il possible de prédire l'activité de nouvelles CDPS ?

Pour répondre à ces questions, nous avons réalisé un inventaire des CDPS potentielles disponibles dans les bases de données au début de l'étude (en mai 2013) et nous avons sélectionné une cinquantaine de CDPS (partie I). Après avoir décrit les CDPS actives parmi celles sélectionnées et les conséquences pour les propriétés générales de cette famille d'enzymes (partie II), je présenterai la mise en évidence de la présence de deux sous-familles de CDPS caractérisées par des signatures fonctionnelles spécifiques (partie III). Je décrirai ensuite la diversité de cyclodipeptides produits (partie IV) avant de détailler l'outil préliminaire de prédiction d'activités que nous avons mis en place (partie V). Le travail décrit dans ces parties a fait l'objet d'un article qui va être publié prochainement dans *Nature Chemical Biology* (Annexe 1). Je terminerai ce chapitre par l'analyse de l'activité de ces enzymes *in vitro* et les conséquences de cette présente étude pour les études ultérieures (partie VI).

# I. Identification par bio-informatique de 80 CDPS potentielles et sélection de 49 CDPS à tester

L'objectif de cette étude bio-informatique est donc d'obtenir un « panorama » des CDPS disponibles dans les bases de données. Nous avons identifié les séquences pouvant correspondre à des gènes de CDPS. Nous avons étudié la similitude et la divergence de ces séquences avec les CDPS connues, avant de sélectionner une cinquantaine de CDPS potentielles pour évaluer leurs activités de synthèse de cyclodipeptides.

## 1. Identification des CDPS potentielles par PSI-BLAST

Dans un premier temps, les séquences des CDPS potentielles ont été identifiées dans la base de données NCBI, en utilisant un algorithme dérivé de BLAST (Basic Local Alignment Search Tool)<sup>399</sup>, le PSI-BLAST (Position-Specific Iterated BLAST)<sup>400</sup>. Le BLAST permet de trouver, dans les bases de données, les séquences ayant des régions similaires à la séquence entrée par l'utilisateur. Le PSI-BLAST, pour sa part, réalise plusieurs itérations de BLAST. A l'issue de chaque itération, une séquence consensus (un profil) de l'ensemble des séquences obtenues (« hits ») est déterminée, et sert de séquence référence à la prochaine itération. L'utilisation de profil de séquence permet une recherche plus sensible d'homologues « lointains » d'une séquence unique. A chaque étape, l'utilisateur a la possibilité de définir les protéines qui serviront à l'établissement du consensus. Dans le cas des CDPS, certaines protéines dont l'identité de séquence était faible ont été ajoutées, car elles présentaient les résidus conservés des CDPS. En revanche, des protéines particulièrement divergentes, pour ce qui est de la taille ou de la présence des résidus conservés, ont été exclues du calcul du profil. Les itérations ont été interrompues lorsque le lot de séquences renvoyées était stable (au bout de 6 ou 7 itérations).

Les PSI-BLAST réalisés donnent ainsi un ensemble d'une centaine de séquences de CDPS potentielles. Parmi ces séquences, certaines sont redondantes car elles sont strictement identiques ou très proches et proviennent de séquençages différents du même organisme ou d'organismes très proches. Dans chacun de ces cas (pour *B. subtilis*, *B. licheniformis*, *B. thurengensis*, *M. canetti*, *M. tuberculosis* et *M. bovis*), une seule séquence a été retenue. Chez *Diplorickeetisia massiliensis*, à deux reprises, deux phases ouvertes de lecture contiguës pourraient correspondre, pour une séquence, à la partie N-terminale et pour l'autre, à la partie C-terminale d'une CDPS. Ces quatre séquences incomplètes ainsi que trois autres séquences partielles ont été retirées du lot de séquences. *In fine*, nous avons retenu 82 séquences de CDPS potentielles.

## 2. Correction des séquences obtenues et création d'un arbre phylogénétique

En réalisant un alignement multiple de ces 82 séquences avec ClustalO (EMBL-EBI), nous nous sommes aperçus qu'une partie de l'extrémité N-terminale semblait manquer pour sept séquences. En effet, le premier résidu des séquences C791\_7200, SsomD4\_010100015357, SJA\_C1-32620, Dmas2\_010100010809, K530\_11977, pc1814 et PABE171\_2006 s'aligne respectivement avec les positions 56, 32, 24, 66, 50, 26 et 81 de AlbC (Figure 53). En examinant la séquence d'acides nucléiques en amont de la séquence codante, nous avons identifié un autre codon initiateur pour chaque séquence. Ces éléments suggèrent que l'annotation du codon initiateur par les algorithmes de détection ne correspond peut-être pas au codon initiateur physiologique. De plus, les parties N-terminales ainsi ajoutées s'alignent de manière satisfaisante avec l'extrémité N-terminale d'AlbC, ce qui est cohérent pour des protéines susceptibles d'appartenir à une même famille.

**a.**

```
AlbC -----MLAGLVPAPDHGMREEILGDRSRLIRQRGEHALIGISAG----- 39
C791_7200 -----
SsomD4_010100015357 -----MCIGVSPF----- 8
SJA_C1-32620 -----MNIGDYGHALIAVSLF----- 16
Dmas2_010100010809 -----
K530_11977 -----
pc1814 -----METEKPVIILVSTHSSFHEKISG 23
PABE171_2006 -----
```

```
AlbC ----NSYFSQKNTVMLLQWAGQRFERTDVVYVDTH-IDEMLIADGRSAQEAERSVKRTLK 94
C791_7200 ----MSAFTGFHFFVDPGP-SAFTLEALGYDPIRAAQKAQRQGN 39
SsomD4_010100015357 ----NSYFSTPRLRRLADWALTRFPGVHFFVDPDA-AAAYTLEALGYDPPARARHKAQRQGG 63
SJA_C1-32620 ----NSYYSTDRITTLVRWAKHTFERSHIPVFDLP-HAYTLSARKGHSPGSRVRRARKEGR 71
Dmas2_010100010809 ----MTDSLQEHYLGLENPNLVPNTNIKRLAHKKGT 30
K530_11977 ----MLLQWAGQHFRERTDVVYVDTH-IDDMLMADGRSAQEAESVKRTLK 15
pc1814 DLKMNAFVSTIRN--HV---KG--KITVLLSDRAHINTMSL----- 57
PABE171_2006 ----MPEEDARREAVRLG 14
```

**b.**

```
AlbC -----MLAGLVPAPDHGMREEILGDRSRLIRQRGEHALIGISAG----- 39
C791_7200-EXT -----MLRPLPLTEHCADPLARAHEHVCIGISPF----- 28
SsomD4_010100015357-EXT -----MFEIPLTERCRGLLPEAVHVCIGVSPF----- 28
SJA_C1-32620-EXT -----MDVTVLSHRALTVTSIRPGDRMNIGDYGHALIAVSLF----- 37
Dmas2_010100010809-EXT -----MELHA-----KINGLE---KHL--KQHPTKRDFSIVNVIS----MDQPSQR 37
K530_11977-EXT -----MLSGLVPALDHSRMREEILGNRGRKIRQRGEHALIGISAG----- 39
pc1814-EXT MNKYTWRKNFTAILSNQKEIIKGL---AFKTEETHWLETEKPVIIILVSTHSSFHEKISG 57
PABE171_2006-EXT ---MNLIAEDFSTRSLLAGERYKAKIAFVSPHTRRNSFEDEPKCFLGVSL----- 48
```

```
AlbC ----NSYFSQKNTVMLLQWAGQRFERTDVVYVDTH-IDEMLIADGRSAQEAERSVKRTLK 94
C791_7200-EXT ----NSYFTVDRIADLARWAMSAFTGFHFFVDPGP-SAFTLEALGYDPIRAAQKAQRQGN 83
SsomD4_010100015357-EXT ----NSYFSTPRLRRLADWALTRFPGVHFFVDPDA-AAAYTLEALGYDPPARARHKAQRQGG 83
SJA_C1-32620-EXT ----NSYYSTDRITTLVRWAKHTFERSHIPVFDLP-HAYTLSARKGHSPGSRVRRARKEGR 92
Dmas2_010100010809-EXT GEKFEAFTRRINLLHKK---RGVRKLVIILTDSLQEHYLGLENPNLVPNTNIKRLAHKKGT 93
K530_11977-EXT ----NSYFSQKNVTMLLQWAGQHFRERTDVVYVDTH-IDDMLMADGRSAQEAESVKRTLK 94
pc1814-EXT DLKMNAFVSTIRN--HV---KG--KITVLLSDRAHINTMSL----- 91
PABE171_2006-EXT ----NRNFTPNRFHSMVEWAARRFEKCSILIGDHI-HRITLQSTRDMPEEDARREAVRLG 103
```

**Figure 53 : Alignement multiple partiel d'AlbC et des CDPS potentielles dont la partie N-terminale est manquante**

Les CDPS sont désignées par le locus de leur séquence codante. L'alignement multiple est obtenu en utilisant ClustalO sur le site EMBL-EBI. a. Les séquences telles qu'elles sont trouvées dans les bases de données sont alignées avec AlbC. b. Les séquences dans leurs versions allongées (-EXT) sont alignées avec AlbC. La sérine catalytique (S37) est représentée en rouge.

Par ailleurs, la séquence encodée par le gène SCAT\_0901 présente une séquence allongée en N-terminale riche en résidus Pro et Ala. En choisissant le codon ATG Met52 comme codon initiateur, nous obtenons une version plus courte du gène dont la protéine s'aligne correctement avec AlbC. (Figure 54).

```

AlbC -----MLAGLVPAPDHGMREE 16
SCAT_0901 MHDNGHRPTAAGVRPGTVQARPAAGTASLTAPAPAPAPAPAPATAPVPAMRSGDFNCR 60
SCAT_0901_SHORT -----MRS GDFNCR 9

AlbC ILGDRSRLIRQRGEHALIGI SAGNSYFSQKNTVMLLQWAGQRFERTDVVYVDTHIDEMLI 76
SCAT_0901 PLTPNCARVVDAGEHAVVGVSLGNSYFRGDI LRALLRWLSGRFAQLD VVIPDSSYRDNLL 120
SCAT_0901_SHORT PLTPNCARVVDAGEHAVVGVSLGNSYFRGDI LRALLRWLSGRFAQLD VVIPDSSYRDNLL 69

```

Figure 54 : Alignement partiel avec AlbC de la CDPS potentielle encodée par SCAT\_0901

L'alignement multiple est obtenu en utilisant ClustalO sur le site EMBL-EBI. La version raccourcie de la séquence est désignée par le suffixe SHORT. La sérine catalytique (S37) est représentée en rouge.

L'alignement multiple des 82 séquences obtenu en utilisant l'ensemble des séquences corrigées a servi à construire un arbre phylogénétique (Figure 55). Cet arbre porte les noms des CDPS colorées en fonction du phylum d'origine afin d'illustrer la diversité de provenance des CDPS potentielles. Plusieurs phyla bactériens sont représentés, notamment les protéobactéries, les actinomycètes et les firmicutes. Quatre séquences en plus de celle de *N. vectensis* proviennent d'organismes eucaryotes. Aucune séquence provenant des archées n'est recensée. De plus, quatre séquences provenant du métagénome ont « émergé » de cette analyse.

### 3. Sélection des CDPS à étudier

Les CDPS connues avant le début de mon étude (au nombre de neuf, cf. Introduction de ce chapitre) ou découvertes pendant cette étude par d'autres équipes (au nombre de deux, Amir\_4627 de *A. mirum*<sup>348</sup> et Ndas\_1148 de *N. dassonvillei*<sup>346</sup>) sont marquées par des points bleus sur l'arbre. Nous nous sommes aperçus que les CDPS dont l'activité était connue, en plus d'être peu nombreuses, sont proches sur l'arbre phylogénétique (Figure 55). Cela signifie que la diversité phylogénétique des CDPS disponibles dans les bases de données n'avait été que peu exploitée. Cette diversité de séquences protéiques pourrait s'accompagner d'une diversité d'activités de synthèse de cyclodipeptides.

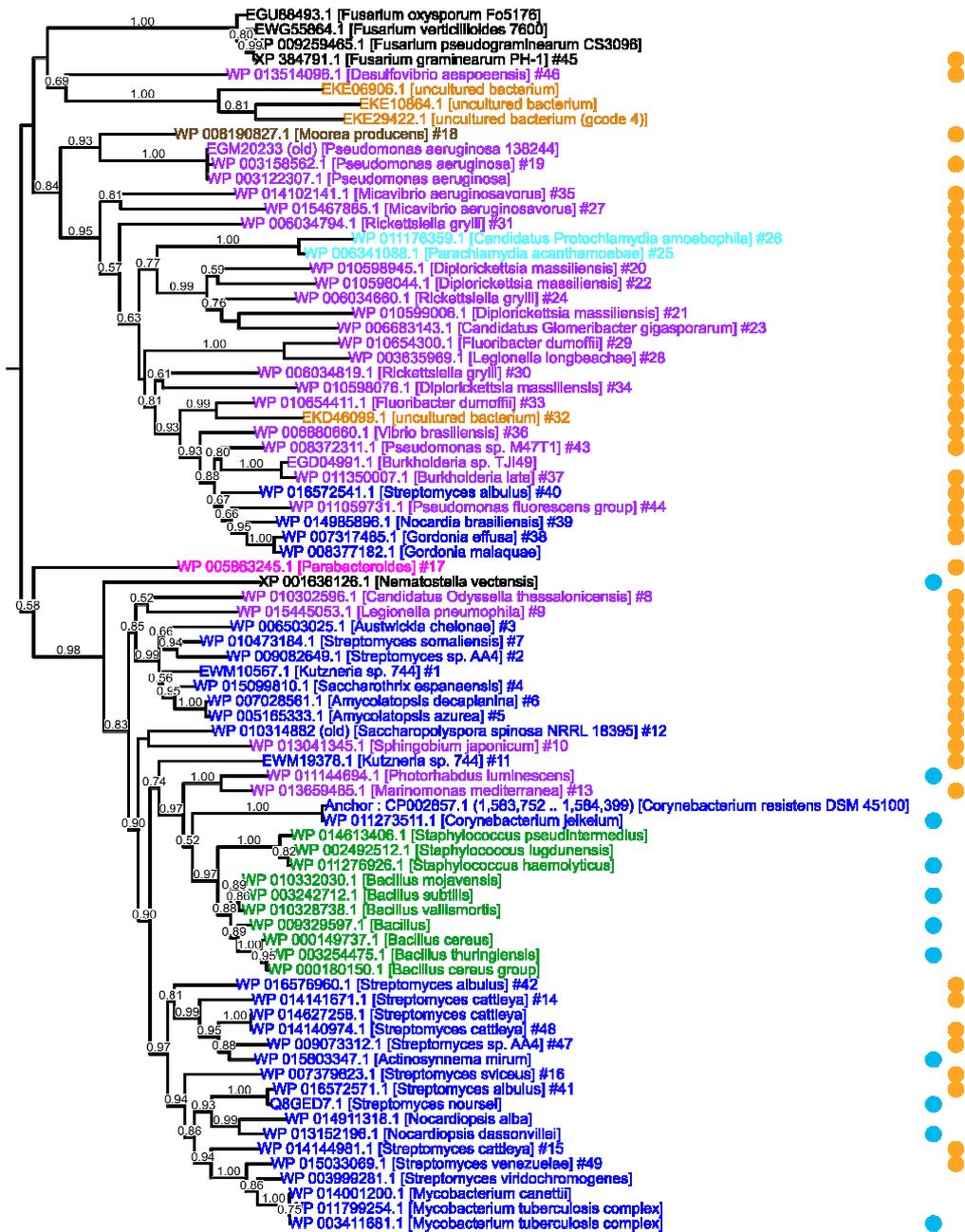


Figure 55 : Arbre phylogénétique des 82 CDPS connues et CDPS potentielles identifiées dans les bases de données par bio-informatique (Mai 2013) : analyse préliminaire et sélection.

Les CDPS sont désignées par leurs numéros d'identifiant dans NCBI et leur organisme d'origine. Les numéros attribués aux CDPS sélectionnées pour l'étude sont ajoutés à la fin de leur nom. Le nom des séquences est coloré en fonction de leur origine taxonomique : protéobactéries en violet, actinobactéries en bleu foncé, chlamydiae en cyan, cyanobactérie en marron, bacteroidetes en fuchsia, firmicutes en vert, eucaryotes en noir (incluant quatre champignons filamenteux et un metazoaire (*Nematostella vectensis* du phylum Cnidaria), les séquences issues du métagénome sont en orange. Les CDPS connues avant mon étude ou découvertes par d'autres équipes pendant celle-ci sont représentées par des points bleus et les 49 CDPS sélectionnées pour l'étude par des points orange.

Ainsi, afin d'agrandir la famille des CDPS et d'explorer ses possibilités de synthèse, 49 CDPS potentielles ont été sélectionnées ; elles sont représentées par des points orange sur l'arbre. Elles ont été choisies de manière à être distribuées sur la totalité de l'arbre et à ne pas être redondantes entre elles. Plus précisément, les CDPS potentielles qui partagent plus de 75 % d'identité de séquence avec une CDPS déjà caractérisée ou avec une CDPS sélectionnée ont été retirées de la sélection (19 séquences). Seules trois séquences ont fait exception à ce critère. Il s'agit de la séquence K530\_11977 (WP\_016572571) de *Streptomyces albulus* qui partage 88% d'identité avec AlbC et des séquences d'*Amycolatopsis azurea* DSM 43854 et d'*Amycolatopsis decaplanina* DSM 44594 qui partagent 93% d'identité entre elles (Figure électronique 3). Par ailleurs, trois séquences issues du métagénome et qui formaient un groupe à l'extrémité d'une longue branche sur l'arbre n'ont pas été testées. Ces 49 CDPS potentielles sont numérotés de 1 à 49 selon le Tableau 14 ci-dessous et sont testées pour leurs activités de synthèse de cyclodipeptides selon la méthodologie mise au point au début de ma thèse et décrite au chapitre précédent. Dans la suite du document, les CDPS potentielles seront désignées par leur numéro.

**Tableau 14 : CDPS sélectionnées pour l'étude, désignées par le locus de leur gène, le numéro d'identifiant au NCBI, l'organisme d'origine et le numéro attribué dans l'étude.**

Numéro	Organisme d'origine	Locus du gène	Protéine (identifiant)
1	<i>Kutzneria</i> sp. 744	KUTG_00871	EWM10567
2	<i>Streptomyces</i> sp. AA4	SSMG_06201	WP_009082649
3	<i>Austwickia chelonae</i> NBRC 105200	AUCHE_08_05160	WP_006503025
4	<i>Saccharothrix espanaensis</i> DSM 44229	BN6_23810	WP_015099810
5	<i>Amycolatopsis azurea</i> DSM 43854	C791_7200	WP_005165333
6	<i>Amycolatopsis decaplanina</i> DSM 44594	H074_03090	WP_007028561
7	<i>Streptomyces somaliensis</i> DSM 40738	SsomD4_010100015357	WP_010473184
8	<i>Candidatus Odysseella thessalonicensis</i> L13 HMO	COdytL_10466	WP_010302596
9	<i>Legionella pneumophila</i> subsp. pneumophila LPE509 plasmid	LPE509_p00072	WP_015445053
10	<i>Sphingobium japonicum</i> UT26S	SJA_C1-32620	WP_013041345
11	<i>Kutzneria</i> sp. 744	KUTG_09682	EWM19378
12	<i>Saccharopolyspora spinosa</i> NRRL 18395	SspiN1_010100037249	WP_010314882
13	<i>Marinomonas mediterranea</i> MMB-1	Marme_0276	WP_013659485
14	<i>Streptomyces cattleya</i> NRRL 8057	SCAT_0901	WP_014141671
15	<i>Streptomyces cattleya</i> NRRL 8057	SCAT_4268	WP_014144981
16	<i>Streptomyces svicens</i> ATCC 29083	SSEG_10527	WP_007379823
17	<i>Parabacteroides</i> sp. 20_3	HMPREF9008_03529	WP_005863245
18	<i>Moorea producens</i>	LYNGBM3L_69800	WP_008190827
19	<i>Pseudomonas aeruginosa</i> PA21_ST175	H123_18841	WP_003158562
20	<i>Diplorickeetsia massiliensis</i> 20B	Dmas2_010100010809	WP_010598945
21	<i>Diplorickeetsia massiliensis</i> 20B	Dmas2_010100011124	WP_010599006
22	<i>Diplorickeetsia massiliensis</i> 20B	Dmas2_010100005805	WP_010598044

Numéro	Organisme d'origine	Locus du gène	Protéine (identifiant)
23	<i>Candidatus</i> Glomeribacter gigasporarum BEG34 WGS	CAGGBEG34_30028	WP_006683143
24	<i>Rickettsiella grylli</i>	RICGR_1344	WP_006034660
25	<i>Parachlamydia acanthamoebae</i> str. Hall's coccus	pah_c026o110	WP_006341088
26	<i>Candidatus</i> Protochlamydia amoebophila UWE25	pc1814	WP_011176359
27	<i>Micavibrio aeruginosavorus</i> EPB	A11S_1546	WP_015467885
28	<i>Legionella longbeachae</i> D-4968	LLB_2968	WP_003635969
29	<i>Fluoribacter dumoffii</i> Tex-KL	FdumT_010100010802	WP_010654300
30	<i>Rickettsiella grylli</i>	RICGR_0139	WP_006034819
31	<i>Rickettsiella grylli</i>	RICGR_1349	WP_006034794
32	Uncultured bacterium ACD_69C00020	ACD_69C00020G0002	EKD46099
33	<i>Fluoribacter dumoffii</i> Tex-KL	FdumT_010100011357	WP_010654411
34	<i>Diplorickettsia massiliensis</i> 20B	Dmas2_010100005975	WP_010598076
35	<i>Micavibrio aeruginosavorus</i> ARL-13	MICA_581	WP_014102141
36	<i>Vibrio brasiliensis</i> LMG 20546	VIBR0546_16848	WP_006880660
37	<i>Burkholderia lata</i>	Bcep18194_C7320	WP_011350007
38	<i>Gordonia effusa</i> NBRC 100432	GOEFS_047_00070	WP_007317485
39	<i>Nocardia brasiliensis</i> ATCC 700358	O3I_025450	WP_014985896
40	<i>Streptomyces albulus</i> CCRC 11814	K530_11825	WP_016572541
41	<i>Streptomyces albulus</i> CCRC 11814	K530_11977	WP_016572571
42	<i>Streptomyces albulus</i> CCRC 11814	K530_34281	WP_016576960
43	<i>Pseudomonas</i> sp. M47T1	PMM47T1_20098	WP_008372311
44	<i>Pseudomonas protegens</i> Pf-5	PFL_1389	WP_011059731
45	<i>Gibberella zeae</i> PH-1	FG04615.1	XP_384791
46	<i>Desulfonivibrio aespoensis</i> Aspo-2	Daes_1150	WP_013514096
47	<i>Streptomyces</i> sp. AA4	SSMG_00951	WP_009073312
48	<i>Streptomyces cattleya</i> NRRL 8057	SCAT_0198	WP_014140974
49	<i>Streptomyces venezuelae</i> ATCC 10712	SVEN_1864	WP_015033069

Grâce à cette étude bio-informatique préliminaire, la diversité des séquences contenues dans les bases de données peut être appréciée. Entre 2005 et 2013, le nombre de séquences susceptibles de coder des CDPS est multiplié par dix. La portion de CDPS dont l'activité est connue est de 13 % et ne couvre absolument pas la diversité phylogénétique rencontrée. Nous pensons que l'étude des 49 CDPS potentielles sélectionnées permettrait d'étendre les connaissances sur les CDPS, tout au moins pour la diversité de cyclodipeptides produits. Cependant, les résultats obtenus nous ont permis d'aller bien au-delà d'une simple diversité d'activités, comme présenté dans les parties suivantes.

## II. 41 nouvelles CDPS sur 49 sont actives : redéfinition des caractéristiques générales de ces enzymes

### 1. 41 des 49 CDPS sélectionnées actives *in vivo*

A l'issue du criblage moyen-débit réalisé pour tester l'activité de synthèse de cyclodipeptides, des 49 CDPS potentielles sélectionnées, 41 CDPS recombinantes se sont révélées actives *in vivo* avec détection d'un ou plusieurs cyclodipeptides dans les surnageants de culture d'*E. coli* ; elles sont repérées par un point fuchsia sur l'arbre phylogénétique (Figure 56). Le détail des cyclodipeptides synthétisés sera donné et discuté ultérieurement (cf. IV. Une plus grande diversité de cyclodipeptides).

Huit des CDPS potentielles (CDPS 11, 12, 15, 32, 34, 35, 45 et 49) n'ont pas démontré d'activité de synthèse de cyclodipeptides dans nos conditions expérimentales, alors qu'elles étaient exprimées (Figure 57). Dans seulement trois cas, certaines particularités de séquences peuvent expliquer cette absence d'activité. Chez les CDPS 12, 15 et 45, on retrouve à la position occupée par la sérine catalytique, Ser37 chez AlbC, respectivement un résidu Ala, Gly et Cys. Il est à noter que le remplacement de Ser par Cys n'est pas nécessairement suffisant pour expliquer l'inactivité de la CDPS 45, car chez AlbC, le variant S37C est capable de synthétiser des cyclodipeptides<sup>361</sup>. Une autre explication proviendrait de l'origine eucaryote de cette CDPS qui pourrait ne pas utiliser comme substrats les aminoacyl-ARNt de *E. coli*.

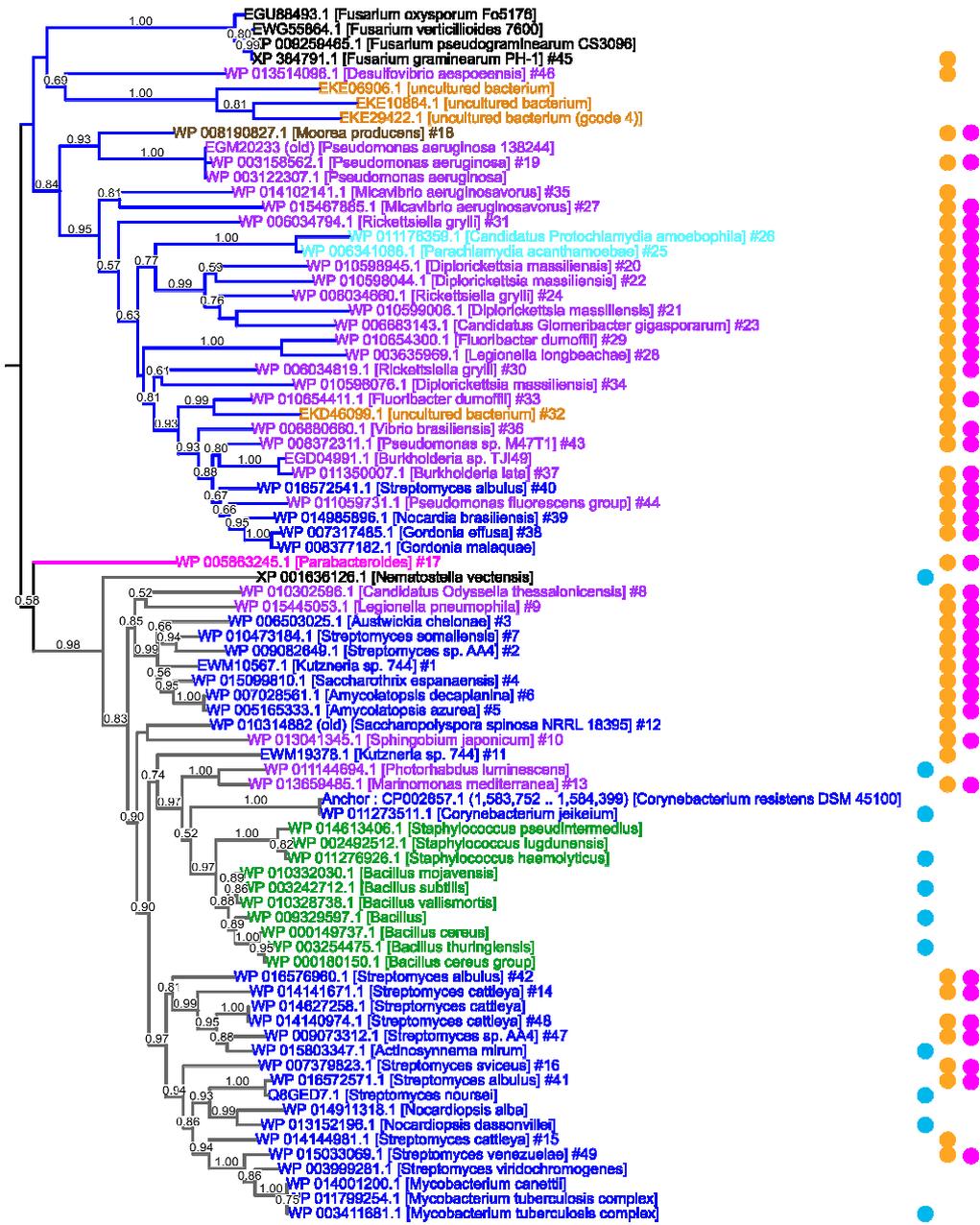


Figure 56 : Arbre phylogénétique des 82 CDPS connus et CDPS potentielles identifiées dans les bases de données par bio-informatique (Mai 2013) : résultats des tests d'activité et sous-familles

Les CDPS sont désignées par leurs numéros d'identifiant dans NCBI et leur organisme d'origine. Les numéros attribués aux CDPS sélectionnées pour l'étude sont ajoutés à la fin de leur nom. Le nom des séquences est coloré en fonction de leur origine taxonomique : protéobactéries en violet, actinobactéries en bleu foncé, chlamydiae en cyan, cyanobactérie en marron, bacteroidetes en fushia, firmicutes en vert, eucaryotes en noir (incluant quatre champignons filamenteux et un metazoaire (*Nematostella vectensis* du phylum Cnidaria), les séquences issues du métagénome sont en orange. Les CDPS connues avant mon étude ou découvertes par d'autres équipes pendant celle-ci sont représentées par des points bleus, les 49 CDPS sélectionnées pour l'étude par des points orange et les CDPS actives par des points fuchsia. Les CDPS peuvent être classées en 3 sous-familles qui correspondent à trois branches sur l'arbre. La NYH correspond à la branche grise, la XYP à la bleue et la SYQ à la rose.

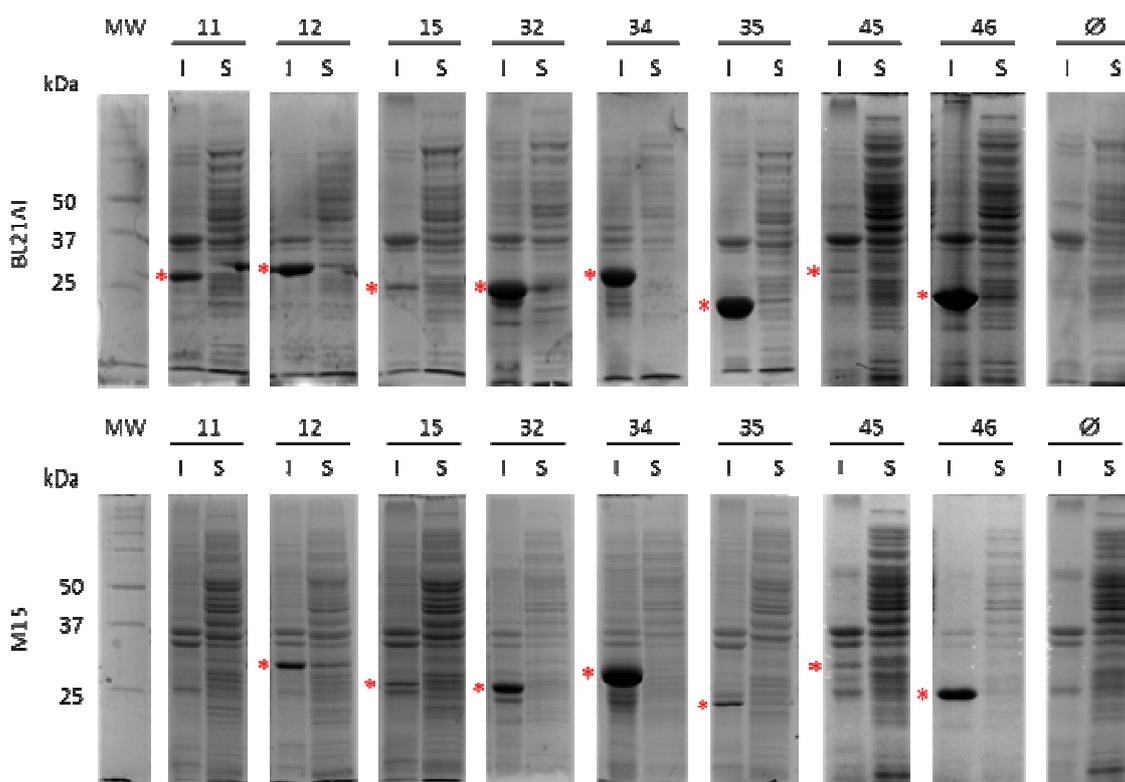


Figure 57 : Analyses SDS-PAGE de l'expression des CDPS potentielles inactives dans nos conditions expérimentales

Les fractions solubles (S) et insolubles (I) des bactéries BL21AI, BL21AI-pREP4 et M15-pREP4 transformées par le plasmide pIJ196 codant une CDPS potentielle ou non (Ø) après 24h de culture sont analysées sur des gels SDS-PAGE à 12% colorés au bleu de Coomassie. Les protéines surexprimées sont indiquées par une étoile rouge.

Avec ces 41 nouvelles CDPS et les 11 CDPS précédemment caractérisées, la famille des CDPS comporte désormais 52 membres. Nous nous sommes demandé si les caractéristiques de cette famille élargie coïncidaient avec celles qui avaient été établies pour la famille initiale.

## 2. Caractéristiques générales des CDPS actives

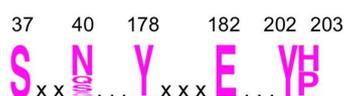
Comme l'indiquait l'étude bio-informatique, les CDPS sont principalement d'origine bactérienne, avec trois phyla particulièrement représentés : les actinomycètes, les firmicutes et les protéobactéries. Un représentant des bacteroidetes, deux représentants des chlamydiae et un représentant des cyanobactéries diversifient quelque peu le tableau des phyla d'origine des CDPS. Aucune nouvelle CDPS d'origine eucaryote n'ayant été confirmée au cours de l'étude, la CDPS de *Nematostella vectensis* reste la seule et unique CDPS eucaryote à ce jour<sup>27</sup>. De plus, la CDPS potentielle provenant d'approche métagénomique ne s'est pas révélée active dans nos conditions expérimentales. (Figure 55 et Figure 56)

La plupart des CDPS ont une taille comprise en 200 et 300 résidus, avec deux exceptions notables (les CDPS 28 et 29) qui présentent un domaine C-terminal additionnel. La grande majorité des nouvelles CDPS ne partagent pas plus de 30 % d'identité de séquence avec les 11 CDPS caractérisées

précédemment. Ce pourcentage est même compris entre 6 et 20 % d'identité pour 22 des nouveaux membres, soit plus de la moitié (Annexe électronique 3). Malgré cette diversité de séquence, une recherche de prédiction de structures secondaires réalisée en utilisant HHPRED (Max Planck Institute) renvoie systématiquement, pour chacune des 41 CDPS actives, les trois CDPS de structures connues dans les trois meilleurs résultats avec un score très élevé (Annexe électronique 4). L'outil HHPRED permet de trouver des similitudes de mode de repliement entre une séquence protéique et les protéines de structure connue présentes dans les bases de données (PDB essentiellement). Ces résultats suggèrent que ces CDPS partagent toutes une homologie de structure.

### 3. Signature fonctionnelle

Nous nous sommes ensuite intéressés aux résidus catalytiques, décrits comme conservés dans les études précédentes (cf. Etude bibliographique V.1.4 Les CDPS utilisent un mécanisme catalytique de type ping-pong). Afin de pouvoir déterminer les positions équivalentes chez les nouvelles CDPS, l'ajustement multiple des 52 CDPS actives a été ajusté manuellement en prenant en compte les informations fournies par les résultats des interrogations HHPRED (prédiction de structures secondaires et similitude de séquences) (Annexe électronique 5). Nous avons défini une signature fonctionnelle des CDPS à partir de la fréquence des résidus retrouvés aux positions correspondant aux résidus catalytiques (Figure 58). Sur cette signature, on observe que seulement deux résidus (S37, Y202, en numérotation d'AlbC) sont strictement conservés et que Y178 et E182 présentent également une bonne conservation. On distingue une fréquence quasi équivalente entre H et P à la position 203 et la dégénérescence de la position 40 avec malgré tout une fréquence importante du résidu N.



**Figure 58 : Signature fonctionnelle des CDPS représentée par un logo correspondant à la fréquence des acides aminés catalytiques aux positions 37, 40, 178, 182, 202 et 203 (en numérotation d'AlbC)**

## III. Des sous-familles de CDPS aux résidus catalytiques différents

### 1. Définition des sous-familles

Les travaux précédents de l'équipe réalisés en collaboration avec l'équipe de Yves Méchulam (Ecole Polytechnique, Palaiseau) ont montré chez AlbC que Y202, un des deux résidus strictement conservés, participe à un réseau de liaisons hydrogène, impliquant également N40 et H203. Ces deux derniers résidus sont définis comme essentiels pour le bon positionnement des boucles catalytiques<sup>364</sup> (cf. Etude bibliographique V.1.4 Les CDPS utilisent un mécanisme catalytique de type ping-pong). Parmi la famille de 52 CDPS, seulement environ la moitié possède la paire canonique « N40, H203 », tandis que

l'autre moitié possède une paire « X40, P203 » (X étant un résidu non conservé). Seules quelques exceptions possèdent des paires différentes. (cf. Discussion, conclusion et perspectives). Nous nous sommes aperçus qu'il y avait une forte corrélation entre la nature de la paire présente, « N40, H203 » ou « X40, P203 » et la position des CDPS sur l'une ou l'autre des deux branches principales observables sur l'arbre phylogénétique (Figure 56, une « branche bleue » et une « branche grise »). Chaque branche de l'arbre regroupe donc les membres d'une sous-famille que nous avons baptisée NYH (« branche grise ») et XYP (« branche bleue »), en référence à la nature des résidus aux positions 40, 202 et 203.

La CDPS 17 est située sur une branche indépendante des deux branches citées précédemment (Figure 56, « branche rose ») et possède une nouvelle paire, à savoir la paire « S40, Q203 ». Cette CDPS pourrait appartenir à une troisième sous-famille que nous avons nommée SYQ mais elle en serait le seul représentant caractérisé.

Pour chaque sous-famille ainsi définie, nous avons établi une signature fonctionnelle (Figure 59) comme précédemment. Nous voyons qu'en dehors des résidus en position 40 et 203 dont la nature définit l'appartenance à la sous-famille et reflète peut-être une manière différente de positionner les résidus catalytiques, les autres résidus (positions 37, 178, 182 et 202) sont bien conservés. Il est à noter que les 11 CDPS précédemment caractérisées appartiennent toutes à la famille NYH.



**Figure 59 : Signature fonctionnelle pour les sous-familles de CDPS (NYH et XYP) et résidus catalytiques du membre de la famille potentielle SYQ**

## 2. Caractéristiques fonctionnelle et structurale des sous-familles

Nous avons regardé de manière plus large la conservation des séquences dans chaque sous-famille en créant un logo sur l'ensemble de la séquence à partir des séquences de la sous-famille NYH et des séquences de la sous-famille XYP (Figure 60). On voit que le nombre de positions présentant des résidus strictement conservés est bien supérieur dans la famille NYH à ce qu'il est dans la famille XYP. Dans la sous-famille NYH, en plus des six résidus catalytiques, six autres résidus sont conservés, avec notamment trois résidus essentiels chez AlbC pour l'activité enzymatique (H31 et Y128) et l'intégrité structurale (W54)<sup>361</sup>. En ce qui concerne la sous-famille XYP, les séquences sont plus dégénérées, car excepté les cinq

résidus conservés de la signature fonctionnelle, aucun autre résidu n'apparaît comme conservé au travers des séquences.

Le patch basique impliqué dans la liaison à l'ARNt du premier substrat<sup>365</sup> (cf. étude bibliographique V.1.3 Les CDPS possèdent une structure réminiscente des aminoacyl-ARNt synthétases et V.1.5 Les bases moléculaires de l'interaction entre les CDPS et leurs substrats) est retrouvé chez tous les membres de la sous-famille NYH et comprend cinq à 11 résidus. Le nombre de résidus définissant ce patch est plus réduit, chez la sous-famille XYP, avec seulement deux à six résidus basiques

D'une manière générale, la sous-famille XYP présente plus de diversité de séquences que la sous-famille NYH. En effet, non seulement elle possède un nombre plus faible de résidus bien conservés, mais il existe aussi des variations sur certains de ces résidus. Ainsi, les CDPS 25 et 26 présentent un motif DxxxQ au lieu du motif canonique YxxxE observé entre les positions Y178 et E182 d'AlbC (Annexe électronique 5). Il a été montré que le résidu E agit comme une base catalytique pour la formation de l'intermédiaire dipeptidyl-enzyme et le résidu Y sert au bon positionnement de l'entité aminoacyle du premier substrat chez AlbC<sup>361,365</sup> (cf. Etude bibliographique V.1.4 Les CDPS utilisent un mécanisme catalytique de type ping-pong). Dans le cas des CDPS 25 et 26, le rôle du résidu E pourrait être assuré par le résidu D et celui du résidu Y par le résidu Q. Il y aurait donc une inversion des deux positions catalytiques.

L'alignement des séquences de ces 52 CDPS fait également apparaître chez certaines CDPS de la sous-famille XYP des insertions et délétions internes par rapport aux CDPS caractérisées précédemment (Annexe électronique 5). En utilisant le programme PDSum sur le site de l'EMBL-EBI et les trois CDPS de structure connue, nous avons généré un diagramme représentant les structures secondaires communes à ces CDPS et nous avons positionné sur ce diagramme les insertions et délétions observées pour les CDPS 23, 25, 26, 28 et 29 (Figure 61). On peut constater que ces modifications, allant de quelques résidus à environ 30 résidus, sont positionnées au niveau de boucles, ce qui est cohérent avec la conservation d'un même mode de repliement comme prédit par HHPRED.

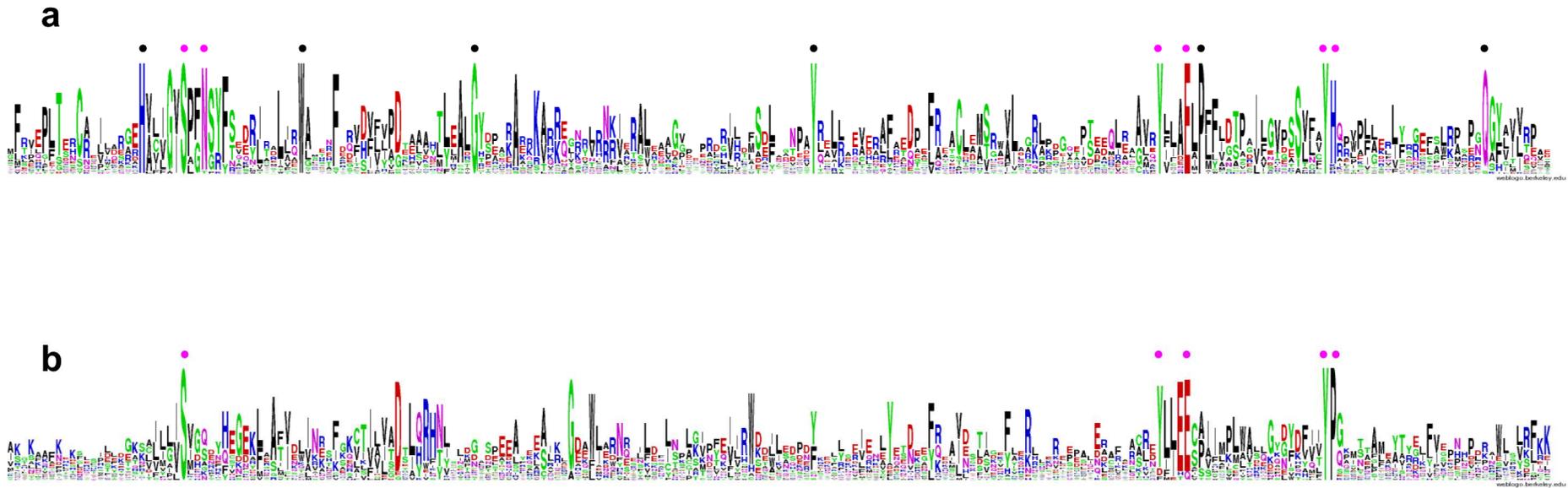


Figure 60 : Résidus catalytiques et résidus conservés dans les sous-familles NYH et XYP

Logos obtenus sur le site Weblogo à partir de l'alignement des 52 CDPS actives en supprimant les zones contenant peu de séquences alignées. Les résidus catalytiques sont marqués par un point fuchsia et les résidus conservés par un point noir. a. Logo obtenu pour la sous-famille NYH (28 séquences, les séquences de jk0923 et Nvec-CDPS2 ont été retirées) b. Logo obtenu pour la sous-famille XYP (22 séquences).

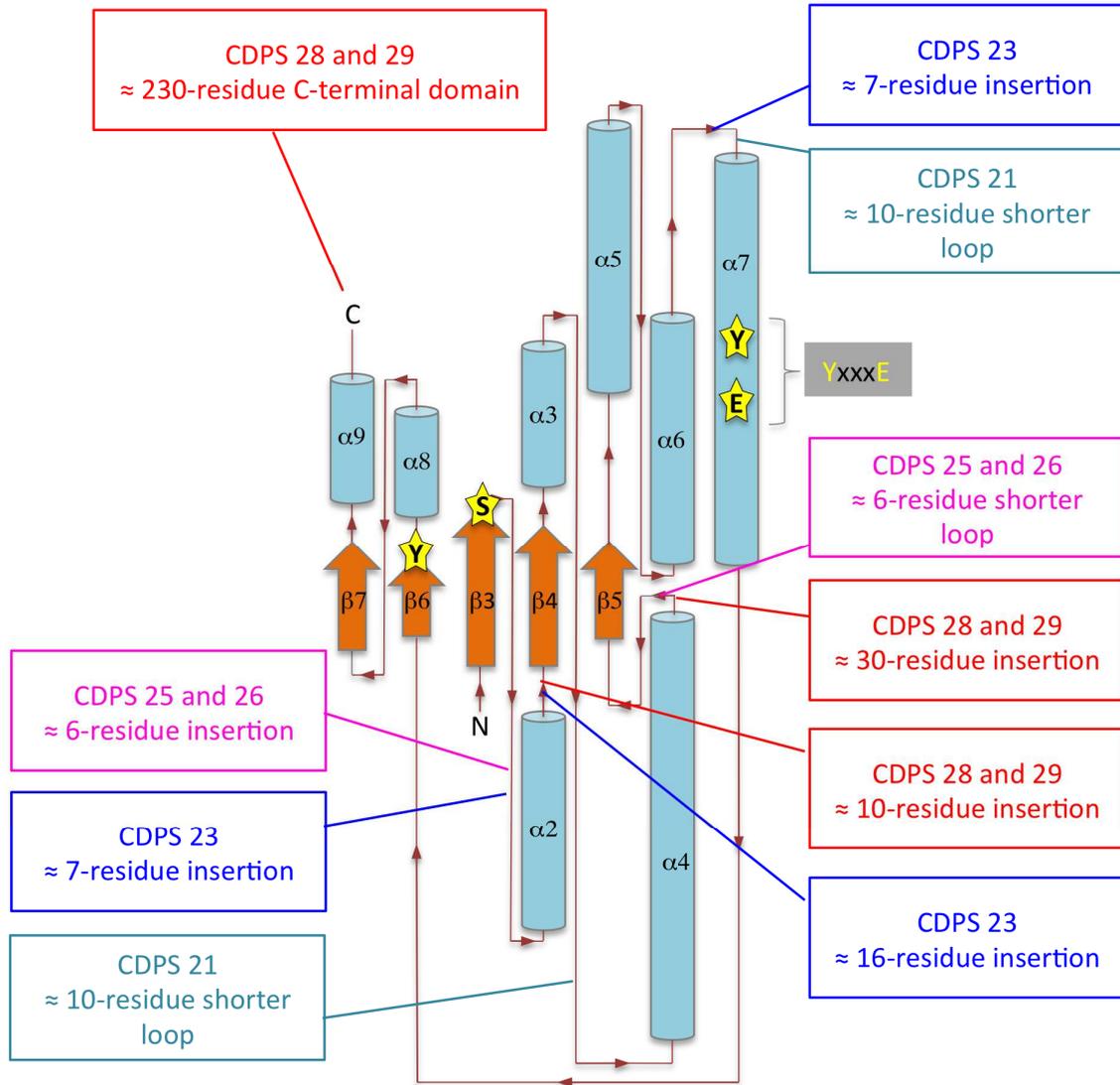


Figure 61 : Insertions et délétions observées chez les CDPS 21, 23, 25, 26, 28 et 29 localisées sur une représentation topologique des éléments de structures secondaires communes aux trois CDPS de structures connues

Les extrémités N-terminales et C-terminales sont indiquées respectivement par la lettre N et la lettre C. Les éléments de structures secondaires sont annotés comme décrits précédemment<sup>345</sup>). Le code couleur suivant a été utilisé : CDPS 21, turquoise ; CDPS 23, bleu ; CDPS 25 et 26, fuchsia ; CDPS 28 et 29, rouge).

Ainsi, les caractéristiques définies à partir des CDPS connues précédemment se restreignent maintenant uniquement aux membres de la sous-famille NYH. La sous-famille XYP introduit une plus grande variabilité dans les éléments que l'on croyait caractéristiques de la famille des CDPS, avec une diminution du nombre des résidus conservés et des résidus constituant le patch basique, une inversion des rôles catalytiques entre deux positions, des insertions ou des délétions de séquence. Par ailleurs, la diversité des caractéristiques propres à chaque sous-famille s'accompagne d'une diversité accrue de cyclodipeptides produits, comme présenté dans la partie suivante.

## IV. Une plus grande diversité de cyclodipeptides

### 1. Méthode pour l'identification des cyclodipeptides

Chaque surnageant de culture de bactéries exprimant une CDPS est analysé sur les deux colonnes chromatographiques (cf. Chapitre I. III. Introduction d'un nouveau mode de séparation chromatographique). Les chromatogrammes obtenus sont comparés à ceux obtenus lors de la culture de bactéries transformées par le plasmide ne contenant aucun gène de CDPS. Par différence, les pics chromatographiques apparaissant pour l'échantillon correspondant à l'expression d'une CDPS sont identifiés. Un exemple des chromatogrammes obtenus est représenté Figure 62.

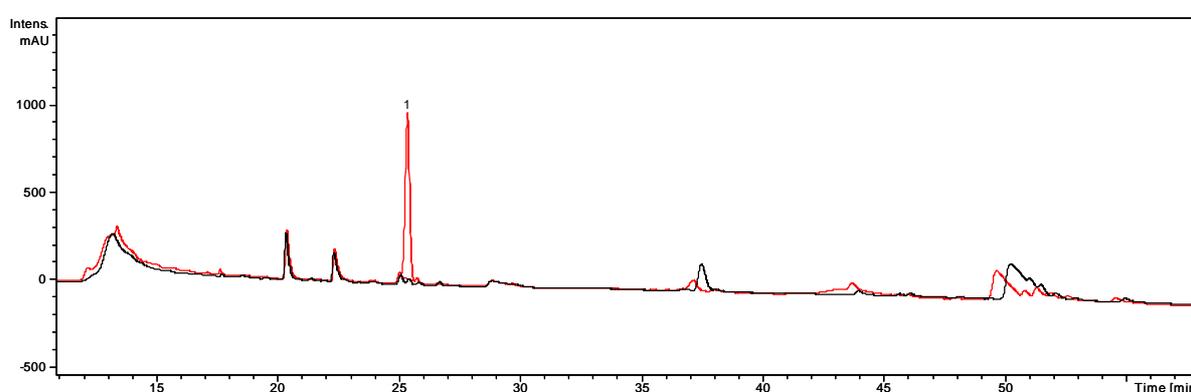


Figure 62 : Comparaison des analyses LC-MS/MS des surnageants de culture de bactéries *E. coli* n'exprimant pas de CDPS (en noir) et exprimant la CDPS 37 (en rouge)

Les chromatogrammes UV enregistrés à  $\lambda = 214$  nm sont représentés.

Les spectres de masse associés à ces nouveaux pics sont ensuite analysés en vue de l'identification des composés correspondants. Grâce aux fragmentations caractéristiques des cyclodipeptides (cf. Etude bibliographique. I.4 L'identification des 2,5-dicétopipérazines), la nature du cyclodipeptide présent est postulée. Cette identification est ensuite confirmée soit par comparaison à un standard chimique (temps de rétention et fragmentation), soit par comparaison à des données de fragmentation en ESI disponibles dans la littérature. Dans les cas où aucun standard n'était disponible et que le cyclodipeptide était produit en quantité suffisante (cas du cyclo(Tyr-Val), cyclo(Gly-Asn), cyclo(Gly-Val), cyclo(Ala-Pro), cyclo(Leu-Glu) et cyclo(L-cystine)), une purification par HPLC semi-préparative à partir de 100 ml de surnageant de culture a été réalisée. (cf. Matériels et méthodes. VI.3.1 A partir de surnageant de culture). Afin d'isoler le cyclodipeptide d'intérêt des autres composés du surnageant de culture, le gradient d'élution a été optimisé pour chaque cyclodipeptide à purifier. (

Tableau 15). Dans ces six cas, la nature du cyclodipeptide a été confirmée par une analyse en composition d'acides aminés et par la détermination de la masse haute résolution du composé (cf. Matériels et

méthodes VI.4 Analyse en composition d'acides aminés et VI.5 Analyse par spectrométrie de masse haute résolution (HRMS)). Pour les composés produits en faible quantité pour lesquels aucun standard chimique n'était disponible, l'identification s'est appuyée uniquement sur la fragmentation MS/MS.

**Tableau 15 : Gradients optimisés pour la purification des cyclodipeptides**

Cyclo(Tyr-Val)			Cyclo(Gly-Asn)			Cyclo(Gly-Val)			Cyclo(Ala-Pro)			Cyclo(Leu-Glu)			Cyclocystine		
t	%A	%B	t	%A	%B												
0	100	0	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0
4	87	13	21	93	7	21	93	7	3	96	4	3	90	10	3	92	8
23	85	15	24	0	100	24	0	100	39	80	20	21	80	20	25	90	10
25	0	100	26	0	100	26	0	100	43	0	100	23	0	100	26	0	100
27	0	100	28	100	0	28	100	0	45	0	100	25	0	100	27	0	100
29	100	0	48	100	0	48	100	0	50	100	0	27	100	0	28	100	0
49	100	0							70	100	0	47	100	0	48	100	0

t en minutes ; Eluant A : 0,1 % TFA dans H<sub>2</sub>O, éluant B : 0,09 % TFA dans acétonitrile/ H<sub>2</sub>O (90/10)

Les résultats présentés dans la suite de cette partie sont donc issus d'une compilation de l'analyse LC-MS/MS d'une centaine d'échantillons de type surnageant de culture sur colonne dC18 et sur colonne Hypercarb, en combinaison avec les analyses LC-MS/MS des standards chimiques correspondants réalisées sur la colonne appropriée.

## 2. Activités de synthèse de cyclodipeptides des CDPS

Nous avons retrouvé des cyclodipeptides dans les surnageants de culture correspondant à l'expression de 41 des 49 CDPS testées. Le nombre de cyclodipeptides retrouvés par surnageant ainsi que la quantité produite, estimée par l'aire du pic sur un chromatogramme enregistré à 214 nm, sont variables. Dans la Figure 63, la détection des cyclodipeptides produits par les bactéries recombinantes exprimant une CDPS est indiquée par des rectangles d'une couleur bleue dont l'intensité varie en fonction de l'estimation de la quantité produite. Ces mêmes données, mais ne tenant compte que des cyclodipeptides majoritaires (représentant plus de 5% des cyclodipeptides totaux), sont résumées dans le Tableau 16. On peut ainsi constater qu'environ un tiers des CDPS nouvellement caractérisées présentent une spécificité stricte. Cette caractéristique ne semble pas être liée à l'appartenance à une sous-famille ou une autre ; elle est probablement liée à la nature des substrats eux-mêmes tant au niveau de la séquence du bras accepteur de l'ARNt qu'au niveau de l'entité aminoacyle qui peut présenter des particularités physico-chimiques. Les autres CDPS caractérisées démontrent une spécificité relâchée dans nos conditions expérimentales, même si un cyclodipeptide est la plupart du temps retrouvé comme produit majoritaire à côté de plusieurs produits cyclodipeptidiques secondaires. Il est à noter que l'expression recombinante des CDPS 22 et 30 a conduit à la synthèse de plusieurs cyclodipeptides majoritaires en quantités équivalentes.

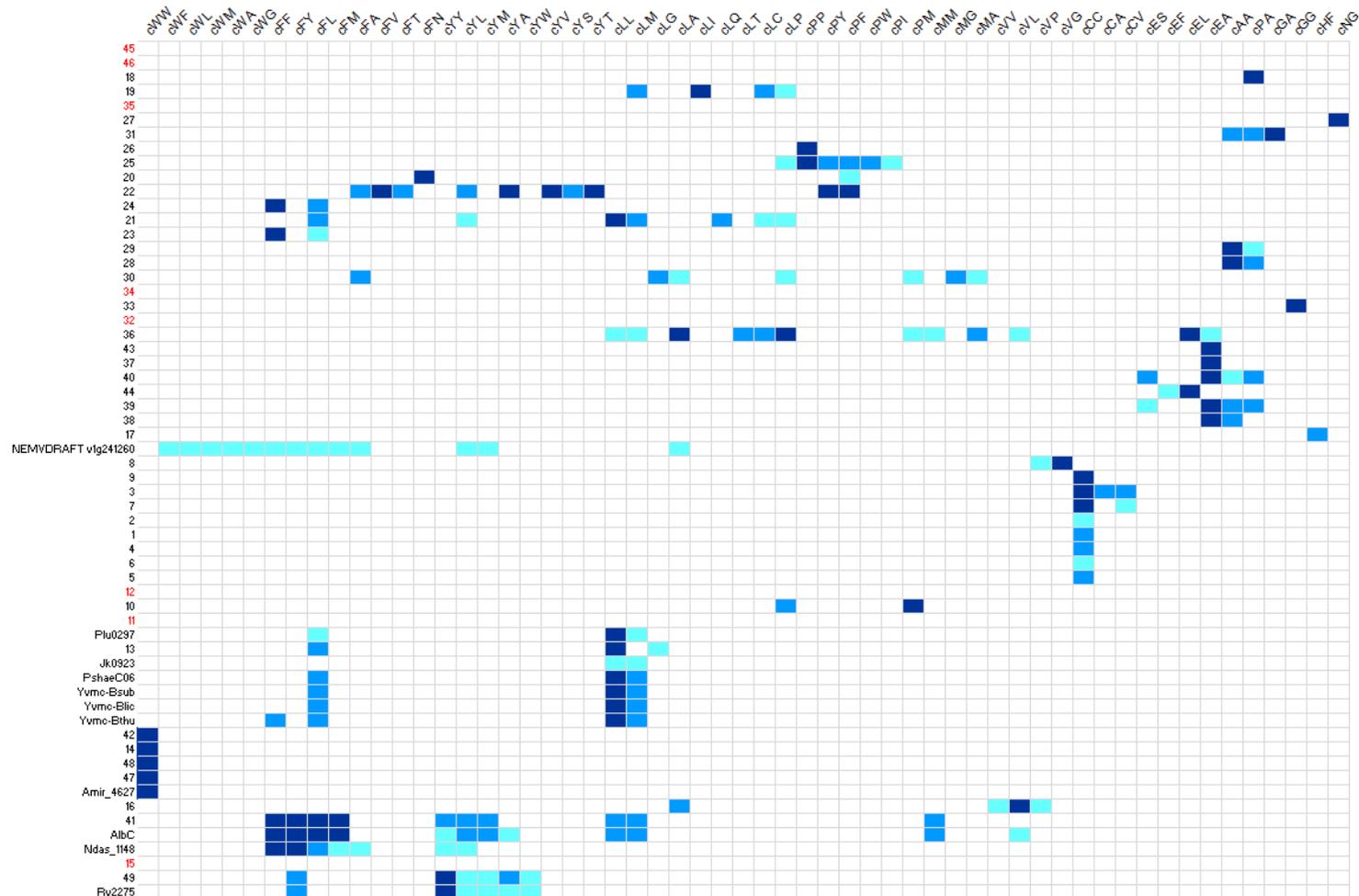


Figure 63 : Cyclodipeptides produits par la famille des CDPS

La présence des cyclodipeptides (ligne supérieure) dans les surnageants de culture de bactéries exprimant chaque CDPS (colonne de gauche) est indiquée par un rectangle bleu. L'intensité de la couleur est liée à la valeur de l'aire sous le pic (A) du chromatogramme enregistré à 214 nm pour les 41 nouvelles CDPS (bleu foncé,  $A > 1000$  ; bleu intermédiaire,  $1000 \geq A > 100$  ; bleu clair,  $100 \geq A > 1$ , d'après l'annexe électronique 2) ou à la quantité de cyclodipeptide synthétisé pour les 11 CDPS précédemment caractérisées (bleu foncé,  $q > 10 \text{ mg.L}^{-1}$  de culture ; bleu intermédiaire,  $10 \geq q > 1$  ; bleu clair,  $1 \geq A > 0,01$ ). Les CDPS sont classées dans l'ordre de l'arbre de la Figure 56 et sont indiquées par leur numéro ou par le nom qui leur a été donné dans des publications précédentes. Les CDPS inactives dans nos conditions sont en rouge.

**Tableau 16 : Cyclodipeptides majoritaires produits par les CDPS nouvellement identifiées**

CDPS	Organisme	Sous-famille de CDPS	Activité <i>in vivo</i> <sup>1</sup>
1	<i>Kutzneria</i> sp. 744	NYH	cCC (100%)
2	<i>Streptomyces</i> sp. AA4	NYH	cCC (100%)
3	<i>Austwickia chelonae</i>	NYH	cCC (63%) cCA (25%) cCV (12%)
4	<i>Saccharothrix espanaensis</i>	NYH	cCC (100%)
5	<i>Amycolatopsis azurea</i>	NYH	cCC (100%)
6	<i>Amycolatopsis decaplanina</i>	NYH	cCC (100%)
7	<i>Streptomyces somaliensis</i>	NYH	cCC (99%)
8	<i>Candidatus</i> <i>Odyssella thessalonicensis</i>	NYH	cGV (99%)
9	<i>Legionella pneumophila</i>	NYH	cCC (100%)
10	<i>Sphingobium japonicum</i>	NYH	cPM (95%)
11	<i>Kutzneria</i> sp. 744	NYH	-
12	<i>Saccharopolyspora spinosa</i>	NYH	-
13	<i>Marinomonas mediterranea</i>	NYH	cLL (98%)
14	<i>Streptomyces cattleya</i>	NYH	cWW (100%)
15	<i>Streptomyces cattleya</i>	NYH	-
16	<i>Streptomyces svicens</i>	NYH	cLV (90%)
17	<i>Parabacteroides</i> sp. 20_3	SYQ	cHF (100%)
18	<i>Moorea producens</i>	XYP	cAP (100%)
19	<i>Pseudomonas aeruginosa</i>	XYP	cIL (93%)
20	<i>Diplorickeetsia massiliensis</i>	XYP	cFN (90%)
21	<i>Diplorickeetsia massiliensis</i>	XYP	cLL (80%)
22	<i>Diplorickeetsia massiliensis</i>	XYP	cYV (30%) cYP (23%) cYT (11%) cYA (10%) cFP (10%) cFV (7%)
23	<i>Candidatus</i> <i>Glomeribacter gigasporarum</i>	XYP	cFF (93%) cFL (7%)
24	<i>Rickettsiella grylli</i>	XYP	cFF (53%) cFL (47%)
25	<i>Parachlamydia acanthamoebae</i>	XYP	cPP (75%)
26	<i>Candidatus</i> <i>Protochlamydia amoebophila</i>	XYP	cPP (100%)
27	<i>Micavibrio aeruginosavorus</i>	XYP	cGN (100%)
28	<i>Legionella longbeachae</i>	XYP	cAA (95%)
29	<i>Fluoribacter dumoffii</i>	XYP	cAA (97%)
30	<i>Rickettsiella grylli</i>	XYP	cMA (23%) cMG (27%)
31	<i>Rickettsiella grylli</i>	XYP	cAG (86%) cAA (14%)
32	Uncultured bacterium ACD_69C00020	XYP	-
33	<i>Fluoribacter dumoffii</i>	XYP	cGG (100%)
34	<i>Diplorickeetsia massiliensis</i>	XYP	-
35	<i>Micavibrio aeruginosavorus</i>	XYP	-
36	<i>Vibrio brasiliensis</i>	XYP	cLE (58%) cLA (20%) cLP (14%)
37	<i>Burkholderia lata</i>	XYP	cEA (100%)
38	<i>Gordonia effusa</i>	XYP	cEA (97%)
39	<i>Nocardia brasiliensis</i>	XYP	cEA (96%)
40	<i>Streptomyces albulus</i>	XYP	cEA (97%)
41	<i>Streptomyces albulus</i>	NYH	cFL (48%) cFF (23%) cFY (19%) cFM (6%)
42	<i>Streptomyces albulus</i>	NYH	cWW (100%)
43	<i>Pseudomonas</i> sp. M47T1	XYP	cEA (100%)
44	<i>Pseudomonas protegens</i>	XYP	cLE (99%)
45	<i>Gibberella zeae</i>	XYP	-
46	<i>Desulfovibrio aespoeensis</i>	XYP	-
47	<i>Streptomyces</i> sp. AA4	NYH	cWW (100%)
48	<i>Streptomyces cattleya</i>	NYH	cWW (100%)
49	<i>Streptomyces venezuelae</i>	NYH	cYY (92 %)

<sup>1</sup> Les cyclodipeptides majoritaires produits par la souche M15-pREP4 exprimant le gène de CDPS correspondant sont indiqués. Les pourcentages entre parenthèses indiquent la proportion que représente chaque cyclodipeptide par rapport à la totalité des cyclodipeptides produits. Ils sont calculés en utilisant les valeurs d'aire sous les pics correspondants du chromatogramme enregistré à 214 nm. Les cyclodipeptides sont rangés par ordre décroissant de

proportion et les cyclodipeptides représentant moins de 5% des produits formés ne sont pas indiqués. cCC indique la détection de cyclo(L-cystine), « - » indique qu'aucun cyclodipeptide n'a été détecté.

### 3. Diversité de cyclodipeptides apportée par les 41 nouvelles CDPS

Intéressons-nous maintenant à la diversité de cyclodipeptides offerte par les nouvelles CDPS, que les cyclodipeptides soit majoritairement produits ou détectés en très faible quantité.

Les 11 CDPS précédemment caractérisées synthétisent 19 cyclodipeptides, de nature plutôt hydrophobe avec des chaînes latérales aromatiques et aliphatiques<sup>26,27,346,348</sup>. Les 41 CDPS nouvellement identifiées synthétisent certains de ces cyclodipeptides mais aussi 35 cyclodipeptides décrits pour la première fois en tant que produits de CDPS (Figure 64a). Plusieurs de ces cyclodipeptides sont constitués d'acides aminés qui n'avaient pas encore été observés dans les produits synthétisés par des CDPS, à savoir l'acide aminé Gly, des acides aminés avec des chaînes latérales de nature hydrophobe (Val, Pro, Ile) et des acides aminés chargés et polaires (Glu, His, Asn, Gln, Ser, Thr, Cys). 17 des 20 acides protéinogéniques sont donc désormais retrouvés dans des cyclodipeptides synthétisés par des CDPS.

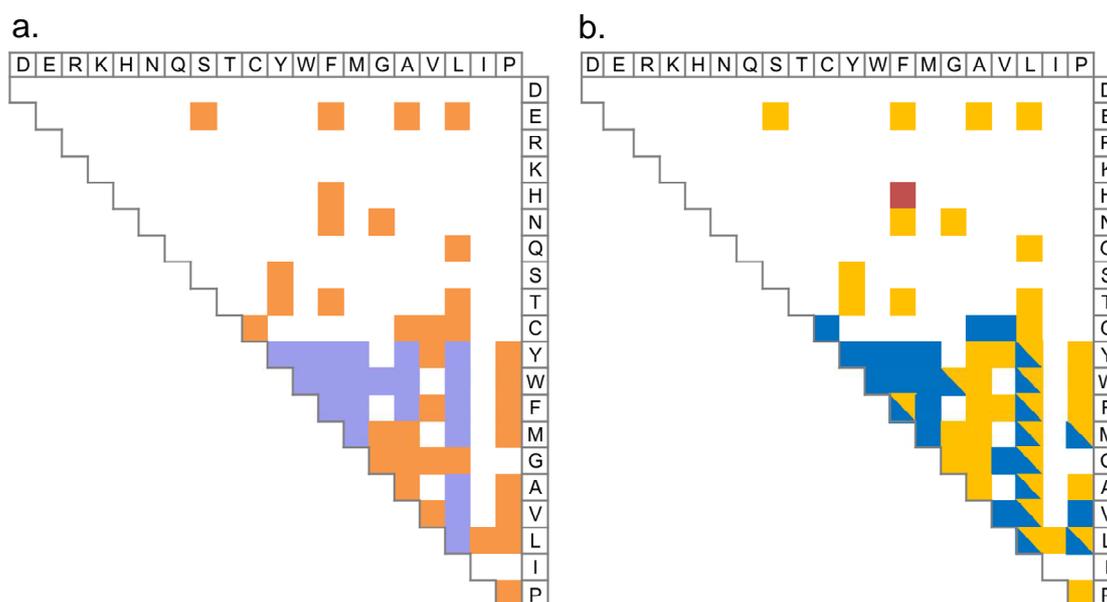


Figure 64 : Diversité de cyclodipeptides produits par les CDPS

Les deux acides aminés constituant les cyclodipeptides sont indiqués, dans le code à une lettre, dans la ligne supérieure et la colonne de droite. La présence d'un carré coloré à l'intersection d'une ligne et d'une colonne indique que le cyclodipeptide correspondant a été détecté dans le surnageant de culture de bactéries exprimant un gène de CDPS. a. Cyclodipeptides produits par les 11 CDPS précédemment caractérisées (en orange) et les 41 CDPS nouvellement identifiées (en orange). b. Cyclodipeptides produits par les membres des sous-familles NYH (en bleu), XYP (en jaune) et le membre SYQ (en rouge foncé).

Nous avons étudié si l'une ou l'autre des sous-familles identifiées est à l'origine de la diversité accrue des cyclodipeptides (Figure 64b). Les membres de la famille NYH participent à l'incorporation dans les cyclodipeptides de dix acides aminés. Dans notre étude, l'unique acide aminé supplémentaire incorporé par cette sous-famille est la cystéine. Les membres de la famille XYP sont, pour leur part, capables de prendre en charge 16 des 20 acides aminés protéinogéniques, à savoir des acides aminés hydrophobes, également pris en charge par la famille NYH, ainsi que des acides aminés polaires et chargés tels que Gln, Thr, Asn, Ser et Glu. Le nombre de cyclodipeptides synthétisés par chaque sous-famille diffère également, avec 39 cyclodipeptides pour la sous-famille XYP contre 26 pour la sous-famille NYH. Cette diversité de produits est à mettre en regard de la diversité de séquences observée chez la famille XYP (cf. III.2 Caractéristiques fonctionnelle et structurale des sous-familles).

Ainsi, la diversité des cyclodipeptides synthétisés par les CDPS a été grandement augmentée en élargissant la famille. Seuls des cyclodipeptides ayant des chaînes latérales chargées positivement ne sont pas retrouvés, comme produit de CDPS. Il ne peut d'ailleurs pas être exclu que les phases stationnaires utilisées pour l'analyse ne permettent pas de retenir ces composés et donc de les détecter. La sous-famille XYP, dont les séquences des membres sont plus dégénérées, est responsable de la synthèse d'une plus grande diversité de cyclodipeptides. Environ un tiers des CDPS présente une spécificité élevée menant à la synthèse d'un seul cyclodipeptide tandis que le reste des CDPS présente une spécificité relâchée menant, dans la plupart des cas, à la synthèse d'un cyclodipeptide majoritaire et de plusieurs cyclodipeptides secondaires.

## V. Vers la prédiction de la spécificité des CDPS

La prédiction de spécificité d'enzymes de voies de biosynthèse constitue un outil d'un intérêt majeur car il permet d'analyser rapidement un grand nombre de séquences sans systématiquement utiliser l'analyse biochimique. Cette approche est indispensable lors de la recherche de nouvelles activités codées dans les génomes (« genome mining »). De plus, elle aide à poser les fondements pour une meilleure compréhension des mécanismes de reconnaissance enzyme/substrat et pour l'utilisation de l'ingénierie d'enzymes afin de créer de nouvelles activités. Cependant, elle nécessite au départ la caractérisation biochimique d'un nombre suffisant d'enzymes et la connaissance des zones d'interaction entre le substrat et l'enzyme. Les travaux réalisés pour la détermination de la spécificité des substrats des domaines d'adénylation des NRPS illustrent bien ces deux aspects<sup>279,401</sup> (cf Etude bibliographique. IV.2 Voies de biosynthèse de dicétopipérazines dépendant des NRPS).

## 1. Prédiction des résidus constituant les poches de liaison des substrats chez les CDPS

Dans le cas des CDPS, l'étude moyen-débit réalisée au cours de mon projet, qui a identifié 41 nouvelles CDPS et les cyclodipeptides qu'elles synthétisent, peut aider à comprendre la spécificité de ces enzymes. Par ailleurs, la détermination de la structure d'AlbC liée à un mime d'un intermédiaire dipeptidyle a permis de montrer qu'AlbC, membre de la sous-famille NYH, possède deux poches de liaison des substrats : une poche P1 qui accommode l'entité aminoacyle du premier substrat et une poche P2 qui accommode celle du deuxième substrat. Les huit résidus constituant la poche P1 et les sept résidus constituant la poche P2 ont ainsi pu être identifiés<sup>364</sup> (cf. Etude bibliographique. V.1.4 Les CDPS utilisent un mécanisme catalytique de type ping-pong) (Figure 65).

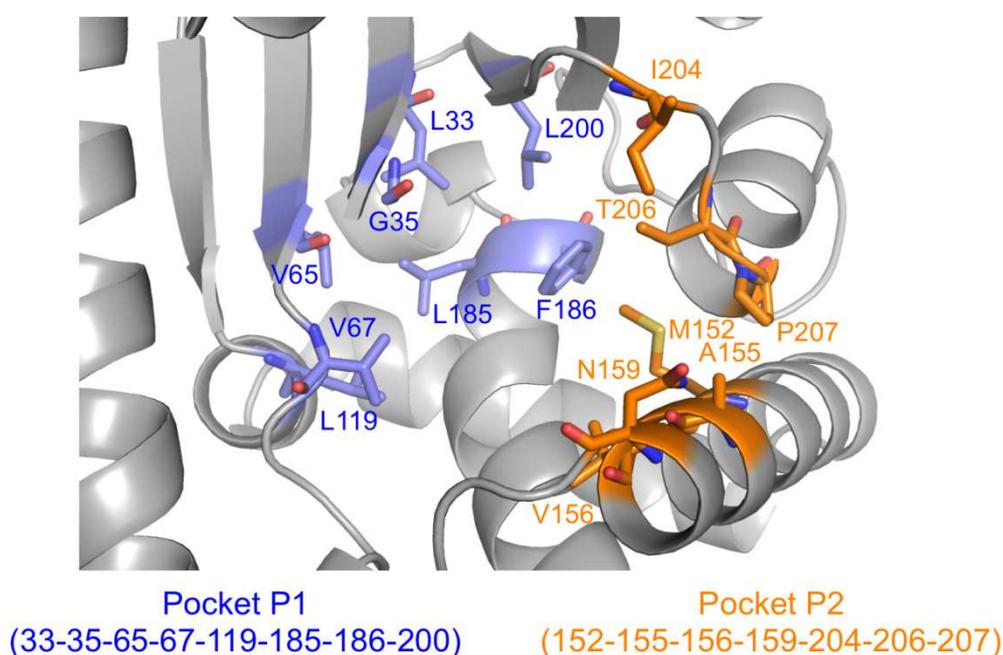


Figure 65 : Résidus constituant les poches P1 et P2

**Vue tridimensionnelle des poches P1 et P2 accommodant respectivement les entités aminoacyles du premier et du second substrat. Les huit résidus constituant la poche P1 sont en bleu et les sept résidus constituant la poche P2 sont en orange.**

La poche P1 est une poche présente dans la structure tridimensionnelle d'AlbC libre<sup>361</sup>, et les résidus qui la constituent sont identiques à ceux identifiés dans la structure d'AlbC liée au mime de dipeptidyle<sup>364</sup>. La poche P2, quant à elle, n'est visible que dans la structure tridimensionnelle d'AlbC liée. Les résidus constituant les poches P1 de Rv2275 et YvmC-Blic sont identifiés car les structures tridimensionnelles de ces enzymes sont connues<sup>362,363</sup>. Dans le cas des poches P2 ainsi que pour les poches P1 et P2 des autres CDPS, la détermination des résidus les constituant ne peut être réalisée que par comparaison de séquences. Comme les CDPS montrent une conservation de la prédiction des éléments de

structures secondaires et du mode de repliement (Annexe électronique 4), nous avons pu comparer les séquences qui ont été ajustées manuellement en prenant en compte les informations fournies par les résultats des interrogations HHPRED (Annexe électronique 5). Cette comparaison a permis de prédire les résidus constituant les poches P1 et P2 de toutes les CDPS actives (Annexe électronique 6). Les résidus identifiés par cette analyse bio-informatique pour la poche P1 de Rv2275 et celle de YvmC-Blic sont en accord avec ceux identifiés avec les deux structures tridimensionnelles correspondantes.<sup>362,363</sup>.

## 2. Les CDPS de même activité montrent des séquences consensus pour les résidus constituant les poches P1 et P2

Avec les prédictions des résidus constituant les poches P1 et P2, nous avons voulu déterminer si la production d'un cyclodipeptide majoritaire est associée à une conservation de certains résidus de ces poches. Nous avons ainsi constitué des groupes de spécificité en s'appuyant, dans un premier temps, uniquement sur l'activité de synthèse du cyclodipeptide majoritaire. Néanmoins, le cas des CDPS synthétisant du cyclo(Leu-Leu) a fait évoluer notre manière de constituer les groupes de CDPS. En effet, des CDPS synthétisant le cyclo(Leu-Leu) sont retrouvées dans la sous-famille NYH et dans la sous-famille XYP. En comparant les résidus constituant les poches chez ces enzymes, il apparaît que certains résidus sont conservés au sein d'une même sous-famille mais ils diffèrent entre les deux sous-familles (Tableau 17). Ceci est particulièrement remarquable pour les résidus prédits pour la poche P2. Ainsi, il semblerait que deux combinaisons différentes de résidus soient possibles pour mener à la synthèse du même cyclodipeptide. Cela pourrait résulter de deux évolutions différentes mais convergentes, car ces CDPS appartiennent à des clades phylogénétiques différents mais ont une même activité. Ainsi, les regroupements des CDPS ayant la même activité de synthèse ont été réalisés indépendamment pour chacune des sous-familles. (Annexe électronique 6).

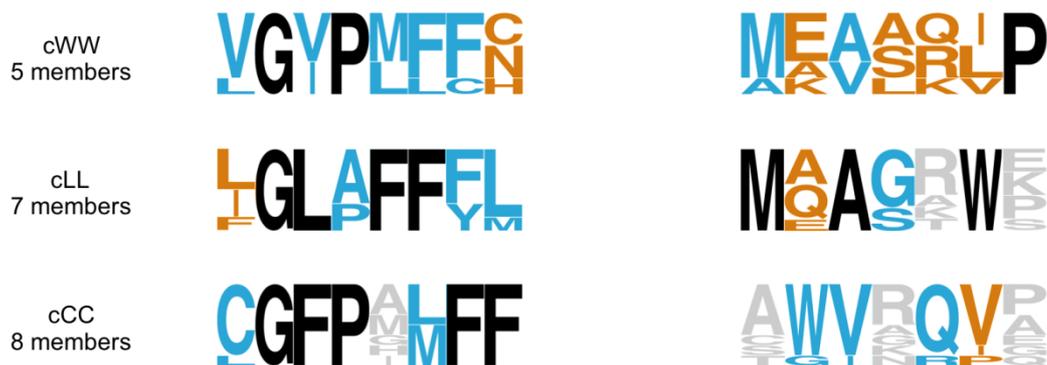
**Tableau 17 : Comparaison des résidus constituant les poches P1 et P2 chez les CDPS synthétisant du cyclo(Leu-Leu)**

Sous-famille	CDPS	Organisme	Résidus poche P1	Résidus poche P2
NYH	YvmC-Blic	<i>B. licheniformis</i>	LGLAFFFL	MAAGRWK
	YvmC-Bthu	<i>B. thurengiensis</i>	IGLAFFFL	MAAGKWE
	YvmC-Bsub	<i>B. subtilis</i>	LGLAFFFL	MEAGRWK
	Plu0297	<i>P. luminescens</i>	LGLPFFYL	MQAGRWS
	Jk0923	<i>C. jeikeium</i>	IGLPIAYL	MAASTWP
	13 (Marme_0276)	<i>M. mediterranea</i>	FGLPFFYL	MQAGRWP
	pSHaeC06	<i>S. haemolyticus</i>	IGLAFFFM	MQASAWK
XYP	21 (Dmas2_11124)	<i>D. massiliensis</i>	IGVSWAIL	LCHKSEL

Ces groupes de spécificité sont constitués d'un nombre variable de membres, allant de un à huit. Lorsqu'un groupe de spécificité contient au moins cinq membres, nous avons établi un motif de séquence

des deux poches en utilisant une représentation sous forme de logo<sup>388</sup>. Cela a été réalisé pour trois groupes de CDPS dans la sous-famille NYH (CDPS synthétisant du cyclo(Leu-Leu), du cyclo(Trp-Trp) et du cyclo(L-cystine)) et un groupe dans la famille XYP (CDPS synthétisant du cyclo(Ala-Glu)). (Figure 66)

#### NYH subfamily



#### XYP subfamily



Figure 66 : Motif de séquence des acides aminés constituant les deux poches P1 et P2 pour les groupes synthétisant le même cyclodipeptide

Les résidus suspectés de constituer les poches P1 et P2 ont été déterminés pour chaque CDPS active (Annexe électronique 6). Des logos correspondant à la représentation graphique de la fréquence des acides aminés aux positions des résidus de P1 et P2 sont déterminés pour les groupes de CDPS appartenant à la même sous-famille et ayant la même activité de synthèse de cyclodipeptide : cWW (CDPSs 14, 42, 47, 48 et Amir\_4627; identité de séquence : 35–55%); cLL (CDPSs 13, jk0923, Plu0297, pSHaeC06, YvmC\_BSU, YvmC\_BLIC et YvmC\_BTHU; identité de séquence : 26–70%); cCC (CDPSs 1, 2, 3, 4, 5, 6, 7 et 9; identité de séquence : 28–93%); cAE (CDPSs 37, 38, 39, 40 et 43; identité de séquence : 40–58%); cXE (CDPSs 36, 37, 38, 39, 40, 43 et 44; identité de séquence : 38–58%); cAX (CDPSs 28, 31, 37, 38, 39, 40 et 43; identité de séquence : 16–58%). A chaque position, la présence d'un seul, de deux, de trois ou plus acides aminés est indiquée respectivement par la couleur noire, bleue, orange ou grise.

Les poches P1 et P2 étant indépendantes<sup>364</sup>, nous avons également pu rassembler les CDPS accommodant le même acide aminé dans une poche, sans tenir compte de l'acide aminé accommodé dans

l'autre poche. Pour réaliser ce type de regroupement, il a fallu déterminer quelle poche accommode quel acide aminé et, pour cela, nous avons utilisé les données acquises sur AlbC. AlbC a une spécificité relâchée et elle produit des cyclodipeptides de type cyclo(Phe-Xaa), Phe étant l'acide aminé systématiquement incorporé et Xaa un acide aminé variable. Il s'avère que le résidu Phe est accommodé dans la poche P1 alors que la poche P2 accommode le résidu variable. Comme AlbC, la plupart des CDPS ont une spécificité relâchée et produisent plusieurs cyclodipeptides de type cyclo(AA1-Xaa), AA1 étant l'acide aminé systématiquement incorporé. En faisant l'hypothèse que ces CDPS fonctionnent toutes selon un mécanisme ping-pong avec formation d'une acyl-enzyme, nous pouvons supposer que AA1 est accommodé dans la poche P1 et Xaa dans la poche P2. Ainsi, chez la sous-famille XYP, les deux CDPS synthétisant du cyclo(Ala-Ala) et les cinq synthétisant du cyclo(Ala-Glu) peuvent être regroupées pour constituer un motif de poche pour P1, qui incorpore dans les deux cas du Ala (l'accommodation de Glu dans P1 ne pouvant pas être considérée du fait de l'hydrophobicité de cette poche). De la même manière, les cinq CDPS produisant du cyclo(Ala-Glu) et les deux produisant du cyclo(Leu-Glu) peuvent être regroupées pour constituer un motif de séquence pour la poche P2 accommodant du Glu (Figure 66).

Plusieurs observations et tendances sont à relever en analysant les résidus des poches et les motifs de séquence. Tout d'abord, la production d'un cyclodipeptide corrèle avec la conservation de certains résidus au sein des poches (Figure 66). Cependant, les positions conservées et leur degré de conservation ne sont pas toujours les mêmes. Par exemple, les CDPS synthétisant du cyclo(Trp-Trp) montrent la stricte conservation de seulement deux positions dans P1 (35 et 67) et une position dans P2 (207), alors que les CDPS synthétisant du cyclo(L-cystine) affichent la stricte conservation de 5 positions dans P1 (35, 65, 67, 186, 200) et aucune position dans P2. On peut noter, dans certains cas, une adéquation entre la nature physico-chimique de l'acide aminé à incorporer et celle de la poche, comme pour la poche P2 des CDPS synthétisant du cyclo(Ala-Glu) qui montre de nombreux résidus polaires plus ou moins conservés alors qu'elle doit accueillir la chaîne latérale du glutamate. Néanmoins, avec les connaissances disponibles actuellement, les compositions des séquences de poche sont insuffisantes pour comprendre pourquoi un acide aminé est accommodé par l'une ou l'autre des poches.

Cependant, on peut prédire l'importance d'une position dans le cas des CDPS synthétisant soit du cyclo(Ala-Glu), soit du cyclo(Leu-Glu). En effet, l'unique différence entre les poches P1 de ces deux groupes de CDPS se situe à la position 35 (Annexe électronique 6). La présence d'une alanine à cette position corrèle avec l'incorporation de Leu alors que la présence d'une leucine corrèle avec l'incorporation de Ala (Tableau 18). La sélection d'un acide aminé par rapport à l'autre pourrait donc se faire en fonction de l'encombrement stérique de la poche. Cet exemple est le seul où une explication rationnelle peut être avancée.

**Tableau 18 : Comparaison des résidus constituant les poches P1 et P2 chez les CDPS synthétisant du cyclo(Leu-Glu) ou du cyclo(Ala-Glu).**

Cyclodipeptides produits	CDPS	Résidus poche P1 <sup>1</sup>	Résidus poche P2
<b>cLE</b>	44 (PFL_1389)	VAVAWIIV	TVFRQMT
<b>cLE cLA cLP cLx</b>	36 (VIBR0546_16848)	IAMGWAI I	TTFRQMS
<b>cAE</b>	37 (Bcep18194_C7320)	LLGAWIII	TVFRKMT
<b>cAE</b>	38 (GOEFS_047_00070)	ILAAWIIV	TRFRQIS
<b>cAE cAA</b>	39 (O3I_35936)	VLAAWIIV	TRFRQIS
<b>cAE</b>	40 (K530_11825)	ILAAWIIV	TKFRQMT
<b>cAE</b>	43 (PMM47T1_20098)	LLAAWIIV	TLFRQMT

<sup>1</sup>Le résidu de la position 35 est repéré en fuchsia.

La pertinence des motifs de séquences est illustrée par l'exemple suivant. La CDPS 8 est proche sur l'arbre phylogénétique du groupe de CDPS formant du cyclo(L-cystine) (CDPS 1 à 7 et 9) alors qu'elle synthétise du cyclo(Gly-Val). Les résidus constituant ses poches ne coïncident absolument pas avec les motifs des poches P1 et P2 du groupe de CDPS produisant du cyclo(L-cystine). Ainsi, si la CDPS 8 n'avait pas été testée au cours de l'étude, l'analyse de la composition des résidus des poches aurait suffi à exclure, à juste titre, cette CDPS du groupe produisant du cyclo(L-cystine).

Ainsi même si, dans la majorité des cas, l'incorporation d'un acide aminé ne peut pas être expliquée rationnellement, des motifs de séquences peuvent être définis pour une activité donnée de synthèse d'un cyclodipeptide, et ce pour un ensemble de CDPS de la même sous-famille. Nous avons par la suite voulu éprouver ces motifs de séquence comme outil prédictif.

### 3. Validation expérimentale du modèle prédictif

En septembre 2014, une nouvelle interrogation des banques de données génomiques a permis d'identifier 200 séquences de CDPS. Les séquences partielles ont été retirées et le codon initiateur de 30 séquences a été corrigé. Cet ensemble de séquences, une fois qu'elles ont été alignées, a servi au calcul d'un nouvel arbre phylogénétique (Annexe électronique 7). Il est à noter que les nouvelles CDPS se répartissent dans l'une ou l'autre des sous-familles identifiées et sont compatibles avec les signatures fonctionnelles définies (cf. III Des sous-familles de CDPS aux résidus catalytiques différents).

Les résidus constituant les poches P1 et P2 des nouvelles CDPS ont été déterminés et ont été comparés aux motifs de séquences établis précédemment (Annexe électronique 8 et Figure 66). Ceci nous a mené à l'identification de CDPS susceptibles d'être membres des groupes synthétisant du cyclo(Leu-Leu), du cyclo(Trp-Trp), du cyclo(L-cystine) ou du cyclo(Ala-Glu) (en couleur sur l'arbre). Afin de vérifier la valeur de ces prédictions, nous avons sélectionné trois nouvelles CDPS dans chacun de ces groupes, de manière à ce que chaque séquence partage le minimum d'identité de séquence avec les CDPS déjà

caractérisées. Ces 12 CDPS potentielles ont été testées pour leur activité de synthèse de cyclodipeptides comme lors de l'étude principale. 10 CDPS se sont révélées actives et le cyclodipeptide majoritairement produit correspondait à celui prédit (Tableau 19). Ainsi, en utilisant le lien existant entre les motifs de séquence de poche et la spécificité des CDPS, nous avons été capables de prédire l'activité de CDPS potentielles, et ce malgré une identité de séquence modérée avec les membres caractérisés du groupe d'appartenance prédit.

**Tableau 19 : Prédiction et validation de l'activité de synthèse de cyclodipeptides des CDPS**

CDPS <sup>1</sup>	Organisme	Résidus de la poche <sup>2</sup>		Prédiction	Identité de séquence <sup>3</sup>	Cyclodipeptides synthétisés <sup>4</sup>
		P1	P2			
50	<i>Streptomyces</i> sp. F12	CGFPWLFF	AWVGQPDV	cCC	34–52%	cCC (100%)
51	<i>Nocardioopsis gilva</i>	CGYPSLFF	AWVGRPEF	cCC	34–52%	-
52	<i>Kutzneria albida</i> DSM 43870	CGFPSSLFF	AWVRQVWF	cCC	32–57%	cCC (100%)
53	<i>Staphylococcus pseudintermedius</i>	IGLAFFFFM	MQASAWE	cLL	26–76%	-
54	<i>Bacillus</i> sp. 171095_106	LGLAFFFFL	MAAGKWR	cLL	30–63%	cLL (56%) cLF (23%) cLM (20%)
55	<i>Bacillus</i> sp. NSP9.1	LGLAFFFFL	MAAGRWK	cLL	29–82%	cLL (52%) cLF (28%) cLM (20%)
56	<i>Streptomyces albus</i>	VGIPMFFC	ATVARLP	cWW	35–56%	cWW (100%)
57	<i>Streptomyces</i> sp. CNH287	VGVPMFFC	MDVAQLP	cWW	39–60%	cWW (100%)
58	<i>Nocardioopsis halophila</i>	VGVPMFFN	MEVARLP	cWW	39–61%	cWW (100%)
59	<i>Xenorhabdus doncetiae</i>	VLAAWIII	TVFRKMT	cAE	37–66%	cAE (92%) cSE (7%)
60	<i>Photorhabdus luminescens</i> BA1	ILAAWIIIV	TIFRQMT	cAE	42–54%	cAE (100%)
61	<i>Mycobacterium neoaurum</i>	ILAAWIIIV	TRFRQIS	cAE	41–69%	cAE (80%) cAP (18%)

<sup>1</sup>Les CDPS sont regroupées en fonction de l'activité prédite : 50–52, cCC; 53–55, cLL; 56–58, cWW; 59–61, cAE.

<sup>2</sup>Les résidus prédits pour les poches P1 et P2 sont indiqués pour chaque CDPS.

<sup>3</sup>Clustal Omega du site EMBL a été utilisé pour déterminer les identités de séquence entre chaque CDPS et les membres du groupe.

<sup>4</sup>« - » : aucun cyclodipeptide détecté. Détails des résultats disponibles dans l'annexe électronique 2.

En plus des CDPS potentielles pour lesquelles une prédiction d'activité a pu être établie, 30% des nouvelles séquences de CDPS identifiées ont des résidus des poches P1 et P2 similaires à ceux de groupes de CDPS qui possèdent peu de membres (< 5) et pour lesquels un motif de séquence n'a pas été établi (Annexe électronique 8). En déterminant les activités de synthèse de ces CDPS potentielles et celles d'autres CDPS qui vont apparaître dans les bases de données, de nouveaux motifs de séquence de poches vont pouvoir être déterminés et venir enrichir l'outil de prédiction mis en place au cours de cette étude.

Dans une approche visant à augmenter la diversité de cyclodipeptides produits par les CDPS, l'utilisation de cet outil prédictif pour trier les CDPS à étudier apparaît extrêmement pertinente. En effet, si l'on exclut les CDPS appartenant à un groupe de sélectivité déterminé et les CDPS présentant de fortes similitudes de séquence pour P1 et P2 avec des CDPS d'activité connue, il nous reste environ seulement 15% des séquences qui sont susceptibles de correspondre à une CDPS possédant une activité originale. Une étude est actuellement en cours dans le laboratoire pour caractériser l'activité de telles CDPS potentielles, orphelines en ce qui concerne la prédiction d'activité.

L'ensemble des résultats présentés jusqu'à présent dans ce chapitre a fait l'objet d'une publication. Les résultats présentés ci-après sont issus d'études biochimiques ultérieures

L'analyse bioinformatique des séquences des CDPS actives a ainsi révélé l'existence d'au moins deux sous-familles chez les CDPS. Toutes les CDPS étudiées précédemment appartiennent à la sous-famille NYH. Les caractéristiques des nouvelles CDPS (NYH et XYP) se sont largement appuyées sur les données obtenues lors de l'étude biochimique d'AlbC. L'étude biochimique d'autres CDPS apporterait des informations sur le fonctionnement des CDPS de groupe et sous-famille différents. Elle permettrait aussi d'obtenir des données qui ne sont pas accessibles avec AlbC. Par exemple, la spécificité relâchée d'AlbC empêche de déterminer des paramètres cinétiques pour l'activité CDPS. AlbC est également sensible à l'agrégation, ce qui interdit son utilisation à forte concentration. Malgré les essais réalisés au laboratoire, aucun complexe entre AlbC et un ligand représentatif de l'un de ces substrats n'a été obtenu. L'étude biochimique de nouvelles CDPS possédant des caractéristiques différentes d'AlbC revêt donc un intérêt particulier pour le laboratoire. Les travaux présentés dans la partie qui suit sont préliminaires mais ont permis d'identifier un certain nombre de difficultés et d'ouvrir la voie à de nouvelles études actuellement menées au sein du laboratoire.

## **VI. Vers une caractérisation biochimique plus approfondie des CDPS : l'étude *in vitro* des CDPS nouvellement identifiées**

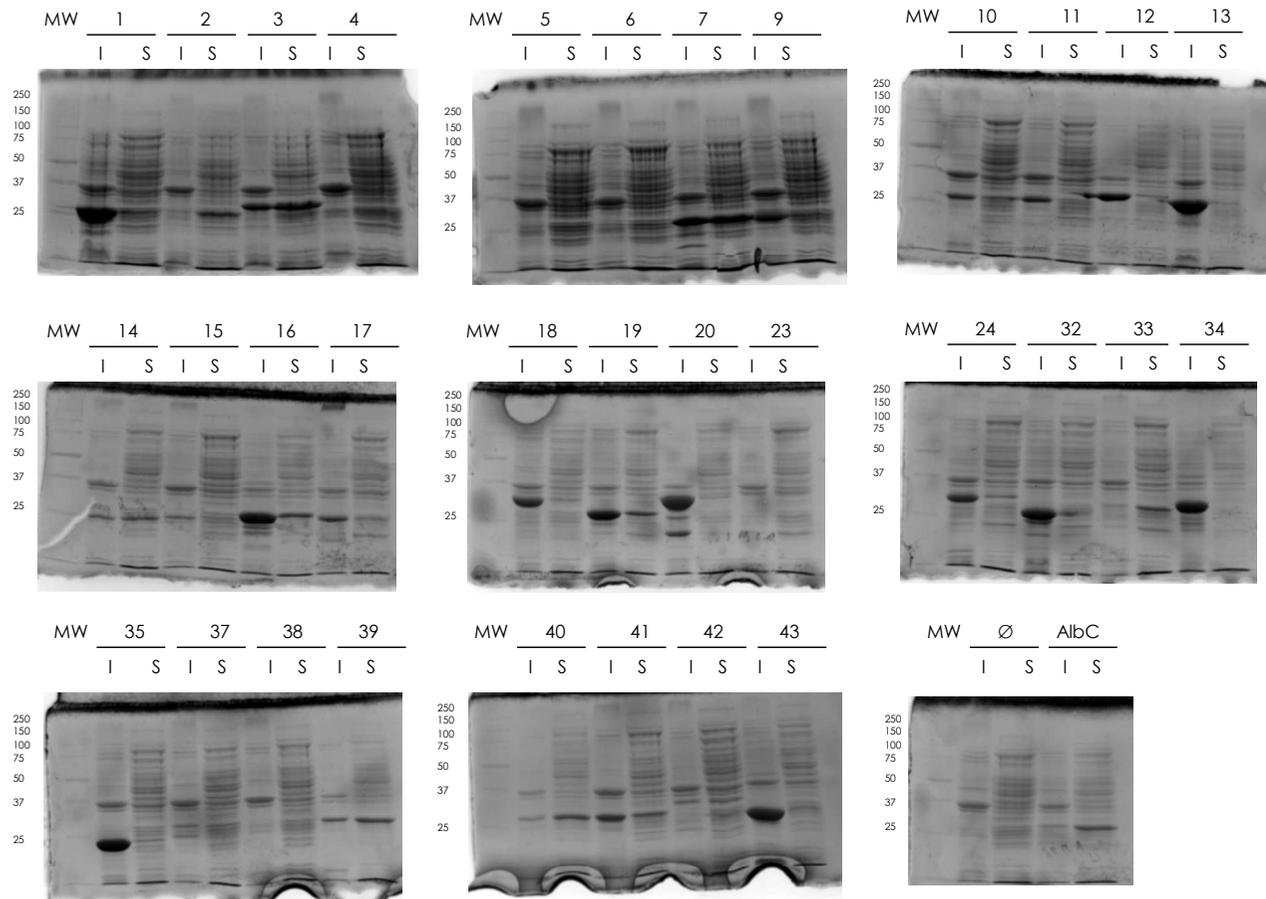
Les résultats décrits précédemment ont révélé de nouvelles caractéristiques des CDPS, que ce soit vis-à-vis de leur activité ou des particularités de séquences. Un certain nombre d'entre elles sont apparues comme des candidats intéressants pour des études biochimiques et structurales. En travail préliminaire à ces études qui se poursuivront au laboratoire, la suite de mon projet a consisté en la production d'un certain nombre de CDPS et l'évaluation de leurs activités *in vitro*.

Je présenterai dans un premier temps la sélection des CDPS que nous avons purifiées, sélection qui utilise notamment les résultats de l'étude *in vivo* (cf. II.1 41 des 49 CDPS sélectionnées actives *in vivo*). Puis, je détaillerai les tests réalisés pour reproduire *in vitro* l'activité observée et les difficultés rencontrées.

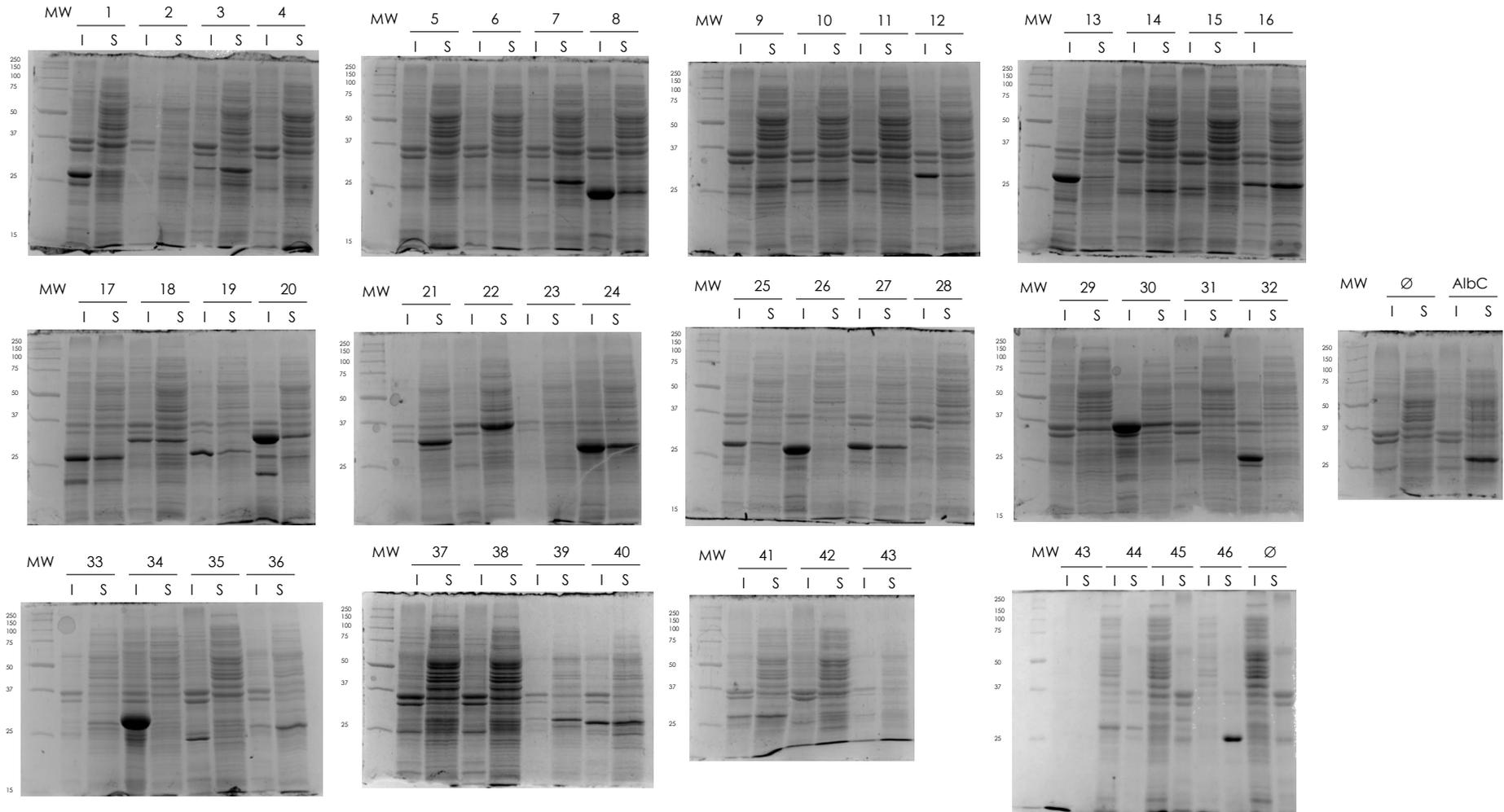
### **1. Sélection et purification de 16 CDPS**

Nous avons choisi des CDPS de manière à avoir une représentation large des activités de synthèse. Entre plusieurs CDPS de même activité, nous avons sélectionné la CDPS retrouvée en quantité la plus importante sous forme soluble lors de l'analyse par SDS-PAGE des contenus protéiques des culots bactériens obtenus durant le criblage *in vivo* des activités. (Figure 67).

a.



b.



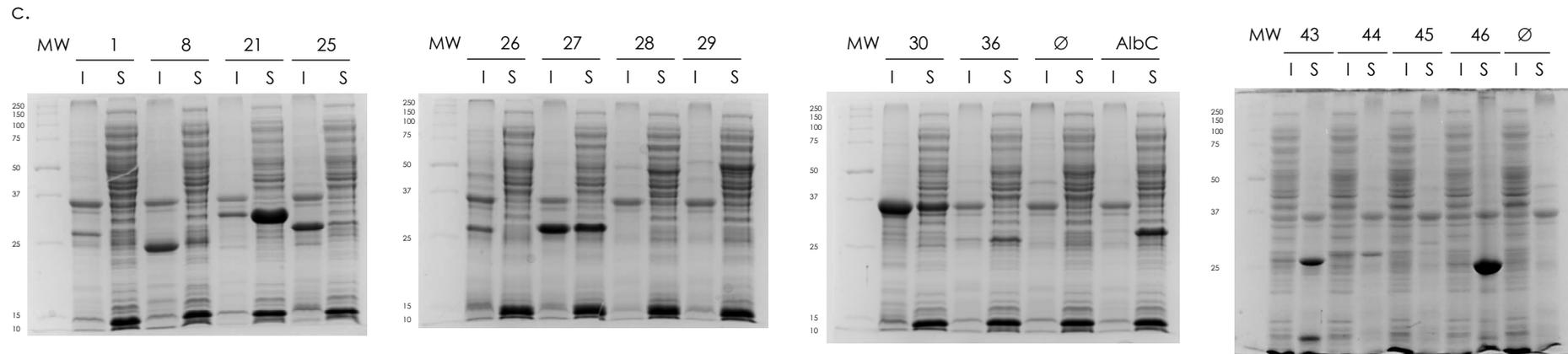


Figure 67 : Analyse SDS-PAGE de l'expression des CDPS et CDPS potentielles 1 à 46

Les fractions solubles (S) et insolubles (I) des bactéries BL21AI (a), et M15-pREP4 (b), BL21AI-pREP4 (c) transformées par le plasmide pIJ196 codant une CDPS potentielle (1à 46) ou non (Ø) après 24h de culture sont analysées sur des gels SDS-PAGE à 12% colorés au bleu de Coomassie.

La production de protéine est évaluée visuellement et une note de 0 à 4 lui est attribuée, 0 correspondant au cas où la bande attendue n'est pas visible et 4 au cas où une bande extrêmement intense est visible (Tableau 20)

**Tableau 20 : CDPS détectées dans les fractions soluble (S) et insoluble (I) préparées depuis des culots bactériens obtenus à l'issue du criblage d'activité de synthèse de cyclodipeptide**

N°CDPS	Cyclodipeptides majoritaires synthétisés	BL21AI			M15-pREP4			BL21AI-pREP4		
		A <sub>600</sub> en fin de culture	Expression protéique <sup>1</sup>		A <sub>600</sub> en fin de culture	Expression protéique <sup>1</sup>		A <sub>600</sub> en fin de culture	Expression protéique <sup>1</sup>	
			I	S		I	S		I	S
1	cCC	4,9	4	1	3,4	3	1	5,0	2	1
2	cCC	2,2	1	2	1,3	?	1	non testé		
3	cCC, cCA	3,4	2	3	2,8	2	3	non testé		
4	cCC	3,9	1	1	3,1	0	0	non testé		
5	cCC	4,2	0	0	3,4	0	0	non testé		
6	cCC	3,3	0	0	3,1	0	0	non testé		
7	cCC	4,2	3	3	3,8	2	3	non testé		
8	cGV	pas de transformants			4,6	4	2	5,3	3	2
9	cCC	4,4	4	1?	2,5	1	1	non testé		
10	cPM	4,6	2	1	2,5	2	2	non testé		
11	-	4,6	2	0	3	0	0	non testé		
12	-	2,6	3	1	2,5	3	1	non testé		
13	cLL	5,1	4	1?	3,5	4	1	non testé		
14	cWW	3,8	2	2	3,4	1	2	non testé		
15	-	4,2	2	1?	3,3	2	0	non testé		
16	cLV	5,4	4	2	3,5	3	3	non testé		
17	cHF	3,5	2	1	2,8	3	3	non testé		
18	cAP	2,9	4	1	3,5	3	3	non testé		
19	cIL	4,2	4	2	1	3	1	non testé		
20	cFN	5,1	4	1	2,6	4	2	non testé		
21	cLL	pas de transformants			2	2	3	4,4	2	4
22	cYV, cYP	pas de transformants			3,1	1?	2	pas de transformants		
23	cFF	3,1	0	1	1,9	0	0	non testé		
24	cFF, cFL	3,7	3	1	2,8	4	2	non testé		
25	cPP	pas de transformants			2,3	3	1	5,3	3	2
26	cPP	pas de transformants			2,1	4	0	4,8	3	0
27	cGN	pas de transformants			2,2	4	3	5,2	3	3
28	cAA	non testé			2,6	?	?	4,6	1	2
29	cAA	non testé			3	?	?	5,2	1	2
30	cMA, cMG	pas de transformants			2,8	4	1	5,4	4	2
31	cAG	pas de transformants			3	0	0	pas de transformants		

N°CDPS	Cyclodipeptides majoritaires synthétisés	BL21AI			M15-pREP4			BL21AI-pREP4		
		A <sub>600</sub> en fin de culture	Expression protéique <sup>1</sup>		A <sub>600</sub> en fin de culture	Expression protéique <sup>1</sup>		A <sub>600</sub> en fin de culture	Expression protéique <sup>1</sup>	
			I	S		I	S		I	S
32	-	4,7	4	0	2	3	0	non testé		
33	cGG	3,2	1	2	2,8	0	1	non testé		
34	-	4,7	4	0	2,3	4	0	non testé		
35	-	5,8	4	1	3,7	1	0	non testé		
36	cLE, cLA	pas de transformants			1,4	1	2	4,1	1	2
37	cEA	3,5	1	1	3,1	0	0	non testé		
38	cEA	3,7	0	0	2,7	0	0	non testé		
39	cEA	1,7	1	2	0,8	0	2	non testé		
40	cEA	1,8	1	2	1,8	2	1	non testé		
41	cFL, cFF	4,3	3	1	3,1	2	2	non testé		
42	cWW	4,9	1	1	3,5	0	0	non testé		
43	cEA	6,3	3	0	0,6	0	0	non mesuré	1	3
44	cLE	non testé			non mesuré	2	1	non mesuré	1	1
45	-	non testé			non mesuré	1	1	non mesuré	0	1
46	-	non testé			non mesuré	0	3	non mesuré	1	4
AlbC	cFL, cFF	3,9	1	2	2,7	1	4	4,9	1	3

<sup>1</sup> I : fraction insoluble ; S : fraction soluble.

Nous avons ainsi sélectionné 16 CDPS (notées en orange dans le Tableau 20) pour leur purification ultérieure. Nous avons appliqué le protocole à grande échelle utilisé dans le laboratoire qui part de 1 l de culture bactérienne. Ce protocole comprend classiquement trois étapes chromatographiques : une colonne d'affinité de type IMAC (Immobilized Metal Affinity Chromatography) chargée en nickel, une colonne d'héparine et une colonne de filtration sur gel<sup>26</sup>. Etant donné le nombre important de protéines à purifier, nous avons décidé, dans un premier temps, d'effectuer uniquement la colonne d'affinité chargée en nickel (il est à noter que les CDPS nouvellement identifiées ont été exprimées depuis le vecteur pIJ196-CDPS et donc produites avec une étiquette 6×His). Cette unique étape de purification permet d'obtenir, à deux exceptions près (CDPS 21 et 20), une fraction enrichie et suffisamment pure de chaque CDPS (Figure 68). Pour la CDPS 21, la fraction est contaminée et présente, de plus, un aspect jaune. La CDPS 20 n'a pas été purifiée dans des quantités suffisantes pour être étudiée dans la suite du travail, malgré deux essais de purification. Pour les autres CDPS purifiées, nous avons obtenu des quantités allant de 2 à 100 mg environ. Par ailleurs, dans nos conditions, cinq des 16 protéines purifiées ont tendance à précipiter, dont deux, les CDPS 19 et 27, très fortement. Un résumé des caractéristiques et de la quantité purifiée pour chaque CDPS est présenté dans le Tableau 21).

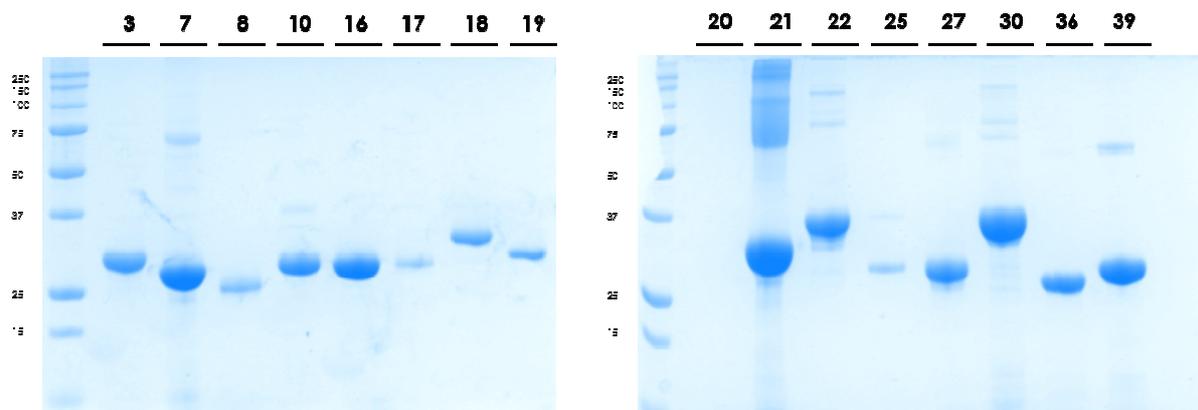


Figure 68 : Analyse SDS-PAGE des CDPS purifiées

Les extraits protéiques obtenus en sortie de colonne d'affinité chargée en nickel sont analysés sur des gel SDS-PAGE à 12% colorés au bleu de Coomassie. Les numéros des CDPS purifiées sont indiqués. Les masses correspondant aux protéines du marqueur de poids moléculaires (piste de gauche sur chaque gel) sont indiquées en kDa.

**Tableau 21 : Résumé des observations réalisées au cours des purifications et quantité finale des CDPS évaluée par absorbance à 280 nm et par Bradford**

CDPS	M (Da)	Expression <sup>1</sup>	A <sub>600</sub> finale de la culture	Solubilité après la colonne HisTrap <sup>2</sup>	Pureté après la colonne HisTrap <sup>3</sup>	V (ml)	Spectroscopie <sup>4</sup>			Bradford <sup>5</sup>	
							[C] (μM)	[C] (mg/ml)	Qté (mg)	[C] (mg/ml)	Qté (mg)
3	28696	2	4	S+++	++	8	60	1,7	13,6	1	8
7	28500	4	4,4	S+++	+	12	90	2,7	32,4	4,8	57,6
8	27288	3	5,1	S+++	++	8	12	0,3	2,4	0,44	3,52
10	28323	1	4,2	P+	+	10	26	0,7	7	1,6	16
16	28167	3	4,1	S++	++	12	64	1,8	21,6	2	24
17	30271	1	5,8	S++	++	8	8	0,25	2	0,18	1,44
18	35391	2	4,1	S+++	++	12	8	0,3	3,6	0,7	8,4
19	29775	3	4,8	P++	++	12	13	0,4	4,8	0,4	4,8
20	35267	4	5,5	S+++	∅	6	/	/	/	0,03	0,18
21	35032	4	3,5	S+ (j)	-	14	230	8	112	0,8	11,2
22	36407	1	4,2	S++	+	12	66	2,4	28,8	3,5	42
25	30885	3	4,3	S++	+	10	9	0,3	3	0,21	2,1
27	28279	4	4	P+++	+	8	45	1,3	10,4	2,4	19,2
30	36321	4	6,8	P+	+	12	265	9,6	115	6,6	79,2
36	28158	3	2	S+++	++	10	77	2,1	21	2,4	24
39	28366	3	1,4	P+	+	10	76	2,1	21	4,6	46

<sup>1</sup>La production des CDPS est évaluée visuellement et une note de 0 à 4 lui est attribuée, 0 correspondant au cas où la bande attendue n'est pas visible et 4 au cas où une bande extrêmement intense est visible. <sup>2</sup>La solubilité en sortie de colonne d'affinité chargée en nickel (HisTrap) est renseignée. La lettre indique si la protéine a tendance à être soluble (S) ou à précipiter (P) et le nombre de « + » indique l'intensité du comportement. <sup>3</sup> La pureté de la protéine obtenue est évaluée à partir de l'analyse SDS-PAGE (Figure 68) (++ bande unique correspondant à la protéine d'intérêt, + bande majoritaire correspondant à la protéine d'intérêt et quelques bandes de contaminants visibles, - protéine insuffisamment purifiée, ∅ pas de bande visible). <sup>4</sup>Pour chaque CDPS, le coefficient d'extinction molaire à 280 nm a été calculé à l'aide de ProtParam sur le site ExPASy. <sup>5</sup>La gamme d'étalonnage a été réalisée en utilisant de l'albumine de sérum bovin.

Trois CDPS supplémentaires (23, 29, 31) purifiées indépendamment au laboratoire, ont été ajoutées à ces 15 CDPS purifiées en quantité suffisante pour la suite de l'étude (en bleu sur le Tableau 20). Ainsi, 18 CDPS au total, provenant de la sous-famille XYP (12), de la sous-famille NYH (5) et le membre SYQ, ont été purifiées en une étape.

## 2. Analyse *in vitro* des activités des CDPS

Afin de vérifier si les CDPS avaient été purifiées sous forme active, nous avons déterminé leur activité *in vitro*. Les CDPS utilisent des acides aminés chargés sur des ARNt comme substrats, or ces substrats sont

particulièrement instables et s'hydrolysent rapidement. Une solution adéquate pour tester l'activité des CDPS consiste à synthétiser les aminoacyl-ARNt (aa-ARNt) directement *in situ* et en continu, dans le mélange réactionnel utilisé pour mesurer l'activité de la CDPS, en mettant en présence des acides aminés, des ARNt et les aminoacyl-ARNt synthétases (aaRS). Nous avons employé deux méthodes pour synthétiser *in situ* les aa-ARNt. La première méthode, facile à mettre en œuvre, utilise un lysat soluble d'*E. coli* qui contient donc, entre autres, les ARNt et les aaRS<sup>26</sup>. La seconde méthode utilise des d'ARNt et des aaRS purifiés disponibles commercialement. Dans un premier temps, nous avons effectué les tests d'activité des CDPS purifiées en utilisant un lysat soluble. Puis, pour les CDPS inactives dans ce premier test, nous avons utilisé les molécules commerciales dans un test couplé.

### **2.1 Mise au point du test avec un lysat soluble de *E. coli*.**

De tels essais utilisant un lysat soluble de *E. coli* avaient été précédemment décrits par le laboratoire<sup>26</sup>. La concentration en CDPS alors utilisée était particulièrement élevée (35  $\mu\text{M}$ ). Les quantités de CDPS dont nous disposions ne permettaient pas d'atteindre ces concentrations pour toutes les CDPS. La quantité de CDPS utilisée dans le test devant être diminuée, nous avons réalisé des essais avec AlbC pour estimer la concentration nécessaire pour obtenir une activité de synthèse de cyclodipeptides. Par ailleurs, étant donné que nous avons arrêté la purification après la colonne d'affinité, dont l'éluion se fait avec un gradient d'imidazole, les CDPS sont obtenues dans un tampon contenant environ 200 mM d'imidazole. L'objectif des tests était donc également d'évaluer si la présence d'imidazole était délétère à l'activité des CDPS.

Trois concentrations d'AlbC ont été testées (1, 5 et 10  $\mu\text{M}$ ) à chaque fois en présence et en absence d'imidazole. Dans chaque cas, les cyclodipeptides cyclo(Phe-Leu), cyclo(Phe-Phe) mais aussi cyclo(Tyr-Leu), cyclo(Met-Met), cyclo(Tyr-Phe), cyclo(Leu-Met), cyclo(Phe-Met) et cyclo(Leu-Leu), caractéristiques de l'activité d'AlbC ont été identifiés par LC-MS/MS dans le milieu réactionnel. Pour comparer l'effet de la concentration en AlbC et de la présence d'imidazole, la production de cyclo(Phe-Leu) est évaluée par l'aire sous le pic chromatographique correspondant (Tableau 22).

**Tableau 22 : Production de cyclo(Phe-leu) dans les différentes conditions testées**

[AlbC]	[Imidazole] <sup>2</sup>	cyclo(Phe-Leu) <sup>1</sup> (unité arbitraire)
0 $\mu$ M	40 mM	0
0 $\mu$ M	0 mM	0
1 $\mu$ M	10 mM	339
1 $\mu$ M	0 mM	381
5 $\mu$ M	50 mM	546
5 $\mu$ M	0 mM	738
10 $\mu$ M	54 mM	654
10 $\mu$ M	0 mM	904

<sup>1</sup>Aire sous le pic d'un chromatogramme HPLC enregistré à 220 nm.

L'imidazole inhibe partiellement l'activité d'AlbC (entre 11 et 28 % d'inhibition en fonction de la concentration finale d'imidazole dans le milieu réactionnel). La production de cyclo(Phe-Leu) est environ multipliée par un facteur deux lorsque la concentration en enzyme est multipliée par 10.

Dans un premier temps, nous avons utilisé 1  $\mu$ M d'enzyme pour tester l'activité des 16 CDPS purifiées, ce qui permet de minimiser tant la quantité de protéine utilisée que la quantité d'imidazole présent dans le milieu réactionnel. Pour les CDPS, dont l'activité n'est pas détectée à cette concentration, l'expérience est reproduite à 10  $\mu$ M d'enzyme. Enfin pour les CDPS pour lesquelles aucun cyclodipeptide n'a été détecté, le test couplé utilisant des ARNt et des aaRS purifiés (cf. Matériels et Méthodes. IX.1.2 Test couplé avec ARNt et aaRS purifiés) est réalisé en se plaçant à 10  $\mu$ M d'enzyme.

## 2.2 Résultats des tests *in vitro* des activités

Sur les 18 CDPS purifiées, l'activité de 13 CDPS a pu être reproduite *in vitro* en utilisant l'un ou l'autre des tests (Tableau 23). Le profil de cyclodipeptides produits dans les tests avec le lysat correspond au profil obtenu lors de l'analyse des surnageants de culture dans le test *in vivo*. *A contrario*, avec le test couplé utilisant des ARNt et des aaRS purifiés, on observe des différences plus importantes même pour une enzyme bien décrite comme AlbC. En effet, le profil de production de cyclodipeptides d'AlbC montre un défaut dans l'incorporation de Phe. Ce défaut est corrigé lorsqu'une PheRS purifiée au laboratoire est ajoutée au lot commercial des aaRS utilisé. De la même manière, Ala n'est pas incorporé dans les cyclodipeptides produits par la CDPS 36, alors que la CDPS est active puisqu'elle produit d'autres cyclodipeptides. Une explication potentielle serait que l'ARNt<sup>Ala</sup> n'est pas chargé efficacement par l'AlaRS. Ceci permettrait d'expliquer pourquoi les CDPS 29, 31 et 39 apparaissent inactives dans ce test, étant donné que tous les cyclodipeptides qu'elles produisent contiennent de l'alanine. La même conclusion peut être faite pour le chargement de l'ARNt<sup>Cys</sup> par la CysRS. En effet, dans nos conditions, la CDPS 3 ne produit pas de cyclo(L-cystéine), alors

qu'elle est active et que ce composé est son produit principal *in vivo*. Ceci expliquerait que la CDPS 7 ne produise pas, dans le test couplé utilisant des ARNt et aaRS purifiés, son unique produit, la cyclo(L-cystine).

**Tableau 23 : Test *in vitro* de l'activité de synthèse de cyclodipeptides<sup>1</sup>**

CDPS	Test avec lysat soluble		Test couplé utilisant des ARNt et aaRS purifiées <sup>2</sup>
	[CDPS]=1 $\mu\text{M}$ <sup>3</sup>	[CDPS]=10 $\mu\text{M}$ <sup>3</sup>	[CDPS]=10 $\mu\text{M}$ <sup>3</sup>
3	-	-	cCV
7	-	-	-
8	cGV	non testé	non testé
10	-	<b>cPM</b>	non testé
16	cLV, cLA	non testé	non testé
17	CHF	non testé	non testé
18	<b>cAP</b>	non testé	non testé
19	<b>cLL</b> , cML	non testé	non testé
21	-	-	cLM, cLL, cLF
22	<b>cYP</b> (55%), <b>cYV</b> (42%) <b>cFP</b> (3%), cFV, cYL, cYA	-	non testé
23	non testé	<b>cFF</b> (97%), <b>cFL</b> (3%)	non testé
25	<b>cPP</b> , cPL, cPF, cPW	non testé	non testé
27	-	-	-
29	non testé	-	-
30	<b>cAF</b> , cAP, cMP	non testé	non testé
31	non testé	-	-
36	<b>cLE</b> , cLA, cVI, cLL, cPL, cLC	non testé	cLE, cLC, cLV
39	-	-	-
AlbC	<b>cFF</b> (41%), <b>cFL</b> (24%), <b>cFY</b> (20%), <b>cFM</b> (14%), cLL, cLM, cYL	<b>cFL</b> (35%), <b>cFF</b> (32%), <b>cYF</b> (22%), <b>cFM</b> (9%), <b>cLM</b> (<1%), cLL, cYL, cYM	<b>cYY</b> (65%), <b>cYL</b> (20%), <b>cYM</b> (11%), <b>cLM</b> (2%), <b>cYF</b> (<1%), cMM, cLL, cFL, cYV

<sup>1</sup>L'absence de détection de cyclodipeptide est représenté par « - ». Les cyclodipeptides détectés par LC-MS/MS dans le milieu réactionnel sont indiqués en utilisant le code à une lettre. Les cyclodipeptides en gras sont ceux pour lesquels une aire sous le pic du chromatogramme enregistré à 214 nm a pu être mesurée. Les autres cyclodipeptides ont été détectés par leur signal EIC spécifique. Lorsque cela était possible, la proportion que représente chaque cyclodipeptide par rapport à la totalité des cyclodipeptides pour lesquels une aire de pic a été mesurée, a été indiquée entre parenthèses. Les cyclodipeptides sont rangés par ordre décroissant de proportion. <sup>2</sup>Les CDPS 21, 27 et 39 ont été purifiées à nouveau pour ce test car la purification précédente n'avait pas mené à l'obtention d'une protéine suffisamment purifiée (21) ou ne précipitant pas (27 et 30). <sup>3</sup>Lorsque la concentration de la protéine ne permettait pas d'atteindre la concentration voulue, la concentration maximale possible dans le test a été utilisée.

Ces études, bien que très préliminaires, ont servi de point de départ à plusieurs études dans le laboratoire (cf. VII. Discussion, conclusion et perspectives). 18 CDPS actives *in vivo* ont été purifiées et pour 13 d'entre elles une activité *in vitro* a été obtenue. Néanmoins, nous avons rencontré plusieurs difficultés au cours de ces études. Plusieurs des CDPS purifiées avaient ainsi une tendance importante à précipiter et probablement à s'agréger. De plus, étant donné le substrat utilisé par les CDPS, reproduire une activité *in vitro* est loin d'être évident. L'instabilité du substrat nécessite, du moins dans le cadre d'une étude de plusieurs CDPS en parallèle, de former le substrat *in situ* et donc de disposer de aaRS actives. Or, le lot de aaRS commercial semble contenir plusieurs enzymes inactives. D'après les informations données par le fournisseur, il semblerait que le lot de aaRS soit uniquement testé vis-à-vis de son activité à charger l'acide aminé Arg sur l'ARNt correspondant. Aucune information n'est donnée sur les autres activités. Si certaines aaRS sont inactives dans les lots commerciaux, il est probable qu'elles soient particulièrement fragiles et donc également inactives dans le lysat soluble que nous avons utilisé. D'autres possibilités sont envisageables telles que l'utilisation de aaRS purifiées au laboratoire ou encore l'utilisation de préparations commerciales destinées à la transcription et la traduction *in vitro*.

## VII. Discussion, conclusion et perspectives

L'étude *in vivo* de 61 CDPS potentielles et l'étude préliminaire *in vitro* sur 18 d'entre elles ont mis en évidence 51 nouvelles CDPS actives, ce qui porte à 62 le nombre de CDPS connues à ce jour. En réalisant une étude bio-informatique s'appuyant sur l'arbre phylogénétique et la détermination de signatures fonctionnelles pour les CDPS, nous avons montré que les CDPS, caractérisées biochimiquement ou non, peuvent être classées dans au moins deux sous-familles phylogénétiques distinctes. Les sous-familles NYH et XYP contiennent la quasi-totalité des membres des CDPS, la sous-famille NYH étant celle qui contient le plus de membres. La CDPS 17 et un autre membre identifié lors de la seconde interrogation des bases de données en septembre 2014 (cf. annexe électronique 7), qui appartiennent tous deux au phylum bacteroidetes, ne se classent pas dans les sous-familles NYH et XYP et ils pourraient constituer une troisième sous-famille SYQ. La manière dont les membres se répartissent dans les sous-familles ne reflète pas nécessairement ce qui serait trouvé dans la nature, car le séquençage des génomes est fortement biaisé en faveur des organismes pathogènes pour les humains, les animaux ou les plantes. Seules quelques CDPS actives ne remplissent pas tous les critères d'appartenance à une sous-famille, notamment chez la famille NYH qui possède quelques membres n'ayant pas la paire canonique « N40, H203 ». Avec le séquençage de nouveaux génomes codant des CDPS, les caractéristiques définies dans cette étude sont amenées à être modifiées ou affinées.

Concernant le fonctionnement des CDPS, les données biochimiques et cristallographiques obtenues avant ma thèse proviennent d'études réalisées sur les membres de la famille NYH. Ces éléments ont été, en

l'absence de données sur les autres sous-familles, extrapolées à toutes les CDPS. Or, la sous-famille XYP, en particulier, ne possède pas certains des éléments précédemment décrits comme conservés, par exemple certains résidus catalytiques ou le patch basique impliqué dans la liaison à l'ARNt. Plusieurs questions se posent sur le fonctionnement des membres de cette nouvelle sous-famille : comment est assuré le rôle dévolu à la paire « N40, H203 » dans la famille XYP ? Comment se fait l'interaction avec le substrat aa-ARNt avec un patch basique réduit voire inexistant ? Pour répondre à ces questions, le laboratoire s'est lancé dans l'étude cristallographique d'un membre de la famille XYP, la CDPS 39. Plusieurs structures ont d'ores et déjà été obtenues et sont en cours d'analyse. Elles devraient notamment permettre de savoir si le résidu P203 est responsable d'une variation structurale qui induit un positionnement différent des résidus catalytiques. Elles devraient également permettre de vérifier si les prédictions des résidus composant les poches P1 et P2 sont exactes ou à affiner. Ceci devrait améliorer notre capacité à prédire l'activité de synthèse des cyclodipeptides.

Les 62 CDPS caractérisées synthétisent 54 cyclodipeptides différents qui contiennent 17 des 20 acides aminés protéinogéniques. Les CDPS ne sont donc pas limitées à la synthèse de cyclodipeptides hydrophobes. La diversité des cyclodipeptides produits par les CDPS peut encore être augmentée. En effet, nous prédisons que 55% des 200 CDPS potentielles synthétisent du cyclo(L-cystine), du cyclo(Leu-Leu), du cyclo(Trp-Trp) ou du cyclo(Ala-Glu). 30 % pourraient être des membres de groupe de spécificité déjà identifiés mais n'ayant pas assez de membres. Il reste donc 15 % de CDPS dont l'activité de synthèse de cyclodipeptides ne peut absolument pas être prédite. Ces CDPS pourraient synthétiser de nouveaux cyclodipeptides jamais identifiés comme produits de CDPS. La caractérisation de nouvelles CDPS pourrait donc enrichir, d'une part, l'outil de prédiction mis en place dans cette étude et, d'autre part, la diversité des cyclodipeptides produits. Dans cette optique, un nouveau lot de CDPS potentielles a été sélectionné et est en cours de caractérisation *in vivo*. Par ailleurs, la diversité des cyclodipeptides produits peut être augmentée en incorporant des acides aminés non naturels. La preuve de concept de cette approche a été réalisée, au laboratoire, en utilisant des CDPS pour incorporer des analogues de proline dans des cyclodipeptides.

La question de la pertinence physiologique de la caractérisation de l'activité des CDPS *in vivo* chez *E. coli* peut se poser. En effet, étant donné que la CDPS est surexprimée et utilise les aa-ARNt de *E. coli*, les cyclodipeptides retrouvés dans les surnageants de culture sont-ils représentatifs des cyclodipeptides produits par la CDPS dans son organisme d'origine ? Les études menées sur AlbC pour identifier les déterminants de la spécificité permettent d'avancer un certain nombre d'arguments en faveur d'une activité chez *E. coli* représentative de l'activité dans l'organisme d'origine. Pour rappel, l'entité aminoacyle est essentielle pour la reconnaissance du premier substrat. L'entité aminoacyle et la paire de bases N<sup>1</sup>-N<sup>72</sup> participent toutes deux à la reconnaissance du second substrat<sup>365</sup>. Or comme la paire N<sup>1</sup>-N<sup>72</sup> est particulièrement conservée dans le monde procaryote (Annexe électronique 9) et que les aa-ARNt sont présents en quantité suffisante dans toutes les cellules vivantes, il est probable que les produits majoritaires des CDPS dans notre étude soient

identiques à ceux produits dans l'organisme d'origine. Par ailleurs, l'activité de quelques CDPS a été caractérisée chez l'organisme d'origine et elle correspond à celle observée lorsque la CDPS est surexprimée chez *E. coli* : cyclo(Leu-Leu) pour YvmC de différents *Bacillus*<sup>26,374</sup>, cyclo(Phe-Leu) pour AlbC de *Streptomyces noursei*<sup>25,35</sup>, et cyclo(Phe-Tyr) pour Ndas\_1148 de *Nocardiopsis dassonvillei*<sup>348</sup>. De plus, les produits majoritaires obtenus *in vivo* chez *E. coli* correspondent également à ceux détectés *in vitro*, les aa-ARNt étant disponibles en excès par rapport à la CDPS<sup>26,348,365</sup> (cf. VI.2 Analyse *in vitro* des activités des CDPS). Néanmoins, nos conditions expérimentales, notamment le niveau d'expression de la CDPS, ne reproduisent probablement pas les conditions rencontrées dans l'organisme d'origine, et il est possible que certains cyclodipeptides détectés en faible quantité ne soient pas synthétisés dans l'organisme d'origine.

Au-delà des données ayant trait à la compréhension du fonctionnement des CDPS, mon travail a permis l'identification de l'activité de synthèse de nombreuses CDPS impliquées dans des voies de biosynthèse menant à des dicétopipérazines plus complexes. Seulement trois voies de biosynthèse utilisant des CDPS étaient caractérisées au début de mon doctorat et impliquaient des cyclodipeptides hydrophobes<sup>24,25,29,367</sup>. Cette étude pose les fondements pour le déchiffrement de nombreuses voies dépendant des CDPS, afin de découvrir toute la diversité encodée par ces voies et d'alimenter les études sur le rôle physiologique de ces molécules chez les hôtes d'origine.

CHAPITRE III:  
MISE EN ÉVIDENCE DE NOUVELLES  
ACTIVITÉS DE MODIFICATION DANS  
LES VOIES DEPENDANT DES CDPS

L'objectif de mon doctorat est d'explorer la diversité moléculaire pouvant être synthétisée par les voies de biosynthèse dépendant des CDPS. Ces voies sont constituées d'une CDPS et éventuellement d'une ou plusieurs enzymes appelées enzymes de modification qui introduisent des modifications chimiques sur le cyclodipeptide formé (cf. Etude bibliographique. V.2 Voies de biosynthèse de dicétopipérazines dépendant des CDPS). L'approche retenue dans mon travail a consisté à étudier, dans un premier temps, un panel varié de CDPS avant de s'intéresser aux étapes de modifications des cyclodipeptides.

Dans le chapitre précédent, nous avons identifié 51 nouveaux membres de la famille des CDPS, redéfini les caractéristiques générales de cette famille et posé les bases d'un nouvel outil de prédiction des activités CDPS. Ce travail a montré notamment que les produits cyclodipeptidiques formés par les CDPS sont extrêmement variés et que le potentiel de synthèse de ces enzymes n'avait pas été révélé par les études précédentes. L'objectif de ce nouveau chapitre est de révéler la diversité d'enzymes de modification qui peuvent être associées à des CDPS et de mettre en évidence de nouvelles activités menant à la synthèse de nouvelles 2,5-dicétopipérazines.

Au début de mon doctorat, seulement trois voies à CDPS avaient été biochimiquement caractérisées. Il s'agit des voies de biosynthèse de l'albonoursine<sup>24,25</sup>, de la mycocyclosine<sup>29</sup> et de l'acide pulcherrimique<sup>367</sup> (cf. Etude bibliographique.V.2 Voies de biosynthèse de dicétopipérazines dépendant des CDPS). Dans la voie de biosynthèse de l'albonoursine découverte chez *S. noursei*, le cyclo(Phe-Leu) synthétisé *via* la CDPS est modifié par une cyclodipeptide oxydase (CDO) qui catalyse une  $\alpha,\beta$ -déshydrogénation des deux chaînes latérales. Dans chacune des deux autres voies, découvertes respectivement chez *M. tuberculosis* et *B. subtilis*, un cytochrome P450 spécifique à chaque voie oxyde le cyclodipeptide. Chez *M. tuberculosis*, il catalyse la formation d'une liaison C-C entre les cycles tyrosyles du cyclo(Tyr-Tyr) tandis que chez *B. subtilis*, il catalyse la N-oxydation du noyau dicétopipérazine du cyclo(Leu-Leu). Les gènes codant les enzymes constituant ces voies de biosynthèse sont regroupés au niveau génomique, ce qui est souvent le cas chez les procaryotes. Des analyses de l'environnement génomique de CDPS potentielles ont déjà permis d'indiquer que certains gènes de CDPS sont regroupés avec des gènes codant potentiellement des enzymes susceptibles de transformer les cyclodipeptides, notamment des cytochromes P450 et des dioxygénases<sup>28,345</sup>. Ces travaux ont constitué le point de départ de l'étude que je vais présenter dans ce chapitre. Cette étude a consisté, dans un premier temps, à réaliser une analyse bioinformatique des environnements génomiques des CDPS procaryotes étudiées au chapitre II. Cette analyse a permis d'effectuer une sélection d'enzymes de modification à étudier selon le système moyen-débit développé au chapitre I. Puis, je présenterai le travail qui a consisté à caractériser biochimiquement plus spécifiquement une voie dépendant de CDPS, ce qui a mené à l'élucidation structurale complète de nouvelles dicétopipérazines.

## **I. Environnement génomique des CDPS : diversité et sélection d'enzymes de modification à étudier**

L'objectif de cette partie est, dans un premier temps, de donner un aperçu rapide de l'environnement génomique rencontré pour les gènes de CDPS tout en focalisant sur les CDPS associées potentiellement à des enzymes de modifications. Dans un second temps, je décrirai les enzymes de modification codées dans l'environnement génomique de gènes de CDPS que nous avons sélectionnées pour étudier leurs activités sur des cyclodipeptides.

### **1. Analyse des environnements génomiques des CDPS**

Le positionnement de chaque CDPS est repéré sur le génome disponible dans les bases de données. La fonction probable des protéines codées par les gènes présents en amont et en aval du gène de CDPS a été identifiée en utilisant le serveur du NCBI et l'outil MicroScope d'annotation des génomes, développé par le LabGem (GénoScope ; CEA, Institut de Génétique ; CNRS, Génétique métabolique, Evry). Cette identification s'appuie sur l'annotation préexistante dans les bases de données, l'identification de domaines conservés (Conserved Domain Database du NCBI) et l'annotation fonctionnelle de protéines homologues repérées par BLAST (NCBI) et par l'outil de visualisation de conservation de l'environnement ou synthétiques (MicroScope) (cf. Matériels et méthodes. I.4 Analyse des environnements génomiques des CDPS). Pour chaque CDPS procaryote active ou inactive dans nos conditions expérimentales, un schéma représentant l'organisation génomique autour du gène de CDPS a été réalisé (Figure 69).

# Famille XYP

Inactive

CDPS 46



Cyclo(Ala-Pro)

CDPS 18



Cyclo(Leu-Ile)

CDPS 19



Inactive

CDPS 35



Cyclo(Gly-Gln)

CDPS 27



Cyclo(Ala-Gly)

CDPS 31



Cyclo(Pro-Pro)

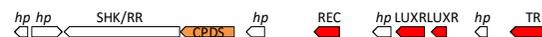
CDPS 26



CDPS 25

Cyclo(Phe-Asn)

CDPS 20



Cyclo(Tyr-Xaa)

CDPS 22



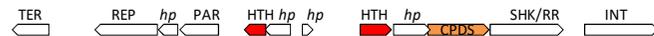
**Cyclo(Phe-Phe)**

CDPS 24



**Cyclo(Leu-Leu)**

CDPS 21



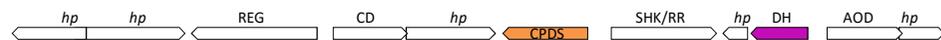
**Cyclo(Phe-Phe)**

CDPS 23

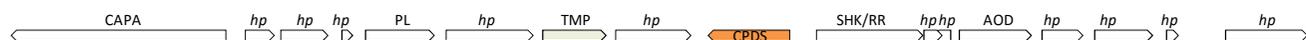


**Cyclo(Ala-Ala)**

CDPS 29



CDPS 28



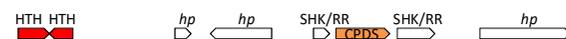
**Cyclo(Met-Xaa)**

CDPS 30



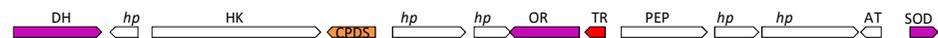
**Inactive**

CDPS 34



**Cyclo(Gly-Gly)**

CDPS 33



**Cyclo(Xaa-Glu)**

CDPS 36



CDPS 43



CDPS 37



CDPS 40



CDPS 44



CDPS 39



CDPS 38



# Famille SYQ

## Cyclo(His-Phe)

CDPS 17



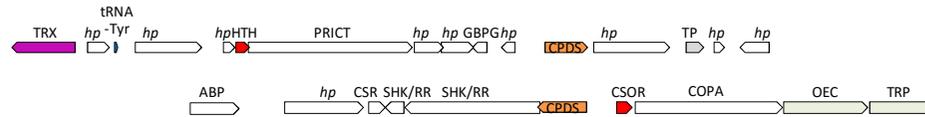
# Famille NYH

## Cyclo(Gly-Val)

CDPS 8

## Cyclo(L-cystine)

CDPS 9



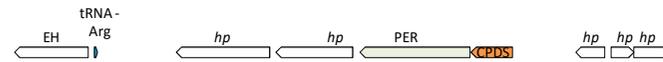
CDPS 3



CDPS 7



CDPS 2



CDPS 1



CDPS 4



CDPS 6

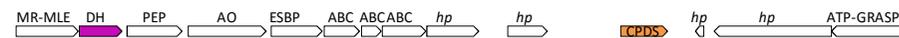


CDPS 5



## Inactive

CDPS 12



## Cyclo(Pro-Met)

CDPS 10



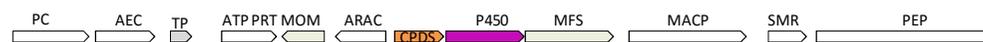
## Inactive

CDPS 11



### Cyclo(Leu-Leu)

CDPS 13



### Cyclo(Trp-Trp)

CDPS 42



CDPS 14



CDPS 48



CDPS 47



### Cyclo(Leu-Val)

CDPS 16



### Cyclo(Phe-Xaa)

CDPS 41



### Inactive

CDPS 15



### Cyclo(Tyr-Tyr)

CDPS 49



Figure 69 : Environnements génomiques des CDPS d'origine procaryote testées pour leur activité de synthèse de cyclodipeptides dans le chapitre II

Les CDPS sont désignées par le numéro qui leur a été attribué et sont classées selon l'ordre de l'arbre phylogénétique de la Figure 56 et donc par sous-famille. Le cyclodipeptide ou le type de cyclodipeptide majoritairement produit par chacune d'entre elles est indiqué au début de chaque ligne. La figure est réalisée à partir des données collectées sur le site NCBI et la plateforme MicroScope. Les gènes sont colorés selon les caractéristiques des protéines qu'ils encodent : Orange, CDPS ; violet, enzymes impliquées dans des réactions d'oxydoréduction ; vert, protéine de la membrane externe ; vert pâle, protéine de la membrane interne ; rouge, régulateur de la transcription ; gris, protéine impliquée dans le métabolisme des sucres. La fonction probable déterminée pour chaque protéine encodée par les gènes représentés est indiquée au dessus de la flèche. aa-tRNA RS : aminoacyl-tRNA synthetase ; ABC : ABC transporter system component ; ABH :  $\alpha/\beta$  hydrolase; ACDH : acyl-CoA dehydrogenase ; ACP : acyl carrier protein ; ACPS : Holo-[acyl-carrier-protein] synthase ; ACS3: acyl-CoA ligase; ABP : ATP-binding protein; ACT : acetyl-CoA acyltransferase ; AE : acyl esterase ; AEC : auxin efflux carrier family protein ; AFD : Adenylate forming domain ; AKR : Aldo-keto reductases ; ALD : Aldolase ; ALKSP : alkaline serine protease ; ALN : allantoinase ; AM : amidase ; AO : amine oxidase ; AOD : acetylornithine deacetylase ; ARSC : arsenate reductase family; ASP : Alkaline shock protein ; AT : acetyl transferases ; AT : Acyl transferase ; ATP GRASP : ATP-dependent carboxylate-amine ligase activity. ; ATP PRT : ATP phosphoribosyltransferase ; BAC : bacitracin synthetase ; CAPA : calcium-transporting ATPase ; CD : carbonic anhydrase ; CE : carboxylesterase ; CH : Chitinase ; CHDC : décarboxylase;

CLPS: ATP-dependent Clp protease adaptor protein ClpS ; COPA : Copper exporting ATPase ; CPS : Carbamoyl phosphate synthetase ; CS : Cysteine synthase; CSOR : copper-sensitive operon repressor CsoR ; CSR : carbon storage regulator ; CYCOX : Cytochrome C oxidase ; DBP : DNA-binding protein; DCDPB : dipeptide-binding protein; DCTP : Divalent-cation tolerance protein ; DG : DNA gyrase ; DH : deshydrogenase; DHDPS : Dihydrodipicolinate synthetase ; DLD : dihydrolipoamide deshydrogenase; DMT : DNA methylase ; DO : ; DPEP : dipeptidase ; ECH : enoyl-CoA hydratase ; EH : epoxide hydrolase ; ENDO : endoribonuclease L-PSP; ESBP : extracellular solute-binding protein family 3 ; EXBD :biopolymer transport protein ExbD/ToIR ; EXO : exonuclease; FAD-OR : fad dependent oxidoreductase; FDX :Ferredoxine ; FIC : filamentation induced by cAMP) ; FK : fructokinase ; FLIB : Flagellin N-methylase ; FMO : flavin-utilizing monooxygenase; FNBP :fibronectin/fibrinogen binding protein ; FOLC : bifunctional folylpolyglutamate synthase/ dihydrofolate synthase; GBPG : glucan biosynthesis protein G ; GIY-YIG : GIY-YIG nuclease domain superfamily ; GTRA : GtrA-like protein ; ; HAD : Haloacid dehalogenase-like hydrolases ; HAD : haloacid deshydrogenase ; HAL : histidine ammonia-lyase ; HK : histidine kinase ; HL : hydrolase; HTH : helix-turn-helix family protein ; ICLR : IclR family transcriptional regulator ; ICLR : IclR-type transcriptional regulator ; IMMT : Inner membrane metabolite transport protein ; INT : integrase; LACI : LacI family transcriptional regulator ; LACT : lactamase; LIP : lipoprotein ; LIPB :Octanoate-[acyl-carrier-protein]-protein-N-octanoyltransferase ; LS : Lipoate synthase; LSLUXR : LuxR family transcriptional regulato ; LYSr : LysR family transcriptional regulator ; MACP Methyl-accepting chemotaxis protein ; MAO : monoamine oxidase ; MARR : MarR family transcriptional regulator; MAT : S-adenosyl-methionine synthetase ; MERR : merR family transcriptional regulator ; MFS : major facilitator super family ; MO : monooxygenase; MOM : major outer membrane protein ; MOTA : MotA/ToIQ/ExbB proton channel ; MPH : metal-dependent phosphohydrolase ; MR-MLE: mandelate racemase/muconate lactonizing-like protein ; MT : methyl transferase ; NDH : NADH2 dehydrogenase ; NHE : sodium/hydrogen antiporter ; NIFU/ ISCU : Iron-sulfur cluster scaffold ; NR : nitroreductase ; NUDIX : Nudix hydrolase; OEC : Outer membrane efflux protein ; OG : 2-oxoglurate and fe(II)-dependant oxygenase ; O-MT : O-méthyl transférse ; OR : oxydoreductase ; PA : Penicillin acylase ; PAD : peptidylarginine deiminase ; PANK : pantothenate kinase; PAR: Chromosome partitioning protein ; PC : pyruvate carboxyltransferase ; PEP : peptidase; PEPC : phosphoenolpyruvate carboxylase ; PER : Permease; PHA : phasin-family protein; PHAS : polyhydroxyalkanoate synthase ; PILT : pilus retraction protein ; PL : phospholipase; PMI : phosphomannose isomerase type I ; PMSR : Peptide methionine sulfoxide reductase; PPAT : Phosphopantetheine adenylyltransferase ; PRICT : primase C 2 (PriCT-2) family protein ; PRX : Peroxidase ; PSP : Phosphoserine phosphatase; PTC : Plastocyanin ; PYR : pyrophosphatase ; RDH : replicative DNA helicase ; REC : Signal receiver domain ; REG : Regulator ; REP : replication protein/replicase protein ; RF : release factors ; RHOD : rhodanese-like protein ; RPSF : RNA polymerase sigma factor; RR : response regulator; RTM : reverse transcriptase /maturaseSBP : siderophore biosynthesis protein ; SC : Solute carrier ; SECS : L-seryl-tRNA selenium transferase ; SH : sucrose-6-phosphate hydrolase ; SHK/RR :sensor box histidine kinase/response regulator; SMR : Smr protein/MutS2 ; SOD :superoxide dismutase, Cu, Zn ; SP: sensor protein ; SPIE : sugar phosphate isomerase/epimerase ; ST : sulfurtransferase ; TAG : DNA-3-methyladenine glycosylase I (3-methyladenine-DNA glycosylase I, constitutive) (TAG I) (DNA-3-methyladenine glycosidase I); TCS : two-component sensor ; TER : terminase; TETR: TetR family transcriptional regulator ; TGM : transglutaminase-related protein ; TMP : transmembrane protein ; TONB : TonB-dependent receptor ; TPR: tetratricopeptide ; TR : Transcriptional regulator ; TRAP : Tripartite ATP-independent Periplasmic Transporter; TRP : Transporter; TRX : thioredoxin ; TRXFD : Thioredoxin (TRX)-like [2Fe-2S] Ferredoxin (Fd) family; TRXR : thioredoxin reductase ; TRYP : Trypsin ; TSP : transposase; UH : urocanate hydratase; XRE : XRE family transcriptional regulator.

Avant de s'intéresser aux voies de biosynthèse potentielles, une analyse rapide des gènes environnants montre que les gènes de CDPS sont souvent à proximité de gènes codant des protéines impliquées dans le transport de molécules à travers la membrane externe (TONB, pour Récepteur dépendant de TonB, en vert) ou à travers la membrane interne (MFS pour « major facilitator super family » ou perméase en vert clair). Une proximité avec les gènes codant des protéines de régulation de la transcription (en rouge) est également à noter. Les gènes de CDPS 27, 20, 22, 24, 21, 29, 28, 30, 34 et 9 sont proches de gènes annotés comme des systèmes de régulation à deux composants (autokinase et régulateur de réponse). Ceci est probablement à mettre en regard du rôle de communication qui est supposé pour les dicétopipérazines (cf. Etude bibliographique. II.1.1 Dicétopipérazines, des molécules de communication). Une autre association intéressante est celle des gènes de CDPS synthétisant du cyclo(L-cystine) avec des gènes codant une protéine de type thiorédoxine : ces protéines contiennent une ou des séquences de type Cys-X-X-Cys, capables d'exister sous forme réduite ou oxydée, et on imagine que l'association avec la CDPS synthétisant le cyclo(L-cystine) pourrait constituer un système senseur d'état d'oxydoréduction.

En ce qui concerne l'appartenance des CDPS à une voie de biosynthèse plus complexe, plusieurs configurations sont relevées. Certains gènes de CDPS actives semblent clairement impliqués dans un groupement de gènes de type de voies de biosynthèse. Le gène de CDPS est alors suivi ou précédé par des gènes codant des protéines identifiées comme appartenant à une famille d'enzymes impliquées dans la biosynthèse de métabolites secondaires. Ces enzymes sont alors codées dans le même sens de lecture que le gène de CDPS. C'est le cas pour les CDPS produisant du cyclo(Xaa-Glu) (CDPS 36, 43, 37, 40, 44 et 39), du cyclo(Leu-Ile) (CDPS 19), du cyclo(Leu-Leu) (CDPS 13), du cyclo(Leu-Val) (CDPS 16), du cyclo(Trp-Trp) (CDPS 42 et 47) et du cyclo(Tyr-Tyr) (CDPS 49). Dans d'autres cas, des gènes d'enzyme de modification potentielle sont retrouvés dans le même sens de lecture mais séparés du gène de CDPS par un autre gène (CDPS 10, 38, 44 et 41). Dans les cas restants, l'appartenance à une voie de biosynthèse est difficile à déterminer, soit car la fonction des gènes juxtaposant et dans le même sens que les gènes de CDPS n'est pas identifiée ou ne semble pas pouvoir introduire de modification sur un cyclodipeptide (par exemple pour les CDPS 48, 14, 8, 9, 3, 7, 2, 1, 4, 6, 5), soit car le gène de CDPS semble isolé et sans relation avec les autres gènes environnants (par exemple pour les CDPS 27, 26, 25, 23). Notons que sur les dix CDPS pour lesquelles nous n'avons pas trouvé d'activité, cinq d'entre elles ne sont associées à aucune protéine particulière au niveau génétique (CDPS 12, 15, 32, 34 et 35).

L'étude présentée ci-après ne concerne que les environnements contenant des voies de biosynthèse probables.

La CDPS 10 semble être associée à une enzyme de type oxydoréductase. D'après les informations du NCBI, l'activité de cette enzyme dépendrait du NADPH. La CDPS 13 est associée au niveau génomique à un cytochrome P450, homologue à 49 % au cytochrome CYP134A1 (*cypX*) de *B. subtilis* responsable de

l'oxydation du cyclo(Leu-Leu) en acide pulcherriminique<sup>367</sup> (cf. Etude bibliographique. V.2.2 Voies de biosynthèse de l'acide pulcherriminique).

De part et d'autre du gène de la CDPS 16, trois gènes codant pour des déshydrogénases sont identifiés. Le gène en amont (dans le sens de lecture de la CDPS) coderait pour une protéine de type 2-hydroxymuconique semialdéhyde déshydrogénase, dépendante du NADPH. Le gène en aval coderait pour une enzyme de type 3-hydroxyisobutyrate déshydrogénase. Les résultats du BLAST effectué sur cette séquence laissent à penser que cette séquence est partielle et qu'une centaine de résidus manque à l'extrémité N-terminale. Enfin, l'oxydoréductase la plus en aval serait une protoporphyrinogène oxydase nécessitant la présence de F420 comme coenzyme. La coenzyme F420 intervient dans les réactions d'oxydoréduction comme donneur d'hydrure. Elle est retrouvée principalement chez les archées et certaines bactéries<sup>402,403</sup>.

Le gène de la CDPS 19 appartient à un groupe de huit gènes, transcrits dans le même sens et proches les uns des autres, rappelant une structure de type opéron. On trouve en amont du gène de CDPS, un gène de phosphosérine phosphatase (PSP) et en aval du gène de CDPS, cinq gènes codant des dioxygénases dépendant du Fe(II) et du 2-oxoglutarate (OG) avec un gène de cytochrome P450 (P450) intercalé. La conservation de cette association chez d'autres souches de *P. aeruginosa* a été analysée grâce à l'outil Microscope (Figure 70). Il s'avère que le gène de CDPS, les cinq gènes de dioxygénases et le gène de P450 sont soit tous absents, soit tous conservés dans cette région. Cette observation fait donc apparaître une relation de conservation de ces gènes au sein d'une même espèce bactérienne et permet par ailleurs d'exclure le gène de PSP de la voie de biosynthèse potentielle.

Dans les environnements génomiques des CDPS produisant du cyclo(Ala-Glu), une organisation récurrente est observée. Elle consiste en un gène de OG suivi par le gène de CDPS puis le gène d'un symporteur de type MFS. Les CDPS correspondantes (37, 38, 39, 40, 43) présentent un fort pourcentage d'identité de l'ordre de 40-45% avec les CDPS synthétisant du cyclo(Leu-Glu) (36 et 44) et pour lesquelles la même organisation chromosomique OG-CDPS-MFS est observée (Figure 71). De plus, l'enchaînement des trois gènes constitue une synthénie dans les sept génomes étudiés. Plus précisément, les OG de ces sept organismes partagent entre 37% et 67% d'identité de séquences, les MFS entre 37 % et 75 % et les CDPS entre 40 et 58 %. (Tableau 24). Il est à remarquer que l'environnement génomique autour de cette synthénie est très variable d'un organisme à l'autre. Alors que le triptyque OG-CDPS-MFS semble isolé chez *S. albulus* (CDPS 40), il est associé avec des protéines variées dans les autres organismes : régulateur transcriptionnel chez *Vibrio brasiliensis* (CDPS 36), transporteurs de type dépendant de TonB et de type ABC chez *Pseudomonas* sp. M47T1 (CDPS 43), deux transporteurs de type dépendant de TonB chez *Burkholderia lata* (CDPS 37), flavine monooxygénase (FMO) chez *Pseudomonas fluorescens* (CDPS 44), transporteurs de type ABC et régulateurs transcriptionnels chez *Nocardia brasiliensis* (CDPS 39), et FMO et transporteurs de type ABC chez *Gordonia effusa* (CDPS 38). Ces environnements, même s'ils peuvent posséder les mêmes annotations, ne présentent pas d'identité de séquence forte entre eux. Dans le cadre de mon doctorat, nous avons été

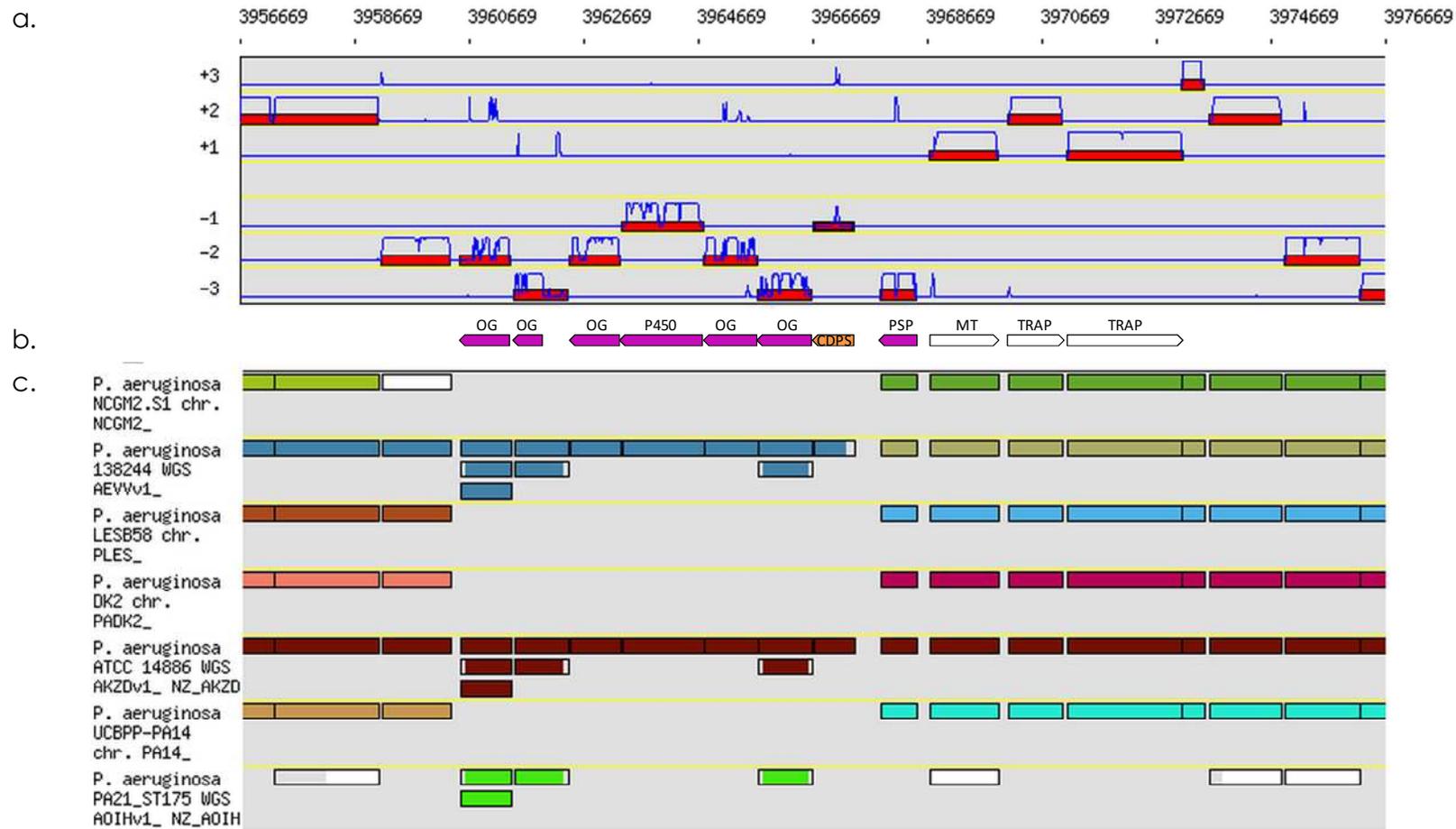


Figure 70 : Environnement génomique de la CDPS 19 et synthèse

a. Représentation cartographique du génome de *P. aeruginosa* PA21\_ST175 de 10 kbp autour du gène de la CDPS 19 (encadré de noir). b. Schéma de l'organisation génomique autour de la CDPS 19 issu de la Figure 69. c. Carte de synthèse obtenue avec les autres génomes bactériens disponibles dans la base de données de MicroScope (PkgDB).

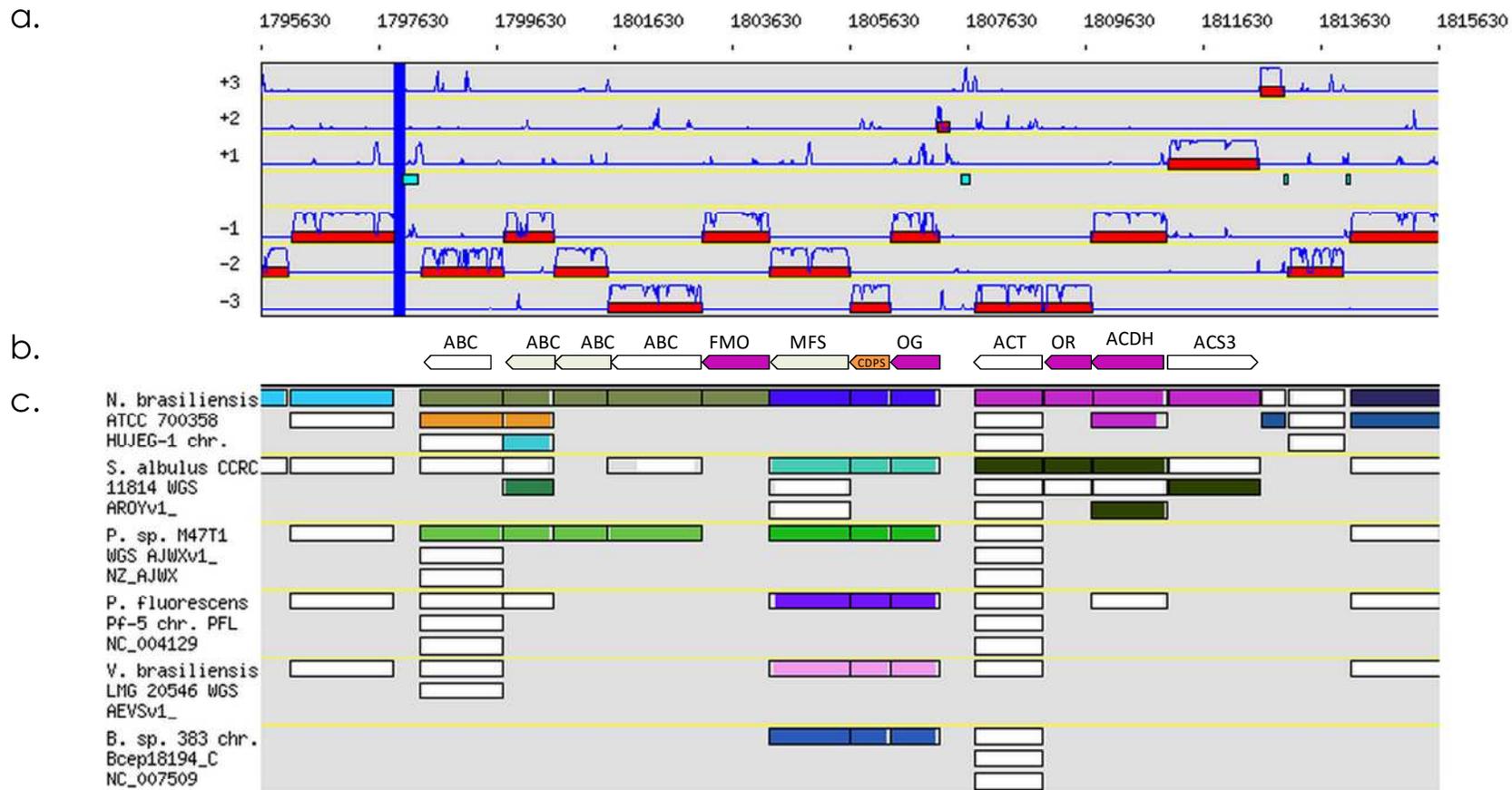


Figure 71 : Synthénie dans les environnements génomiques des CDPS produisant du cyclo(Xaa-Glu)

a. Représentation cartographique du génome de *G. effusa* de 10 kbp autour du gène de la CDPS 38 (encadré de noir). b. Schéma de l'organisation génomique autour de la CDPS 38 issu de la Figure 69. c. Carte de synthénie obtenue avec les autres génomes bactériens contenant les CDPS produisant du cyclo(Xaa-Glu).

particulièrement intéressés par la présence de gènes codant des FMO, protéines susceptibles de catalyser une transformation sur des petites molécules comme les dicétopipérazines. Ces FMO sont associées avec les CDPS 38 et 44 qui synthétisent respectivement du cyclo(Ala-Glu) et du cyclo(Leu-Glu). Elles ne présentent pas de similitudes de séquences entre elles et ne s'ajoutent donc pas à la synthèse formée par le triptyque OG-CDPS-MFS (Figure 71).

**Tableau 24 : Pourcentage d'identité des séquences entre les séquences des protéines conservées chez *G. effusa*, *N. brasiliensis*, *S. albulus*, *P. sp. M47T1*, *P. fluorescens*, *V. brasiliensis* et *B. lata* (sp. 383).**

<i>G. effusa</i>	<i>N. brasiliensis</i>	<i>S. albulus</i>	<i>P. sp. M47T1</i>	<i>P. fluorescens</i>	<i>V. brasiliensis</i>	<i>B. lata</i> (sp. 383)		
100	66,14	44,71	40,24	38,76	36,43	40,94	OG	<i>G. effusa</i>
100	58,04	49,09	49,32	44,64	42,53	42,38	CDPS	
100	53,74	47,71	40,04	41,87	37,32	38,44	MFS	
	100	46,72	45,23	40,94	34,24	43,08	OG	<i>N. brasiliensis</i>
	100	48,23	47,37	46,54	42,51	43,2	CDPS	
	100	47,61	43,35	44,18	34,77	34,08	MFS	
		100	47,79	53,75	41,2	47,79	OG	<i>S. albulus</i>
		100	52,15	45,62	44,69	46,12	CDPS	
		100	52,94	53,06	37,47	40,49	MFS	
			100	50	39,15	60	OG	<i>P. sp. M47T1</i>
			100	42,17	43,78	43,54	CDPS	
			100	74,88	45,07	41,96	MFS	
				100	40,31	49	OG	<i>P. fluorescens</i>
				100	40,83	41,15	CDPS	
				100	45,97	42,96	MFS	
					100	37,74	OG	<i>V. brasiliensis</i>
					100	42,79	CDPS	
					100	37,41	MFS	
						100	OG	<i>B. lata</i> (sp. 383)
						100	CDPS	
						100	MFS	

Le gène de la CDPS 41 est associé avec un gène codant probablement une nitroréductase. Les gènes des CDPS 42 et 49 sont, pour leur part, suivis par un gène de P450 et celui de la CDPS 47 est associé à un gène codant une méthyltransférase.

## 2. Sélection des voies de biosynthèse à étudier

Nous avons sélectionné 18 enzymes de modification appartenant aux voies de biosynthèse potentielles identifiées au cours de l'analyse bioinformatique, pour étudier *in vivo* leur activité de modification de cyclodipeptide (Tableau 25). Le système à deux plasmides permettant la coproduction d'une CDPS et d'une enzyme de modification dans une même souche d'*E. coli* sera utilisée (cf. Chapitre I. I.1 Choix et mise en place du système de clonage) et permettra de détecter si l'enzyme de modification produite dans cette souche a une activité sur le cyclodipeptide produit par la CDPS à laquelle elle est génétiquement associée dans l'organisme d'origine. En premier lieu, comme pour la sélection des CDPS, les enzymes dont les séquences sont partielles et non corrigibles ne sont pas retenues (déshydrogénase en aval de la CDPS 16). Les enzymes annotées comme nécessitant des coenzymes non présentes chez *E. coli* ne sont également pas sélectionnées (oxydoréductase dépendant du F420, cytochrome P450 nécessitant généralement des chaînes de transfert d'électrons spécifiques). Parmi les enzymes restantes, la priorité a été donnée aux enzymes conservées chez plusieurs souches ou organismes (voie associée à la CDPS 19 produisant du cyclo(Ile-Leu) et voies associées aux CDPS produisant du cyclo(Xaa-Glu)). Les FMO présentes chez *S. albulus* et *G. effusa* ont été sélectionnées malgré leur manque de conservation, pour examiner si elles constitueraient, chez ces organismes uniquement, une voie différente. Les enzymes associées à des CDPS produisant du cyclo(Phe-Xaa) comme la CDPS 41 n'ont également pas été choisies car des modifications de ces cyclodipeptides ont déjà été décrites<sup>24,25</sup>. Le cytochrome P450 de *P. aeruginosa* associée à la CDPS 19 a été maintenu dans la sélection, car il fait partie d'une voie plus complexe. Nous l'avons coexprimé avec la CDPS correspondante, au cas où le système flavodoxine/flavodoxine réductase présent chez *E. coli* pourrait constituer un partenaire d'oxydoréduction approprié pour fournir les électrons nécessaires au cycle catalytique du P450, comme cela a déjà été décrit pour certains P450<sup>404</sup>. Par contre, nous n'avons pas essayé de coexprimer d'autres partenaires redox éprouvés pour les P450 (putidarédoxine/putidarédoxine réductase ; adrénodoxine/adrénodoxine réductase), et nous n'avons pas non plus recherché dans le génome d'origine la présence de partenaires potentiels d'oxydoréduction.

**Tableau 25 : Enzymes de modification sélectionnées pour l'étude moyen-débit *in vivo***

Lettre attribuée	Locus enzyme de modification	Type d'enzyme	CDPS associée
A	SJA_c1-32600	Oxydoréductase	10
B	SSEG_06807	Déshydrogénase	16
C	H123_18836	Dioxygénase dépendante du Fe(II) et du 2-oxoglutarate	19
D	H123_18831	Dioxygénase dépendante du Fe(II) et du 2-oxoglutarate	19
E	H123_18826	P450	19
F	H123_18821	Dioxygénase dépendante du Fe(II) et du 2-oxoglutarate	19
G	H123_18816	Dioxygénase dépendante du Fe(II) et du 2-oxoglutarate	19
H	H123_18811	Dioxygénase dépendante du Fe(II) et du 2-oxoglutarate	19
I	VIBR0546_16853	Dioxygénase dépendante du Fe(II) et du 2-oxoglutarate	36
J	Bcep18194_C7321	Dioxygénase dépendante du Fe(II) et du 2-oxoglutarate	37
K	GOEFS_047_00080	Dioxygénase dépendante du Fe(II) et du 2-oxoglutarate	38
L	GOEFS_047_00050	Monooxygénase à flavine	38
M	O3I_025455	Dioxygénase dépendante du Fe(II) et du 2-oxoglutarate	39
N	K530_11830_BIS	Dioxygénase dépendante du Fe(II) et du 2-oxoglutarate	40
O	PMM47T1_20103	Dioxygénase dépendante du Fe(II) et du 2-oxoglutarate	43
P	PFL_1388	Dioxygénase dépendante du Fe(II) et du 2-oxoglutarate	44
Q	PFL_1391	Monooxygénase à flavine	44
R	SSMG_00950	Méthyltransférase	47

Dans la suite du document, les enzymes de modification seront désignées par la lettre qui leur a été attribuée dans le Tableau 25.

## II. Mise en évidence d'activités de modification

18 enzymes de modification potentielles ont été sélectionnées pour tester leur activité de modification de cyclodipeptide. Une étude bibliographique sur ces enzymes hors cytochrome P450 a été réalisée dans le but

de recenser les modifications chimiques qu'elles peuvent catalyser et ainsi aider à l'identification des modifications obtenues éventuellement sur les cyclodipeptides. J'ai choisi de présenter les données bibliographiques recueillies avant d'exposer les résultats de l'étude expérimentale.

## **1. Enzymes de modification sélectionnées et modifications potentielles opérées**

### **1.1 Déshydrogénase/réductase à chaîne moyenne**

L'oxydoréductase potentielle A Tableau 25 est annotée comme une déshydrogénase/réductase à chaîne moyenne (MDR pour « medium chain dehydrogenase/reductase »). La fonction prédite de cette enzyme est la déshydrogénation des prostaglandines. La superfamille des MDR regroupe de nombreuses sous-familles de déshydrogénases, comme par exemple les alcool-déshydrogénases (ADH), les formaldéhyde-déshydrogénases, les quinone-réductases, les leucotriène-déshydrogénase/prostaglandine-réductase. Ces enzymes se présentent sous forme de monomères, dimères, trimères ou tétramères dont les sous-unités sont composées d'environ 350 résidus, ce qui les distingue des déshydrogénases/réductases à chaîne courte (SDR pour « short chain dehydrogenase/reductase ») dont les sous-unités sont composées de seulement 250 résidus environ. Une autre distinction entre ces deux familles provient de la dépendance au zinc. Les SDR sont toujours indépendantes au zinc tandis que les MDR peuvent avoir aucun, un ou deux centres métalliques nécessaires à leur activité. Les MDR sont constituées de deux domaines : le domaine C-terminal qui présente un repliement de type Rossmann et qui permet la liaison à la coenzyme NAD(P) et le domaine N-terminal qui contient le site catalytique. Ces enzymes réalisent soit des oxydations d'alcools en aldéhydes ou en cétones et des oxydations d'aldéhydes en acides, soit des réductions d'aldéhyde en alcool. Certaines ont également montré des activités contre des quinones et des diamides. La variété des substrats des membres de cette superfamille montre leur importance dans le métabolisme<sup>405</sup>.

### **1.2 Aldéhyde déshydrogénases**

La déshydrogénase B est annotée comme appartenant à la superfamille des aldéhydes déshydrogénases (ALDH) et, plus précisément, comme une 2-hydroxyomuconique semialdéhyde déshydrogénase. Les ALDH sont principalement connues pour être impliquées dans le métabolisme d'aldéhydes physiologiques et pathologiques, en ayant un rôle de détoxification. Elles catalysent l'oxydation des aldéhydes en acides carboxyliques<sup>406</sup>. Elles sont organisées en trois domaines : un domaine de liaison au NAD(P) et au cofacteur, un domaine catalytique et un domaine appelé « bridging domain »<sup>407</sup>. En plus de l'oxydation des aldéhydes, plusieurs autres fonctions catalytiques (hydrolyse d'ester, réduction de nitrate et antioxydant indirect (par la production de NAD(P)H)) et non-catalytiques (antioxydant direct, absorption de rayonnement UV, liaison à des endobiotiques ou xénobiotiques) sont reconnues pour ces enzymes<sup>408</sup>.

### 1.3 Dioxygénases dépendant du Fe(II) et du 2-oxoglutarate

Les dioxygénases dépendantes du Fe(II) et du 2-oxoglutarate (OG) (enzymes C, D, F à K, M à P) sont des enzymes impliquées dans de nombreux processus biologiques comme la réparation des acides nucléiques, la biosynthèse d'antibiotiques et de produits végétaux, la modification de chaînes latérales de protéines, le métabolisme des lipides ou la dégradation de nombreux composés<sup>409</sup>. Ces enzymes présentent toutes une structure composée de feuillettes  $\beta$  organisés en « gâteau roulé » (jelly-roll) qui possèdent le motif His<sup>1</sup>-X-Asp/Glu-X<sub>n</sub>-His<sup>2</sup>, dont les trois résidus conservés sont impliqués, avec le 2-oxoglutarate (2OG), dans la chélation du FeII essentiel à l'activité des OG (Figure 72).

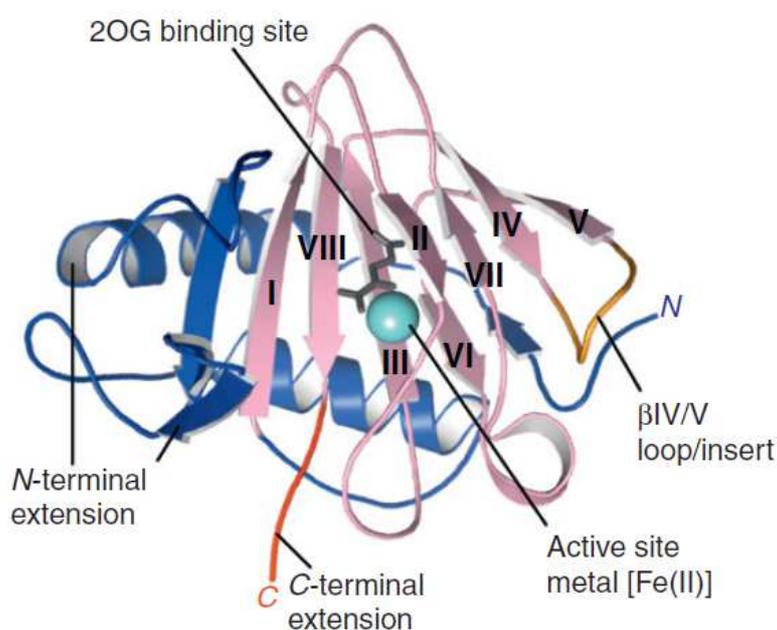


Figure 72 : Structure tridimensionnelle des dioxygénases dépendant du Fe(II) et du 2-oxoglutarate, adaptée de AlkB (*Escherichia coli* alkylation repair protein B (PDB : 3BIE))

Les différentes régions définies dans la littérature<sup>410</sup> sont indiquées : les éléments de structures secondaires appartenant au motif « gâteau roulé » sont en rose, les éléments le précédant sont en bleu, l'insert entre les feuillettes IV et V est en orange et l'extension C-terminale en rouge. Reproduit de <sup>410</sup>

La fixation du 2OG précède celle du substrat. Un complexe ternaire composé de l'enzyme, du 2OG et du substrat est alors formé. La décarboxylation oxydante du 2OG produisant du CO<sub>2</sub> et du succinate mène à l'obtention d'une espèce réactive Fe(IV)-oxo (FeIV=O) qui réalise l'oxydation du substrat. Cette oxydation à deux électrons fait passer le Fe(IV) à l'état Fe(II)<sup>410,411</sup> (Figure 73). Dans le cadre de la biosynthèse, l'oxydation réalisée peut être de différents types : hydroxylation, déshydrogénation, épimérisation, oxydation d'aldéhyde en acide carboxylique, cyclisation, expansion de cycle ou encore un réarrangement oxydatif<sup>409,410</sup>. La OG

intervenant dans la voie de la clavamate (clavamate synthase, CAS) catalyse trois étapes oxydantes et illustre la versatilité de ces enzymes<sup>409</sup>. Les acides aminés (proline, arginine, asparagine, leucine et isoleucine notamment) sont des substrats courants de ces enzymes et sont généralement transformés en hydroxy-aminoacides<sup>412</sup>. Ces activités d'hydroxylation sont stéréosélectives et régiosélectives sur l'un des trois sites de la chaîne latérale : C-3, C-4 ou C-5. Deux autres familles d'enzymes liées par la séquence ou la structure aux OG, les isopénicilline-N-synthases (IPNS) et les 1-aminocyclopropane-1-carboxylate oxydase (ACCO), présentent des activités similaires mais n'utilisent pas le 2OG comme cosubstrat. Il est à noter que l'un des domaines conservés identifié sur les enzymes sélectionnées appartient aux IPNS.

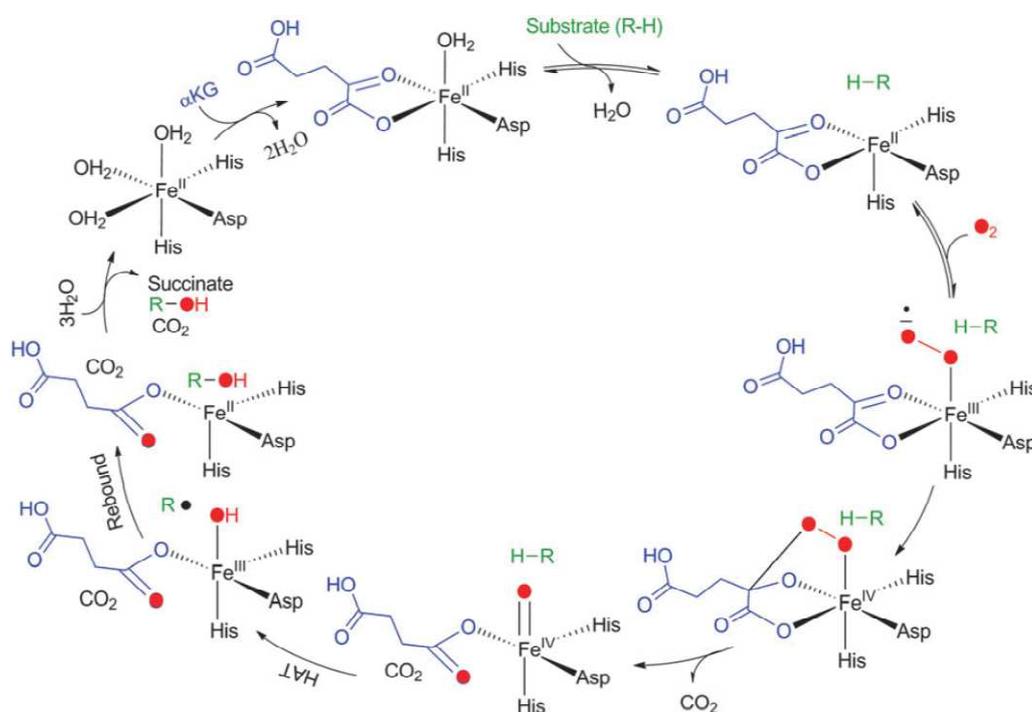


Figure 73 : Mécanisme du cycle catalytique des dioxygénases dépendant du Fe(II) et du 2-oxoglutarate (bleu) menant à la monoxydation du substrat (vert) et la formation de succinate et CO<sub>2</sub>

Les différents états d'oxydation du centre métallique sont indiqués. Les résidus de l'enzyme impliqués dans la chélation du Fe sont indiqués en noir. Reproduit de<sup>411</sup>.

#### 1.4 Monoxygénase à flavine

Les enzymes L et Q sont annotées dans NCBI comme des monoxygénases à flavine (FMO pour « Flavin-utilizing MonoOxygenase »). Les FMO interviennent dans des processus biologiques variés comme le catabolisme de produit endogènes ou exogènes et la biosynthèse d'hormones, de vitamines ou d'antibiotiques. Elles utilisent un cofacteur flavinique de type flavine mononucléotide (FMN) ou flavine adénine dinucléotide

(FAD). Ce sont des dérivés de la riboflavine (vitamine B<sub>2</sub>) et c'est leur cycle isoalloxazine, capable de réaction à un ou deux électrons, qui leur confère leurs propriétés oxydoréductrices exploitées par les enzymes à flavines (Figure 74)<sup>413</sup>.

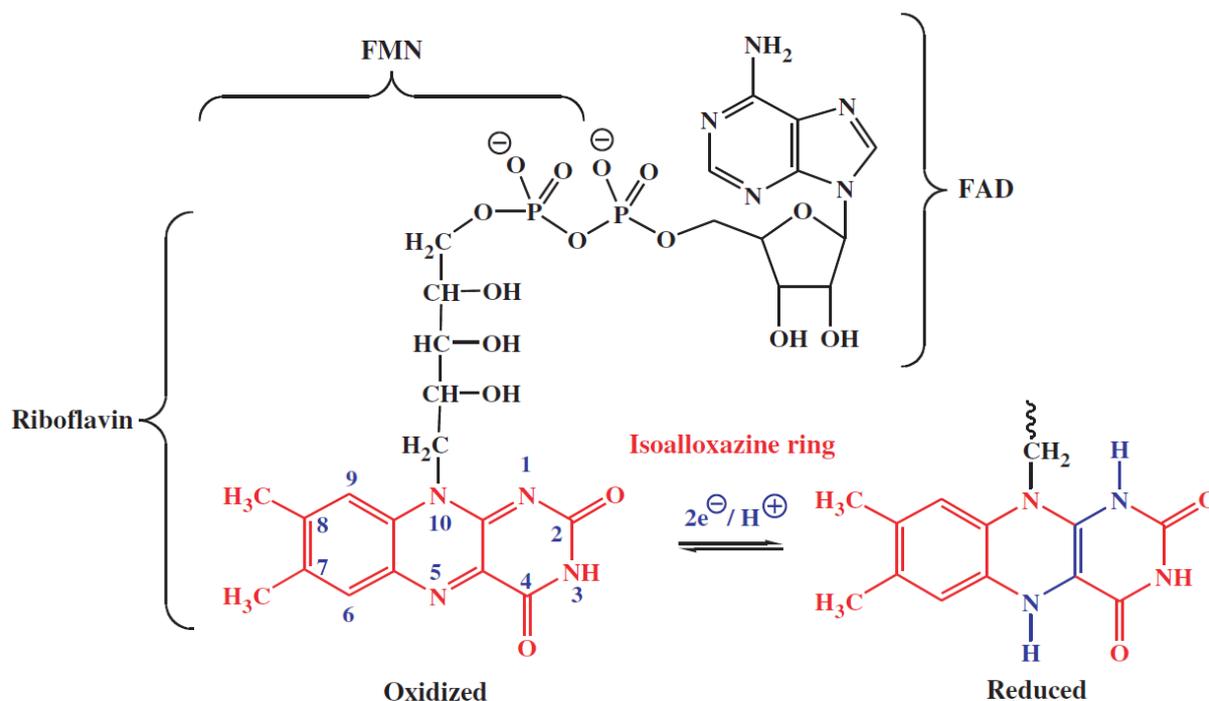


Figure 74 : Structures de la riboflavine, FMN et FAD

Le cycle isoalloxazine est représenté sous forme oxydée et réduite à deux électrons (rouge et bleu). La numérotation pour le cycle isoalloxazine est indiquée sur le cycle oxydé. Reproduit de<sup>413</sup>.

Les FMO utilisent l'oxygène moléculaire et catalysent l'incorporation d'un de ses atomes d'oxygène dans un substrat tandis que le second atome d'oxygène est réduit en eau. Dans la majorité des cas, un adduit stable entre le cofacteur flavinique réduit et le dioxygène est formé transitoirement. Cette espèce peroxyde réagit ensuite avec un électrophile ou un nucléophile, ce qui entraîne le clivage de la liaison O-O, l'incorporation d'un atome d'oxygène dans un substrat et la régénération de la flavine oxydée. Les FMO sont réputées pour être chimio-, régio- et énantiosélective. Les FMO sont réparties en huit groupes selon leur fonctionnement et leur structure (Tableau 26). Les groupes A et B regroupent les FMO qui utilisent le NAD(P)H en tant que donneur d'électron externe pour réduire le cofacteur flavinique. Ils se distinguent l'un de l'autre par le fait que les FMO du groupe B lient le NAD(P) pendant toute la catalyse. Les FMO des groupes C et F nécessitent une réductase comme partenaire protéique pour fournir une flavine réduite

(FADH<sub>2</sub> ou FMNH<sub>2</sub>). Enfin, les FMO des groupes G et H réduisent les flavines par l'oxydation du substrat. Concernant les réactions catalysées par les FMO, elles sont particulièrement variées : hydroxylations, oxydations de type Bayer-Villiger, sulfoxydations, époxydations et halogénations<sup>414</sup>.

**Tableau 26 : Classification des FMO adapté de<sup>414</sup>**

Groupe	Nombre de protéines impliquées <sup>1</sup>	Cofacteur	Donneur d'électron	Repliement de la protéine	Réaction
A	1 (MO)	FAD	NAD(P)H	Rossmann de type glutathion réductase avec zone de liaison pour le FAD	Hydroxylation (ortho ou para de composé phénolique) Sulfoxydation
B	1 (MO)	FAD	NAD(P)H	Rossmann, zone de liaison FAD et NAD(P)H	Oxydation de Bayer-Villiger Hétéro atome oxydation N-hydroxylation Décarboxylation oxydante
C	2 (MO + réductase à flavine dépendant du NAD(P)H)	FMN	FMNH <sub>2</sub>	TIM barrel (tonneau de type triose phosphate isomérase)	Désulfonation Oxydation de Bayer-Villiger Sulfoxydation Epoxydation, Hydroxylation Emission de lumière
D	2 (MO+ réductase à flavine dépendant du NAD(P)H)	FMN/ FAD	FMNH <sub>2</sub> / FADH <sub>2</sub>	Acyl-CoA déshydrogénase	Aromatique hydroxylation N-hydroxylation
E	2 (MO+ réductase à flavine dépendant du NAD(P)H) ou 1 (fusion)	FAD	FADH <sub>2</sub>	Rossmann de type glutathion réductase	Epoxydation
F	2 (MO+ réductase à flavine dépendant du NAD(P)H)	FAD	FADH <sub>2</sub>	Rossmann de type glutathion réductase avec zone de liaison pour le FAD	Chloration, bromation
G	1 (MO à flavoprotéine interne)	FAD	Substrat aminoacide	Monoamine oxydase (oxydase d'acides aminés I)	Décarboxylation oxydante
H	1 (MO à flavoprotéine interne)	FMN	Substrat	TIM barrel	Décarboxylation et dénitrification oxydante

<sup>1</sup>MO : monoxygénase

### 1.5 Méthyltransférase

L'enzyme R est annotée comme une méthyltransférase (MT) dépendante de la S-adénosylméthionine (SAM) de classe I. Les MT dépendant de la SAM sont des enzymes qui catalysent le transfert d'un groupement

méthyle depuis la SAM sur une grande variété de substrats allant de petits métabolites à des biomacromolécules comme des protéines ou de l'ADN, en générant de la *S*-adénosyl-homocystéine (SAH) comme sous-produit<sup>415</sup>. Elles sont regroupées en cinq grandes classes en fonction de leurs caractéristiques structurales, la classe I étant celle possédant le plus de membres<sup>416</sup>. Les MT de classe I présentent un repliement de type Rossmann et agissent selon un mécanisme de type S<sub>N</sub>2. La méthylation peut avoir lieu sur les centres nucléophiles O, N, C, S, Se ou As d'un grand nombre de métabolites primaires et secondaires dans des voies cataboliques ou de biosynthèse. Quelques MT de protéines ou de l'ADN font également partie de cette classe I. La classe II se caractérise par la présence d'un domaine de réactivation de type méthionine synthase (MetH) ; la classe III regroupe les précorrène-4-méthyltransférases, la classe IV les méthyltransférases de l'ARN de la famille SPOUT. Enfin, les MT possédant un domaine SET forme la classe V qui regroupe notamment les MT agissant sur les résidus lysine<sup>415</sup>. De plus, une dernière classe de MT agissant selon un mécanisme radicalaire a été identifiée<sup>417</sup>.

La recherche de MT pour la biocatalyse est un domaine particulièrement stratégique car la méthylation est connue pour améliorer les propriétés des composés bioactifs. Ce phénomène appelé « l'effet méthyle » permet par exemple d'augmenter la lipophilicité, et ainsi la biodisponibilité des composés, ou de les protéger de la dégradation enzymatique *in vivo*<sup>418,419</sup>. Or, la méthylation par voie chimique utilise de l'iodométhane, réactif peu recommandable d'un point de vue toxicologique et environnemental. La méthylation par une voie enzymatique représente donc une alternative intéressante bien que encore relativement onéreuse.

L'identification réalisée par informatique permet d'annoter de manière hypothétique la fonction de chaque protéine. Ces annotations sont donc à considérer avec beaucoup de précaution et cette étude bibliographique permet seulement de faire des hypothèses sur le type de modification qui peut être produite par chaque enzyme. Cependant, elle nous montre également que les fonctions prédites peuvent regrouper des enzymes aux activités diverses ou agissant sur des substrats variés. De plus, dans une même famille, les mécanismes catalytiques impliqués peuvent également diverger, certains utilisant un cofacteur, une coenzyme ou encore un partenaire protéique dont d'autres n'ont pas besoin. L'étude expérimentale doit permettre d'identifier les enzymes actives sur les cyclodipeptides produits par les CDPS auxquelles sont associées.

## **2. Identification de nouvelles activités de modification**

Une souche d'*E. coli* BL21AI est cotransformée avec deux plasmides, l'un portant le gène de la CDPS et l'autre portant le gène de l'enzyme de modification potentielle associée. Dans le cas d'une voie de biosynthèse impliquant plusieurs enzymes de modification, chacune d'entre elles a été testée en association

avec la CDPS correspondante, indépendamment des autres enzymes de modification. Des colonies transformantes ont été obtenues dans tous les cas. Les enzymes A à Q ont été testées dans le système de culture moyen-débit mis au point au chapitre I. Pour l'activité de l'enzyme R, les cultures ont été réalisées en utilisant le système de culture en erlenmeyer classique (cf. Matériels et méthodes. V.1 Culture en erlenmeyer). Les surnageants de cultures obtenus sont analysés par LC-MS/MS. Les chromatogrammes obtenus sont comparés à ceux obtenus lorsque seule la CDPS est exprimée chez *E. coli* : la disparition de pics correspondant à des cyclodipeptides précédemment identifiés et l'apparition de nouveaux pics ont été recherchées.

Pour les enzymes A, B, E, F, L et Q, les profils de production de cyclodipeptides par les souches *E. coli* sont identiques, que la CDPS soit exprimée seule ou en association avec l'une de ces enzymes de modification potentielles. De plus, le pic correspondant au cyclodipeptide produit par la CDPS a une aire équivalente en présence et en l'absence de l'enzyme de modification, ce qui suggère une absence de consommation de ce cyclodipeptide. La production de chacune des enzymes de modification a été vérifiée dans les fractions soluble et insoluble par gel SDS-PAGE. Parmi les six enzymes qui n'ont pas montré d'activité dans notre test, l'enzyme Q est la seule pour laquelle aucune bande compatible avec son expression, même de faible intensité, n'ait été détectée (Figure 75 et Figure 76). Les enzymes A, B, E, F, L, bien que produites en quantité visible dans la fraction soluble, ne semblent pas avoir d'activité sur les cyclodipeptides produits par la CDPS coexprimée. Aucune raison évidente ne nous est apparue pour les enzymes A et B (déshydrogénase de chaîne moyenne et aldéhyde déshydrogénase), si ce n'est la possibilité que le cyclodipeptide produit ne soit pas substrat. L'enzyme E est un P450 et l'absence d'activité peut s'expliquer par l'absence d'une chaîne de transfert d'électrons appropriée.

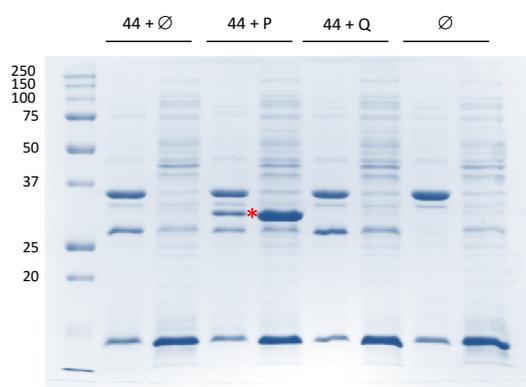
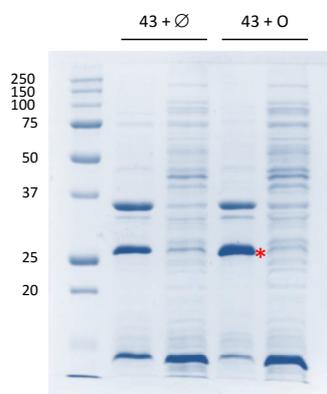
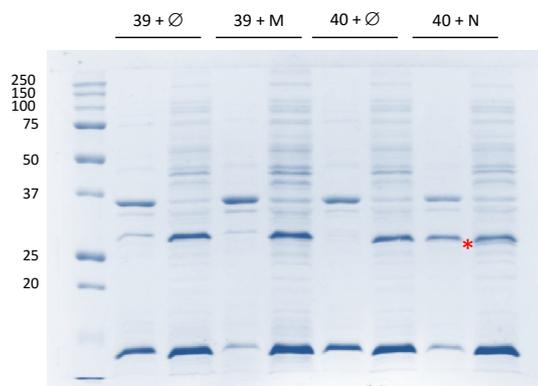
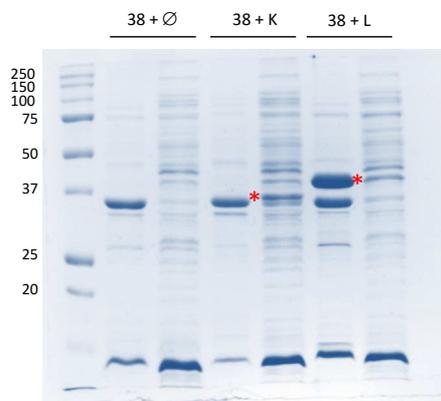
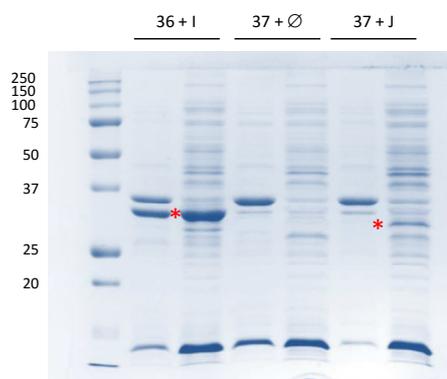
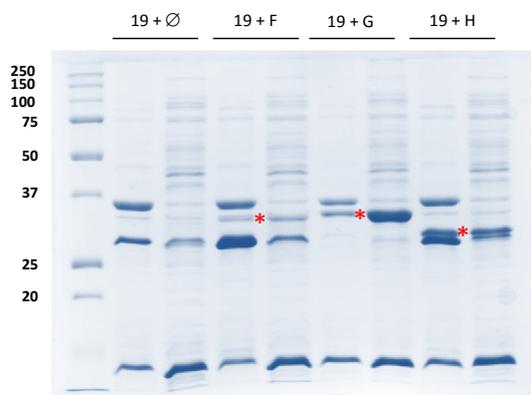
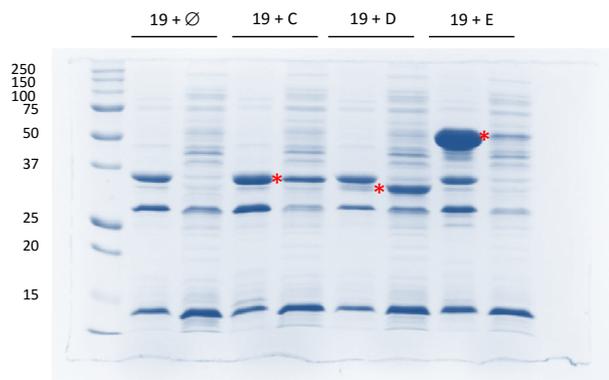
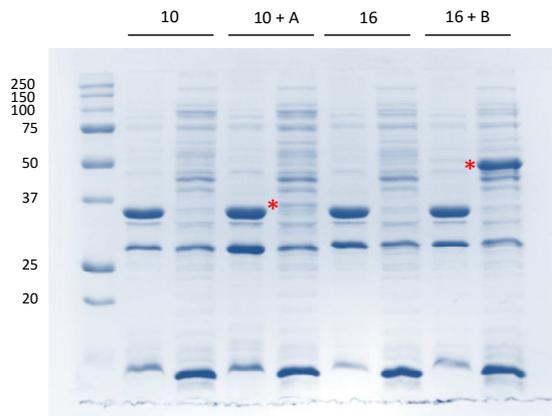


Figure 75 : Analyse SDS-PAGE de la production des enzymes de modification, hors méthyltransférase R

Analyse sur des gels SDS-PAGE à 12% colorés au bleu de Coomassie des fractions insoluble à gauche et soluble à droite provenant des culots bactériens à l'issue de 24 h de culture de souche de BL21AI co-transformée par le pIJ196 codant pour la CDPS indiquée au-dessus de la piste et le pIJ194 codant pour l'enzyme indiquée au-dessus de la piste. Un témoin exprimant la CDPS mais n'exprimant pas l'enzyme de modification est réalisée (n° CDPS + Ø) pour chaque CDPS (sauf pour la CDPS 36 qui n'a pas donné de transformant). Un témoin n'exprimant ni CDPS, ni enzyme de modification (Ø) est également ajouté sur le dernier gel. La bande correspondant à l'enzyme de modification est signalée par un astérisque rouge entre la bande de l'insoluble et du soluble. Le poids moléculaire des enzymes de modification est compatible avec la taille apparente de ces bandes.

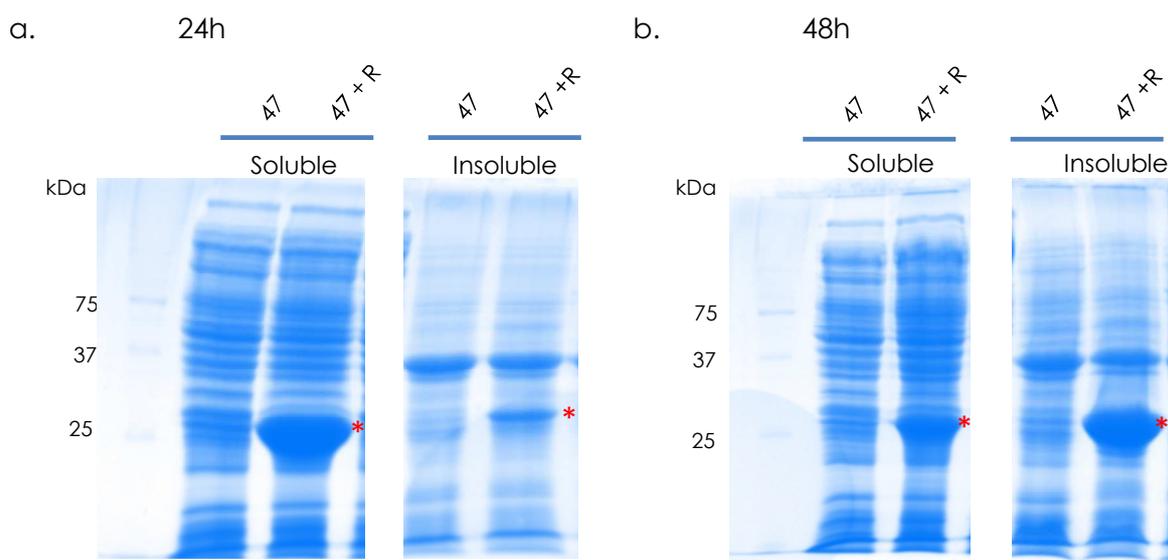


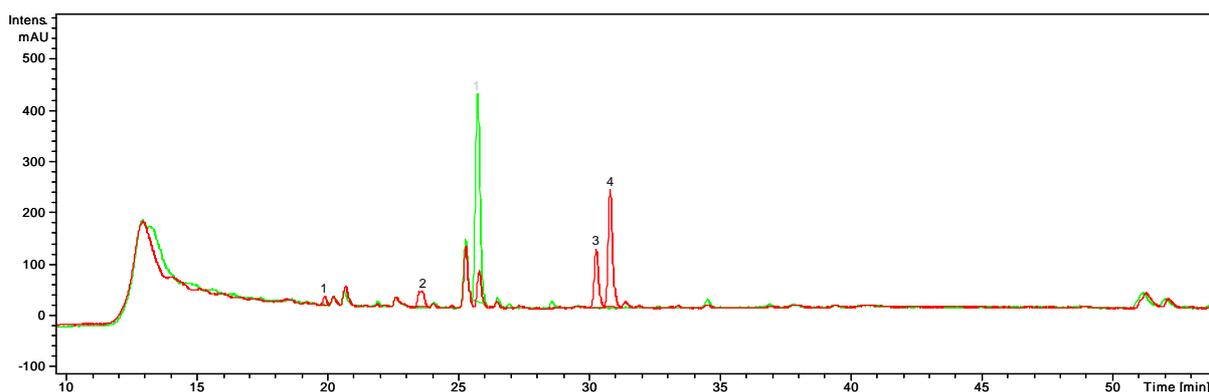
Figure 76 : Analyse SDS-PAGE de la production de méthyltransférase R

Les fractions soluble et insoluble provenant des culots bactériens à l'issue de 24 et 48 h de culture de souche de BL21AI co-transformée par le pIJ196 codant pour la CDPS 47 et le pIJ194 codant pour la méthyltransférase R ou non sont analysées sur des gels SDS-PAGE à 12% colorés au bleu de Coomassie. La bande correspondant à la méthyltransférase (M = 24,47 kDa) est signalée par un astérisque rouge.

Les enzymes L et Q sont des monooxygénases à flavine. L'inactivité des FMO dans nos conditions expérimentales peut être due au fait que leur substrat n'est pas le cyclodipeptide produit par la CDPS mais le cyclodipeptide modifié par l'action de la dioxygénase (cf. ci-dessous). Afin de vérifier cette hypothèse, nous avons cultivés des bactéries *E. coli* co-exprimant la CDPS, la dioxygénase et la FMO en utilisant le système de plasmide à deux vecteurs. Pour ce faire, le gène de la dioxygénase K et celui de dioxygénase P ont été implémentés dans le pIJ196 contenant respectivement le gène de CDPS 38 et le gène de CDPS 44. Les surnageants de culture de bactéries co-transformées, d'une part avec le vecteur pIJ196 codant pour la CDPS et la dioxygénase et d'autre part avec le vecteur pIJ194 codant pour la FMO associée ont été analysés par LC-MS/MS. A nouveau, aucun nouveau composé n'a été identifié dans le surnageant par rapport au cas où seules la CDPS et la dioxygénase sont co-exprimées (cf. ci-dessous). Les interrogations HHPRED avec chacune des

séquences de ces FMO donnent avec un score de probabilité de 100% les 18 mêmes protéines, mais dans un ordre différent. Ces protéines présentent toutes le même Pfam (PF00296). Plusieurs d'entre elles peuvent être classées dans le groupe C des monooxygénases à flavine qui regroupe des enzymes possédant un repliement de type tonneau TIM et nécessitant une réductase de flavine pour fournir un cofacteur FMNH<sub>2</sub> de manière spécifique. Certaines sont décrites comme nécessitant le cofacteur flavinique F420, cofacteur synthétisé seulement chez certains microorganismes et absent chez *E. coli*. L'absence d'un cofacteur flavinique fonctionnel (FMNH<sub>2</sub> délivré de manière spécifique ou F420) pourrait expliquer l'absence d'activité. Des essais de purification de la FMO L ont été entrepris au laboratoire pour essayer d'identifier si l'enzyme était co-purifiée avec un cofacteur flavinique. Les résultats préliminaires que nous avons obtenus indiquent que la FMO purifiée ne présente pas de spectre visible caractéristique de la présence d'un cofacteur flavinique. L'absence d'activité des FMO peut s'expliquer par l'absence d'un partenaire approprié pour fournir le cofacteur flavinique réduit. Cependant, il ne peut pas être exclu que ces FMO n'appartiennent aux voies de biosynthèse de dicétopipérazines.

Pour les 12 autres enzymes, des activités de modification de cyclodipeptides sont repérées. Parmi ces enzymes, la production d'une seule d'entre elle, l'enzyme M, n'est pas visualisée correctement sur les gels SDS-PAGE (Figure 75). A partir des données de spectrométrie de masse obtenues, nous avons tenté d'identifier les cyclodipeptides modifiés responsables des nouveaux pics. Pour chaque pic, la fragmentation MS/MS obtenue a été comparée avec celle obtenue pour les cyclodipeptides majoritairement produits par la CDPS. Dans le cas de l'enzyme J, par exemple, la co-expression de cette enzyme avec la CDPS 37 provoque la disparition du pic principal de cyclo(Ala-Glu) et l'apparition de quatre nouveaux pics (notés de 1 à 4) sur les chromatogrammes (Figure 77). La fragmentation MS/MS de l'ion majoritaire est étudiée pour chaque pic.



**Figure 77 : Comparaison de l'analyse LC-MS/MS du surnageant de culture de bactéries *E. coli* exprimant la CDPS 37 en vert et co-exprimant la CDPS 37 et l'enzyme de modification potentielle J (en rouge)**

Les analyses sont effectuées comme décrit dans le Chapitre II. IV.1 Méthode pour l'identification des cyclodipeptides. Le pic correspondant à l'éluion du cyclo(Ala-Glu) porte le n° 1 en gris et les pics

correspondant à l'éluion de quatre nouveaux composés portent les n° 1 à 4 en noir). Le chromatogramme UV enregistré à  $\lambda = 214$  nm est représenté.

Pour le pic 2, la fragmentation MS<sup>2</sup>, MS<sup>3</sup> et MS<sup>4</sup> de l'ion majoritaire à  $m/z$  199 est comparée à celle du cyclo(Ala-Glu) ( $m/z$  201) (Figure 78 et Figure 79). Les pertes caractéristiques de -28 et -45, liées à la fragmentation du cycle 2,5-dicétopipérazine (cf. Etude bibliographique I.4 L'identification des 2,5-dicétopipérazines) sont retrouvées de la même manière dans les deux molécules et les fragmentations sont superposables pour la MS<sup>2</sup> et la MS<sup>3</sup> avec un simple décalage de 2 uma. Cela nous amène à faire l'hypothèse que ce composé est un dérivé du cyclo(Ala-Glu) et que, si cet ion est l'ion moléculaire, une modification de type déshydrogénation pourrait être à l'origine de l'écart de  $m/z$  observée.

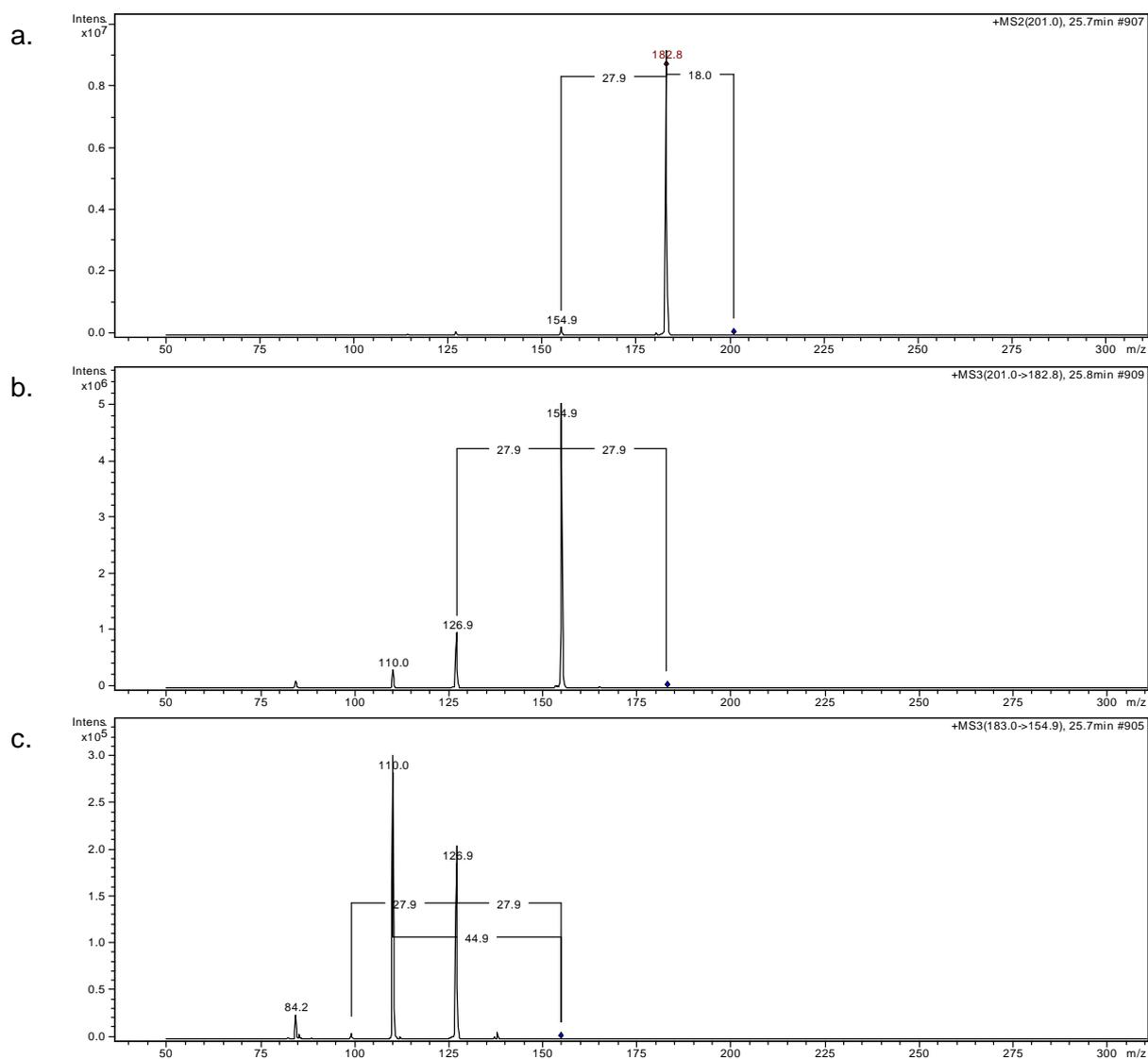
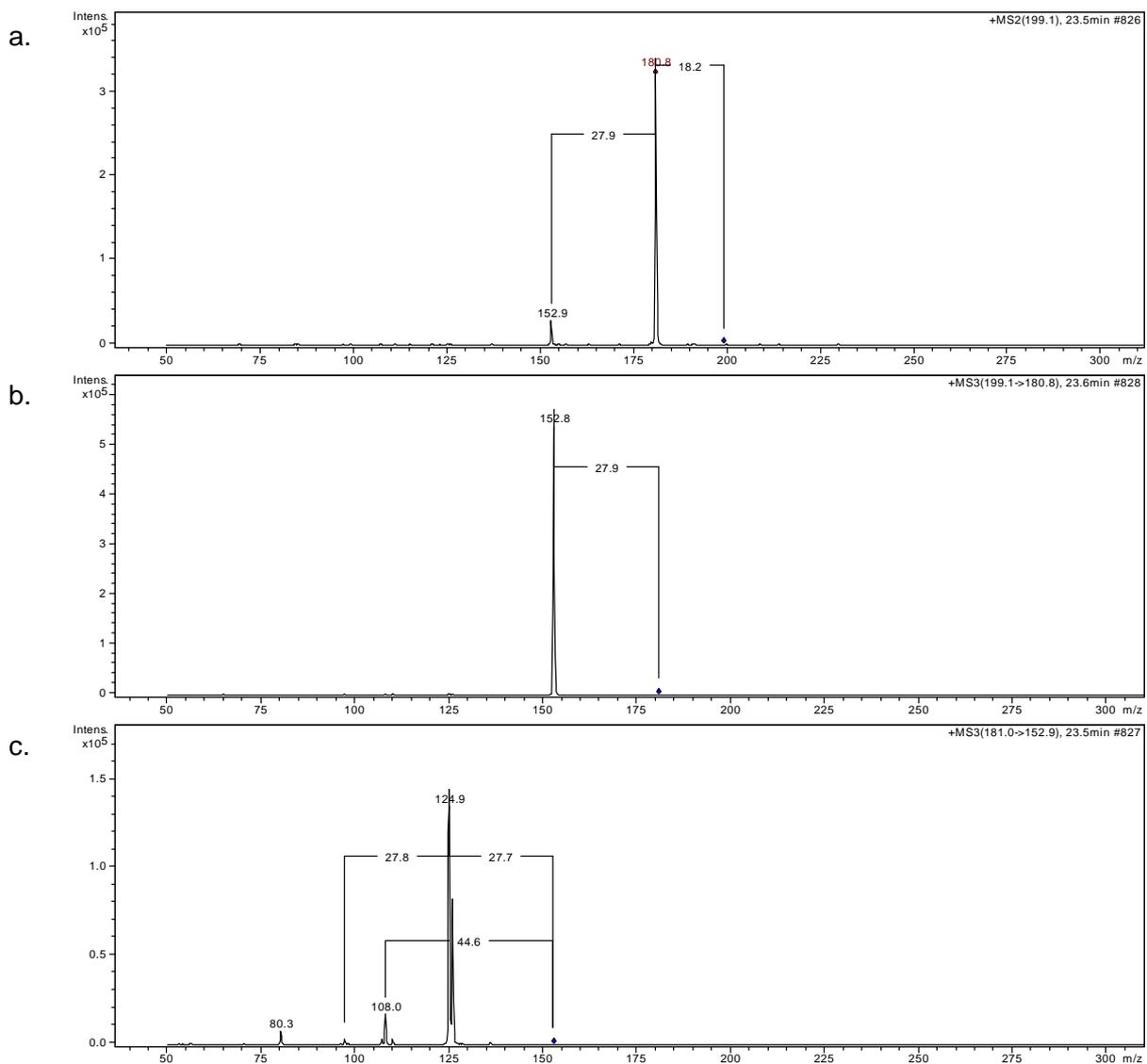


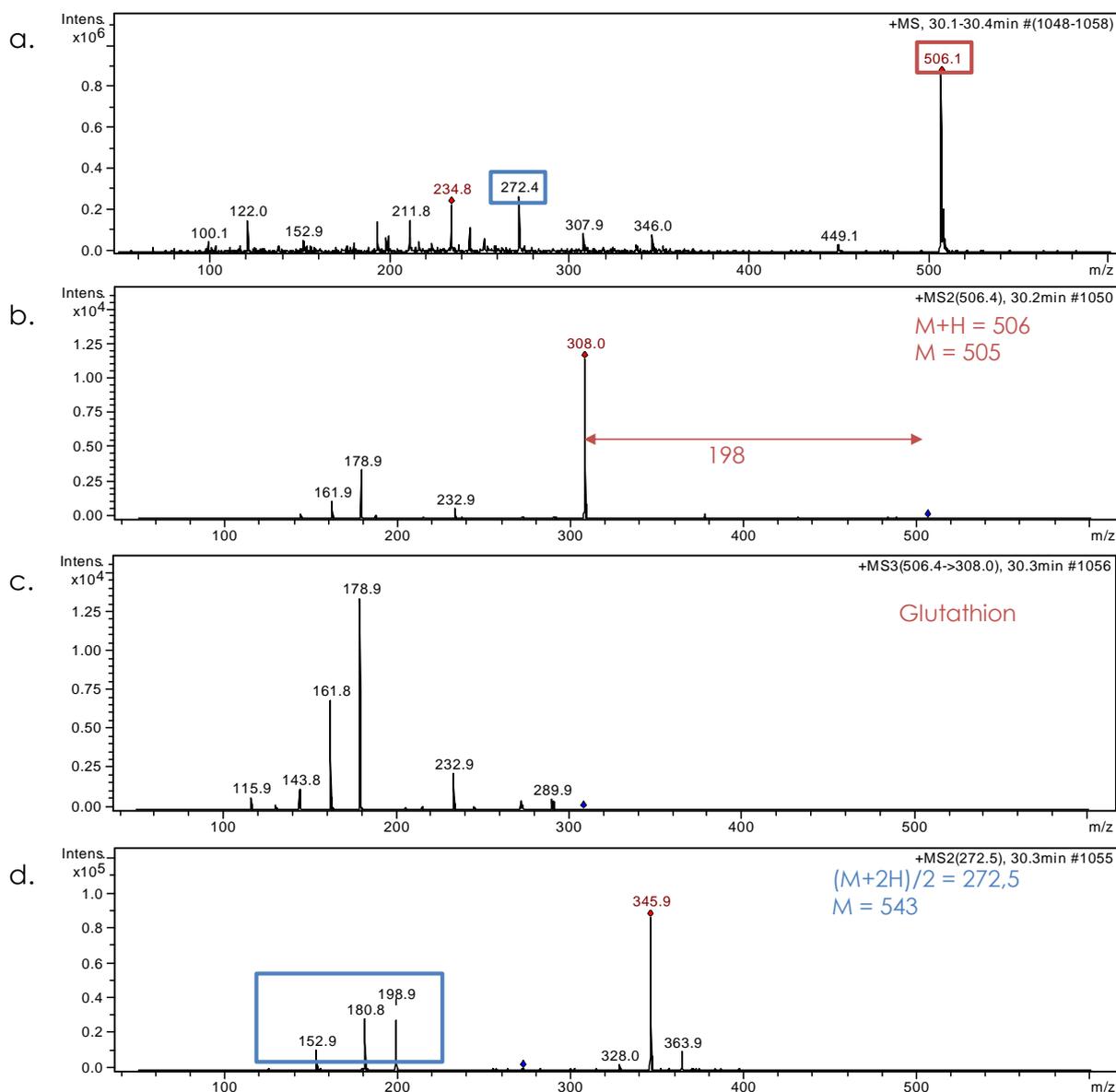
Figure 78 : Fragmentation MS<sup>2</sup> (a.), MS<sup>3</sup> (b.) et MS<sup>4</sup>(c.) du cyclo(Ala-Glu) ( $m/z$  201,  $t_R = 23,5$  min (Hybercarb))



**Figure 79 : Fragmentation MS<sup>2</sup> (a.), MS<sup>3</sup> (b.) et MS<sup>4</sup>(c.) du composé inconnu de  $m/z$  199 et  $t_R=23,5$  min (Hybercarb)**

Pour les pics 3 et 4, deux ions principaux sont retrouvés :  $m/z$  506 et  $m/z$  272,5. La fragmentation MS2 du premier provoque une perte de 198, ce qui donne un ion produit de  $m/z$  308 (Figure 80). Les données de la fragmentation de cet ion produit (Figure 80c.) servent à interroger la base de données de spectrométrie de masse MassBank (<http://massbank.ufz.de/MassBank/>). Le composé de  $m/z$  308 est identifié comme du L-glutathion. L'ion  $m/z$  272,5 est dichargé et sa fragmentation fait apparaître notamment les ions 199, 180, 153 déjà présents dans l'analyse du composé du pic 2. Ces données suggèrent que les composés responsables

des pics 3 et 4 sont des adduits de glutathion sur un dérivé du cyclodipeptide. L'analyse des fragmentations associées au pic 1 n'a pas permis d'identifier une relation avec le cyclo(Ala-Glu).



**Figure 80 : Données MS<sup>n</sup> sous le pic 3 à t<sub>R</sub> = 30,3 min (Hybercarb). Données MS (a.). Fragmentation MS<sup>2</sup> (506) (b.), MS<sup>3</sup> (506 → 308) (b.) et MS<sup>2</sup> (272,5) (c.)**

Pour chaque pic apparaissant en présence d'une enzyme de modification, la même analyse est réalisée, Un numéro de molécule est attribué sous la forme IJ + numéro à chaque composé identifié. Les résultats obtenus sont présentés dans le tableau suivant. Excepté dans le cas de la méthyltransférase, l'apparition de ces composés est corrélée avec une disparition ou une forte diminution du cyclodipeptide produit par la CDPS associée.

**Tableau 27 : Composés apparaissant lorsque les enzymes de modification sélectionnées sont co-exprimées avec la CDPS correspondante (résultats complets disponibles dans l'annexe électronique 2, feuille 3)**

Cyclodipeptide	CDPS et enzyme de modification	Nom	t <sub>R</sub> (min) <sup>1</sup>	m/z	Modification probable
Cyclo(Ile-Leu) m/z 227, t <sub>R</sub> =43 min	19 + C	IJ 1	33,6*	243, 225	?
		IJ 2	36,6*	532	Adduit glutathion sur cyclo(Ile-Leu) déshydrogéné
	19 + D	IJ 3	19*	275	3 hydroxylations
		IJ 4	21,8*	259	2 hydroxylations
		IJ 5	22,4*	259	2 hydroxylations
		IJ 6	22,9*	259	2 hydroxylations
		IJ 7	25,5*	243	1 hydroxylation
		IJ 8	26,1*	243	1 hydroxylation
		IJ 9	32,2*	243	1 hydroxylation
	19 + G	IJ 10	30,3*	243	1 hydroxylation
		IJ 11	32,1*	257	2 hydroxylations + déshydrogénation
	19 + H	IJ 12	28,4*	257	2 hydroxylations + déshydrogénation
Cyclo(Leu-Glu), m/z 243 ; 36,2 min	36 + I	IJ 13	33,6	241	Déshydrogénation
		IJ 14	35,5	241	Déshydrogénation
		IJ 15	39,7	548, 293,7	Adduit glutathion sur cyclo(Leu-Glu) déshydrogéné
		IJ 16	40,2	548, 293,7	Adduit glutathion sur cyclo(Leu-Glu) déshydrogéné
	44 + P	IJ 15	39,7	548, 293,7	Adduit glutathion sur cyclo(Leu-Glu) déshydrogéné
		IJ 16	40,2	548, 293,7	Adduit glutathion sur cyclo(Leu-Glu) déshydrogéné
Cyclo(Ala-Glu), m/z 201; 25,5min	37 + J	IJ 17	19,9	179	Pas apparenté au cyclo(Ala-Glu)

Cyclodipeptide	CDPS et enzyme de modification	Nom	t <sub>R</sub> (min) <sup>1</sup>	m/z	Modification probable
		IJ 18	23,6	199	Déshydrogénation
		IJ 19	30,3	506, 272,5	Adduit glutathion sur cyclo(Ala-Glu) déshydrogéné
		IJ 20	30,8	506, 272,5	Adduit glutathion sur cyclo(Ala-Glu) déshydrogéné
	38 + K	IJ 17	20	179	Pas apparenté au cyclo(Ala-Glu)
		IJ 18	23,7	199	Déshydrogénation
		IJ 19	30,4	506, 272,5	Adduit glutathion sur cyclo(Ala-Glu) déshydrogéné
		IJ 20	30,9	506, 272,5	Adduit glutathion sur cyclo(Ala-Glu) déshydrogéné
	39 + M	IJ 17	20	179	Pas apparenté au cyclo(Ala-Glu)
		IJ 18	23,7	199	Déshydrogénation
		IJ 19	30,4	506, 272,5	Adduit glutathion sur cyclo(Ala-Glu) déshydrogéné
		IJ 20	31	506, 272,5	Adduit glutathion sur cyclo(Ala-Glu) déshydrogéné
	40 + N	IJ 17	20	179	Pas apparenté au cyclo(Ala-Glu)
		IJ 18	23,7	199	Déshydrogénation
		IJ 19	30,5	506, 272,5	Adduit glutathion sur cyclo(Ala-Glu) déshydrogéné
		IJ 20	31	506, 272,5	Adduit glutathion sur cyclo(Ala-Glu) déshydrogéné
	43 + O	IJ 17	20	179	Pas apparenté au cyclo(Ala-Glu)
		IJ 18	23,7	199	Déshydrogénation
		IJ 19	30,5	506, 272,5	Adduit glutathion sur cyclo(Ala-Glu) déshydrogéné
		IJ 20	31	506, 272,5	Adduit glutathion sur cyclo(Ala-Glu) déshydrogéné
Cyclo(Trp-Trp), m/z 373; 41,1 min	47 + R	IJ 21	44,4*	387	Méthylation

<sup>1</sup> Par défaut le temps de rétention pour une analyse sur Hypercarb. \* temps de rétention pour une analyse dC18,

Pour les voies contenant les CDPS 37, 38, 39, 40, ou 43 produisant le cyclo(Ala-Glu), le même profil de production de composés que celui présenté pour 37+J (Figures 77, 79 et 80), est retrouvé pour les associations CDPS + OG, à savoir la production de IJ18, IJ19 et IJ20.

La CDPS 19 produisant du cyclo(Ile-Leu) est associée avec six enzymes de modification potentielles, comprenant un P450 trouvé inactif (voir ci-dessus) et cinq OG. Une seule des OG (enzyme F) est inactive sur le cyclodipeptide cyclo(Ile-Leu). Pour les quatre autres OG (enzymes C, D, G et H), une activité sur le cyclo(Ile-Leu) est observée. Les enzymes D et G sont versatiles, notamment l'enzyme D qui produit sept composés différents. Au vue des  $m/z$  de ces composés, il est probable que l'enzyme accomplisse trois hydroxylations à des positions différentes et que les produits formés contiennent une, deux ou trois de ces hydroxylations. Aucun des produits hydroxylés ou déshydratés par l'une des enzymes n'est retrouvé produit par une autre dioxygénase de la voie, les temps de rétention étant tous différents. L'enzyme C produit également un composé retrouvé sur forme d'adduit avec le glutathion (IJ2). Les fragmentations retrouvées pour IJ2 sont similaires à celle observées pour IJ19 et IJ20 produits par l'enzyme J, mais décalées de  $m/z$  26 (écart entre la masse du cyclo(Ala-Glu) produit par la CDPS 37 et celle du cyclo(Ile-Leu) produit par la CDPS 19).

Les CDPS 36 et 44 produisant du cyclo(Leu-Glu) ont été associées respectivement aux enzymes de modification I et P. Ces associations 36+I et 44+P ne donnent pas exactement les mêmes produits. Les adduits glutathion potentiels sont également retrouvés dans les deux cas, avec cette fois un décalage de  $m/z$  42 par rapport aux fragmentations observées dans le cas de la modification par l'enzyme J (écart entre la masse du cyclo(Ala-Glu) produit par la CDPS 37 et celle du cyclo(Leu-Glu) produit par les CDPS 36 et 44). Mais des produits probablement déshydrogénés apparaissent avec l'association 36+I, alors qu'ils ne sont pas présents avec l'association 44+P.

La CDPS 47 produisant du cyclo(Trp-Trp) a été associée avec la méthyltransférase R. Cette association permet de mettre en évidence l'apparition d'un composé potentiellement méthylé bien que cette transformation soit peu efficace (3 % de conversion) malgré la surexpression de la méthyltransférase observé dans la fraction soluble (Figure 76).

Dans tous les cas, excepté celui de la méthyltransférase R, les modifications proposées ne peuvent pas être attestées avec une simple étude LC-MS/MS. Les fragmentations obtenues ne permettent ni de confirmer la nature, ni d'identifier la localisation de la modification. Pour la méthyltransférase R, la fragmentation et certaines données de la littérature publiées concomitamment à l'étude de cette enzyme permettent de proposer une localisation de la méthylation. Des études complémentaires permettent de proposer le tropisme qui guide la réaction.

### 3. Le cas de la méthyltransférase R

Dans le cas de la méthyltransférase R, une analyse poussée de la fragmentation MS<sup>2</sup> et MS<sup>3</sup> du composé méthylé ( $m/z$  387) a permis d'émettre une hypothèse sur la localisation de la méthylation. Les pics visibles dans la fragmentation du cyclo(Trp-Trp) ont été attribués grâce à des données de la littérature<sup>62</sup> (Figure 81 et Figure 82). Les ions correspondant à un fragment contenant le cycle dicétopipérazine présentent tous un incrément de masse de 14 uma (+CH<sub>3</sub> -H), alors que les fragments correspondant à la chaîne latérale du Trp restent inchangés ( $m/z$  130). Ceci semble indiquer que la méthylation a lieu sur le cycle DKP.

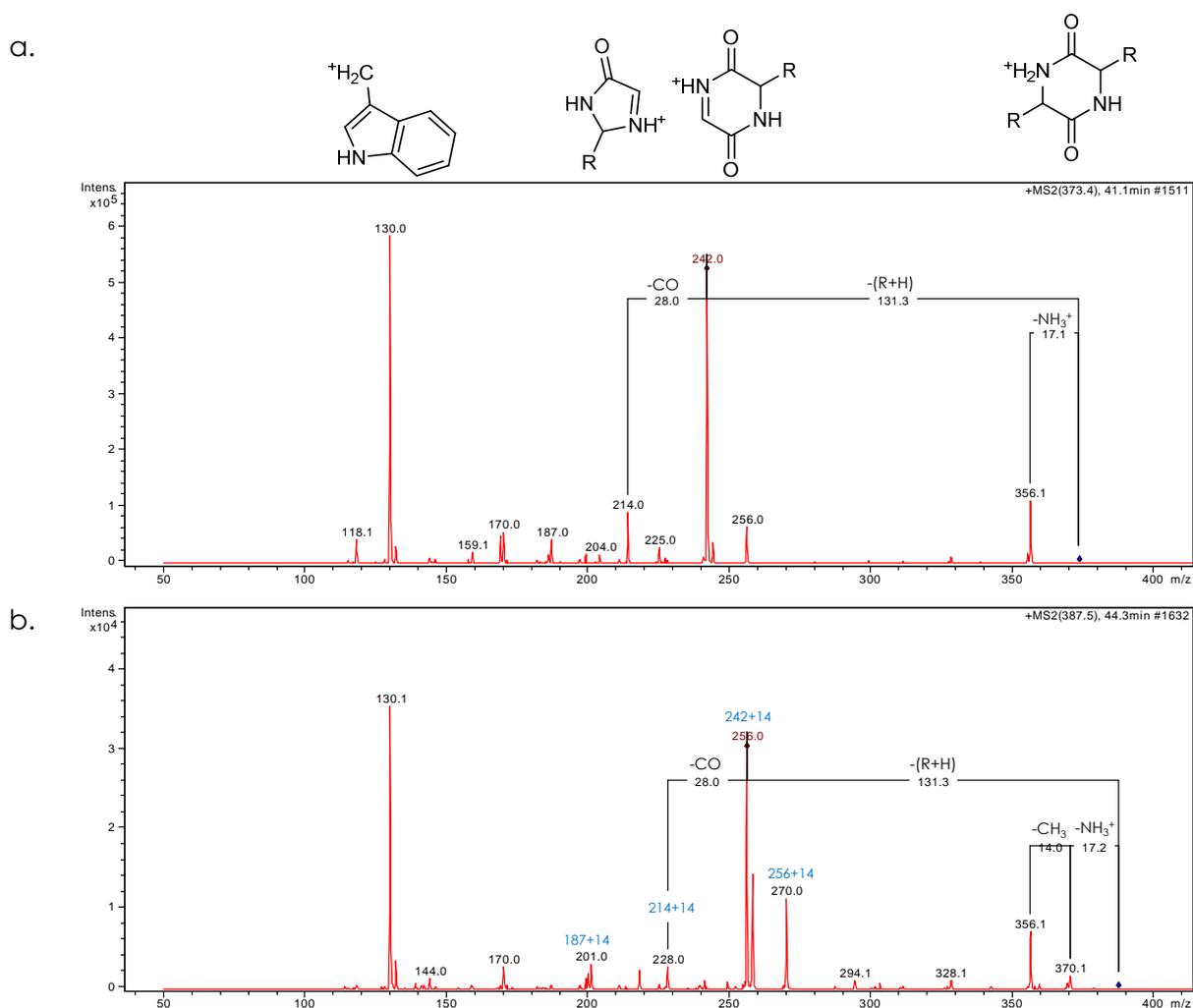
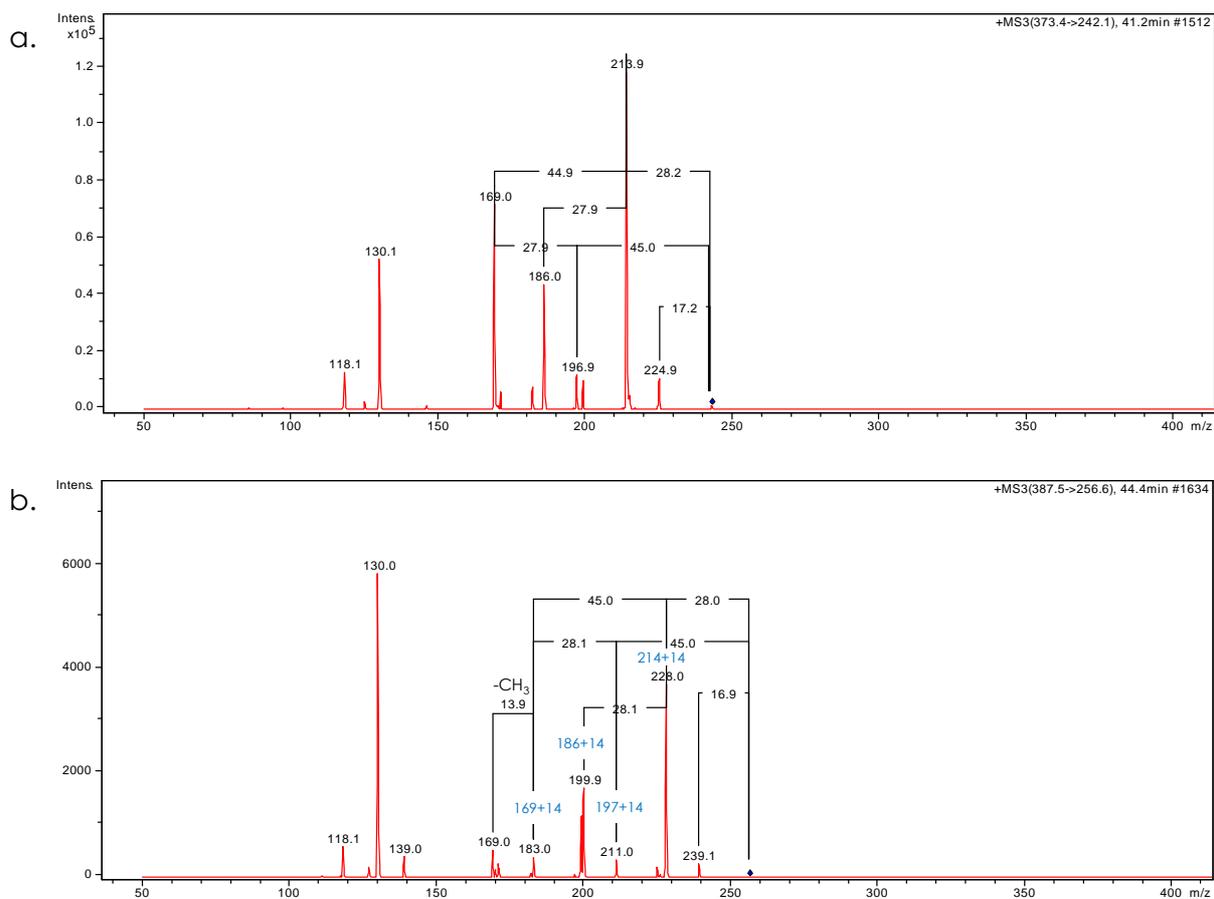


Figure 81 : Fragmentation MS<sup>2</sup> du cyclo(Trp-Trp) (a.) et cyclo(Trp-Trp) méthylé (b.)



**Figure 82 : Fragmentation MS<sup>3</sup> (373→242) du cyclo(Trp-Trp) (a.) et MS<sup>3</sup> (387→256) du cyclo(Trp-Trp) méthylé (b.)**

Par ailleurs, afin de tester la capacité de l'enzyme R à méthyler d'autres cyclodipeptides, des bactéries co-exprimant la méthyltransférase R avec des CDPS différentes de la CDPS 47 ont été cultivées. L'activité de l'enzyme R a été testée sur les cyclodipeptides produits par les CDPS AlbC de *S. noursei* et Rv2275 de *M. tuberculosis* selon le même protocole que précédemment. Les composés ayant un  $m/z$  correspondant au  $m/z$  d'un cyclodipeptide produit par la CDPS incrémenté de 14 uma ont été spécifiquement recherchés (cf. Annexe électronique 2). Dans les surnageants de cultures de bactéries co-exprimant Rv2275 et la méthyltransférase R, un composé correspondant au cyclo(Tyr-Tyr) méthylé est retrouvé. Les ions immonium Tyr méthylés et non méthylés sont alors identifiés sur la fragmentation MS<sup>2</sup> du composé. Dans le cas de l'association de AlbC avec l'enzyme R, la méthylation du cyclo(Phe-Phe) et du cyclo(Phe-Leu) a également lieu. Le dernier est retrouvé à l'état de traces alors qu'il est le cyclodipeptide majoritairement produit. Ceci indique que la méthyltransférase R est active sur d'autres cyclodipeptides que le cyclo(Trp-Trp) mais qu'elle semble avoir une préférence pour les substrats aux chaînes aromatiques. La fragmentation de ces deux

cyclodipeptides méthylés montre notamment un ion immonium Phe méthylé dans le cas du cyclo(Phe-Phe) ( $m/z$  134), et un ion immonium Leu méthylé dans le cas du cyclo(Phe-Leu) ( $m/z$  100). Cela suggère que la méthyltransférase agit préférentiellement du côté de la chaîne latérale la moins encombrée. En regroupant les données de fragmentations obtenues sur les différents cyclodipeptides, nous pouvons faire l'hypothèse que la méthylation a lieu sur le cycle dicétopipérazine des cyclodipeptides avec des chaînes aromatiques, avec une régiosélectivité pour la chaîne la moins encombrée dans le cas d'un hétérocyclodipeptide. La position de méthylation la plus probable est alors sur l'un des azotes du cycle dicétopipérazine.

Cette hypothèse a été confirmée par un article<sup>348</sup> publié concomitamment aux études menée sur la méthyltransférase R. Dans cet article, Marahiel *et al.* mettent en évidence la voie de biosynthèse du cyclo(Trp-Trp) simplement et doublement méthylé chez *Actinosynema mirum*. Cette voie fait intervenir deux enzymes, la CDPS Amir\_4627, homologue à la CDPS 47 (cf. Figure 56) et la méthyltransférase Amir\_4628, homologue à la méthyltransférase R (46 % d'identité de séquence). La méthylation simple et double qu'ils observent a lieu sur les azotes du cycle. Par ailleurs, lorsque l'activité de méthylation est testée sur les cyclodipeptides produits par AlbC, quatre cyclodipeptides contenant tous un résidu Phe se sont révélés substrats ; ce sont par ordre d'efficacité de méthylation, le cyclo(Phe-Phe), le cyclo(Phe-Tyr), le cyclo(Phe-Met) et le cyclo(Phe-Leu). De plus, les seuls cyclodipeptides pour lesquels une double méthylation a été observée sont des composés dont les deux chaînes latérales sont aromatiques ((cyclo(Phe-Phe) et cyclo(Phe-Tyr)). Ces résultats sont en accord avec nos résultats qui montraient une préférence de la méthyltransférase pour les cyclodipeptides avec des chaînes aromatiques. En revanche, Marahiel *et al.* en concluent que la méthylation a lieu sur l'azote en  $\alpha$  de la chaîne aromatique, ce qui ne correspond pas à nos observations, la méthylation étant portée par l'ammonium du résidu Leu dans le cas du cyclo(Phe-Leu) méthylé.

Ainsi, pour la méthyltransférase R, une analyse poussée de la fragmentation du cyclo(Trp-Trp) méthylé, étayée par les données issues de la caractérisation biochimique d'une enzyme homologue par une autre équipe, a permis de confirmer la *N*-méthylation du cycle dicétopipérazine. Dans tous les autres cas, des analyses supplémentaires sur les composés formés sont nécessaires pour réaliser une caractérisation structurale complète. Dans le cadre de mon travail de thèse, il n'était pas envisageable de caractériser les composés issus de l'activité de toutes les enzymes décrites ci-dessus. Nous avons alors choisi de concentrer nos efforts sur les enzymes des voies dont les CDPS synthétisent du cyclo(Ala-Glu). Ces voies présentent l'avantage d'être relativement simples comparativement à la voie associée à la CDPS 19, car elles ne sont constituées que d'une CDPS et d'une enzyme de modification (la FMO associée à CDPS 38 n'ayant pas montré d'activité). De plus, le cyclo(Ala-Glu) est un cyclodipeptide possédant une chaîne latérale polaire et chargée et aucune modification sur ce type d'acide aminé n'a encore été décrite, pour des voies dépendant des CDPS, dans la littérature.

### III. Enzymes de modification du cyclo(Ala-Glu) : activité *in vitro* et caractérisation structurale des produits

En collaboration avec l'équipe du professeur Olivier Lequin (Laboratoire de Biomolécules, Université Pierre et Marie Curie, Paris) et celle du professeur Gérard Bolbach (Plateforme de spectrométrie de masse et protéomique, Université Pierre et Marie Curie, Paris), nous avons décidé de réaliser la caractérisation structurale des produits formés par les OG J, K, M, N O (qui sont respectivement associées aux CDPS 37, 38, 39, 40 et 43 produisant du cyclo(Ala-Glu)), en combinant des approches de spectrométrie de masse à haute résolution (HRMS) et de résonance magnétique nucléaire (RMN). L'analyse RMN des composés nécessite la purification d'environ 1 mg de chaque composé à identifier. Nous avons donc décidé de produire ces cyclodipeptides modifiés *in vitro* par réaction enzymatique et de les purifier par HPLC préparative. Le passage à *l'in vitro* a été envisagé car cette approche permet en général d'obtenir des produits en plus grande quantité et meilleure pureté que par une approche *in vivo*. Cette production par voie enzymatique nécessite la purification des enzymes de modification. Il s'avère que les enzymes J, K, M, N O sont produites en grande quantité (Figure 75), que le cyclo(Ala-Glu) est disponible commercialement et que l'activité des OG est bien décrite dans la littérature. Cependant, dans un premier temps, nous avons dû reproduire la réaction de transformation enzymatique *in vitro*. A cette occasion, nous souhaitons vérifier notamment la dépendance de l'activité de l'enzyme au Fe(II) et au 2-oxoglurate, car l'étude bioinformatique a relevé que certains domaines de l'enzyme se rapprochaient de domaine conservé chez les IPNS, qui sont des enzymes proches des (Fe(II)/2OG)DO mais indépendantes du 2-oxoglutarate.

#### 1. Reproduction *in vitro* de l'activité enzymatique d'une OG

Afin de reproduire *in vitro* l'activité de modification opérée par les OG des voies à CDPS produisant du cyclo(Ala-Glu), l'une de ces enzymes a été purifiée.

##### 1.1 Purification de l'enzyme K

Pour faciliter la purification, une étiquette 6×His est ajoutée à l'extrémité N-terminale de la séquence des enzymes de modification J, K, M, N et O (cf. Matériels et méthodes III.1.3 Ajout d'une étiquette 6-Histidine). Ensuite, nous avons étudié si cette étiquette ne modifiait pas l'activité des enzymes, en réalisant un test *in vivo* comme précédemment lors de la coexpression avec la CDPS correspondante (cf. II.2 Identification de nouvelles activités de modification). Les profils chromatographiques obtenus sont identiques à ceux obtenus lorsque les enzymes sont exprimées sans étiquette 6×His. L'activité de ces enzymes n'est donc pas modifiée par l'ajout d'une étiquette 6×His à l'extrémité N-terminale de l'enzyme.

Parmi les enzymes J, K, M et N, l'enzyme K est choisie pour être purifiée car l'analyse des contenus protéiques des culots bactériens obtenus à l'issue du premier criblage *in vivo* a montré qu'elle était produite en quantité et de manière soluble (Figure 75).

L'enzyme K a été purifiée au laboratoire en utilisant une IMAC chargée en nickel (cf. Matériels et méthodes. VIII.2 Purification des dioxygénases dépendant du Fe(II) et du 2-oxoglutarate (réalisée par Emmanuel Favry), Figure 83). Après dialyse et concentration, 1,8 ml de protéine purifiée à environ 36 mg/ml sont obtenues soit 67 mg de protéine à partir d'un litre de culture. (Figure 84). La pureté, estimée par gel SDS-PAGE et coloration au bleu de Coomassie montre que l'enzyme K constitue plus de 95 % du mélange protéique.

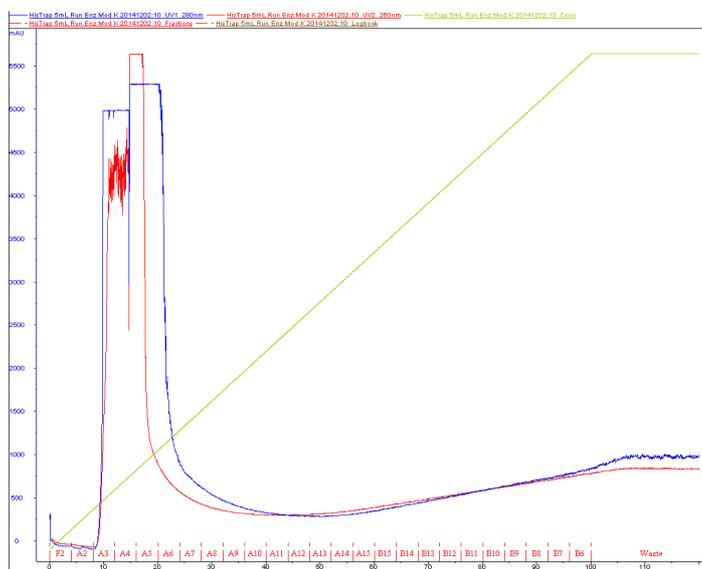


Figure 83 : Purification de la protéine K sur une IMAC chargée en nickel

Chromatogrammes enregistrés à 280 nm (bleu) et 260 nm (rouge) obtenus lorsque l'extrait soluble est purifié par chromatographie d'affinité. L'échantillon a été chargé sur une colonne HisTrap HP de 5 ml (GE-Healthcare BioSciences) et la protéine a été éluée avec un gradient linéaire d'imidazole de 20mM à 1 M (en vert).

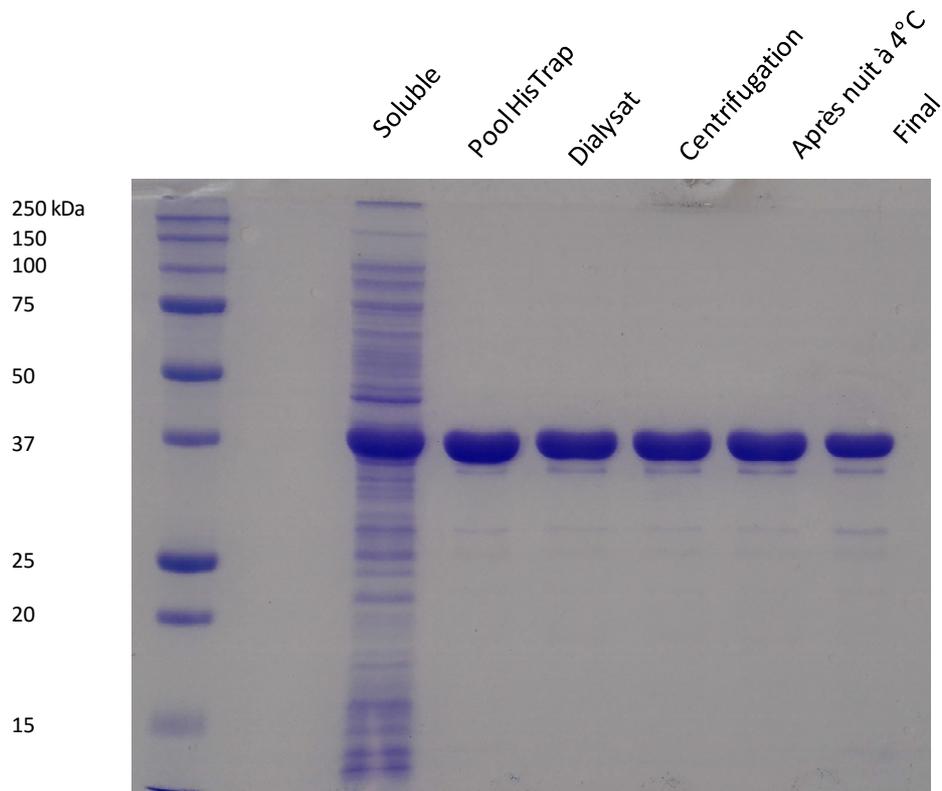


Figure 84 : Analyse SDS-PAGE des différentes étapes de la purification

L'extrait protéique soluble a été purifié sur colonne d'affinité HisTrap de 5 ml (GE-Healthcare-BioSciences), les fractions contenant la protéine ont été rassemblées et dialysées. A l'issue de la dialyse, la protéine purifiée a été centrifugée, avant d'être stockée une nuit à 4°C puis concentrée avec une unité de concentration de type Amicon (seuil 10000 Da).

## 1.2 Caractérisation de l'activité *in vitro*

### 1.2.1 Produits obtenus *in vitro*

Le test d'activité réalisé est inspiré de celui décrit par Mazmouz *et al.*<sup>396</sup> Il met en présence l'enzyme K purifiée, le substrat le cyclo(Ala-Glu), et des éléments décrits comme nécessaires à la réaction, notamment le Fe<sup>2+</sup> et le 2-oxoglutarate (cf. Matériel et méthodes. IX.2 Dioxygénases dépendant du Fe(II) et du 2-oxoglutarate). Il conduit à la production de six produits numérotés de PB1 à PB6 (Figure 85 et Tableau 28).

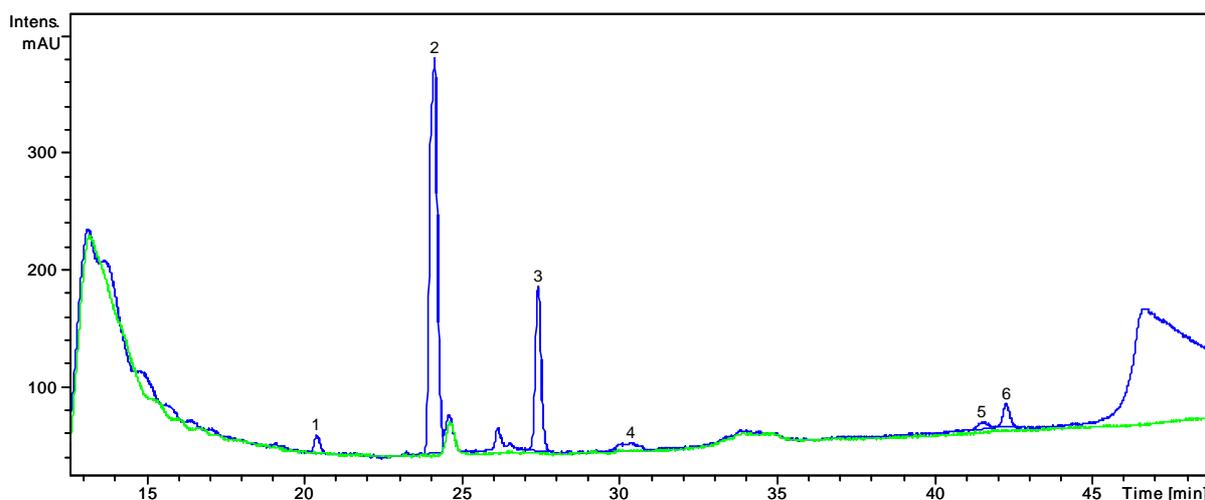


Figure 85 : Analyse LC-MS/MS du milieu réactionnel en présence (en bleu) et en l'absence (en vert) de cyclo(Ala-Glu)

Chromatogramme enregistré à 241 nm obtenu lorsque 50  $\mu$ l du milieu réactionnel sont injectés sur une colonne Hypercarb (ThermoScientific) et élué avec le gradient linéaire 1 (cf. matériel et méthode). Le pic non numéroté à 26 min correspond au cEA non consommé.

Tableau 28 : Produits de la réaction catalysée par l'enzyme K sur le cyclo(Ala-Glu)

Nom	$t_R$ (min)	$m/z$ majoritaire	Remarque
PB1	20,4	119	Acide succinique
PB2	24,2	199	IJ18
PB3	27,4	199	Nouveau composé
PB4	30,3	217	Nouveau composé
PB5	41,4	?	
PB6	42,2	199	Nouveau composé

Sur les six produits identifiés, seulement deux sont connus. Le composé PB1 correspond à l'acide succinique, sous-produit de la réaction identifié grâce à un standard. Il provient de la dégradation, lors de la réaction du 2-oxoglutarate en  $CO_2$  et en acide succinique. Cette identification a été confirmée par la comparaison de son temps de rétention à celui d'un standard chimique. Le composé PB2, produit majoritaire, correspond d'un point de vue du temps de rétention et de la fragmentation au composé IJ18 identifié *in vivo*. De nouveaux composés non détectés *in vivo* apparaissent *in vitro*. Il pourrait s'agir de la forme libre des composés retrouvés sous forme d'adduit avec le glutathion ou de composés non produits *in vivo*. L'apparition de ces produits est spécifique de l'activité de l'enzyme et de la présence de cyclo(Ala-Glu) (Figure 85).

### 1.2.2 Etude cinétique préliminaire

L'apparition de ces six composés a été suivie depuis l'initiation de la réaction puis régulièrement pendant 1 h (Figure 86). 15 % du cyclo(Ala-Glu) disparaît en une minute puis, sa consommation est régulière jusqu'à atteindre 85 % du cyclo(Ala-Glu) initiale au bout de 1 h. Parmi les différents composés apparaissant, le composé PB2 est le composé majoritaire à partir de 10 min et son apparition semble relativement linéaire dans le temps. Le composé PB3 apparaît rapidement dès la première minute et sa concentration dans le milieu réactionnel reste constante au cours de l'heure observée. Enfin les composés PB4, PB5 et PB6 sont minoritaires et apparaissent lentement.

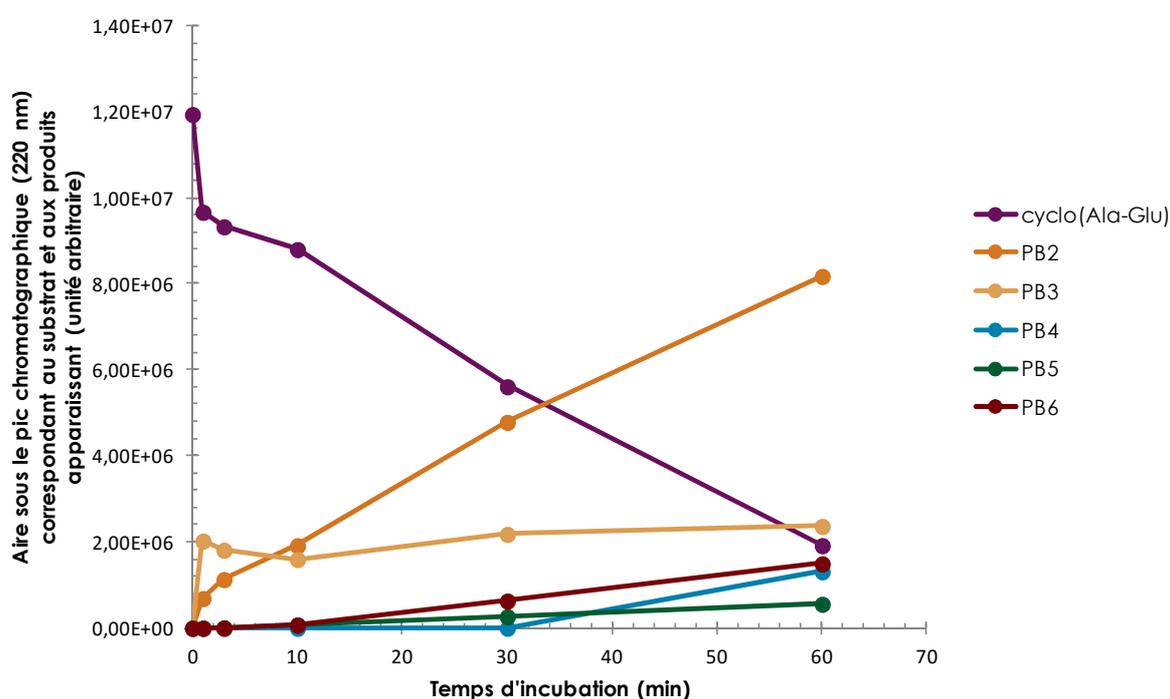


Figure 86 : Suivi pendant 60 min de la consommation du cyclo(Ala-Glu) et de la formation des composés PB2, PB3, PB4, PB5 et PB6 catalysés par la dioxygénase K

Hepes 50 mM pH 7,5 ; DTT 1 mM ; L-ascorbate 2 mM ; 2-oxoglutarate 2 mM ; FeSO4 0,05 mM ; catalase 0,1 mg/ml ; cyclo(Ala-Glu) 1 mM ; dioxygénase K 5  $\mu$ M ; Incubation à 30 °C. Analyse de 50  $\mu$ l de milieu réactionnel par HPLC sur Hypercarb (220 nm).

### 1.2.3 Dépendance aux cofacteurs

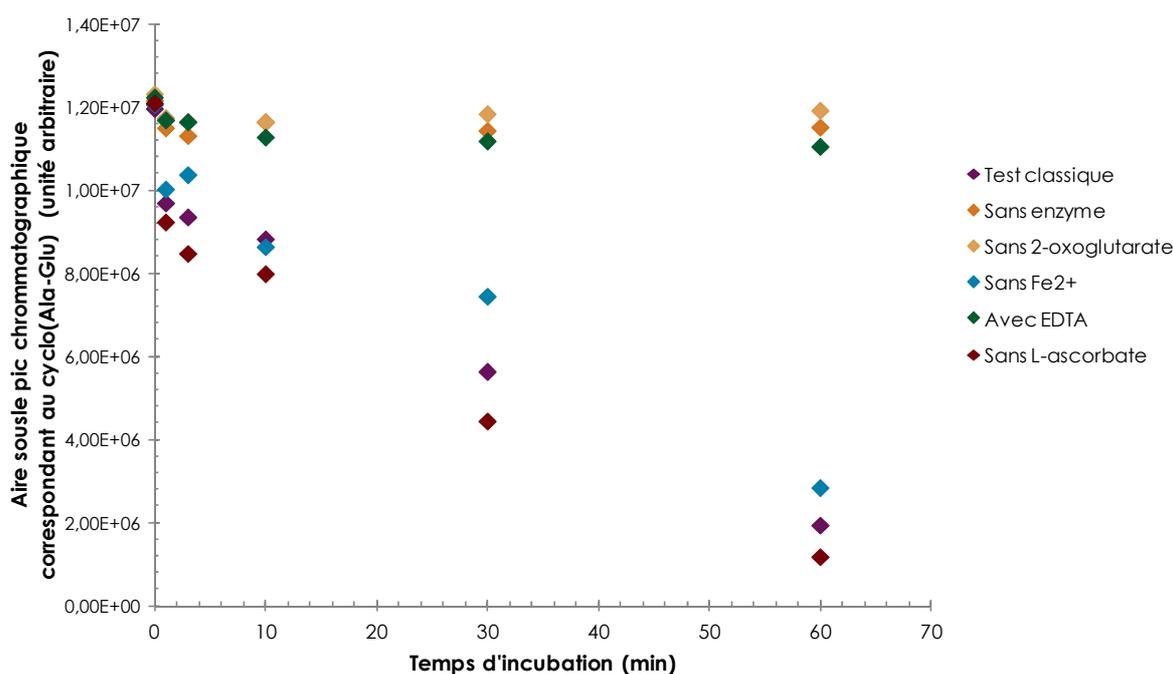
Nous nous sommes ensuite intéressés à la dépendance de l'activité de l'enzyme vis-à-vis des éléments introduits dans le test d'activité et notamment à la dépendance aux cofacteurs, celle-ci étant une signature de l'appartenance à la famille des OG. La catalase et le DTT sont décrits comme ayant une action protectrice sur

l'enzyme tandis que le L-ascorbate est réputé pour augmenter grandement l'activité des OG<sup>420</sup>. Différentes conditions expérimentales détaillées dans le Tableau 29 ont été testées. La consommation du cyclo(Ala-Glu) a été quantifiée en mesurant l'aire sous la courbe du pic correspondant sur le chromatogramme enregistré à 220 nm lorsque 50 µl de milieu réactionnel ont été analysés par HPLC (Figure 87).

**Tableau 29 : Conditions testées pour mesurer l'effet de la présence de cofacteurs, d'additifs et d'inhibiteurs**

Condition	Composants du milieu réactionnel					Activité <sup>1</sup>
	Enzyme 5 µM	2-oxoglutarate 2 mM	Fe <sup>2+</sup> 0,05 mM	EDTA 25 mM	L-ascorbate 2 mM	
Test classique	+	+	+	-	+	84 %
Sans enzyme	-	+	+	-	+	5 %
Sans 2-oxoglutarate	+	-	+	-	+	3 %
Sans Fe <sup>2+</sup>	+	+	-	-	+	77%
Avec EDTA	+	+	+	+	+	10%
Sans L-ascorbate	+	+	+	-	-	90 %

<sup>1</sup> Proportion de cyclo(Ala-Glu) consommé à t=60min



**Figure 87 : Effets de la présence de cofacteurs, d'additifs et d'inhibiteurs sur la consommation de cyclo(Ala-Glu) par la dioxygénase K**

Hepes 50 mM pH 7,5 ; DTT 1 mM ; cyclo(Ala-Glu) 1 mM ; Les différentes conditions testées sont résumées dans le Tableau 29 ; Incubation à 30 °C. Analyse de 50 µl de milieu réactionnel par HPLC sur Hypercarb (220 nm).

Cette étude révèle que l'enzyme K dépend strictement de la présence du 2-oxoglutarate. L'enzyme dépend également de la présence de  $Fe^{2+}$ , puisque lorsque ce dernier est complexé par un excès d'EDTA, aucune activité n'est observée. Une activité résiduelle est observée lorsque le milieu réactionnel n'est pas supplémenté avec du  $Fe^{2+}$ . Il semblerait donc que l'enzyme soit purifié en partie du moins avec son cofacteur ferreux. Contrairement à ce qui était attendu, l'absence de L-ascorbate semble améliorer quelque peu l'efficacité de la réaction, tout au moins au terme du temps observation.

L'enzyme K purifiée est donc active *in vitro* et elle produit de l'acide succinique, du IJ18 (=PB2) comme produit majoritaire et au moins trois autres produits minoritaires (PB3, PB4 et PB6). L'apparition de ces produits et la consommation du substrat cyclo(Ala-Glu) est dépendante de la présence de 2-oxoglutarate et de la disponibilité du  $Fe^{2+}$ . Ainsi cette enzyme appartient bien à la famille des OG et non aux IPNS qui sont indépendantes du 2-oxoglutarate. Nous souhaitons désormais élucider les structures des composés dont la formation est catalysée par l'enzyme K *in vivo* et *in vitro*, en combinant une approche HRMS et RMN.

## 2. Elucidation structurale des dérivés du cyclo(Ala-Glu)

L'élucidation structurale des dérivés du cyclo(Ala-Glu) modifié par l'enzyme K est réalisée en combinant les données obtenues lors de l'analyse de ces composés par spectrométrie de masse à haute résolution et par résonance magnétique nucléaire.

### 2.1 Données HRMS

La collecte de données  $MS^n$  à haute résolution sur les composés à identifier permet de mesurer la masse exacte des ions moléculaires et leurs ions produits. Ceci permet de remonter non seulement à la formule brute de la molécule mais également à celle de ses fragments, donnant ainsi des informations d'ordre structurale. A ce stade de l'étude, nous pensons que l'analyse par  $MS^n$  à haute résolution pourrait permettre de proposer une structure pour les molécules IJ18(=PB2), IJ19, IJ20, PB3, PB4 et PB6.

Les fractions correspondantes à ces composés ont été collectées en sortie de colonne HPLC. 100 $\mu$ l de surnageant de culture de bactéries *E. coli* co-exprimant la CDPS 38 et l'enzyme de modification K ont été injectés pour collecter les produits IJ18, IJ19 et IJ20. Pour les produits PB2, PB3, PB4 et PB6, la bioconversion a été obtenue en utilisant les conditions classiques du test *in vitro* décrit précédemment dans lequel la concentration initiale en substrat a été portée à 4 mM et l'incubation a été prolongée jusqu'à 14 h.

Les fractions ont ensuite été analysées en injection directe en utilisant un spectromètre de masse ESI-Orbitrap dans le laboratoire de Gérard Bolbach. Le cyclo(Ala-Glu) commercial a également été analysé afin de disposer d'une référence pour aider à l'identification des composés. Les masses à haute résolution mesurées pour les ions majoritaires et les formules brutes associées de chaque composé sont résumées dans le tableau

suivant. Dans certains cas, il n'est pas possible de discerner quel est l'ion moléculaire parmi les ions majoritaires observés.

**Tableau 30 : Masse haute résolution et formule brute probable pour les dérivés du cyclo(Ala-Glu)**

Composé	Ions précurseur		
	$m/z^1$	Formule brute <sup>2</sup>	Delta ppm
cyclo(Ala-Glu)	201,08697	C <sub>8</sub> H <sub>13</sub> O <sub>4</sub> N <sub>2</sub>	-0,067
IJ18	217,08185	C <sub>8</sub> H <sub>13</sub> O <sub>5</sub> N <sub>2</sub>	-0,221
	199,07125	C <sub>8</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	-0,419
IJ19	506,15520	C <sub>18</sub> H <sub>28</sub> O <sub>10</sub> N <sub>5</sub> S	-0,120
	272,55478	?	
IJ20 <sup>3</sup>	506,16	C <sub>18</sub> H <sub>28</sub> O <sub>10</sub> N <sub>5</sub> S	
	272,56	?	
PB2	217,08209	C <sub>8</sub> H <sub>13</sub> O <sub>5</sub> N <sub>2</sub>	0,885
	199,07149	C <sub>8</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	0,787
PB3	199,07169	C <sub>8</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	1,792
PB4	262,14038	C <sub>10</sub> H <sub>20</sub> O <sub>5</sub> N <sub>3</sub>	2,414
	217,08239	C <sub>8</sub> H <sub>13</sub> O <sub>5</sub> N <sub>2</sub>	-2.266
PB6	199,07176	C <sub>8</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	2.093

<sup>1</sup>Il s'agit d'ion monochargé sauf pour  $m/z$  272,56 qui est dichargé. Les ions peu intenses sont en gris. <sup>2</sup>Les calculs de formules brutes ont été réalisés avec le logiciel Thermo Xcalibur 2.1 (Qual Browser). Les formules brutes et les fragments proposés sont des hypothèses effectuées à partir des calculs réalisés par le logiciel. <sup>3</sup>Les masses hautes résolutions des deux ions précurseurs majoritaires ne sont pas disponibles car le spectre MS<sup>1</sup> n'a pas été enregistré pour IJ20 ; il était équivalent à celui de IJ19. La formule brute correspondante à  $m/z$  506,16 (en gris) est déduite de celle déterminée pour IJ19.

Les résultats précédemment obtenus en LC-MS/MS au laboratoire, nous avaient amenés à conclure que le composé IJ18 (issu du surnageant de culture) et le composé PB2 (issue de la biotransformation *in vitro*) étaient identiques. Les analyses de masse à haute résolution font apparaître que l'ion moléculaire du PB2 et du IJ18 serait à  $m/z$  217 et non  $m/z$  199. L'ion à  $m/z$  217 est vraisemblablement particulièrement fragile et il se peut qu'une déshydratation ait lieu en source. La fraction provenant de la réaction *in vitro* est plus concentrée que celle provenant du surnageant de culture, cela explique que l'ion à  $m/z$  217 ne soit que peu visible pour le composé IJ18. En recherchant spécifiquement un ion à  $m/z$  217 dans les données de LC-MS/MS enregistrées au laboratoire, la présence de cet ion, bien qu'il fût en faible quantité, a pu être confirmée.

Des données de MS<sup>n</sup> ont également été collectées et sont regroupées en annexe 2. Néanmoins celles-ci ne permettent pas d'obtenir des informations d'ordre structural comme cela était espéré.

Si les analyses en HRMS n'ont pas permis d'élucider la structure des dérivés du cyclo(Ala-Glu), elles ont permis cependant de déterminer la masse haute résolution de leur ion moléculaire et donc leur formule brute. Pour élucider les structures de ces composés, une analyse par RMN est donc nécessaire.

## 2.2 Analyses RMN

Pour obtenir suffisamment de produit pour ces analyses, le test *in vitro* est amélioré afin de produire des quantités suffisantes de dérivés du cyclo(Ala-Glu). Afin d'optimiser la réaction de biotransformation catalysée par l'enzyme K, les concentrations en enzyme, en substrat et les temps d'incubation sont augmentés. Des concentrations en substrat de 5, 10 et 20 mM et des concentrations de 10 et 20  $\mu$ M en enzyme sont testées. La concentration en substrat ne peut pas être supérieure à 20 mM. En effet, la solution mère de cyclo(Ala-Glu) est préparée, en se plaçant à la limite de solubilité dans du DMSO (400 mM). Comme le DMSO peut dénaturer l'enzyme, la concentration finale en DMSO ne doit pas dépasser 5% du volume final, ce qui correspond à une concentration en cyclo(Ala-Glu) de 20 mM. En 21 h d'incubation, la totalité du cyclo(Ala-Glu) est consommée lorsque la concentration en enzyme est de 10 ou 20  $\mu$ M, quelle que soit la concentration en substrat testée. La condition retenue met en présence 10  $\mu$ M d'enzyme K et 20 mM de cyclo(Ala-Glu). La concentration de 2-oxoglutarate est augmentée de manière à se placer en excès molaire (1,2) par rapport au substrat. La réaction est réalisée dans 5 ml, ce qui permet de transformer 20 mg de cyclo(Ala-Glu). Les produits alors obtenus sont purifiés par HPLC préparative en conservant le gradient linéaire analytique puis lyophilisés. Les produits blancs obtenus sont ensuite analysés par RMN au Laboratoire des Biomolécules (UPMC, UMR 7203, Olivier Lequin & Isabelle Correia).

L'acquisition des données et l'attribution des signaux RMN sur le cyclo(Ala-Glu) et les composés PB2, PB3, PB4 et PB6 ont été réalisées en collaboration avec Isabelle Correia. Les signaux <sup>1</sup>H, <sup>13</sup>C et <sup>15</sup>N ont été attribués à partir d'expériences 1D <sup>1</sup>H, 1D <sup>13</sup>C, d'expériences 2D homonucléaires (<sup>1</sup>H-<sup>1</sup>H COSY) et hétéronucléaires (2D <sup>1</sup>H-<sup>13</sup>C HSQC, 2D <sup>1</sup>H-<sup>13</sup>C HMBC et 2D <sup>1</sup>H-<sup>15</sup>N HSQC). L'attribution des différents signaux est regroupée dans le Tableau 31 et les spectres 1D <sup>1</sup>H, 2D <sup>1</sup>H-<sup>13</sup>C HMBC ainsi qu'une sélection de spectres pertinents pour la détermination des différentes structures sont disponibles en annexe 3.

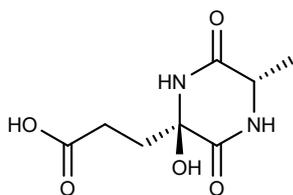
Le composé PB2 s'avère être un mélange de deux diastéréoisomères (2R,5S) et (2S,5S) du cyclo(Ala-Glu) hydroxylé sur le carbone C $\alpha$  de la chaîne Glu (Figure 88). Cette identification s'appuie notamment sur la disparition du signal du H $\alpha$  de Glu à  $\delta$ =3,88 ppm par rapport au spectre du cyclo(Ala-Glu) et l'apparition du signal à  $\delta$ =6,5 ppm correspondant au H de l'hydroxyle. Cette hypothèse est corroborée par l'apparition du carbone quaternaire à  $\delta$ =80,7 ppm pour PB2a et  $\delta$ =80,4 ppm pour PB2b, corrélant en HMBC avec les

protons H $\beta$ , H $\beta'$  de Glu, et le proton NH<sup>Ala</sup>, et correspondant au C $\alpha$  Glu sur lequel l'hydroxylation a eu lieu. L'expérience 2D <sup>1</sup>H-<sup>1</sup>H NOESY réalisée sur les composés PB2 a permis de différencier le diastéréoisomère (2R,5S) PB2a du diastéréoisomère (2S,5S) PB2b. Le spectre NOESY montre en effet la proximité spatiale des protons H $\alpha$  Ala et H $\beta$  Glu du composé PB2b.

La fraction PB3 est également un mélange de deux diastéréoisomères (5R,8S) et (5S,8S) d'un dérivé du cyclo(Ala-Glu) dans lequel la chaîne latérale du Glu est cyclisée sur le C $\alpha$  *via* une fonction ester (Figure 88). En effet, deux signaux caractéristiques du cyclo(Ala-Glu) ne sont pas retrouvés sur les composés PB3. Il s'agit du signal élargi à  $\delta=12,1$  ppm correspondant au proton de la fonction acide de la chaîne latérale Glu ainsi que le signal à  $\delta=3,88$  ppm du proton du C $\alpha$  Glu. Le carbone quaternaire C $\alpha$  Glu apparaît à  $\delta=89,9$  ppm pour PB3a et  $\delta=89,4$  ppm pour PB3b et corrèle par des couplages <sup>3</sup>J<sup>1</sup>H-<sup>13</sup>C avec le proton amide NH<sup>Ala</sup> et les protons H $\gamma$  de Glu. L'attribution de la configuration pour les composés PB3 a et b a été réalisée par analogie avec celle du composé PB2 car aucune corrélation de type NOE probante n'a été observée.

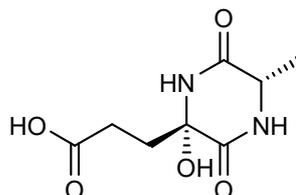
Le composé PB4 est un composé acyclique, correspondant à une ouverture du cycle dicétopipérazine. La présence d'un groupement amide primaire sur le résidu alanine permet de localiser l'ouverture du cycle au niveau de la liaison N-C $\alpha$  du résidu Glu et de proposer la structure ci-dessous (Figure 88). L'ouverture du cycle crée sur le C $\alpha$  Glu, une fonction carbonyle, ce qui explique l'important déblindage de ce carbone quaternaire ( $\delta=197,2$  ppm) qui corrèle en HMBC avec les protons H $\beta$  et H $\gamma$ .

Enfin le composé PB6 correspond au cyclo(Ala-Glu) déshydrogéné sur les carbones  $\alpha$  et  $\beta$  de la chaîne latérale du Glu (Figure 88), comme le montrent notamment l'absence du signal sur le spectre 1D <sup>1</sup>H correspondant au H $\alpha$  et la présence d'un seul H $\beta$  pour le Glu. De plus, le déplacement chimique du C $\beta$  Glu ( $\delta=109,1$  ppm) observé sur le spectre 1D <sup>13</sup>C est compatible avec l'apparition d'une double liaison. Le carbone quaternaire C $\alpha$  Glu apparaît ( $\delta=128,6$  ppm) dans la même gamme de déplacement chimique et corrèle par des couplages <sup>3</sup>J<sup>1</sup>H-<sup>13</sup>C aux protons H $\gamma$  et NH<sup>Ala</sup> (cf. spectre HMBC en annexe 3). La conformation de la double liaison (Z ou E) n'a pas pu être déterminée.



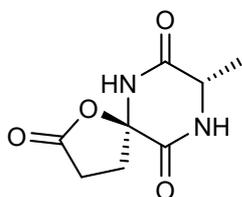
3-((2*R*,5*S*)-2-hydroxy-5-methyl-3,6-dioxopiperazin-2-yl)propanoic acid

PB2a (*R,S*) (45,2 %)



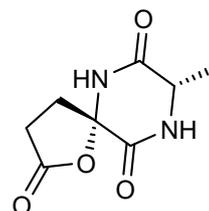
3-((2*S*,5*S*)-2-hydroxy-5-methyl-3,6-dioxopiperazin-2-yl)propanoic acid

PB2b (*S,S*) (54,8%)



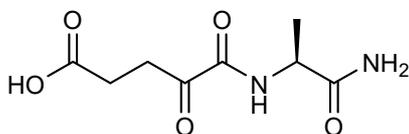
(5*R*,8*S*)-8-methyl-1-oxa-6,9-diazaspiro[4.5]decane-2,7,10-trione

PB3a (*R,S*) (50%)



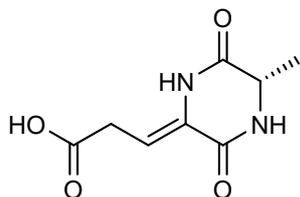
(5*S*,8*S*)-8-methyl-1-oxa-6,9-diazaspiro[4.5]decane-2,7,10-trione

PB3b (*S,S*) (50%)



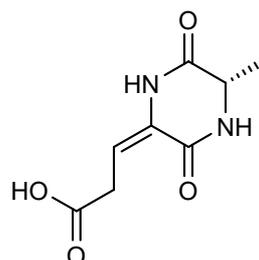
(*S*)-5-((1-amino-1-oxopropan-2-yl)amino)-4,5-dioxopentanoic acid

PB4



(*S,Z*)-3-(5-methyl-3,6-dioxopiperazin-2-ylidene)propanoic acid

PB6 Z



(*S,E*)-3-(5-methyl-3,6-dioxopiperazin-2-ylidene)propanoic acid

PB6 E

Figure 88 : Structures des composés PB2a, PB2b, PB3a, PB3b, PB4 et PB6 (*Z* ou *E*) déterminées par RMN

Tableau 31 : Attribution des signaux RMN des composés en solutions (DMSO-d<sub>6</sub>, 30 °C)

Composés	Attribution RMN		
	RMN <sup>1</sup> H (500 MHz, DMSO)	RMN <sup>13</sup> C (125 MHz, DMSO)	RMN <sup>15</sup> N (50 MHz, DMSO)
c[EA]	δ 12.1 (br s, 1H, εH Glu), δ 8.14 (s, 1H, H <sup>N</sup> Ala), δ 8.08 (s, 1H, H <sup>N</sup> Glu), δ 3.90 (q, J = 6.8, 1H, Hα Ala), δ 3.88 (q, J = 6.8, 1H, Hα Glu), δ 2.33, 2.28 (m, 2H, Hγγ' Glu), δ 1.94 (m, 1H, Hβ Glu), δ 1.86 (m, 1H, Hβ' Glu), δ 1.26 (d, J = 7.3, 3H, CH <sub>3</sub> β Ala)	δ 173.8 (Cδ Glu), δ 169.0 (C'Ala), δ 167.6 (C' Glu), δ 53.2 (Cα Glu), δ 49.7 (Cα Ala), δ 29.2 (Cγ Glu), δ 27.7 (Cβ Glu), δ 18.9 (Cβ Ala)	δ 121.1 (N <sup>H</sup> Ala) δ 115.7 (N <sup>H</sup> Glu)
Pb2a (2R, 5S) (45.2 %) Pb2b (2S, 2S) (54.8 %)	δ 12.0 (br s, 1H, εH Glu), δ 8.56 (s, 1H, H <sup>N</sup> Glu <b>2a</b> ), δ 8.49 (s, 1H, H <sup>N</sup> Glu <b>2b</b> ), δ 8.14 (d, J = 2.6, 1H, H <sup>N</sup> Ala <b>2b</b> ), δ 8.13 (s, 1H, H <sup>N</sup> Ala <b>2a</b> ), δ 6.5 (br s, 1H, OHα Glu), δ 4.01 (q, J = 6.8, 1H, Hα Ala <b>2a</b> ), δ 3.81 (dq, J = 6.8, J = 2.9, 1H, Hα Ala <b>2b</b> ), δ 2.17 (m, 4H, Hγγ' Glu <b>2a, 2b</b> ), δ 2.12 (m, 1H, Hβ Glu <b>2b</b> ), δ 2.10 (m, 1H, Hβ Glu <b>2a</b> ), δ 1.86 (m, 2H, Hβ' Glu <b>2a, 2b</b> ), δ 1.35 (d, J = 7.1, 3H, CH <sub>3</sub> β Ala <b>2b</b> ), δ 1.24 (d, J = 6.8, 3H, CH <sub>3</sub> β Ala <b>2a</b> )	δ 173.84 (Cδ Glu <b>2a</b> ), δ 174.78 (Cδ Glu <b>2b</b> ), δ 170.5 (C'Ala <b>2a</b> ), δ 170.0 (C'Ala <b>2b</b> ), δ 167.1 (C' Glu <b>2a</b> ), δ 166.7 (C' Glu <b>2b</b> ), δ 80.7 (Cα Glu <b>2a</b> ), δ 80.4 (Cα Glu <b>2b</b> ), δ 50.8 (Cα Ala <b>2b</b> ), δ 49.5 (Cα Ala <b>2a</b> ), δ 33.2 (Cβ Glu <b>2b</b> ), δ 32.4 (Cβ Glu <b>2a</b> ), δ 29.1 (Cγ Glu <b>2a, 2b</b> ), δ 20.4 (Cβ Ala <b>2b</b> ), δ 17.6 (Cβ Ala <b>2a</b> )	δ 135.3 (N <sup>H</sup> Glu <b>2b</b> ) δ 135.1 (N <sup>H</sup> Glu <b>2a</b> ) δ 118.2 (N <sup>H</sup> Ala <b>2b</b> ) δ 117.4 (N <sup>H</sup> Ala <b>2a</b> )
Pb3a (5R, 8S) (50 %) Pb3b (5S, 8S) (50 %)	δ 9.43 (s, 1H, H <sup>N</sup> Glu <b>3a</b> ), δ 9.38 (s, 1H, H <sup>N</sup> Glu <b>3b</b> ), δ 8.72 (br s, 1H, H <sup>N</sup> Ala <b>3b</b> ), δ 8.70 (s, 1H, H <sup>N</sup> Ala <b>3a</b> ), δ 4.17 (q, J = 6.9, 1H, Hα Ala <b>3a</b> ), δ 3.97 (dq, J = 7.2, J = 3.2, 1H, Hα Ala <b>3b</b> ), δ 2.82 (m, 1H, Hβ Glu <b>3a</b> ), δ 2.77 (m, 1H, Hβ Glu <b>3b</b> ), δ 2.74 (m, 4H, Hγγ' Glu <b>3a, 3b</b> ), δ 2.16 (m, 2H, Hβ' Glu <b>3a, 3b</b> ), δ 1.37 (d, J = 7.6, 3H, CH <sub>3</sub> β Ala <b>3b</b> ), δ 1.29 (d, J = 6.9, 3H, CH <sub>3</sub> β Ala <b>3a</b> )	δ 174.83, 174.78 (Cδ Glu <b>3a, 3b</b> ), δ 170.3 (C'Ala <b>3a</b> ), δ 169.9 (C'Ala <b>3b</b> ), δ 163.8 (C' Glu <b>3a</b> ), δ 163.2 (C' Glu <b>3b</b> ), δ 89.9 (Cα Glu <b>3a</b> ), δ 89.4 (Cα Glu <b>3b</b> ), δ 50.9 (Cα Ala <b>3b</b> ), δ 49.5 (Cα Ala <b>3a</b> ), δ 30.3 (Cβ Glu <b>3b</b> ), δ 29.5 (Cβ Glu <b>3a</b> ), δ 28.9, 28.8 (Cγ Glu <b>3a, 3b</b> ), δ 21.1 (Cβ Ala <b>3b</b> ), δ 17.3 (Cβ Ala <b>3a</b> )	δ 131.8 (N <sup>H</sup> Glu <b>3b</b> ) δ 131.6 (N <sup>H</sup> Glu <b>3a</b> ) δ 119.5 (N <sup>H</sup> Ala <b>3b</b> ) δ 117.9 (N <sup>H</sup> Ala <b>3a</b> )
Pb4	δ 8.31 (d, J = 7.7, 1H, H <sup>N</sup> Ala), δ 7.42 (s, 1H, H <sup>N</sup> Glu), δ 7.10 (s, 1H, H <sup>N</sup> Glu), δ 4.21 (quin., J = 7.4, 1H, Hα Ala), δ 3.0 (m, 2H, Hββ' Glu), δ 2.48 (t, J = 6.5, 2H, Hγγ' Glu), δ 1.28 (d, J = 7.0, 3H, CH <sub>3</sub> β Ala)	δ 197.2 (Cα Glu), δ 173.6 (Cδ Glu), δ 173.4 (C'Ala), δ 160.3 (C' Glu), δ 48.1 (Cα Ala), δ 32.0 (Cβ Glu), δ 27.5 (Cγ Glu), 18.0 (Cβ Ala)	δ 116.1 (N <sup>H</sup> Ala) δ 104.0 (N <sup>H2</sup> Glu)
Pb6	δ 10.16 (br s, 1H, H <sup>N</sup> Glu), δ 8.28 (s, 1H, H <sup>N</sup> Ala), δ 5.84 (t, J = 7.6, 1H, Hβ Glu), δ 4.06 (qd, J = 7.0, J = 2.0, 1H, Hα Ala), δ 3.21 (d, J = 7.6, 2H, Hγ Glu), δ 1.31 (d, J = 7.0, 3H, CH <sub>3</sub> β Ala)	δ 172.1 (Cδ Glu), δ 167.3 (C'Ala), δ 158.8 (C' Glu), δ 109.1 (Cβ Glu), δ 128.6 (Cα Glu), δ 50.3 (Cα Ala), δ 31.3 (Cγ Glu), 19.9 (Cβ Ala)	δ 118.2 (N <sup>H</sup> Ala) Not obs. (H <sup>N</sup> Glu)

La détermination de ces structures s'est appuyée sur les formules brutes proposées grâce aux données de HRMS. Lorsque deux ions moléculaires étaient proposés, seul l'un d'entre eux corrélait avec les données RMN. Les ions moléculaires de chaque composé sont donc observés aux  $m/z$  217 pour PB2a et PB2b,  $m/z$  199 pour PB3a et PB3b,  $m/z$  217 pour PB4 et  $m/z$  199 pour PB6.

L'enzyme K est donc responsable *in vitro* de la synthèse d'au moins six composés dérivés du cyclo(Ala-Glu) dont la structure a été déterminée. Dans toutes les structures moléculaires déterminées, la modification (hydroxylation, cyclisation, déshydratation ou ouverture oxydante) a lieu sur le carbone C $\alpha$  du Glu. L'enzyme K est donc régiosélective de cet atome du cycle. La présence de deux diastéréoisomères pour les composés de type PB2 et PB3 semble indiquer que l'enzyme n'est pas stéréospécifique, même s'il conviendrait de vérifier que l'épimérisation du centre stéréogène n'a pas lieu pendant les étapes de purification et de traitement de l'échantillon, notamment lors du traitement acide.

#### IV. Discussion, conclusion et perspectives

Ce chapitre s'est intéressé à la diversité des enzymes de modification impliquées dans les voies de biosynthèse dépendant des CDPS, avec comme objectif l'identification de nouvelles activités de modification de cyclodipeptides. Dans un premier temps, grâce à une analyse bioinformatique des environnements génomiques des CDPS étudiées au chapitre précédent, des gènes codant pour des enzymes susceptibles de faire partie d'une voie de biosynthèse de dicétopipérazines ont été identifiés. Puis, 18 enzymes ont été sélectionnées pour étudier leur activité de modification de cyclodipeptides à l'aide du protocole de criblage moyen-débit décrit dans le chapitre I. Sur ces 18 enzymes, 12 se sont révélées actives *in vivo* tandis que six n'ont pas montré d'activité de modification de cyclodipeptides dans nos conditions expérimentales. Les raisons de ce manque d'activité peuvent être multiples. Premièrement, les enzymes peuvent nécessiter des cofacteurs ou des co-enzymes qui ne sont pas disponibles chez *E. coli*. L'étude bioinformatique et bibliographique préalable nous a permis de retirer de notre étude les enzymes dont il était certain qu'elles ne pourraient pas fonctionner dans notre test. Cependant, les annotations de fonction ne permettent pas toujours de prédire correctement l'activité des enzymes et *a fortiori* les cofacteurs ou les co-enzymes nécessaires. Ainsi pour les FMO codées par les gènes à proximité de la CDPS 38 et 44, leur manque d'activité pourrait provenir de l'absence de partenaire protéique permettant de leur fournir le cofacteur flavinique réduit. Deuxièmement, l'appartenance à une voie de biosynthèse dépendant des CDPS n'est que supposée par l'étude bioinformatique et bibliographique. Ainsi, il se peut que notre test ait associé des CDPS avec des enzymes dont la fonction n'est pas de modifier les cyclodipeptides. A nouveau, l'exemple des FMO peut être repris pour illustrer ce cas de figure, d'autant plus que l'étude des synthénies avait montré que les FMO ne s'ajoutaient pas à la synthénie formée par la dioxygénase, la CDPS et le transporteur MFS. Enfin, d'une manière générale, notre stratégie qui ne consiste pas à rechercher les enzymes responsables d'une molécule naturelle isolée, mais qui part des environnements génomiques pour

identifier de nouvelles molécules montre peut-être quelque peu ses limites. En effet, dans cette approche, l'identification de nouvelles voies de biosynthèse repose uniquement sur le regroupement des gènes codant pour les enzymes de cette voie. Or, dans les voies dépendant des NRPS, il existe plusieurs cas dans lesquels les gènes d'une même voie de biosynthèse ne sont pas co-localisés sur le génome (cf. Etude bibliographique. IV.2 Voies de biosynthèse de dicétopipérazines dépendant des NRPS)

Néanmoins, malgré tous ces écueils, 12 enzymes ont montré des activités de biosynthèse dans notre test. Pour la méthyltransférase R, les données de fragmentation étaient suffisamment riches pour nous permettre de postuler l'emplacement de la méthylation, qui a été confirmée par des données de la littérature publiées concomitamment à nos expériences. Dans les autres cas, les données de fragmentation des composés obtenus, ont uniquement permis de faire une hypothèse sur le type de modification ayant eu lieu sur le cyclodipeptide. L'élucidation complète de la structure de certains de ces cyclodipeptides modifiés (IJ18 = PB2) a d'ailleurs révélé que certaines de ces hypothèses étaient inexactes. En effet, il avait été postulé que le composé IJ18 résultait d'une déshydrogénation du cyclo(Ala-Glu), car l'analyse avait été réalisée sur l'ion fragment  $m/z$  199, qui était majoritaire et non sur l'ion moléculaire  $m/z$  217, qui était peu intense du fait vraisemblablement de sa fragilité en source. D'autres déshydrogénations du même type avaient été supposées pour certains composés (IJ11, IJ12, IJ13, IJ14). À la lumière du cas du composé IJ18=PB2, une nouvelle analyse des données de LC-MS/MS a visé à rechercher spécifiquement des ions moléculaires peu intenses. Dans le cas des composés IJ11 ( $m/z$  257) et IJ12 ( $m/z$  257), aucun ion à  $m/z$  275 (257+18), n'a été retrouvé. Pour les composés IJ13 ( $m/z$  241) et IJ14 ( $m/z$  241), un faible signal à  $m/z$  259 (241+18) est détecté mais sort à peine du bruit de fond. Ainsi, en fonction de la position et de la stabilité des groupements chimiques apportés par les enzymes de modifications, l'élucidation des structures des cyclodipeptides modifiées par LC-MS/MS peut être envisageable ou se révéler particulièrement ardue.

Les difficultés liées à l'élucidation de la structure des cyclodipeptides modifiés par les dioxygénases nous ont amené à concentrer notre étude sur les dioxygénases agissant sur le cyclo(Ala-Glu) et, plus précisément, sur l'enzyme K associée à la CDPS 38. La dioxygénase K fait partie de la famille des dioxygénases dépendantes du Fe(II) et du 2-oxoglutarate. *In vitro*, cette enzyme est responsable de la consommation du cyclo(Ala-Glu) et de la formation d'au moins six composés : les diastéréoisomères PB2a et PB2b, les diastéréoisomères PB3a et PB3b, PB4 et PB6. L'enzyme K est régiosélective du C $\alpha$  de l'acide aminé Glu mais ne semble pas stéréosélective. La diversité des dérivés du cyclo(Ala-Glu) (hydroxylé, déshydrogéné, cyclisé, linéaire) obtenus nous interroge sur la formation de ces composés : les modifications observées sont-elles toutes catalysées par l'enzyme ? Sont-elles réalisées séquentiellement ou indépendamment l'une des autres ? Sont-elles issues d'une évolution spontanée d'un des composés transformés par l'enzyme ? De plus, lequel de ces composés recèle une pertinence biologique ? Pour répondre à toutes ces questions, une étude biochimique plus exhaustive de l'enzyme K doit être menée afin de déterminer notamment l'ordre d'apparition des différents composés et leur relation. De plus, des

études complémentaires permettront de tester la capacité de l'enzyme K à prendre en charge d'autres substrats cyclodipeptides.

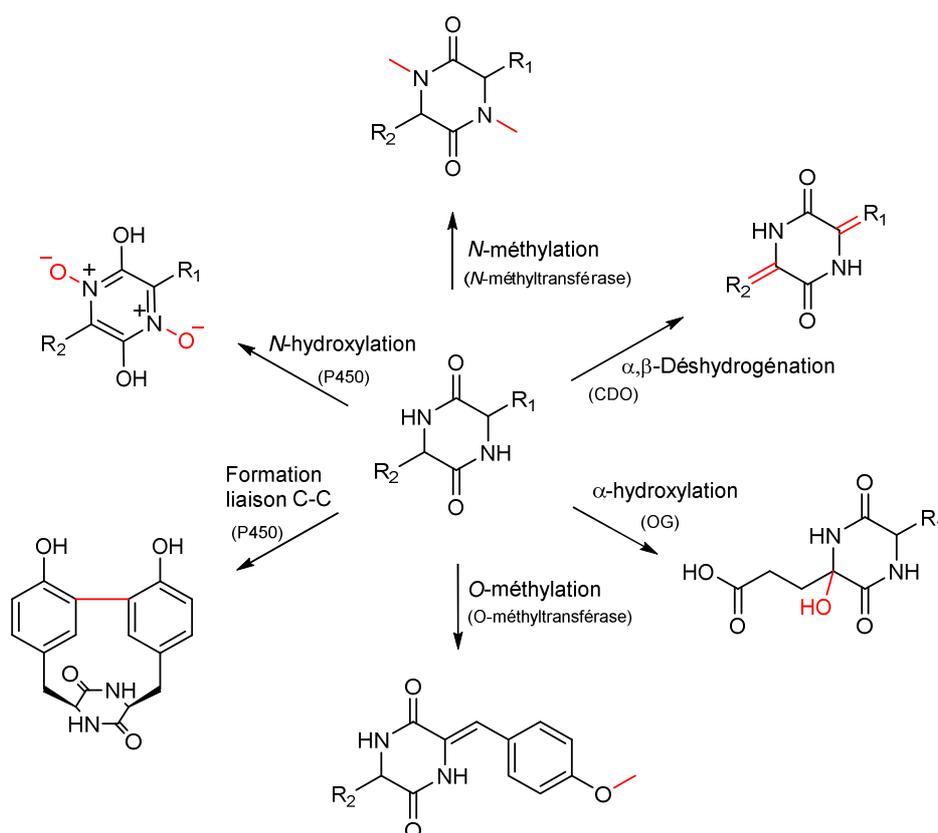
*In vivo*, l'association de dioxygénases avec des CDPS produisant du cyclo(Xaa-Glu) a mené à l'identification d'adduit glutathion sur des composés probablement déshydrogénés. Même si l'élucidation complète de la nature de ces composés n'a pas pu être réalisée, la présence de ces adduits pose la question de leur signification biologique. Des adduits glutathion sur des dicétopipérazines sont retrouvés dans une autre voie de synthèse de dicétopipérazines. Il s'agit de la voie de biosynthèse de la gliotoxine dans laquelle, suite à une hydroxylation en position C3 et C6 du cycle, une glutathion S-transférase catalyse le transfert de deux entités glutathion sur le cycle dicétopipérazine. Cette étape introduit des atomes de soufre qui permettent la formation du pont disulfure au-dessus du cycle dicétopipérazine (cf Etude bibliographique. IV.2.1. Voies de biosynthèse de la gliotoxine et des épithiodioxopipérazines : les gènes de la NRPS et des enzymes de modifications sont regroupés en un même cluster). Étant donné la régiosélectivité de l'enzyme K pour la position C $\alpha$  Glu, c'est-à-dire pour C3 ou C6, un parallèle peut être réalisé entre les deux voies. Dans notre étude de l'enzyme K, la voie de biosynthèse a été reconstituée de manière hétérologue chez *E. coli*, or aucune enzyme homologue à la glutathion S-transférase (GliG) d'*A. fumigatus* n'est retrouvée dans le génome de *E. coli* BL21. Néanmoins, nous pouvons faire l'hypothèse que la formation de l'adduit glutathion peut être catalysée par un autre type de glutathion S-transférase ou avoir lieu spontanément. Dans les deux cas, une question reste en suspens : le couplage à un glutathion est-il une étape d'une voie de biosynthèse visant à produire un composé plus complexe d'intérêt biologique ou, s'agit-il d'une voie de détoxification mise en place par la souche d'*E. coli* productrice ? En effet, le couplage d'une molécule xénobiotique ou endobiotique à un glutathion par une glutathion S-transférase permet de produire un composé moins toxique et plus hydrophile, et donc plus facilement éliminable<sup>421</sup>.

Les travaux présentés dans ce chapitre ont donc permis de déchiffrer une voie de biosynthèse en caractérisant le type d'activité de modification de la dioxygénase K. Cette voie étant homologue à cinq autres voies dont les CDPS produisent du cyclo(Ala-Glu) et deux voies dont les CDPS produisent du cyclo(Leu-Glu), les informations obtenues sur l'enzyme K peuvent être transposées aux dioxygénases des autres voies. Des travaux préliminaires au laboratoire ont permis de mettre en évidence que l'enzyme I associée à la CDPS 36 produisant du cyclo(Leu-Glu) et l'enzyme J associée à la CDPS 37 produisant du cyclo(Ala-Glu) présentaient le même type d'activité, ce qui confirme cette hypothèse. Cette étude a également permis de cerner les difficultés qui pourront être rencontrées lorsque l'étude de la voie associée à la CDPS 19 sera envisagée. Cette voie particulièrement riche en dioxygénase promet de proposer un vaste choix d'activités de modification sur le cyclo(Ile-Leu).

Au cours de mon doctorat, l'équipe de Mohamed Marahiel a, en plus de la voie de biosynthèse du cyclo(Trp-Trp) simplement et doublement méthylé, publié le déchiffrement d'une autre voie de biosynthèse dépendant des CDPS. Il s'agit d'une voie identifiée chez *N. dassonvillei* responsable de la biosynthèse de

nocazines. Les nocazines sont des composés cyclodipeptidiques comprenant deux acides aminés aromatiques avec au moins une  $\alpha$ - $\beta$ -déshydrogénation et une *O*- ou *N*-méthylation. Cette voie fait intervenir une CDPS synthétisant des produits de type cyclo(Phe-Xaa), cyclo(Phe-Tyr) et cyclo(Phe-Phe) étant les produits majoritaires (cf. Etude bibliographique. V.1.1 Les CDPS, une nouvelle famille d'enzymes qui synthétisent principalement des cyclodipeptides hydrophobes). La cyclodipeptide oxydase (CDO) codée par *Ndas\_1146/1147* transforme partiellement tous les cyclodipeptides synthétisés par la CDPS, sauf un, en leur analogue mono- ou di-déshydrogéné. Enfin, la *O*-méthyltransférase *Ndas\_1149* introduit une *O*-méthylation sur la chaîne latérale déshydrotyrosyle de composés mono- ou di-déshydrogénés.

Ainsi, en ajoutant l'activité des dioxygénases dépendantes du Fe(II) et du 2-oxoglutarate, le panel des activités de modification de cyclodipeptides contient désormais 6 types de modification (Figure 89). Les enzymes de modification connues dans les voies à CDPS catalysent des modifications sur le cycle dicétopipérazine (*N*-hydroxylation, *N*-méthylation, *C*-hydroxylation et  $\alpha,\beta$ -déshydrogénation) ou des modifications sur les chaînes latérales (*O*-méthylation de tyrosyle ou formation de liaison C-C entre deux tyrosyles). Notre étude a donc permis d'enrichir la diversité de dicétopipérazines produites et accessibles par voies enzymatiques.



**Figure 89 : Activités de modification de cyclodipeptides connues dans les voies de biosynthèse dépendant des CDPS**

La modification chimique est colorée en rouge sur la dicétopipérazine obtenue. La modification est indiquée le long de la flèche et l'enzyme la catalysant est précisé entre parenthèses. L'activité de modification peut dépendre de la nature des substituants R1 et R2.

# CONCLUSION GÉNÉRALE

Les produits naturels sont une richesse encore trop peu exploitée. Comme présenté en introduction, l'expression de clusters de gènes de voies de biosynthèse, identifiés par des approches d'exploitation des données du génome, constitue un levier puissant pour découvrir de nouvelles molécules naturelles. Cette approche a été particulièrement utilisée dans le cas des voies de biosynthèse dépendant de PKS et/ou NRPS : en l'espace d'environ 25 ans, nous sommes passés de la découverte de l'organisation modulaire de gènes de mégacomplexes enzymatiques<sup>422,423</sup> à la connaissance de plusieurs milliers de composés synthétisés par ces voies et à une meilleure compréhension des mécanismes sous-jacents à leur synthèse. Ces travaux s'inscrivent dans des perspectives bien plus larges qui sont d'explorer la diversité chimique des produits naturels et de pouvoir connecter une molécule aux gènes qui la codent<sup>8</sup>. Parmi les voies de biosynthèse originales de métabolites secondaires, celles dépendant des CDPS sont impliquées dans la synthèse du noyau dicétopipérazine qui est au cœur de nombreuses molécules d'intérêt pharmacologique. Récemment découvertes, les CDPS constituent une famille d'enzymes pour lesquelles, à l'instar des PKS et NRPS il y a 25 ans, les voies dans lesquelles elles sont impliquées pourraient réserver un potentiel de biosynthèse de nouvelles molécules. C'est dans ce contexte que j'ai débuté mon travail doctoral en essayant, à l'échelle de mon laboratoire, d'explorer la diversité moléculaire synthétisée par ces voies de biosynthèse de dicétopipérazines dépendant des CDPS.

Dans un premier temps, j'ai développé un protocole permettant d'étudier parallèlement plusieurs dizaines d'enzymes de ces voies (CDPS et enzyme de modification). Ce protocole permet notamment de coexprimer de manière hétérologue chez *E. coli* une CDPS et une ou plusieurs enzymes de modification ainsi que de détecter une large gamme de cyclodipeptides et de dérivés de cyclodipeptides. Dans un second temps, par une approche combinant bio-informatique et biochimie, j'ai identifié 51 nouvelles CDPS actives, ce qui a permis de redéfinir les caractéristiques générales de la famille des CDPS (en ce qui concerne la taille et les résidus conservés notamment). Cette étude a également mis en évidence que les CDPS se classent en au moins deux sous-familles phylogénétiques distinctes et que l'une d'entre elle, la sous-famille XYP, présente une plus grande diversité de séquences. Les CDPS sont désormais capables d'incorporer 17 des 20 acides aminés protéinogéniques et les cyclodipeptides produits par les CDPS présentent des propriétés physico-chimiques variées (hydrophobe, polaire, chargé). A partir des données collectées, nous avons proposé les bases d'un outil de prédiction des activités de synthèse de cyclodipeptides pour les CDPS. La conception de cet outil se rapproche de celui utilisé pour la prédiction des domaines d'adénylation des NRPS développé par Marahiel *et al*<sup>278-280</sup>. Cet outil devra par la suite être affiné et enrichi par la caractérisation biochimique de nouvelles CDPS pour espérer disposer un jour d'un outil aussi puissant que ceux disponibles pour les NRPS. Dans un troisième et dernier temps, en combinant à nouveau des approches bioinformatiques et biochimiques, j'ai identifié huit nouvelles voies de biosynthèse de CDPS. Sept d'entre elles font intervenir des dioxygénases associées à des CDPS produisant du cyclo(Ile-Leu), du cyclo(Leu-Glu) et du cyclo(Ala-Glu). Une de ces voies, provenant de *Gordonia effusa*, associant la CDPS 38 et dioxygénase K a été biochimiquement caractérisée. Cette voie mène à la synthèse d'au moins trois composés *in vivo* et 6 composés *in vitro*. La dioxygénase K est une

enzyme régiosélective, mais *a priori* non stéréosélective qui catalyse *in vitro* principalement la C-hydroxylation du carbone  $\alpha$  de la chaîne latérale Glu du cyclo(Ala-Glu).

Cette étude s'inscrit donc dans les nouvelles approches d'identification de molécules naturelles s'appuyant sur l'exploitation des données du génome. En reconstituant des voies de biosynthèse de manière hétérologue chez *E. coli* et en comparant les profils de productions de métabolites dans les surnageants de culture par LC-MS/MS, nous avons identifié de nouvelles molécules biosynthétisées. Dans le cas de molécules plus complexes que des cyclodipeptides constitués d'acides aminés protéinogéniques, nous avons été contraints de revenir à des techniques d'identification de molécules plus classiques passant par l'isolement et la caractérisation structurale par RMN. Ces techniques coûteuses en temps nous ont obligé à nous concentrer sur une seule des voies identifiées et ont fait chuter le débit de l'analyse de moyen à faible débit. Ainsi les voies impliquant des CDPS produisant du cyclo(Ile-Leu) et du cyclo(Leu-Glu) n'ont pas pu être biochimiquement caractérisées et devront être étudiées de la même manière afin de pouvoir confirmer la nature des produits qu'elle produisent.

En augmentant le nombre d'enzymes des voies dépendant des CDPS dont l'activité enzymatique est caractérisée ou, pour les CDPS, prédite grâce au nouvel outil de prédiction de spécificité dont nous avons posé les fondements, nous augmentons le nombre d'outils enzymatiques disponibles pour générer de la diversité moléculaire. Cependant, la diversité atteignable par les CDPS reste limitée, et ce pour plusieurs raisons. La nature du substrat de ces enzymes constitue un des principaux freins à l'expansion de la diversité chimique de biosynthèse offerte par ces enzymes. Les aa-ARNt sont essentiellement chargés dans la cellule avec un des 20 acides aminés protéinogènes de la série L, ce qui fait que la diversité de substrats possibles reste bien éloignée de celle observée avec les NRPS. Néanmoins, des essais réalisés dans le laboratoire ont montré l'incorporation de certains analogues d'acides aminés protéinogènes par les CDPS, mais ils requièrent le chargement de cet analogue sur l'ARNt par l'aaRS en absence de l'acide aminé naturel et des conditions particulières de croissance pour la synthèse *in vivo*. De plus, en commençant à comprendre le code de spécificité des CDPS, il devient envisageable de façonner rationnellement les poches P1 et P2 des CDPS pour leur permettre d'incorporer une plus grande variété d'acides aminés naturels ou non naturels.

Il est à noter que les cyclodipeptides sont synthétisables par différentes voies chimiques, et la variété de précurseurs existant a permis de générer une grande diversité de molécules. Cependant, une des limites de la diversité de molécules synthétisées chimiquement réside souvent dans la difficulté à reproduire la complexité des architectures des molécules naturelles. Si le noyau dicétopipérazine est faiblement modifié dans un cyclodipeptide, il en est autrement pour les molécules naturelles issues des voies de biosynthèse dépendant des CDPS. A titre d'exemple, la mycocyclosine, qui résulte de l'action d'un P450 sur un cyclodipeptide, est une molécule qui nécessite pour sa synthèse chimique totale la synthèse d'une iodo-dicétopipérazine, la protection et la déprotection d'un groupement amine et des groupements phénols, et l'utilisation de solvants toxiques comme le toluène et de catalyseur onéreux

comme le palladium<sup>424</sup>. Du fait de la complexité de sa voie de synthèse chimique, la découverte de ce produit aurait été hautement improbable avec des approches de chimie combinatoire par exemple. Par ailleurs, les conditions réactionnelles utilisées sont bien éloignées de l'idéal de la chimie verte, qui favorise notamment l'économie d'atome et l'utilisation de solvant non toxique. Ainsi, il est fort probable que les enzymes de modification qui composent des voies de biosynthèse dépendant des CDPS soit à la fois la clé de la diversité chimique de ces dicétopipérazines et à la fois un outil de biotransformation que la synthèse chimique peine à égaler.

Les résultats de mon étude permettent désormais d'envisager de la biosynthèse combinatoire - résultant en de nouvelles associations de CDPS et d'enzyme de modification, différentes de celles retrouvées dans les génomes - pour biosynthétiser des dicétopipérazines non naturelles. La spécificité des enzymes de modification est ici une clé importante. Plus la spécificité d'une enzyme de modification est relâchée, comme CDO de *S. noursei* par exemple, plus elle permettra d'être utilisée largement pour modifier des cyclodipeptides variées. A l'inverse, des enzymes de modification très spécifiques, comme CYP121 de *M. tuberculosis*, ont un potentiel plus limité en biosynthèse combinatoire. Pour les autres enzymes de modification, la spécificité semble être guidée par au moins une des chaînes latérales du substrat, laissant la possibilité de faire varier l'autre chaîne. Dans tout les cas des approches de mutagenèse ou d'évolution dirigée pourraient permettre de diversifier les substrats acceptés.

De plus, la synthèse chimique pourrait également être une source non négligeable de substrats, sur lesquels l'activité d'enzymes de modification pourrait être évaluée. On trouve souvent parmi les voies de biosynthèse dépendant des CDPS des enzymes de type OG et MT, dont l'activité peut être suivie par un simple test colorimétrique adaptable au format microplaque<sup>425,426</sup>. Ce type d'approche contribuerait également à générer de nouvelles dicétopipérazines. Ainsi, la génération de diversité moléculaire pourrait passer par des voies mixtes utilisant synthèse chimiques et biologiques en exploitant le point fort de chaque approche.

Un aspect important de mon travail de thèse est qu'il ouvre la voie pour une meilleure compréhension du rôle physiologique des dicétopipérazines. En effet, même si des activités biologiques ont été observées pour certaines des dicétopipérazines synthétisées par les voies dépendant des CDPS, aucun rôle physiologique n'est actuellement connu. Ainsi, la mise en évidence de la synthèse de certains cyclodipeptides par une CDPS associée à des systèmes de régulation à deux composants peut amener à tester l'action de ces cyclodipeptides ou d'analogues directement sur la physiologie des organismes. Dans des voies de biosynthèse plus complexes comprenant des enzymes de modification, il faudra déchiffrer l'action de ces enzymes et rechercher la production des dicétopipérazines correspondantes dans les milieux de culture. Ces recherches pourraient avoir tout leur intérêt, notamment dans le cas des organismes pathogènes codant une CDPS.

# ANNEXES

**ANNEXE 1: EPREUVES DE L'ARTICLE ACCEPTE DANS NATURE  
CHEMICAL BIOLOGY**

# Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity

Isabelle B Jacques<sup>1,3</sup>, Mireille Moutiez<sup>1,3,4</sup>, Jerzy Witwinowski<sup>2,4</sup>, Emmanuelle Darbon<sup>2</sup>, Cécile Martel<sup>2</sup>, Jérôme Seguin<sup>1,3</sup>, Emmanuel Favry<sup>1,3</sup>, Robert Thai<sup>1</sup>, Alain Lecoq<sup>1</sup>, Steven Dubois<sup>1</sup>, Jean-Luc Pernodet<sup>2</sup>, Muriel Gondry<sup>1,3\*</sup> & Pascal Belin<sup>1,3\*</sup>

**Cyclodipeptide synthases (CDPSs) constitute a family of peptide bond-forming enzymes that use aminoacyl-tRNAs for the synthesis of cyclodipeptides. Here, we describe the activity of 41 new CDPSs. We also show that CDPSs can be classified into two main phylogenetically distinct subfamilies characterized by specific functional subsequence signatures, named NYH and YYP. All 11 previously characterized CDPSs belong to the NYH subfamily, suggesting that further special features may be yet to be discovered in the other subfamily. CDPSs synthesize a large diversity of cyclodipeptides each made up of 17 proteinogenic amino acids. The identification of several CDPSs having the same specificity led us to determine specificity sequence motifs that, in combination with the phylogenetic distribution of CDPSs, provide a first step toward being able to predict the cyclodipeptides synthesized by newly discovered CDPSs. The determination of the activity of ten more CDPSs with predicted functions constitutes a first experimental validation of this predictive approach.**

CDPSs constitute a family of enzymes that divert aminoacyl-tRNAs (aa-tRNAs) from the ribosomal machinery to catalyze the formation of various cyclodipeptides, the precursors of many secondary metabolites with interesting biological activities<sup>1</sup>. Crystallographic structures of three CDPSs are now available, and indicate that they share a common architecture reminiscent of the catalytic domain of class Ic aminoacyl-tRNA synthetases (aaRSs)<sup>2–4</sup>. They use a ping-pong catalytic mechanism initiated by the binding of the first aa-tRNA. The aminoacyl moiety of this substrate is transferred to a conserved active site serine to generate an aminoacyl-enzyme intermediate<sup>2–5</sup>. This intermediate reacts with the aminoacyl moiety of the second aa-tRNA to form a dipeptidyl-enzyme. Finally, this second intermediate undergoes an intramolecular cyclization through the involvement of a conserved tyrosine, leading to the cyclodipeptide product<sup>6</sup>. Two recent studies on the CDPS AlbC give clues as to how CDPSs and their two substrates interact<sup>6,7</sup>. The two aa-tRNAs bind at different sites: the aminoacyl moiety of the first aa-tRNA is accommodated in pocket P1—structurally similar to the amino acid-binding pocket in class Ic aaRSs—and its tRNA moiety interacts with a patch of basic residues on the amphipathic helix  $\alpha_4$ ; the aminoacyl moiety of the second aa-tRNA is accommodated in a wider pocket, P2, with its tRNA moiety interacting with residues that belong to the flexible loop  $\alpha_6$ – $\alpha_7$ . The specificity of recognition of the first substrate is determined by its aminoacyl moiety, whereas that of the second substrate depends on both its aminoacyl moiety and its tRNA sequence, especially the N<sup>1</sup>–N<sup>72</sup> base pair of the acceptor stem. Only ten bacterial CDPSs and one eukaryotic CDPS have been biochemically characterized<sup>1,5,8,9</sup>. Most show a degree of substrate promiscuity but incorporate essentially five hydrophobic amino acids—F, L, Y, M and W—into cyclodipeptides. Iterative PSI-BLAST searches have identified numerous putative CDPSs<sup>10–12</sup>, and many of them diverge from the known conserved features of the family. To provide a better description of the CDPS family, we first report the identification and characterization of 41 new active members and propose a classification for the CDPS family. Then, we describe

how our data support a predictive tool for CDPS specificity. Finally, we demonstrate the reliability of prediction for few selected groups by characterizing ten additional CDPSs of unknown function.

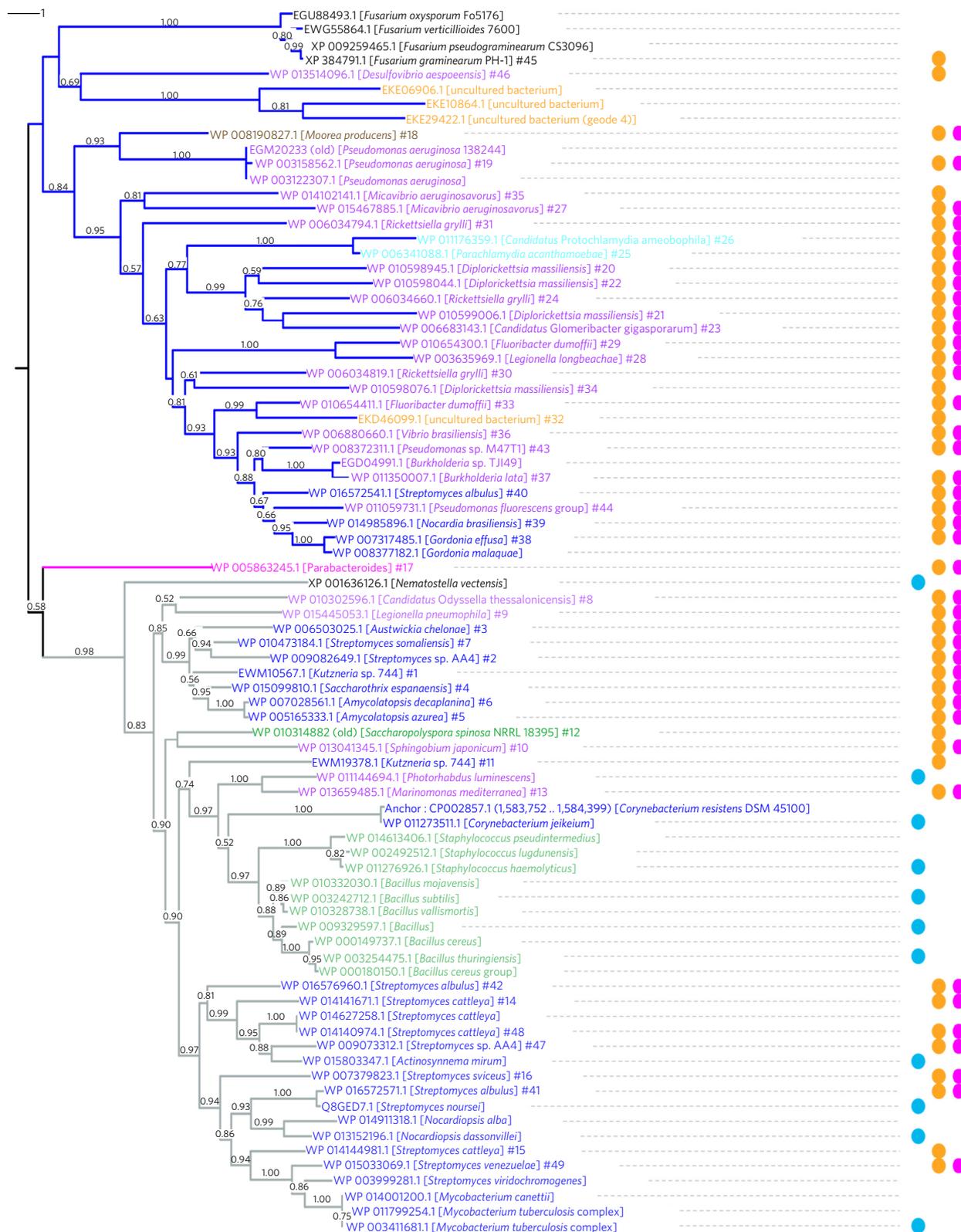
## RESULTS

### Search for active CDPSs identified 41 new members

We updated the list of putative CDPSs by searching the National Center for Biotechnology Information (NCBI) database. We retrieved about 80 new sequences corresponding to putative CDPSs (May 2013) and used these to construct a phylogenetic tree (Fig. 1). We selected 49 sequences representative of the whole sequence set (designated CDPS 1–49; Supplementary Results, Supplementary Data Set 1, Fig. 1) and investigated whether the selected putative CDPSs can synthesize cyclodipeptides. We cloned their genes in a homemade expression vector and produced the proteins in *Escherichia coli* strains, as previous work showed that cyclodipeptides can be recovered in culture supernatants upon CDPS expression in this host<sup>1</sup>. We implemented a medium-throughput method for CDPS expression in *E. coli*, recovery of culture supernatants for cyclodipeptide detection by LC/MS/MS, and bacterial fractionation for protein content analysis by SDS-PAGE (Supplementary Fig. 1). We found that 41 of the 49 recombinant CDPS candidates tested had cyclodipeptide-synthesizing activity (Fig. 1, Supplementary Data Set 2, Table 1).

We found that eight of the putative CDPSs were inactive, although they were effectively produced (Supplementary Fig. 2). Sequences analysis revealed several possible explanations for this absence of activity. CDPSs 12 and 15 lack the catalytic serine residue (replaced by A and G, respectively). In CDPS 45 this residue is replaced by a cysteine; this is, however, not sufficient to rule it out from being an active CDPS as the variant AlbC S37C is able to synthesize cyclodipeptides<sup>3</sup>. CDPS 45 is also eukaryotic, and therefore *E. coli* may lack an appropriate set of tRNA substrates for it to act on. We found no obvious explanations for the lack of activity of the other candidates.

<sup>1</sup>CEA, iBiTec-S, Service d'Ingénierie Moléculaire des Protéines, Gif-sur-Yvette, France. <sup>2</sup>Université Paris-Sud, Centre National de la Recherche Scientifique (CNRS), Unité Mixte de Recherche (UMR) 8621, Institut de Génétique et Microbiologie, Orsay, France. <sup>3</sup>Present address: CEA, iBiTec-S, Service de Biologie Intégrative et Génétique Moléculaire, UMR 9198, Gif-sur-Yvette, France. <sup>4</sup>These authors contributed equally to this work. \*e-mail: pascal.belin@cea.fr or muriel.gondry@cea.fr



**Figure 1 | Phylogenetic tree of known CDPSs and putative CDPSs retrieved from bioinformatics searches in databases (May 2013).** The set of sequences was curated: seven partial sequences were removed and the start of eight sequences manually corrected (see **Supplementary Figs. 6 and 7**). CDPSs are named by their protein accession numbers in NCBI and their host organisms; the numbers given to the CDPSs studied herein are added at the end of the name. CDPS names are colored according to their taxonomic origin: Proteobacteria, purple; Actinobacteria, dark blue; Chlamydiae, cyan; Cyanobacteria, brown; Bacteroidetes, fuchsia; Firmicutes, green; Eukaryotes, black (including 4 fungi and 1 metazoa (XP\_001636126.1 [*Nematostella vectensis*])). CDPSs from metagenomic approaches are colored orange. The CDPS members can be classified into three subfamilies that clearly separate into three main branches on the tree; the branches classified as NYH, XYP and SYQ are in gray, blue and fuchsia, respectively. The 11 known CDPSs, the 49 putative CDPSs studied herein, and the 41 new active CDPSs are labeled with cyan, orange, and magenta dots, respectively.

**Table 1 | The major cyclodipeptides produced by the newly identified CDPSs**

CDPS	Species	CDPS subfamily	<i>In vivo</i> activity <sup>a</sup>
1	<i>Kutzneria</i> sp. 744	NYH	cCC (100%)
2	<i>Streptomyces</i> sp. AA4	NYH	cCC (100%)
3	<i>Austwickia chelonae</i>	NYH	cCC (63%), cCA (25%), cCV (12%)
4	<i>Saccharothrix espanaensis</i>	NYH	cCC (100%)
5	<i>Amycolatopsis azurea</i>	NYH	cCC (100%)
6	<i>Amycolatopsis decaplanina</i>	NYH	cCC (100%)
7	<i>Streptomyces somaliensis</i>	NYH	cCC (99%)
8	<i>Candidatus Odysella thessalonicensis</i>	NYH	cGV (99%)
9	<i>Legionella pneumophila</i>	NYH	cCC (100%)
10	<i>Sphingobium japonicum</i>	NYH	cPM (95%)
11	<i>Kutzneria</i> sp. 744	NYH	-
12	<i>Saccharopolyspora spinosa</i>	NYH	-
13	<i>Marinomonas mediterranea</i>	NYH	cLL (98%)
14	<i>Streptomyces cattleya</i>	NYH	cWW (100%)
15	<i>Streptomyces cattleya</i>	NYH	-
16	<i>Streptomyces sviveus</i>	NYH	cLV (90%)
17	<i>Parabacteroides</i> sp. 20_3	SYQ	cHF (100%)
18	<i>Moorea producens</i>	XYP	cAP (100%)
19	<i>Pseudomonas aeruginosa</i>	XYP	cIL (93%)
20	<i>Diplorickettsia massiliensis</i>	XYP	cFN (90%)
21	<i>Diplorickettsia massiliensis</i>	XYP	cLL (80%)
22	<i>Diplorickettsia massiliensis</i>	XYP	cYV (30%), cYP (23%), cYT (11%), cYA (10%), cFP (10%), cFV (7%)
23	<i>Candidatus Glomeribacter gigasporarum</i>	XYP	cFF (93%) cFL (7%)
24	<i>Rickettsiella grylli</i>	XYP	cFF (53%) cFL (47%)
25	<i>Parachlamydia acanthamoebae</i>	XYP	cPP (75%)
26	<i>Candidatus Protochlamydia amoebophila</i>	XYP	cPP (100%)

### Different catalytic residues for the two CDPS subfamilies

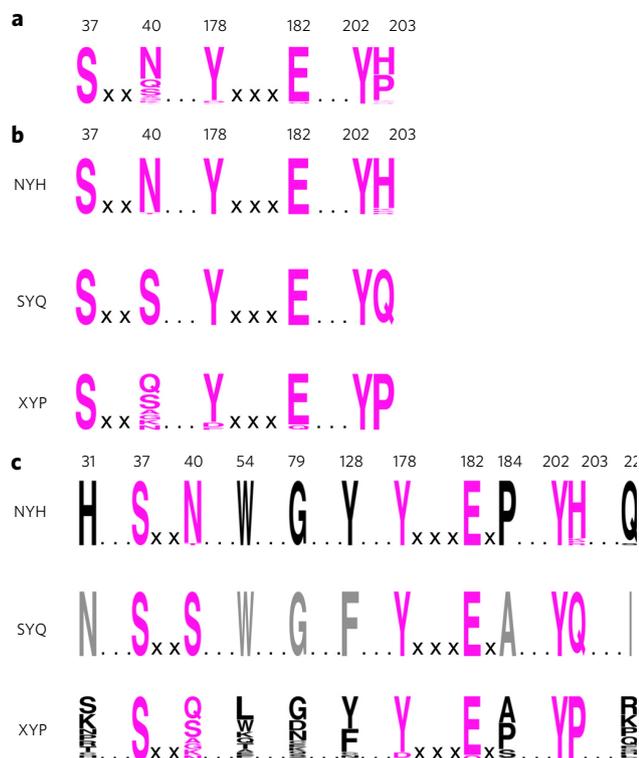
Thus, the CDPS family now includes 52 proteins (the 11 previously characterized CDPSs plus the 41 active CDPSs characterized in this study), which allows re-examination of the features of this family. Most known CDPSs are found in three bacterial phyla, Actinobacteria, Firmicutes and Proteobacteria, and a few are found in Bacteroidetes (1), Chlamydiae (2) and Cyanobacteria (1). Only one active CDPS has been found in a eukaryotic phylum (Cnidaria)<sup>5</sup>, and no CDPSs have been found in Archaea (Fig. 1). CDPSs are typically 200–300 residues long. Most of the newly confirmed CDPSs share less than 30% amino acid sequence identity with any of the 11 previously characterized CDPSs, and 22 share only between 6% and 20% sequence identity with them (Supplementary Data Set 3). However, HHPRED searches with each of the 41 new CDPS sequences retrieved systematically as best hits the three CDPSs of known structure (Supplementary Data Set 4). We used the secondary-structure predictions and sequence similarities identified with HHPRED to adjust the multiple sequence alignment of the 52 CDPSs (Supplementary Data Set 5). The six residues previously identified as catalytic residues are mostly conserved (N40, Y178,

**Table 1 | Continued**

CDPS	Species	CDPS subfamily	<i>In vivo</i> activity <sup>a</sup>
27	<i>Micavibrio aeruginosavorus</i>	XYP	cGN (100%)
28	<i>Legionella longbeachae</i>	XYP	cAA (95%)
29	<i>Fluoribacter dumoffii</i>	XYP	cAA (97%)
30	<i>Rickettsiella grylli</i>	XYP	cMA (23%), cMG (27%)
31	<i>Rickettsiella grylli</i>	XYP	cAG (86%), cAA (14%)
32	Uncultured bacterium ACD_69C00020	XYP	-
33	<i>Fluoribacter dumoffii</i>	XYP	cGG (100%)
34	<i>Diplorickettsia massiliensis</i>	XYP	-
35	<i>Micavibrio aeruginosavorus</i>	XYP	-
36	<i>Vibrio brasiliensis</i>	XYP	cLE (58%), cLA (20%), cLP (14%)
37	<i>Burkholderia lata</i>	XYP	cEA (100%)
38	<i>Gordonia effusa</i>	XYP	cEA (97%)
39	<i>Nocardia brasiliensis</i>	XYP	cEA (96%)
40	<i>Streptomyces albulus</i>	XYP	cEA (97%)
41	<i>Streptomyces albulus</i>	NYH	cFL (48%), cFF (23%), cFY (19%), cFM (6%)
42	<i>Streptomyces albulus</i>	NYH	cWW (100%)
43	<i>Pseudomonas</i> sp. M47T1	XYP	cEA (100%)
44	<i>Pseudomonas protegens</i>	XYP	cLE (99%)
45	<i>Gibberella zeae</i>	XYP	-
46	<i>Desulfovibrio aespoeensis</i>	XYP	-
47	<i>Streptomyces</i> sp. AA4	NYH	cWW (100%)
48	<i>Streptomyces cattleya</i>	NYH	cWW (100%)
49	<i>Streptomyces venezuelae</i>	NYH	cYY (92%)

<sup>a</sup>The major cyclodipeptides produced by recombinant strain M15 pREP4 expressing the corresponding CDPS are given; cyclodipeptides are ranked according to the peak area on chromatograms recorded at 214 nm (see Supplementary Data Set 2); percentages in brackets indicate the proportion of each cyclodipeptide and were determined from peak areas on chromatograms recorded at 214 nm; cyclodipeptides that made up <5% of the total are not indicated; cCC refers to the detection of cyclocystine in bacterial culture supernatants: when a CDPS synthesizes cCC, it is recovered in its oxidized form, i.e., cyclocystine, in the supernatant; -, no cyclodipeptide detected.

E182, H203) or even strictly conserved (S37, Y202) (AlbC numbering, used throughout the paper)<sup>1,3,6</sup>, allowing a functional sequence signature for the CDPS enzymes to be defined (Fig. 2a). We previously showed that the strictly conserved catalytic residue Y202 participates in a hydrogen bond network that also involves residues N40 and H203, which are essential for the accurate positioning of the two loops bearing the catalytic residues in AlbC<sup>6</sup>. We observed that about half of the CDPSs do indeed have the 'N40, H203' pair, whereas others have an 'X40, P203' pair (X being a nonconserved residue) and a very small number have other pairs (see Discussion section). Examination of the phylogenetic tree of CDPSs shows an early divergence into two different branches (Fig. 1). Comparative analysis between CDPS distribution in the tree and their sequences shows a strong correlation between the distribution and the presence of one or the other pair in the sequences. CDPSs clearly divide into two subfamilies or classes that we named 'NYH' and 'XYP' according to the 'X40, Y202, X203' sequence. CDPS 17 has a 'S40, Q203' pair and is also clearly separate from the other CDPSs in the phylogenetic tree, suggesting that it belongs to a probable third subfamily that we named 'SYQ'. Each subfamily can be associated with a



**Figure 2 | CDPS subfamilies.** (a) Functional sequence signature for CDPSs shown with sequence logos corresponding to the frequency plot of the catalytic amino acids at positions 37, 40, 178, 182, 202 and 203 (AlbC numbering). (b) Functional subsequence signatures for the CDPS subfamilies. Subsequence signatures are indicated for the NYH and XYP subfamilies and for CDPS 17, which is the lone member of a probable third subfamily, SYQ. (c) Subsequence signatures for the CDPS subfamilies with catalytic and conserved residues taken into account. Conserved residues, indicated in black for the NYH and XYP subfamilies and in gray for the probable SYQ subfamily (containing only one known member), are shown at positions 31, 54, 79, 128, 184 and 225 (AlbC numbering). The sequence of the eukaryotic NYH member, XP\_001636126.1 [*Nematostella vectensis*], was removed from the set used to define the signatures because it contains modifications probably corresponding to phylum.

functional sub-sequence signature (Fig. 2b)—whose differences are likely to reflect different ways of positioning catalytic residues—and possesses further distinguishing features. Notably, the 11 previously characterized CDPSs all belong to the NYH subfamily.

The NYH subfamily has six conserved residues in addition to the six catalytic residues (Supplementary Data Set 5; Supplementary Fig. 3), including three residues identified in AlbC as being essential for enzyme activity (H31 and Y128) or structural integrity (W54)<sup>3</sup>. The XYP subfamily has a more degenerate sub-sequence, with only five catalytic residues (and no noncatalytic residues) conserved. The single SYQ sequence has half of the noncatalytic conserved residues of the NYH subfamily (Fig. 2c). The basic patch involved in the binding of the tRNA moiety of the first substrate<sup>7</sup> is found in all members of the NYH subfamily and contains 5–11 basic residues. This patch is much smaller in the XYP class, with only 2–6 basic residues. The XYP subfamily is generally far more diverse than the NYH subfamily, with variants of even the highly conserved catalytic sequence YxxxE found: CDPSs 25 and 26 have the sequence DxxxQ and not YxxxE (Supplementary Data Set 5). The role of the E residue, which acts as a catalytic base essential for generating the dipeptidyl-AlbC intermediate, and that of the Y residue in positioning the aminoacyl moiety of the first substrate in AlbC<sup>3,6</sup> could be fulfilled by D and Q, respectively. This provides support for the

idea that the positioning of the catalytic residues is probably not the same in the different subfamilies. The XYP subfamily also contains members containing sequence insertions up to 30 residues long between predicted secondary structures, and deletions of a few residues, mostly in predicted  $\alpha$  helices (CDPSs 21, 23, 25, 26, 28 and 29) (Supplementary Fig. 4). In addition, CDPSs 28 and 29 contain an additional ~230-residue C-terminal domain of unknown function, suggesting the possibility of a CDPS domain in a multifunctional protein (Supplementary Fig. 4).

### An expanded diversity of cyclodipeptides

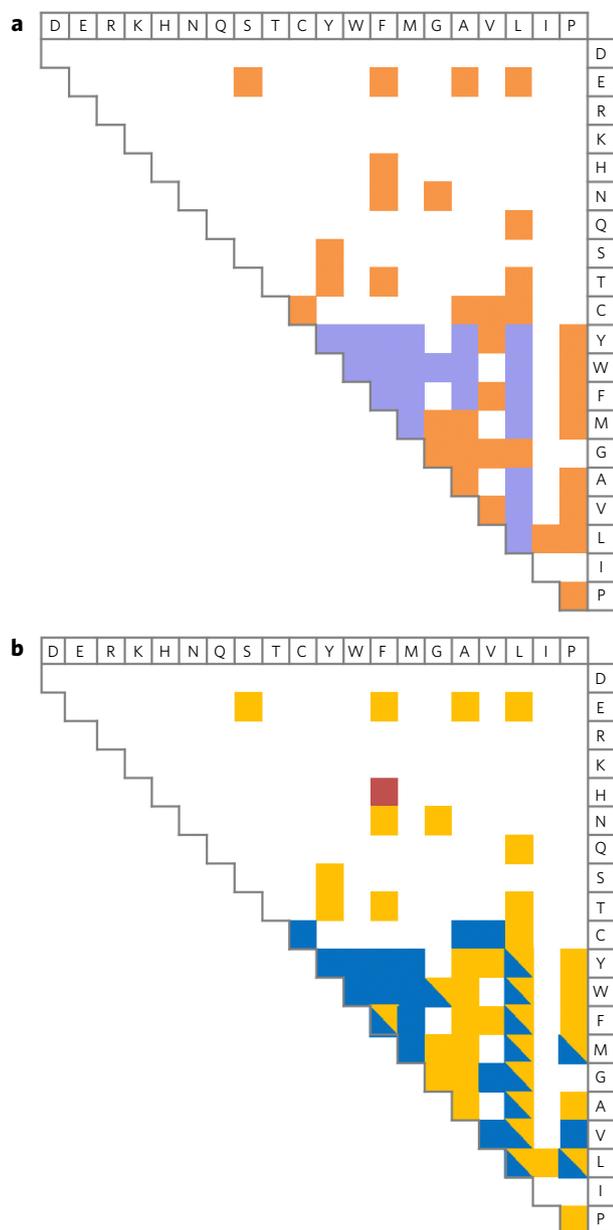
The 19 cyclodipeptides synthesized by the 11 NYH enzymes known before this study are mostly hydrophobic, with aromatic and aliphatic side chains<sup>1,5,8,9</sup>. Here, we identify 35 cyclodipeptides that have not previously been described as CDPS products. Many of these contain amino acids not previously observed in products of CDPS activity, including G, amino acids with hydrophobic side chains (V, P, I) and polar and charged amino acids (E, H, N, Q, S, T, C) (Fig. 3a). Consequently, CDPSs now incorporate 17 of 20 proteinogenic amino acids into cyclodipeptides.

We looked at how cyclodipeptide diversity is distributed among CDPS subfamilies. NYH members incorporate ten different amino acids into cyclodipeptides, with C being the only additional amino acid used by these new NYH subfamily CDPSs. XYP members use a larger set of amino acids—16 of the 20 proteinogenic amino acids—including those used by NYH members and also the polar and charged residues Q, T, N, S and E (Fig. 3b). Consistent with this diversity of amino acid utilization, the number of cyclodipeptides synthesized differs between the families, with 26 and 39 cyclodipeptides found for NYH and XYP members, respectively. This greater product diversity may be related to the larger diversity of sequences within the XYP subfamily (Supplementary Fig. 3).

The previously characterized NYH members show a degree of substrate promiscuity, with Amir\_4628 being the only one synthesizing specifically one cyclodipeptide: cWW<sup>8</sup>. About a third of the newly characterized CDPSs show high substrate specificity (Table 1; Supplementary Fig. 5). These high specificities are not related to their membership in one or other subfamily; they are probably a consequence of the nature of the substrates with either a particular tRNA sequence—the N<sup>1</sup>-N<sup>72</sup> base pair is important for recognition of the second substrate and, although well conserved in the prokaryotic world, it displays some differences (Supplementary Data Set 9)—or an aminoacyl moiety with specific physical properties or chemical properties<sup>7</sup>. The other newly characterized CDPSs demonstrated promiscuity toward different amino acids in our experimental conditions (Table 1; Supplementary Data Set 2; Supplementary Fig. 5), but most synthesized primarily one cyclodipeptide and other cyclodipeptides in smaller amounts.

### Toward prediction of CDPS specificity

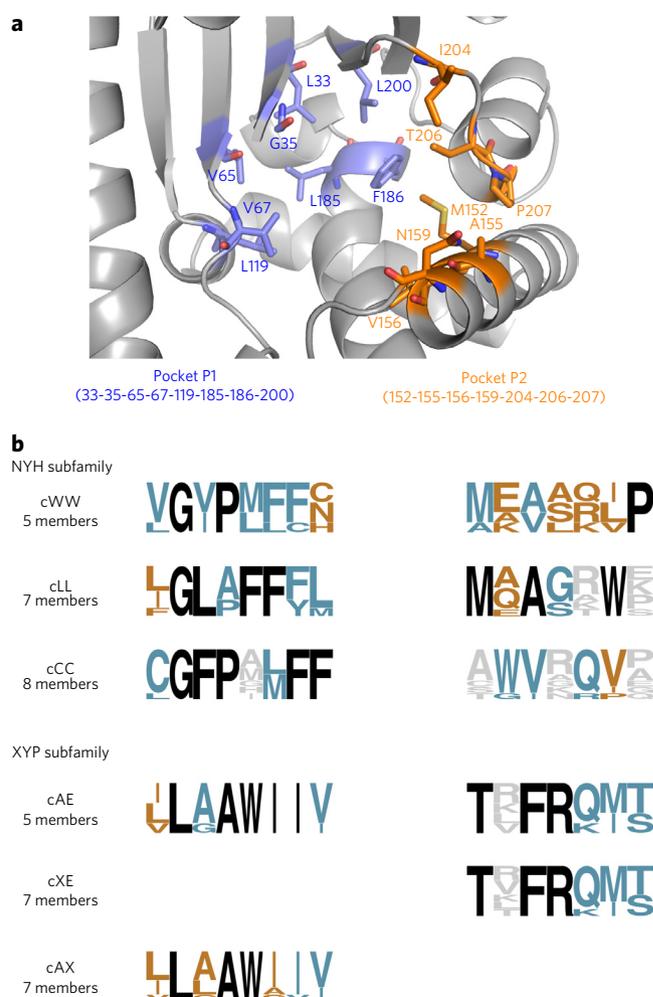
Substrate prediction requires biochemical characterization of a sufficient number of enzymes and the knowledge of the substrate-binding pocket determinants; this was illustrated by the pioneering works on the substrate specificity of the adenylation domains of nonribosomal peptide synthetases<sup>13,14</sup>. The large number of active CDPSs described herein and the identification of the cyclodipeptides they produce can be exploited in a similar way to help understand CDPS specificity. The NYH subfamily member AlbC has two binding pockets, P1 and P2, accommodating the aminoacyl moieties of the first and second aa-tRNA; there are eight residues lining P1 and seven others lining P2 (ref. 6) (Fig. 4a). We used the multiple sequence alignment of CDPSs manually adjusted with HHPRED (Supplementary Data Set 5) to identify the residues of the active CDPSs corresponding to those lining P1 and P2 of AlbC (Supplementary Data Set 6). We predicted, for all active CDPSs, secondary structures similar to those observed in the crystal



**Figure 3 | Diversity of cyclodipeptides produced by the CDPS subfamilies.**

**(a)** Cyclodipeptides produced by the 11 previously characterized CDPSs (in cyan) and the newly identified CDPSs (in magenta). The two amino acids constituting the cyclodipeptides are indicated in the one-letter code at the top and right. The presence of a colored square at the intersection of a row and a column indicates that the corresponding cyclodipeptide was detected in the culture supernatant of bacteria expressing a CDPS. **(b)** Cyclodipeptides produced by the NYH (in blue), XYP (in yellow) and SYQ (in purple) members. The presence and nature of cyclodipeptides are indicated as described in **a**.

structures of AlbC, YmC and Rv2275, and we assumed that the positions of residues lining P1 and P2 were conserved (**Supplementary Data Set 4**). CDPSs possessing similar activities, such as cLL-synthesizing enzymes, are present in both NYH and XYP subfamilies. For CDPSs with the same activities, the residues constituting P1 and P2 appear well conserved within a subfamily, but show significant differences between subfamilies, indicating that different combinations of amino acids are used to synthesize the same product, in agreement with separate evolution. Thus, CDPSs of the



**Figure 4 | Sequence logos of the amino acids constituting the two binding pockets, P1 and P2, for groups of CDPSs synthesizing the same cyclodipeptides.**

**(a)** Three-dimensional view of the binding pockets P1 and P2 accommodating the aminoacyl moieties of the first and second aa-tRNA in AlbC<sup>6</sup>. The eight residues constituting P1 are in blue and the seven constituting P2 are in magenta. **(b)** Residues suspected to delineate P1 and P2 were determined for each CDPS (**Supplementary Data Set 6**). Logo sequences corresponding to the frequency plot of amino acids at positions in P1 and P2 were generated for groups of CDPSs belonging to the same subfamily and having similar cyclodipeptide-synthesizing activity: cWW (CDPSs 14, 42, 47, 48 and Amir\_4627: 35–55% sequence identity); cLL (CDPSs 13, jk0923, Plu0297, pSHaeC06, YmC\_BSU, YmC\_BLIC and YmC\_BTHU: 26–70% sequence identity); cCC (CDPSs 1, 2, 3, 4, 5, 6, 7 and 9: 28–93% sequence identity); cAE (CDPSs 37, 38, 39, 40 and 43: 40–58% sequence identity); cXE (CDPSs 36, 37, 38, 39, 40, 43 and 44: 38–58% sequence identity); cAX (CDPSs 28, 31, 37, 38, 39, 40 and 43: 16–58% sequence identity). At each position, the presence of a single, two, three or more amino acids is indicated in black, blue, orange or gray, respectively.

same subfamily and having the same activity appear to constitute a specific group (**Supplementary Data Set 6**). For groups that were sufficiently populated, that is had at least five members, we created sequence logos to represent sequence motifs of the P1 and P2 pockets. This was performed for four groups: those with cLL-, cWW-, cCC- and cAE-synthesizing members (**Fig. 4b**). As P1 and P2 pockets are independent<sup>6</sup>, we considered combining the groups accommodating the same residue in one of the two pockets, regardless the residue accommodated in the other pocket. However, the

**Table 2 | Prediction and validation of the cyclodipeptide-synthesizing activities of CDPSs**

CDPS <sup>a</sup>	Species	Pocket residues <sup>b</sup>		Prediction	Sequence identity <sup>c</sup>	Cyclodipeptides synthesized <sup>d</sup>
		P1	P2			
50	<i>Streptomyces</i> sp. F12	CGFPWLFF	AWVGPQDV	cCC	34–52%	cCC (100%)
51	<i>Nocardiopsis gilva</i>	CGYPSLFF	AWVGRPEF	cCC	34–52%	–
52	<i>Kutzneria albida</i> DSM 43870	CGFPSLFF	AWVRQVWF	cCC	32–57%	cCC (100%)
53	<i>Staphylococcus pseudintermedius</i>	IGLAFFFM	MQASAWE	cLL	26–76%	–
54	<i>Bacillus</i> sp. 171095_106	LGLAFFFL	MAAGKWR	cLL	30–63%	cLL (56%), cLF (23%), cLM (20%)
55	<i>Bacillus</i> sp. NSP9.1	LGLAFFFL	MAAGRWK	cLL	29–82%	cLL (52%), cLF (28%), cLM (20%)
56	<i>Streptomyces albus</i>	VGIPMFFC	ATVARLP	cWW	35–56%	cWW (100%)
57	<i>Streptomyces</i> sp. CNH287	VGVP MFFC	MDVAQLP	cWW	39–60%	cWW (100%)
58	<i>Nocardiopsis halophila</i>	VGVP MFFN	MEVARLP	cWW	39–61%	cWW (100%)
59	<i>Xenorhabdus doucetiae</i>	VLAAWIII	TVFRKMT	cAE	37–66%	cAE (92%), cSE (7%)
60	<i>Photorhabdus luminescens</i> BA1	ILAAWIIV	TIFRQMT	cAE	42–54%	cAE (100%)
61	<i>Mycobacterium neoaurum</i>	ILAAWIIV	TRFRQIS	cAE	41–69%	cAE (80%), cAP (18%)

<sup>a</sup>CDPSs are grouped according to the prediction of activity: 50–52, cCC; 53–55, cLL; 56–58, cWW; 59–61, cAE. <sup>b</sup>The predicted residues of the P1 and P2 pockets are indicated for each CDPS. <sup>c</sup>Clustal Omega at the EMBL website was used to determine sequence identities between each CDPS and members of the predicted group (as defined in Fig. 4b). <sup>d</sup>–, no cyclodipeptide detected.

Q13 identification of the site involved in amino acid binding is a key concern for CDPSs. We can use our knowledge of AlbC to infer the roles of P1 and P2 in the other CDPSs. Most CDPSs are as promiscuous as AlbC and produce several cyclodipeptides of general formulae c(AA<sub>1</sub>-X), where AA<sub>1</sub> is the preferred amino acyl and X is any of the incorporated amino acids. Assuming a similar mechanism for all CDPSs—sequential ping-pong mechanism with acyl enzyme intermediate formation—we can suggest amino acids accommodated by P1 and P2 for each CDPS, with the preferred amino acyl moiety binding to P1. Thus, combination was possible for cAA- and cAE- synthesizing groups, for which P1 accommodates A (its hydrophobicity excludes the accommodation of E). The resulting cAX-synthesizing group was sufficiently populated (seven members) to create a sequence logo for P1. Similarly, we combined the cAE- and cLE-synthesizing groups to give the cXE-synthesizing group (seven members) and create a sequence logo of P2 (Fig. 4b). The production of any particular cyclodipeptide correlates with the conservation at specific positions in sequence motifs. However, with the knowledge currently available, sequence motif composition is insufficient to provide a complete understanding of why a particular aminoacyl is incorporated, although some pieces of information emerge. For example, for the cLE- and cAE-synthesizing groups (Supplementary Data Set 6), the discrimination between A and L may be based on the second residue of P1 (position 35), for which the presence of an alanine or a leucine correlates with the synthesis of cLE or cAE, respectively.

Sequence motifs allow discrimination between CDPSs that are clustered on the tree but have different cyclodipeptide-synthesizing activities. For example, in the same NYH subfamily, CDPS 8 is clustered on the phylogenetic tree with cCC-synthesizing CDPSs even though it produces cGV (Fig. 1; Table 1). However, its P1 and P2 residues differ substantially from those in the sequence motif of the cCC-synthesizing CDPSs (Supplementary Data Set 6).

We investigated whether sequence motifs could be used to predict the specificity of putative CDPSs. We searched databases for new putative CDPSs at the time of writing (September 2014) and retrieved about 200 further sequences. The set of sequences was curated to remove partial sequences, and about 30 sequences with N-terminal truncations were completed as for the experiment set. We then constructed a new phylogenetic tree representing the extended putative CDPS family (Supplementary Data Set 7). All new putative CDPSs were distributed into the three subfamilies and fulfilled the functional subsequence signatures described

above. We next compared the residues constituting P1 and P2 of all the new putative CDPSs to the sequence motifs described above (Fig. 4b). This led us to the identification of CDPSs likely to be members of the cLL-, cWW-, cCC- and cAE-synthesizing groups. We selected three proteins for each group, sharing as little sequence identity as possible with the characterized CDPSs of the group, and determined their cyclodipeptide-synthesizing activities. We found ten CDPSs to be active, and in all cases the main cyclodipeptides produced were as predicted (Table 2; Supplementary Data Set 2). Using the link between sequence motif and CDPS specificity, we were able to predict the activity of putative CDPSs, in spite of the only moderate sequence identity between these CDPSs and members of the predicted group (Table 2). Among the 200 newly identified putative CDPSs, about 30% have P1- and P2-constituting residues very similar to those of poorly populated groups (those with fewer than five members). These putative CDPSs are listed with the appropriate groups in Supplementary Data Set 8. Accumulating sequence and biochemical data should lead to the identification of additional sequence motifs corresponding to as yet undescribed cyclodipeptide-synthesizing groups.

## DISCUSSION

Our study reveals that all CDPS members—whether or not they are biochemically characterized—can be classified into two main, phylogenetically distinct subfamilies. The NYH and YYP subfamilies contain almost all CDPS members, with the NYH subfamily being the most numerous. We identified only CDPS 17 and one additional member, both belonging to the same Bacteroidetes phylum (Supplementary Data Set 7), as belonging to a probable third subfamily named SYQ. This distribution between subfamilies may not accurately reflect that in nature, as the choice of genomes for sequencing is heavily biased toward organisms pathogenic to humans, animals and plants.

Only very few active CDPS members do not comply with all criteria for subfamily membership. In particular, several NYH members have a noncanonical pair X40-X203. Availability of new CDPS sequences is likely to reveal whether the atypical CDPSs form relevant subgroups in the different subfamilies and to what extent their study may reveal previously unknown features of CDPSs.

Current understanding of CDPSs has mostly come from the crystallographic structures of three NYH members determined so far<sup>2–4,6</sup> and from related biochemical studies<sup>7</sup>. The identification of YYP and SYQ members shows that the two residues forming

the pair 'N40, H203' and previously demonstrated to be essential for the catalysis in NYH members are not conserved. Although the different catalytic steps of the reaction are probably the same throughout the whole CDPS family, this observation means that alternative solutions have been found to fulfill the roles of these two residues. Structural characterization of XYP members will be especially informative as the P203 residue may induce structural variations around the catalytic loops as compared to NYH members, and affect both the positioning of catalytic residues and the composition of amino acids delineating the binding pocket P2 accommodating the aminoacyl moiety of the second aa-tRNA substrate<sup>6</sup>. Structural characterization associated with biochemical studies will probably also improve our ability to predict the cyclodipeptides synthesized by new CDPSs.

The CDPSs characterized to date synthesize 54 different cyclodipeptides, containing 17 of the 20 proteinogenic amino acids. About 55% of the 200 new putative CDPSs are predicted to synthesize cCC, cLL, cWW or cAE (**Supplementary Data Set 7**), and about 30% may be members of already identified but poorly populated cyclodipeptide-synthesizing groups. By implication, therefore, the remaining putative CDPSs (about 15%) may produce different cyclodipeptides. These observations clearly show that CDPSs are able to synthesize a large variety of cyclodipeptides.

The physiological relevance of the substrate specificities of CDPSs determined upon overexpression in *E. coli* is an issue. We previously identified the specificity determinants for the two substrates of the model CDPS AlbC: the aminoacyl moiety is essential for the recognition of the first substrate, and both the aminoacyl moiety and the N<sup>1</sup>-N<sup>72</sup> base pair of the tRNA moiety participate in the recognition of the second substrate<sup>6,7</sup>. As (i) the sequence of the tRNA moiety does not matter for the binding of the first substrate, (ii) the N<sup>1</sup>-N<sup>72</sup> base pair is strongly conserved in prokaryotic sequences (**Supplementary Data Set 9**) and (iii) aminoacyl-tRNAs are present in significant quantities in all living cells, it is likely that the main product(s) of CDPS activity observed in our study are similar to (or the same as) those produced in the native organisms. Indeed, the activities of a few CDPSs in their native organisms have been characterized, and they correspond to the main products observed upon overexpression in *E. coli*: cLL for YvmC from *Bacillus* species<sup>11,15</sup>, cFL for AlbC from *Streptomyces noursei*<sup>16,17</sup> and cFY for Ndas\_1148 from *Nocardia d'assonvillei*<sup>9</sup>. The main products observed upon CDPS overexpression in *E. coli* also correspond to the CDPS activity detected *in vitro*, with the aminoacyl-tRNAs being in excess with respect to the CDPS<sup>17–9</sup>. However, the experimental conditions used in our study probably do not reflect the conditions, and particularly the level of expression of the CDPS, in the native organism. Although these conditions do not appear to influence the nature of the main cyclodipeptides produced, they may cause the production of minor products that are not synthesized *in vivo* in the native organism.

CDPSs are generally genetically associated with cyclodipeptide-tailoring enzymes in biosynthetic pathways dedicated to the synthesis of diketopiperazines<sup>11,18</sup>. Only five CDPS-dependent pathways have been fully characterized so far, and the incorporation of hydrophobic amino acids into cyclodipeptides and their modification by oxidation and methylation reactions have been described<sup>8,9,17,19–21</sup>. The identification of the cyclodipeptide-synthesizing activities of many CDPSs provides valuable data for unraveling a large number of CDPS-dependent pathways and for elucidating and describing the chemical diversity encoded by these pathways.

In conclusion, our work improves knowledge of the CDPS family and gives clues for deciphering CDPS-dependent biosynthetic pathways and their engineering with the aim of producing novel natural products.

Received 20 October 2014; accepted 5 June 2015;  
published online XX XX 2015

## METHODS

Methods and any associated references are available in the [online version of the paper](#).

## References

- Gondry, M. *et al.* Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. *Nat. Chem. Biol.* **5**, 414–420 (2009).
- Vetting, M.W., Hegde, S.S. & Blanchard, J.S. The structure and mechanism of the *Mycobacterium tuberculosis* cyclodityrosine synthetase. *Nat. Chem. Biol.* **6**, 797–799 (2010).
- Sauguet, L. *et al.* Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis. *Nucleic Acids Res.* **39**, 4475–4489 (2011).
- Bonnefond, L. *et al.* Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog. *Proc. Natl. Acad. Sci. USA* **108**, 3912–3917 (2011).
- Seguin, J. *et al.* Nonribosomal peptide synthesis in animals: the cyclodipeptide synthase of *Nematostella*. *Chem. Biol.* **18**, 1362–1368 (2011).
- Moutiez, M. *et al.* Unravelling the mechanism of non-ribosomal peptide synthesis by cyclodipeptide synthases. *Nat. Commun.* **5**, 5141 (2014).
- Moutiez, M. *et al.* Specificity determinants for the two tRNA substrates of the cyclodipeptide synthase AlbC from *Streptomyces noursei*. *Nucleic Acids Res.* **42**, 7247–7258 (2014).
- Giessen, T.W., von Tesmar, A.M. & Marahiel, M.A. A tRNA-dependent two-enzyme pathway for the generation of singly and doubly methylated ditryptophan 2,5-diketopiperazines. *Biochemistry* **52**, 4274–4283 (2013).
- Giessen, T.W., von Tesmar, A.M. & Marahiel, M.A. Insights into the generation of structural diversity in a tRNA-dependent pathway for highly modified bioactive cyclic dipeptides. *Chem. Biol.* **20**, 828–838 (2013).
- Aravind, L., de Souza, R.F. & Iyer, L.M. Predicted class-I aminoacyl tRNA synthetase-like proteins in non-ribosomal peptide synthesis. *Biol. Direct* **5**, 48 (2010).
- Belin, P. *et al.* The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways. *Nat. Prod. Rep.* **29**, 961–979 (2012).
- Giessen, T.W. & Marahiel, M.A. The tRNA-dependent biosynthesis of modified cyclic dipeptides. *Int. J. Mol. Sci.* **15**, 14610–14631 (2014).
- Stachelhaus, T., Mootz, H.D. & Marahiel, M.A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999).
- Challis, G.L., Ravel, J. & Townsend, C.A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**, 211–224 (2000).
- Tang, M.R., Sternberg, D., Behr, R.K., Sloma, A. & Berka, R.M. Use of transcriptional profiling & bioinformatics to solve production problems. *Ind. Biotechnol. (New Rochelle N.Y.)* **2**, 66–74 (2006).
- Fukushima, K., Yazawa, K. & Arai, T. Biological activities of albonoursin. *J. Antibiot. (Tokyo)* **26**, 175–176 (1973).
- Lautru, S., Gondry, M., Genet, R. & Pernodet, J.L. The albonoursin gene cluster of *S. noursei*: biosynthesis of diketopiperazine metabolites independent of nonribosomal peptide synthetases. *Chem. Biol.* **9**, 1355–1364 (2002).
- Gu, B., He, S., Yan, X. & Zhang, L. Tentative biosynthetic pathways of some microbial diketopiperazines. *Appl. Microbiol. Biotechnol.* **97**, 8439–8453 (2013).
- Gondry, M. *et al.* Cyclic dipeptide oxidase from *Streptomyces noursei*. Isolation, purification and partial characterization of a novel amino acyl alpha,beta-dehydrogenase. *Eur. J. Biochem.* **268**, 1712–1721 (2001).
- Belin, P. *et al.* Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **106**, 7426–7431 (2009).
- Cryle, M.J., Bell, S.G. & Schlichting, I. Structural and biochemical characterization of the cytochrome P450 CypX (CYP134A1) from *Bacillus subtilis*: a cyclo-L-leucyl-L-leucyl dipeptide oxidase. *Biochemistry* **49**, 7282–7296 (2010).

## Acknowledgments

This work was supported by the CEA, the CNRS, the Paris-Sud University and a grant from the French National Research Agency (ANR 2010/Blan 1501 01) to M.G. and J.-L.P. I.B.J. was supported by a doctoral fellowship from the CEA. The Service d'Ingénierie Moléculaire des Protéines is member of the Laboratory of Excellence LERMIT. We warmly thank V. Dive, head of the Service d'Ingénierie Moléculaire des Protéines, for his continuous support and encouragement throughout this work. We thank O. Lespinet for advice about the building of phylogenetic trees, D. Vallenet, M. Stam and M. Sorokina for helpful discussion on bioinformatics, and A. Ponties for technical assistance in cloning experiments. We are indebted to L. Beuvier for skillful assistance in mass data analysis. We thank F. Fenaile for kindly performing the experiments using the Orbitrap mass spectrophotometer. We thank P. Kessler and O. Lequin for skillful assistance in NMR experiments.

Q16

**Author contributions**

M.G. obtained funding. M.M., J.-L.P., M.G. and P.B. developed the hypothesis and designed the study. I.B.J., M.M. and P.B. performed bioinformatic analyses. I.B.J., J.W., E.D. and C.M. performed cloning experiments and prepared culture supernatants. I.B.J. performed LC/MS/MS analysis. I.B.J. and M.M. analyzed MS/MS data. I.B.J., J.S., E.F. and P.B. purified cyclodipeptides from culture supernatants. S.D. performed amino acid composition analyses. A.L. and S.D. chemically synthesized cyclodipeptides. R.T. performed high-resolution mass spectrometry. I.B.J., M.M., J.W., E.D., J.S., J.L.P., M.G. and P.B. analyzed and discussed the results. I.B.J., M.M., M.G. and P.B. prepared

the draft manuscript. All of the authors participated in the production of the final version of the manuscript.

**Competing financial interests**

The authors declare no competing financial interests.

**Additional information**

Supplementary information is available in the [online version of the paper](#). Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence should be addressed to P.B. or M.G.

## ONLINE METHODS

**Chemicals, bacterial strains and plasmids.** All chemicals were from Sigma-Aldrich unless otherwise stated. We used *Escherichia coli* DH5 $\alpha$  obtained from Invitrogen for cloning experiments. *E. coli* BL21AI and M15 were obtained from Invitrogen and Qiagen, respectively.

Plasmid pQE60 (Qiagen: ColE1 origin of replication, ampicillin resistance) is an expression vector that carries an IPTG-inducible promoter for protein expression. Plasmid pREP4 (Qiagen: p15A origin of replication, kanamycin resistance) is a repressor plasmid that allows the constitutive expression at high levels of the *lac* repressor for tight control of IPTG-inducible promoters. For CDPS expression, we constructed pIJ196, a pQE60/pREP4 hybrid vector that carries the IPTG-inducible promoter and allows strong expression of the LacI repressor. Briefly, a 1.5 kb *Bst*Z171-SmaI fragment carrying the *lacI* gene and its promoter was excised from pREP4 and ligated into the *Bst*Z171 restriction site of pQE60 such that replication from ColE1 and transcription from *lacI* promoter are in the same direction on the plasmid (**Supplementary Fig. 1**). The cloning sites in the resulting plasmid, pIJ196, were checked by sequencing (MWG-Eurofins). Restriction enzymes were from Thermo Scientific. Standard methods were used for DNA manipulations<sup>22</sup>.

**Bioinformatics. Identification of the new putative CDPSs.** Sequence databases at NCBI and Broad Institute were screened for genes encoding putative CDPSs: PSI-Blast searches<sup>23,24</sup> were conducted with an *E*-value threshold of 0.005 and using sequences of several biochemically characterized CDPSs as query. After each PSI-Blast run, hits irrelevant to a CDPS sequence were omitted for the next run. Five or six iterative runs were performed and we obtained ~90 different hits. Sequences encoded by database entries C791\_7200, SsomD4\_010100015357, SJA\_C1-32620, Dmas2\_010100010809, K530\_11977 and pc1814 showed first residues aligning with positions 56, 32, 24, 66, 50 and 26 of AlbC, respectively (**Supplementary Fig. 6**), suggesting possible start codon misannotation. Examination of the 5' surrounding DNA sequence of each corresponding coding sequence led us to identify an alternative start codon located upstream, leading to longer amino acid sequences that correctly matched the N-terminal part of AlbC. Thus, these extended versions were selected for CDPS activity screening. The PSI-Blast hit SCAT\_0901 encodes an amino acid sequence with an extra N-terminal part that has no homology with AlbC and is rich in proline and alanine residues (**Supplementary Fig. 7**). Examination of the coding sequence indicated the presence of an ATG triplet coding Met52. As Met52 encoded by SCAT\_0901 aligned with Ala8 of AlbC, we decided to select the shortened version starting at Met52 for CDPS activity screening. Seven other partial sequences were removed from the set.

**Tree calculation.** Tree of sequences available in May 2013 (**Fig. 1**): After PSI-Blast searches and a set of relevant sequences had been obtained as described above, we generated a multiple sequence alignment using Muscle<sup>25</sup> integrated into Seaview<sup>26</sup>. This alignment was not manually adjusted, but the accurate alignment of catalytic and conserved residues was checked. The tree was calculated using the PhyML program (v 3.1)<sup>26</sup> based on the maximum-likelihood method. The LG substitution model and the NNI tree searching operation were selected. Reliability of internal branching was assessed using the aLRT test (SH-Like). The iTOL software was used for graphical representation and edition of the phylogenetic tree<sup>27,28</sup>. Tree of sequences available at the time of writing (September 2014) (**Supplementary Data Set 7**): After PSI-Blast searches performed as described above, we obtained about 300 hits. We corrected sequences truncated at the N terminus when possible and we removed sequences truncated at the C terminus, redundant sequences and sequences lacking the catalytic serine (especially putative CDPS 12 and 15). The sequences of CDPSs 26 and 25 were added to the set. Multiple sequence alignment was performed with Muscle<sup>25</sup> integrated into Seaview<sup>26</sup> and was manually adjusted to align the conserved and catalytic residues. Insertions or domains in C terminus which did not align were removed from the multiple alignment. The tree was constructed as described above.

**Selection of a set of 49 sequences.** The putative CDPSs that shared more than 75% sequence identity with an already characterized CDPS or with another selected CDPS were excluded from the selection (19 sequences), with the exception of CDPS 41 (88% identity with AlbC), and CDPSs 5 and 6 (93% identity with each other) (**Supplementary Data Set 3**). Three sequences originating from metagenomes and likely to contain sequence errors were also excluded. The final set of 49 proteins selected for CDPS activity screening is given in **Supplementary Data Set 1**.

**Selection of a set of 12 additional sequences.** To assess the predictive value of sequence logos of P1 and P2 pockets determined for the four groups of CDPSs synthesizing cWW, cLL, cCC and cAE, respectively (**Fig. 4b**), we choose 12 CDPS sequences predicted to belong to one of the four groups (three per group) as listed in **Supplementary Data Set 7**. The selected sequences clustered with members of the group on the phylogenetic tree (**Supplementary Data Set 7**) and displayed sequence motifs for P1 and P2 that are consistent with the sequence logos for the corresponding group. Furthermore, the CDPSs were selected to display as little sequence identity as possible with the biochemically characterized CDPSs of the corresponding group. The final set of the 12 additional proteins is given in **Supplementary Data Set 1**.

**HHPred Prediction.** HHPred at the Max Planck Institute website was used to predict secondary structures for each putative CDPS by searching the RCSB Protein Data Bank (PDB) using default parameters (multiple sequence alignment generation method, HHblits; *E*-value threshold for multiple sequence alignment generation, 10<sup>-3</sup>; alignment mode, local)<sup>29</sup>.

**Topology diagrams.** Topology diagrams of the three structurally characterized CDPSs were obtained with PDBSum at the EMBL-EBI website, and a topology diagram was deduced by conserving secondary structures found in all CDPS crystal structures.

**Sequence logos.** The sequence logos of CDPSs consist of stacks of symbols representing amino acids in one letter code<sup>30</sup>. They were obtained at the WebLogo website<sup>31</sup>. Each stack presents a single position in the sequence. The height of the symbols within the stack indicates the relative frequency of each amino acid in the sequence.

**CDPSs genes.** Synthetic genes encoding CDPSs 1–46 and optimized for expression in *E. coli* were obtained from GeneArt. They were designed to have an *Nco*I restriction site containing the ATG start codon and a *Bgl*II restriction site located downstream from the last codon of the coding sequence. If required for further DNA manipulations, an alanine-encoding codon was introduced after the start codon. The synthetic genes were provided in GeneArt-specific cloning vectors. Their sequences are given in **Supplementary Data Set 1**. Genes encoding CDPSs 47–49 were obtained by PCR amplification of chromosomal DNA from *Streptomyces* sp. AA4 (now referred to as *Amycolatopsis* sp. AA4), *Streptomyces cattleya* NRRL 8057 and *Streptomyces venezuelae* ATCC 10712, respectively. Chromosomal DNA was prepared as follows. Mycelium was disrupted using glass beads and a FastPrep-24 instrument (MP Biomedicals) in the presence of 400  $\mu$ l of a phenol/chloroform/isoamyl alcohol mixture (25:24:1 v/v/v) and 400  $\mu$ l of water. The aqueous phase was collected after centrifugation (16 000 g, 10 min) and precipitated by incubation for 30–60 min at –20 °C with 800  $\mu$ l of propan-2-ol and 120  $\mu$ l of 3 M sodium acetate. The pellet was centrifuged (16,000g, 15 min), washed with 70% ethanol and resuspended in TE buffer containing RNase at 50  $\mu$ g/ml. Phusion High Fidelity DNA Polymerase (ThermoScientific) and the GC protocol provided in the manufacturer's instructions were used for PCR amplification; betaine was added to the mix to the final concentration of 1 M. Oligonucleotides used for PCR amplification were obtained from Integrated DNA Technologies and are described in **Supplementary Table 1**. The PCR cycle was as follows: 98 °C for 30 s; 30 cycles of 98 °C for 10 s, 72 °C for 60 s; 72 °C for 10 min. After incubation with Taq DNA polymerase to add 3' adenine overhangs to each end (Qiagen; according to manufacturer instructions), PCR products were ligated into pGEM-T Easy (Promega). Plasmid DNA was prepared from positive transformants and verified by DNA sequencing (Beckman Coulter Genomics). Sequences of the amplicons corresponding to CDPS 47–49 are given in **Supplementary Data Set 1**.

Synthetic genes encoding the 12 additional CDPSs (numbered 50–61) and optimized for expression in *E. coli* were obtained from GeneArt. They were designed as described above. Their sequences are given in **Supplementary Data Set 1**.

CDPS coding sequences were then inserted between the *Nco*I and *Bgl*II restriction sites of pIJ196 for protein expression in *E. coli*. All recombinant plasmids were prepared using DH5 $\alpha$  bacteria and verified by DNA sequencing of the promoter and CDPS-encoding regions (Eurofins-MWG).

**Expression of CDPSs in medium-throughput format.** Initially, we worked with strain BL21AI in which protein production can be triggered using controlled culture medium appropriate for autoinduction. We did not obtain recombinant clones following transformation with pIJ196-derived plasmids

encoding CDPS 8, 21, 22, 25, 26, 27, 30, 31 and 36. We reasoned that this may have been due to toxicity associated with the CDPS-encoding plasmids and that better control of the repression of CDPS expression was necessary. Indeed, transformation of BL21AI harboring the repressor plasmid pREP4 with the nine plasmids led to appropriate transformants in all cases except for plasmids encoding CDPS 22 and 31. All CDPS-encoding plasmids were derived from pQE60, whose supplier recommends the use of strain M15 pREP4 to alleviate toxicity problems. Accordingly, we transformed M15 pREP4 cells with each of the CDPS-encoding plasmids, and successfully obtained transformants. Finally, we used BL21AI and BL21AI pREP4 and M15 pREP4 to express the recombinant putative CDPSs.

Each recombinant putative CDPS was expressed in *E. coli* M15pREP4 and BL21AI (harboring or not harboring pREP4) from the corresponding pIJ196-derived plasmid. Bacteria were cultured in 10 ml 24-well plates with round-bottomed wells (Dutscher Scientific) containing 2 ml of the appropriate growth medium, covered with a hydrophobic porous film (VWR), and shaken at 200 rpm. Starter cultures were M9-derived minimum medium supplemented with trace elements and vitamins<sup>1</sup>, 200 µg/ml ampicillin, 25 µg/ml kanamycin 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, the starter culture was used to inoculate (1/50) the same M9-derived minimum medium except that glucose was replaced by 0.5% glycerol for M15 bacteria or by a combination of 0.5% glycerol, 0.05% glucose and 0.02% lactose for BL21AI bacteria<sup>32</sup>. M15 bacteria were grown at 37 °C until the OD<sub>600</sub> reached 0.6, and expression of the putative CDPS was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG, 2 mM final concentration). Cultivation was continued for 24 h at 20 °C. BL21AI bacteria were grown in an autoinduced medium<sup>32</sup> and thus did not need the addition of IPTG for CDPS expression. After inoculation of the expression cultures, BL21AI bacteria were grown at 37 °C for 3.5 h, and transferred to 20 °C for 20.5 h. At the end of cultivation, cells were pelleted by centrifugation of the plates. The supernatants were collected, acidified (2% TFA final concentration) and frozen at -20 °C. Bacterial pellets were stored at -80 °C until fractionation and protein content analysis.

Cell pellets were thawed on ice, and suspended in 400 µl of ice-cold lysis buffer (150 mM NaCl, 0.1% Triton X100, 0.5 mg/ml lysozyme, 10 µM phosphoramidon, 1 mM phenylmethylsulfonyl fluoride (PMSF), 100 mM Tris HCl, pH 8.0)<sup>33</sup>. After 1.5 h on ice, MgCl<sub>2</sub> (10 mM final concentration) and 5 units of benzonase (Sigma-Aldrich) were added and the samples incubated for 1 h in an ice-cold bath. The soluble protein fraction was separated from the insoluble fraction by centrifugation (45 min, 3000 g, 4 °C), and both fractions were analyzed by SDS-PAGE using the Mini-Protean 3 Dodeca Cell system from Bio-Rad.

**Cyclodipeptide identification.** 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 an Atlantis dC18 column (4.6 mm × 150 mm, 3 µm, 100 Å, Waters) or a Hypercarb column (4.6 mm × 150 mm, 5 µm, 250 Å, ThermoScientific) developed over 50 min with the linear gradient 1 for CDPSs 1–47 (0% to 50% (v/v)) or the linear gradient 2 for CDPSs 48–49 (15% to 65% (v/v)) (solvent A: 0.1% (v/v) formic acid in H<sub>2</sub>O, solvent B: 0.1% (v/v) formic acid in acetonitrile/H<sub>2</sub>O (90/10), flow rate, 0.6 ml/min). Previous work showed that the Atlantis dC18 3µm 4.6 × 150 mm column is effectively separates cyclodipeptides containing at least one residue with a hydrophobic side chain<sup>1</sup>. However, we observed that these HPLC conditions were unsatisfactory for the retention of less hydrophobic cyclodipeptides (Supplementary Table 2). Retention on the column is important in our screening assay involving HPLC coupled to mass spectrometry because of a lag phase (around 7 min) before mass data acquisition. We therefore tested the Hypercarb column (Thermo Scientific), described to be effective for the separation of polar compounds<sup>34</sup>. We indeed observed substantial retention of the cyclodipeptides tested on the Hypercarb column (Supplementary Table 2). However, the analysis on Hypercarb of highly hydrophobic compounds such as cFF resulted in peaks with long tails, probably due to strong retention on the column. Therefore, we analyzed the bacterial culture supernatants with both the Atlantis dC18 column and the Hypercarb column. Positive electrospray ionization and mass analysis were optimized for the detection of compounds in the range of natural cyclodipeptides as previously described<sup>1,20</sup>.

Cyclodipeptides were detected by both their *m/z* value and their daughter-ion spectra<sup>35–41</sup>. Their identities were confirmed in various ways. For most, the nature of the detected cyclodipeptides was unambiguously established either by comparison with data from literature (cPV and cPP<sup>35</sup>; cWW<sup>8</sup>; cLL, cLM, cFL, cMM, cFM, cYM, cFF, cYF, cYA, cYL and cYY<sup>1</sup>) or by comparison with authentic standards purchased from Bachem (cGG, cGA, cAA, cLG, cVV, cEA, cPL, cPM, cFP, cYS, cYP, cHF) or from Sigma (cLA, cLV) or chemically synthesized in our laboratory (cFA, cFV, cFT, cFN, and cFE; see below). The cyclodipeptides cYV, cGN, cGV, cAP, cLE and cyclo(*l*-cystine) were purified from culture supernatants of recombinant *E. coli* expressing CDPS 22, 27, 8, 18, 36 and 9, respectively. cYV was purified using a Purospher STAR RP-18 endcapped LiChroCART 250-10 column (10 mm × 250 mm, 5 µm, 120 Å, Merck); cGN, cGV, cAP, cLE and cyclocystine were purified with a Hypercarb column (10 mm × 150 mm, 5 µm, 250 Å, Thermo Scientific). The HPLC purifications were carried out at 4.75 ml/min using buffer A consisting of 0.1% TFA in water, buffer B of 90% acetonitrile in water with 0.09% TFA, and the gradients depicted in the Supplementary Table 3. Purified cyclodipeptides were obtained at purity >90% as estimated by UV chromatograms of analytical HPLC recorded at 220 nm (cYV: Atlantis dC18, 4.6 × 150 mm, 3 µm, 100 Å (Waters), linear gradient of acetonitrile in water from 0% to 50% (1% per min) in 0.1% TFA at a flow rate of 0.6 ml/min; cGN, cGV, cAP, cLE and cyclocystine: Hypercarb column 4.6 mm × 150 mm, 5 µm, 250 Å (Thermo Scientific), the same gradients as those used for purification, but at a flow rate of 0.6 ml/min). They were analyzed by determination of amino acid composition as below. No clear data could be obtained for cyclocystine. Purified cyclodipeptides were also analyzed by high-resolution mass spectrometry using either MALDI-TOF/TOF as previously described<sup>1</sup> (cYV, cGV, cAP, cLE) or a high-resolution/high-mass-accuracy LTQ-Orbitrap instrument (Thermo Scientific) with a resolution set at 30,000 at *m/z* 400 (cyclocystine and cGN). HRMS (*m/z*): cYV, [MH<sup>+</sup>] calculated for C<sub>14</sub>H<sub>19</sub>N<sub>3</sub>O<sub>3</sub>, 263.1396, found 263.1393; cGN, [MH<sup>+</sup>] calculated for C<sub>6</sub>H<sub>10</sub>N<sub>3</sub>O<sub>3</sub>, 172.0722, found 172.0713; cGV, [MH<sup>+</sup>] calculated for C<sub>7</sub>H<sub>13</sub>N<sub>3</sub>O<sub>2</sub>, 157.0977, found 157.0974; cAP, [MH<sup>+</sup>] calculated for C<sub>8</sub>H<sub>13</sub>N<sub>2</sub>O<sub>2</sub>, 169.0977, found 169.0966; cLE, [MH<sup>+</sup>] calculated for C<sub>11</sub>H<sub>19</sub>N<sub>2</sub>O<sub>2</sub>, 243.1345, found 243.1324; cyclocystine, [MH<sup>+</sup>] calculated for C<sub>6</sub>H<sub>9</sub>N<sub>2</sub>O<sub>2</sub>S<sub>2</sub>, 205.0100, found 205.0096. Finally, the identity of ten cyclodipeptides produced in only small amounts (cCA, cMG, cCV, cMA, cLT, cLC, cSE, cCE, cLQ and cPW) and one produced in larger amounts (cYT) was not investigated beyond MS/MS analyses.

**Chemical synthesis of cFA, cFV, cFT, cFN and cFE.** Published procedures were used for the chemical synthesis of cFA, cFV, cFT, cFN, and cFE<sup>42</sup>. They were purified by semi-preparative RP-HPLC (Purospher STAR RP-18 endcapped LiChroCART, 10 mm × 250 mm, 5 µm, 120 Å, Merck) using a linear gradient of acetonitrile in water from 0% to 50% (1% per min) in 0.1% trifluoroacetic acid at a flow rate of 4.75 ml/min. After lyophilization of the samples, a white powder was observed for all compounds. Purified cyclodipeptides were obtained at purity >95% as estimated by UV chromatograms of analytical RP-HPLC recorded at 220 nm (Atlantis dC18, 4.6 × 150 mm, 3 µm, 100 Å, Waters; linear gradient of acetonitrile in water from 0% to 50% (1% per min) in 0.1% formic acid at a flow rate of 0.6 ml/min). The purified cyclodipeptides were characterized by determination of amino acid composition under standard conditions: samples (100 µM solution in 1% DMSO) were vacuum dried and sealed in glass tubes using the PicoTag system (Waters) and hydrolyzed under the vapor phase of 6N HCl with a crystal of phenol for 17h at 110 °C. The hydrolyzed sample was dissolved in 20–50 µl of Milli-Q water, and 5–20 µl of the resulting HCl hydrolysate (containing a minimum of 200 pmol of each amino acid) was analyzed and quantified by ninhydrin derivatization on an aminoTac JLC-500/V amino-acid analyzer (JEOL). Standard amino acid solutions were used for calibration at the beginning of each analysis series. Purified, chemically synthesized cyclodipeptides were also characterized by high-resolution mass spectrometry using MALDI-TOF/TOF as previously described<sup>1</sup> and NMR spectroscopy. The NMR experiments were recorded on a Bruker Avance 250 spectrometer. The lyophilized samples were dissolved in DMSO-*d*<sub>6</sub> (Eurisotop) and spectra were recorded at 25 °C. The data from NMR experiments were processed and analyzed with Bruker TOPSPIN 2.0 program.

<sup>1</sup>cFA. <sup>1</sup>H NMR (250 MHz, DMSO): δ 8.13 (s, 1H, H<sup>N</sup> F or A), 8.02 (s, 1H, H<sup>N</sup> F or A), 7.32–7.13 (m, 5H, H<sup>δ-ε-ζ</sup> F), 4.17 (m, 1H, H<sup>α</sup> F), 3.61 (qt, J = 7, 1.5 Hz, 1H, H<sup>α</sup> A), 3.13 (dd, J = 13.5, 3.75 Hz, 1H, H<sup>β</sup> F), 2.85 (dd, J = 13.5, 5 Hz, 1H, H<sup>β</sup> F),

0.45 (d,  $J = 7$  Hz, 3H, H <sup>$\beta$</sup>  A); <sup>13</sup>C NMR (62.8 MHz, DMSO):  $\delta$  167.7 (C<sup>F</sup> or A), 165.8 (C<sup>F</sup> or A), 136.1 (C<sup>F</sup>), 130.4 (C <sup>$\delta$</sup>  F), 128.1 (C <sup>$\epsilon$</sup>  F), 126.7 (C <sup>$\zeta$</sup>  F), 55.4 (C <sup>$\alpha$</sup>  F), 49.7 (C <sup>$\alpha$</sup>  A), 38.3 (C <sup>$\beta$</sup>  F), 19.7 (C <sup>$\beta$</sup>  A). HRMS ( $m/z$ ): [MH<sup>+</sup>] calculated for C<sub>12</sub>H<sub>15</sub>N<sub>2</sub>O<sub>3</sub>, 219.1133, found 219.1143.

cFV. <sup>1</sup>H NMR (250 MHz, DMSO):  $\delta$  8.13 (s, 1H, H<sup>N</sup> F or V), 7.92 (s, 1H, H<sup>N</sup> F or V), 7.27–7.14 (m, 5H, H <sup>$\delta$ - $\zeta$</sup>  F), 4.21 (m, 1H, H <sup>$\alpha$</sup>  F), 3.53 (m, 1H, H <sup>$\alpha$</sup>  V), 3.15 (dd,  $J = 13.5, 4.25$  Hz, 1H, H <sup>$\beta$</sup>  F), 2.86 (dd,  $J = 13.5, 5$  Hz, 1H, H <sup>$\beta$</sup>  V), 1.72 (m, 1H, H <sup>$\beta$</sup>  V), 0.64 (d,  $J = 7.1$  Hz, 3H, H <sup>$\gamma$</sup>  V), 0.24 (d,  $J = 6.8$  Hz, 3H, H <sup>$\gamma$</sup>  V); <sup>13</sup>C NMR (62.8 MHz, DMSO):  $\delta$  166.6 (C<sup>F</sup> or V), 166.4 (C<sup>F</sup> or V), 136.3 (C<sup>F</sup>), 130.4 (C <sup>$\delta$</sup>  F), 128.0 (C <sup>$\epsilon$</sup>  F), 126.5 (C <sup>$\zeta$</sup>  F), 59.2 (C <sup>$\alpha$</sup>  V), 55.0 (C <sup>$\alpha$</sup>  F), 37.8 (C <sup>$\beta$</sup>  F), 31.0 (C <sup>$\beta$</sup>  V), 18.3 (C <sup>$\gamma$</sup>  V), 16.2 (C <sup>$\gamma$</sup>  V). HRMS ( $m/z$ ): [MH<sup>+</sup>] calculated for C<sub>14</sub>H<sub>19</sub>N<sub>2</sub>O<sub>2</sub>, 247.1446, found 247.1444.

cFT. <sup>1</sup>H NMR (250 MHz, DMSO):  $\delta$  8.05 (d,  $J = 2.75$  Hz, 1H, H<sup>N</sup> F or T), 7.89 (d,  $J = 2.75$  Hz, 1H, H<sup>N</sup> F or T), 7.32–7.16 (m, 5H, H <sup>$\delta$ - $\zeta$</sup>  F), 5.00 (s, 1H, H<sup>OH</sup> T), 3.93 (m, 1H, H <sup>$\alpha$</sup>  F), 3.69 (m, 1H, H <sup>$\beta$</sup>  T), 3.52 (t,  $J = \sim 3$  Hz, 1H, H <sup>$\alpha$</sup>  T), 3.16–3.01 (ABX, 2H, H <sup>$\beta$</sup>  F, H <sup>$\beta$</sup>  T), 0.96 (d,  $J = 6.5$  Hz, 3H, H <sup>$\gamma$</sup>  T); <sup>13</sup>C NMR (62.8 MHz, DMSO):  $\delta$  167.4 (C<sup>F</sup> or T), 166.7 (C<sup>F</sup> or T), 137.4 (C<sup>F</sup>), 129.8 (C <sup>$\delta$</sup>  F), 128.2 (C <sup>$\epsilon$</sup>  F), 126.4 (C <sup>$\zeta$</sup>  F), 67.2 (C <sup>$\beta$</sup>  T), 60.3 (C <sup>$\alpha$</sup>  T), 55.9 (C <sup>$\alpha$</sup>  F), 40.8 (C <sup>$\beta$</sup>  F), 19.7 (C <sup>$\gamma$</sup>  T). HRMS ( $m/z$ ): [MH<sup>+</sup>] calculated for C<sub>13</sub>H<sub>17</sub>N<sub>2</sub>O<sub>3</sub>, 249.1239, found 249.1236.

cFN. <sup>1</sup>H NMR (250 MHz, DMSO):  $\delta$  8.10 (s, 1H, H<sup>N</sup> F or N), 7.62 (s, 1H, H<sup>N</sup> F or N), 7.32–7.17 (m, 6H, H <sup>$\delta$ - $\zeta$</sup>  F, H<sup>NH2</sup> N), 6.94 (s, 1H, H<sup>NH2</sup> N), 4.19 (m, 1H, H <sup>$\alpha$</sup>  F), 4.02 (m, 1H, H <sup>$\alpha$</sup>  N), 3.10 (dd,  $J = 13.75, 4.75$  Hz, 1H, H <sup>$\beta$</sup>  F), 2.92 (dd,  $J = 13.75, 5.25$  Hz, 1H, H <sup>$\beta$</sup>  F), 2.26 (dd,  $J = 16, 4$  Hz, 1H, H <sup>$\beta$</sup>  N), 1.386 (dd,  $J = 16, 8.75$  Hz, 1H, H <sup>$\beta$</sup>  N); <sup>13</sup>C NMR (62.8 MHz, DMSO):  $\delta$  171.6 (C<sup>N</sup>), 166.9 (C<sup>F</sup> or N), 166.4 (C<sup>F</sup> or N), 136.4 (C<sup>F</sup>), 130.1 (C <sup>$\delta$</sup>  F), 128.1 (C <sup>$\epsilon$</sup>  F), 126.7 (C <sup>$\zeta$</sup>  F), 55.2 (C <sup>$\alpha$</sup>  F), 51.1 (C <sup>$\alpha$</sup>  N), 38.1 (C <sup>$\beta$</sup>  F or N), 38.1 (C <sup>$\beta$</sup>  F or N). HRMS ( $m/z$ ): [MH<sup>+</sup>] calculated for C<sub>13</sub>H<sub>16</sub>N<sub>3</sub>O<sub>3</sub>, 262.1192, found 262.1186.

cFE. <sup>1</sup>H NMR (250 MHz, DMSO):  $\delta$  11.91 (s, 1H, H<sup>COOH</sup> E), 8.21 (s, 1H, H<sup>N</sup> F or E), 8.10 (s, 1H, H<sup>N</sup> F or E), 7.28–7.14 (m, 5H, H <sup>$\delta$ - $\zeta$</sup>  F), 4.19 (s, 1H, H <sup>$\alpha$</sup>  F), 3.69 (m, 1H, H <sup>$\alpha$</sup>  E), 3.14 (dd,  $J = 13.25, 3.5$  Hz, 1H, H <sup>$\beta$</sup>  F), 2.84 (dd,  $J = 13.25, 4.75$  Hz, 1H, H <sup>$\beta$</sup>  F), 1.66 (m, 2H, H <sup>$\gamma$</sup>  E), 1.32–0.97 (m, 2H, H <sup>$\beta$</sup>  E); <sup>13</sup>C NMR (62.8 MHz, DMSO):  $\delta$  173.9 (C <sup>$\delta$</sup>  E), 166.7 (C<sup>F</sup> or E), 166.4 (C<sup>F</sup> or E), 136.1 (C <sup>$\gamma$</sup>  F), 130.3 (C <sup>$\delta$</sup>  F), 128.1 (C <sup>$\epsilon$</sup>  F), 126.8 (C <sup>$\zeta$</sup>  F), 55.32 (C <sup>$\alpha$</sup>  F), 53.0 (C <sup>$\alpha$</sup>  E), 38.1 (C <sup>$\beta$</sup>  F), 28.6 (C <sup>$\beta$  or  $\gamma$</sup>  E), 28.5 (C <sup>$\beta$  or  $\gamma$</sup>  E). HRMS ( $m/z$ ): [MH<sup>+</sup>] calculated for C<sub>14</sub>H<sub>17</sub>N<sub>2</sub>O<sub>4</sub>, 277.1188, found 277.1187.

22. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 2001).
23. Altschul, S.F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
24. Schäffer, A.A. *et al.* Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.* **29**, 2994–3005 (2001).

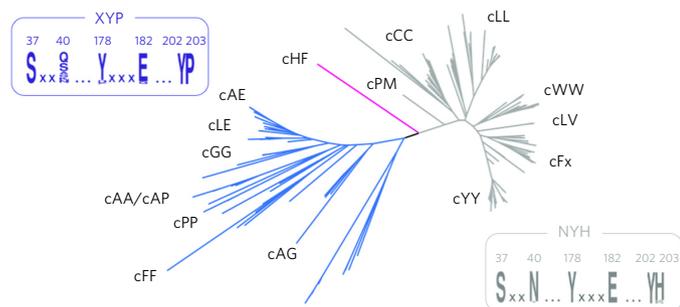
25. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
26. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).
27. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
28. Letunic, I. & Bork, P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* **39**, W475–W478 (2011).
29. Remmert, M., Biegert, A., Hauser, A. & Soding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2012).
30. Schneider, T.D. & Stephens, R.M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990).
31. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
32. Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
33. Braud, S. *et al.* Dual expression system suitable for high-throughput fluorescence-based screening and production of soluble proteins. *J. Proteome Res.* **4**, 2137–2147 (2005).
34. Bajad, S.U. *et al.* Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *J. Chromatogr. A* **1125**, 76–88 (2006).
35. Chen, Y.-H., Liou, S.-E. & Chen, C.-C. Two-step mass spectrometric approach for the identification of diketopiperazines in chicken essence. *Eur. Food Res. Technol.* **218**, 589–597 (2004).
36. Falick, A.M., Hines, W.M., Medzihradzky, K.F., Baldwin, M.A. & Gibson, B.W. Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **4**, 882–893 (1993).
37. Johnson, R.S., Martin, S.A., Biemann, K., Stults, J.T. & Watson, J.T. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Anal. Chem.* **59**, 2621–2625 (1987).
38. Papayannopoulos, I.A. The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrom. Rev.* **14**, 49–73 (1995).
39. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11**, 601 (1984).
40. Stark, T. & Hofmann, T. Structures, sensory activity, and dose/response functions of 2,5-diketopiperazines in roasted cocoa nibs (*Theobroma cacao*). *J. Agric. Food Chem.* **53**, 7222–7231 (2005).
41. Armirotti, A., Millo, E. & Damonte, G. How to discriminate between leucine and isoleucine by low energy ESI-TRAP MSn. *J. Am. Soc. Mass Spectrom.* **18**, 57–63 (2007).
42. Jeedigunta, S., Krenisky, J.M. & Kerr, R.G. Diketopiperazines as advanced intermediates in the biosynthesis of Ecteinascidins. *Tetrahedron* **56**, 3303–3307 (2000).

## Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity

Isabelle B Jacques, Mireille Moutiez, Jerzy Witwinowski, Emmanuelle Darbon, Cécile Martel, Jérôme Seguin, Emmanuel Favry, Robert Thai, Alain Lecoq, Steven Dubois, Jean-Luc Pernodet, Muriel Gondry & Pascal Belin

### EDITORIAL SUMMARY

**AOP:** Cyclodipeptide synthases use amino acid-loaded tRNAs as substrates to form cyclic peptide dimers. Biochemical and bioinformatic analyses now show that these enzymes use a broad range of substrates that can be predicted with newly defined functional motifs.



**ANNEXE 2 : DÉTAILS DES RÉSULTATS HRMS OBTENUS SUR LE CYCLO(ALA-GLU) ET LES COMPOSÉS IJ18, IJ19, IJ20, PB2, PB3, PB4 ET PB6.**

Composé	Ions précurseurs		Ions produits		Fragment neutre	Type de fragmentation	
	$m/z^1$	Formule brute <sup>2</sup>	$m/z^1$	Formule brute <sup>2</sup>			
cyclo(Ala-Glu)	<b>201.08697</b>	C <sub>8</sub> H <sub>13</sub> O <sub>4</sub> N <sub>2</sub>	183.07639	C <sub>8</sub> H <sub>11</sub> O <sub>3</sub> N <sub>2</sub>	H <sub>2</sub> O	MS2	CID
	183.07639	C <sub>8</sub> H <sub>11</sub> O <sub>3</sub> N <sub>2</sub>	155.08124	C <sub>7</sub> H <sub>11</sub> O <sub>2</sub> N <sub>2</sub>	CO	MS3	CID
	155.08124	C <sub>7</sub> H <sub>11</sub> O <sub>2</sub> N <sub>2</sub>	127.08635	C <sub>6</sub> H <sub>11</sub> ON <sub>2</sub>	CO	MS4	CID
			110.05977	C <sub>6</sub> H <sub>8</sub> ON	HCONH <sub>2</sub>	MS4	CID
			84.04407	C <sub>4</sub> H <sub>6</sub> ON	C <sub>3</sub> H <sub>5</sub> ON	MS4	HCD
			82.06483	C <sub>3</sub> H <sub>8</sub> ON	C <sub>2</sub> H <sub>3</sub> ON	MS4	HCD
IJ18	<b>217.08185</b>	C <sub>8</sub> H <sub>13</sub> O <sub>5</sub> N <sub>2</sub>				MS1	
	<b>199.07125</b>	C <sub>8</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	181.06056	C <sub>8</sub> H <sub>9</sub> O <sub>3</sub> N <sub>2</sub>	H <sub>2</sub> O	MS2	CID
			153.06561	C <sub>7</sub> H <sub>9</sub> O <sub>2</sub> N <sub>2</sub>	H <sub>2</sub> O + CO	MS2	CID
	153.06557	C <sub>7</sub> H <sub>9</sub> O <sub>2</sub> N <sub>2</sub>	126.05472	C <sub>6</sub> H <sub>9</sub> O <sub>2</sub> N	CN ?	MS3	HCD
			125.07070	C <sub>6</sub> H <sub>9</sub> ON <sub>2</sub>	CO	MS3	HCD
		108.04412	C <sub>6</sub> H <sub>6</sub> ON	HCONH <sub>2</sub>	MS3	HCD	
IJ19	<b>506.15522</b>	C <sub>18</sub> H <sub>28</sub> O <sub>10</sub> N <sub>5</sub> S	308.0912	C <sub>10</sub> H <sub>18</sub> O <sub>6</sub> N <sub>3</sub> S	C <sub>8</sub> H <sub>10</sub> O <sub>4</sub> N <sub>2</sub>	MS2	CID
	<b>272.55485</b>	?	199.00	C <sub>8</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	?	MS2	CID
	199.00	C <sub>8</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	181.06063	C <sub>8</sub> H <sub>9</sub> O <sub>3</sub> N <sub>2</sub>	H <sub>2</sub> O	MS3	CID
			153.06572	C <sub>7</sub> H <sub>9</sub> O <sub>2</sub> N <sub>2</sub>	H <sub>2</sub> O + CO	MS3	CID
	153.06572	C <sub>7</sub> H <sub>9</sub> O <sub>2</sub> N <sub>2</sub>	125.07078	C <sub>6</sub> H <sub>9</sub> ON <sub>2</sub>	CO	MS5	CID
IJ20 <sup>3</sup>	<b>506.06</b>	C <sub>18</sub> H <sub>28</sub> O <sub>10</sub> N <sub>5</sub> S	308.09137	C <sub>10</sub> H <sub>18</sub> O <sub>6</sub> N <sub>3</sub> S	C <sub>8</sub> H <sub>10</sub> O <sub>4</sub> N <sub>2</sub>	MS2	CID
	308.09137	C <sub>10</sub> H <sub>18</sub> O <sub>6</sub> N <sub>3</sub> S	233.05916	C <sub>8</sub> H <sub>13</sub> O <sub>4</sub> N <sub>2</sub> S	C <sub>2</sub> H <sub>5</sub> O <sub>2</sub> N	MS3	CID
			179.04849	C <sub>5</sub> H <sub>11</sub> O <sub>6</sub> N <sub>2</sub> S	C <sub>5</sub> H <sub>7</sub> N	MS3	CID
			162.02194	C <sub>5</sub> H <sub>8</sub> O <sub>3</sub> NS	C <sub>5</sub> H <sub>10</sub> O <sub>3</sub> N <sub>2</sub>	MS3	CID
	<b>272.56</b>	?	346.03803	?	?	MS2	CID
			199.07131	C <sub>8</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	?	MS2	CID
			181.06070	C <sub>8</sub> H <sub>9</sub> O <sub>3</sub> N <sub>2</sub>	?	MS2	CID
			153.06575	C <sub>7</sub> H <sub>9</sub> O <sub>2</sub> N <sub>2</sub>	?	MS2	CID
			272.25848	?	?	MS2	HCD
			153.06577	C <sub>7</sub> H <sub>9</sub> O <sub>2</sub> N <sub>2</sub>	?	MS2	HCD
	199.07131	C <sub>8</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	153.06577	C <sub>7</sub> H <sub>9</sub> O <sub>2</sub> N <sub>2</sub>	H <sub>2</sub> O + CO	MS3	HCD
			125.07083	C <sub>6</sub> H <sub>9</sub> ON <sub>2</sub>	H <sub>2</sub> O + CO + CO	MS3	HCD
			108.04424	C <sub>6</sub> H <sub>6</sub> ON	HCONH <sub>2</sub> + H <sub>2</sub> O + CO	MS3	HCD
	181.0607	C <sub>8</sub> H <sub>9</sub> O <sub>3</sub> N <sub>2</sub>	153.06575	C <sub>7</sub> H <sub>9</sub> O <sub>2</sub> N <sub>2</sub>	CO	MS3	HCD
			126.05485	C <sub>6</sub> H <sub>8</sub> O <sub>2</sub> N	CO + CN ?	MS3	HCD
		125.07083	C <sub>6</sub> H <sub>9</sub> ON <sub>2</sub>	CO + CO	MS3	HCD	

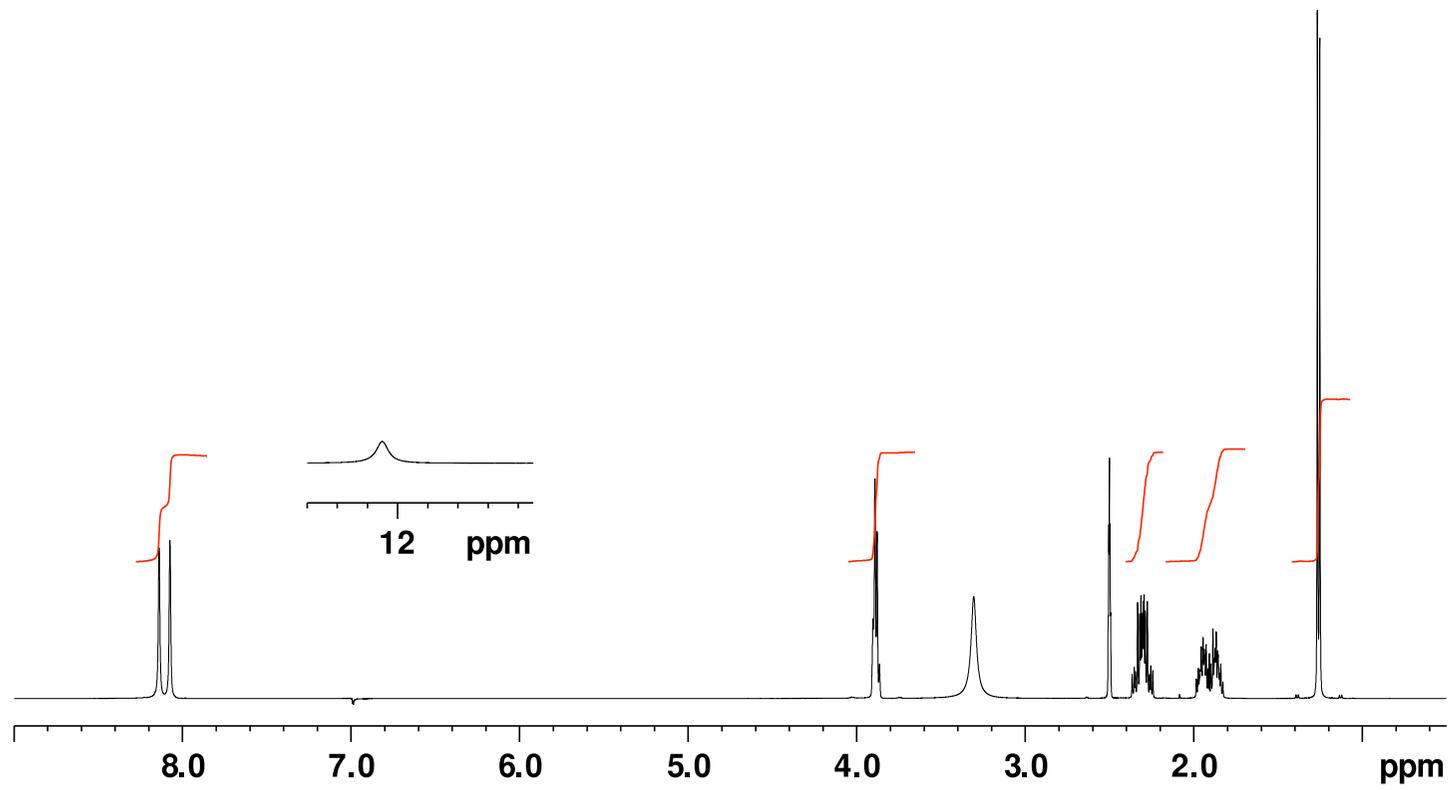
			108.04424	C <sub>6</sub> H <sub>6</sub> ON	HCONH <sub>2</sub> + CO	MS3	HCD
PB2	<b>217.08213</b>	C <sub>8</sub> H <sub>13</sub> O <sub>5</sub> N <sub>2</sub>	200.05531	C <sub>8</sub> H <sub>10</sub> O <sub>5</sub> N <sub>2</sub>	NH <sub>3</sub>	MS2	CID
			199.07137	C <sub>8</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	H <sub>2</sub> O	MS2	CID
			172.06041	C <sub>7</sub> H <sub>10</sub> O <sub>4</sub> N	HCONH <sub>3</sub>	MS2	CID
			154.04987	C <sub>7</sub> H <sub>8</sub> O <sub>3</sub> N	HCONH <sub>2</sub> + H <sub>2</sub> O	MS2	CID
			154.04984	C <sub>7</sub> H <sub>8</sub> O <sub>3</sub> N	HCONH <sub>2</sub> + H <sub>2</sub> O	MS2	HCD
			126.05489	C <sub>6</sub> H <sub>8</sub> O <sub>2</sub> N	CO	MS2	HCD
	<b>199.07153</b>	C <sub>8</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	181.06075	C <sub>8</sub> H <sub>9</sub> O <sub>3</sub> N <sub>2</sub>	H <sub>2</sub> O	MS2	CID
			153.06581	C <sub>7</sub> H <sub>9</sub> O <sub>2</sub> N <sub>2</sub>	H <sub>2</sub> O + CO	MS2	CID
	181.06075	C <sub>8</sub> H <sub>9</sub> O <sub>3</sub> N <sub>2</sub>	153.06588	C <sub>7</sub> H <sub>9</sub> O <sub>2</sub> N <sub>2</sub>	CO	MS3	CID
			125.07093	C <sub>7</sub> H <sub>9</sub> O <sub>3</sub> N <sub>2</sub>	CO + CO	MS3	CID
	153.06581	C <sub>7</sub> H <sub>9</sub> O <sub>3</sub> N <sub>2</sub>	125.07085	C <sub>6</sub> H <sub>9</sub> ON <sub>2</sub>	CO	MS3	CID
			108.04426	C <sub>6</sub> H <sub>6</sub> ON	HCONH <sub>2</sub>	MS3	CID
	125.07085	C <sub>6</sub> H <sub>9</sub> ON <sub>2</sub>	108.04432	C <sub>6</sub> H <sub>6</sub> ON	NH <sub>3</sub>	MS5	CID
		80.04932	C <sub>5</sub> H <sub>6</sub> N	CO	MS5	CID	
PB3	199.07	C <sub>8</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	181.06098		H <sub>2</sub> O	MS2	CID
			153.07		H <sub>2</sub> O + CO	MS2	CID
	181.06	C <sub>7</sub> H <sub>9</sub> O <sub>3</sub> N <sub>2</sub>	153.06602	C <sub>7</sub> H <sub>9</sub> O <sub>2</sub> N <sub>2</sub>	CO	MS3	HCD
			126.05503	C <sub>6</sub> H <sub>8</sub> O <sub>2</sub> N		MS3	HCD
			125.07102	C <sub>6</sub> H <sub>9</sub> ON <sub>2</sub>	CO + CO	MS3	HCD
			108.04439	C <sub>6</sub> H <sub>6</sub> ON	HCONH <sub>2</sub> + CO	MS3	HCD
PB4	<b>262.14</b>	??	217.0822	C <sub>8</sub> H <sub>13</sub> O <sub>5</sub> N <sub>2</sub>		MS2	CID
			200.0556	C <sub>8</sub> H <sub>10</sub> O <sub>5</sub> N		MS2	CID
	<b>217.08239</b>	C <sub>8</sub> H <sub>13</sub> O <sub>5</sub> N <sub>2</sub>	200.05561	C <sub>8</sub> H <sub>10</sub> O <sub>5</sub> N	NH <sub>3</sub>	MS2	CID
			172.06061	C <sub>7</sub> H <sub>10</sub> O <sub>4</sub> N	HCONH <sub>2</sub>	MS2	CID
			154.05001	C <sub>7</sub> H <sub>8</sub> O <sub>3</sub> N	HCONH <sub>2</sub> + H <sub>2</sub> O	MS2	CID
	200.06	C <sub>8</sub> H <sub>10</sub> O <sub>5</sub> N	172.06057	C <sub>7</sub> H <sub>10</sub> O <sub>4</sub> N	CO	MS3	CID
			154.04997	C <sub>7</sub> H <sub>8</sub> O <sub>3</sub> N	H <sub>2</sub> O + CO	MS3	CID
			126.05499	C <sub>6</sub> H <sub>8</sub> O <sub>2</sub> N	H <sub>2</sub> O + CO + CO	MS3	CID
	200.06	C <sub>8</sub> H <sub>10</sub> O <sub>5</sub> N	154.04998	C <sub>7</sub> H <sub>8</sub> O <sub>3</sub> N	H <sub>2</sub> O + CO	MS3	HCD
			126.055	C <sub>6</sub> H <sub>8</sub> O <sub>2</sub> N	H <sub>2</sub> O + CO + CO	MS3	HCD
		101.02327	C <sub>4</sub> H <sub>5</sub> O <sub>3</sub>	C <sub>4</sub> H <sub>5</sub> O <sub>2</sub> N	MS3	HCD	
		98.05998	C <sub>3</sub> H <sub>8</sub> ON	H <sub>2</sub> O + CO + CO + CO	MS3	HCD	
PB6	199.07176	C <sub>8</sub> H <sub>11</sub> O <sub>4</sub> N <sub>2</sub>	181.061002	C <sub>8</sub> H <sub>9</sub> O <sub>3</sub> N <sub>2</sub>	H <sub>2</sub> O	MS2	CID
			153.06605	C <sub>7</sub> H <sub>9</sub> O <sub>2</sub> N <sub>2</sub>	H <sub>2</sub> O + CO	MS2	CID
	153.066	C <sub>7</sub> H <sub>9</sub> O <sub>2</sub> N <sub>2</sub>	126.05503	C <sub>6</sub> H <sub>8</sub> O <sub>2</sub> N	CN ?	MS4	CID

	125.07101	C <sub>6</sub> H <sub>9</sub> ON <sub>2</sub>	CO	MS4	CID
	108.0444	C <sub>6</sub> H <sub>6</sub> ON	HCONH <sub>2</sub>	MS4	CID
	97.07603	C <sub>3</sub> H <sub>9</sub> N <sub>2</sub>	CO + CO	MS4	CID
	80.04939	C <sub>5</sub> H <sub>6</sub> N	HCONH <sub>2</sub> + CO	MS4	CID

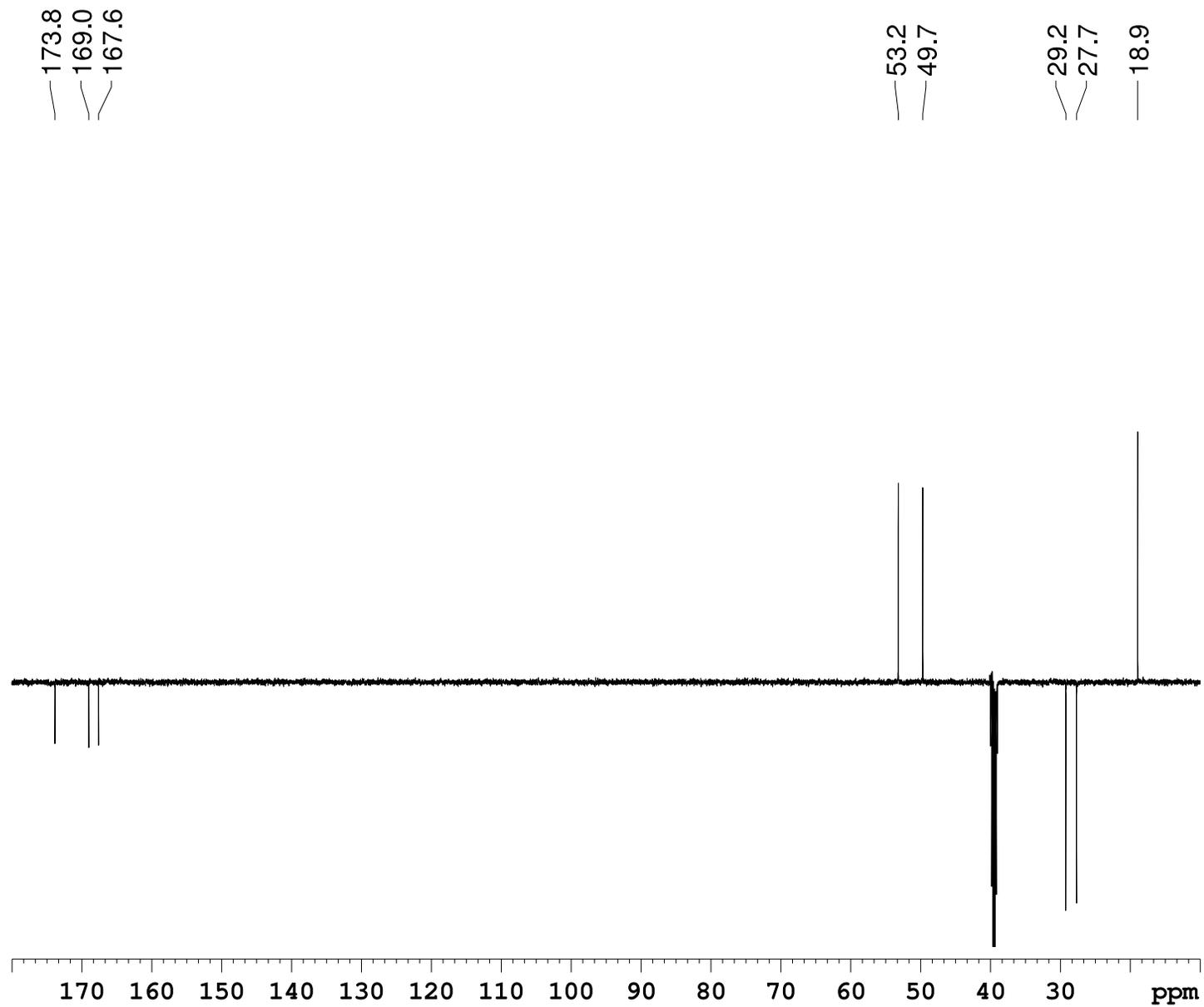
<sup>1</sup>Il s'agit d'ion monochargé sauf pour  $m/z$  272,56 qui est dichargé. Les ions en gras sont les ions moléculaires probables. <sup>2</sup>Les calculs de formules brutes ont été réalisés avec logiciel Thermo Xcalibur 2.1 (Qual Browser). Les formules brutes et les fragments proposés sont des hypothèses effectuées à partir des calculs réalisés par le logiciel. Lorsqu'il existe une forte incertitude sur les formules brutes ou les fragments proposés, les données sont en gris. <sup>3</sup>Les masses hautes résolutions des deux ions précurseurs majoritaires ne sont pas disponibles car le spectre MS1 n'a pas été enregistré pour IJ20 ; il était équivalent à celui de IJ19. La formule brute correspondante à  $m/z$  506,16 (en gris) est déduite de celle déterminée pour IJ19.

**ANNEXE 3: SÉLECTION DE SPECTRES RMN AYANT SERVI A L'ÉLUCIDATION STRUCTURALE DE PB2, PB3, PB4 ET PB6.**

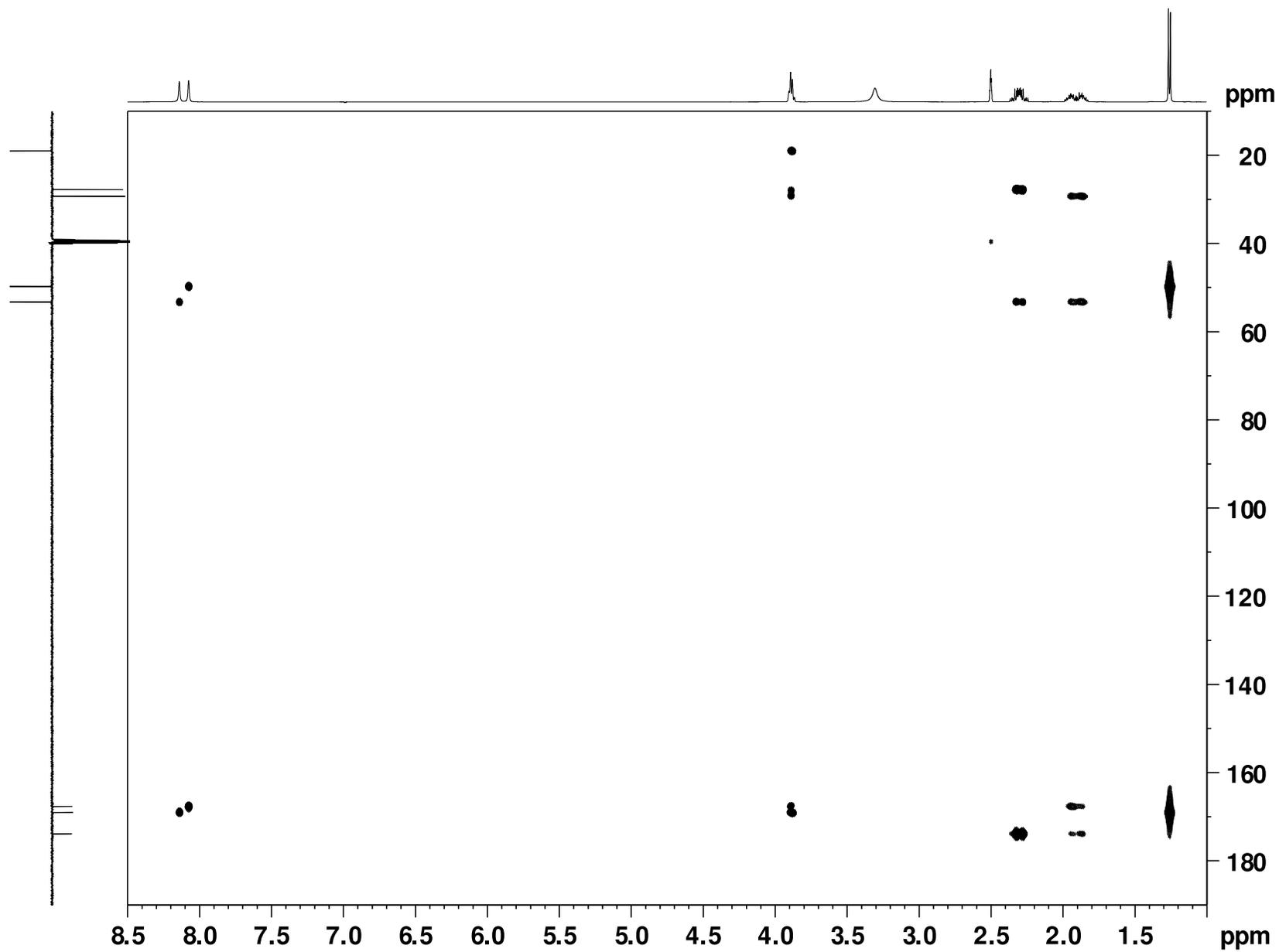
cEA (DMSO-d6, 30 °C)  
Spectre 1D 1H



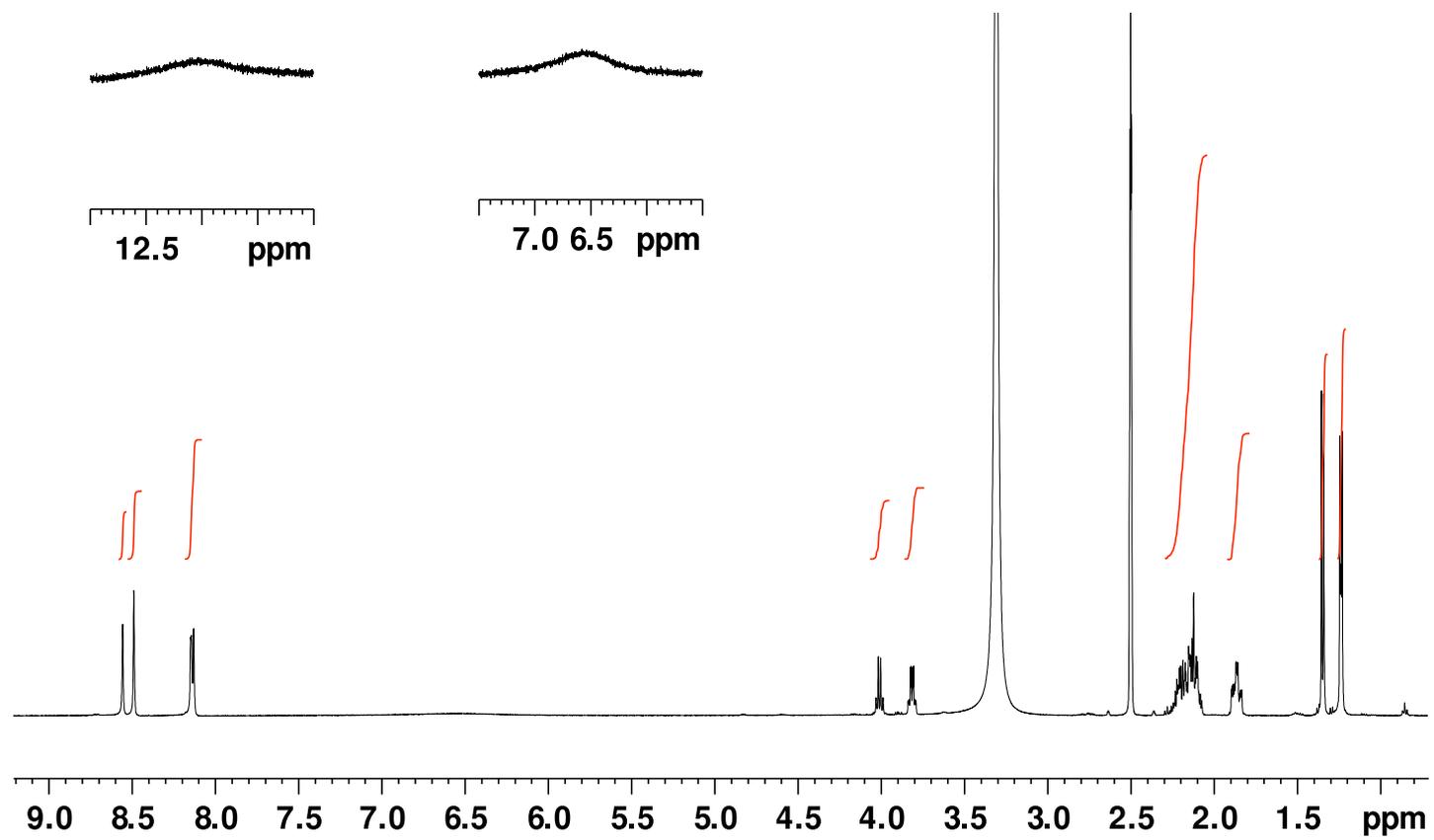
cEA (DMSO-d6, 30 °C)  
Spectre 1D 13C (DEPTQ)



cEA (DMSO-d6, 30°C)  
Spectre 2D HMBC 1H-13C



PB2 (DMSO-d6, 30°C)  
Spectre 1H



PB2 (DMSO-d6, 30 °C)  
DEPTQ 13C

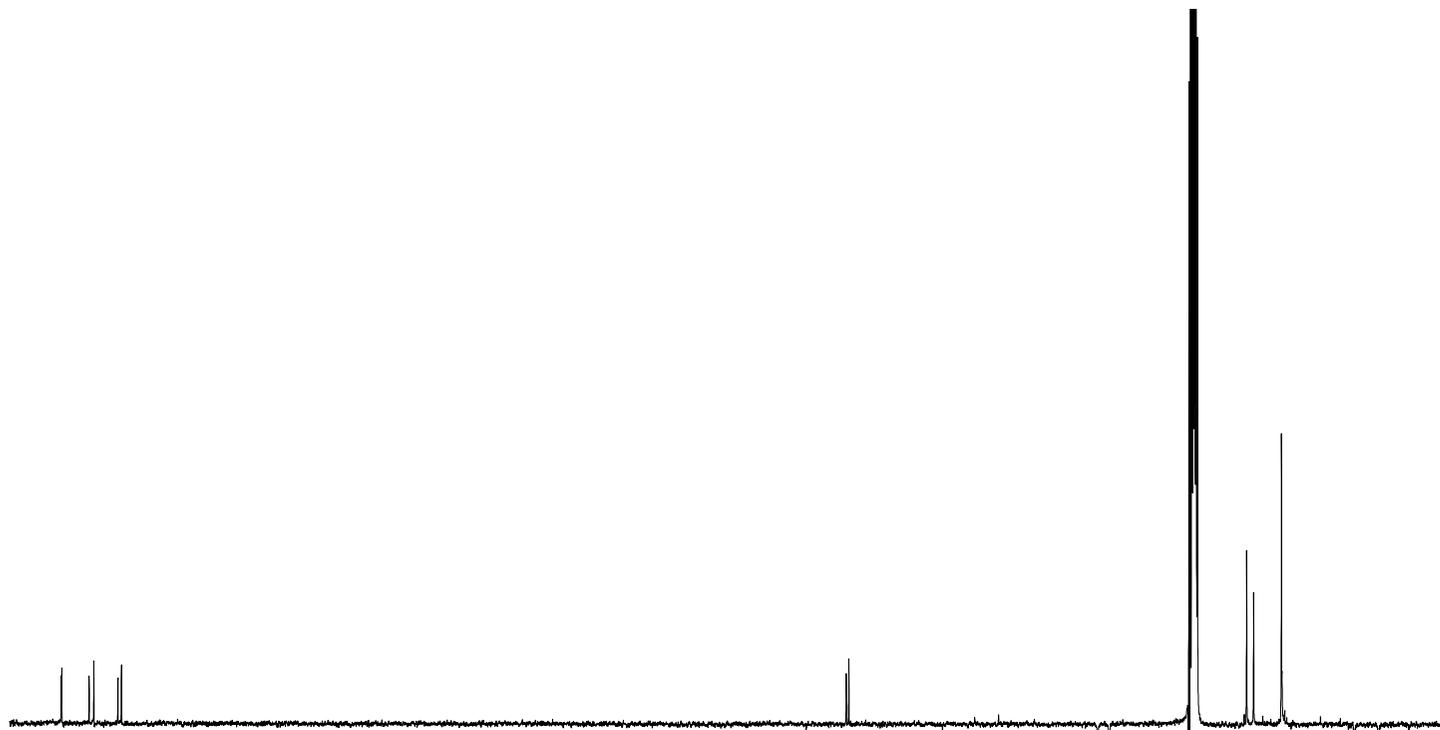
173.85  
173.78  
170.54  
169.98  
167.12  
166.70

80.68  
80.39

50.85  
49.49

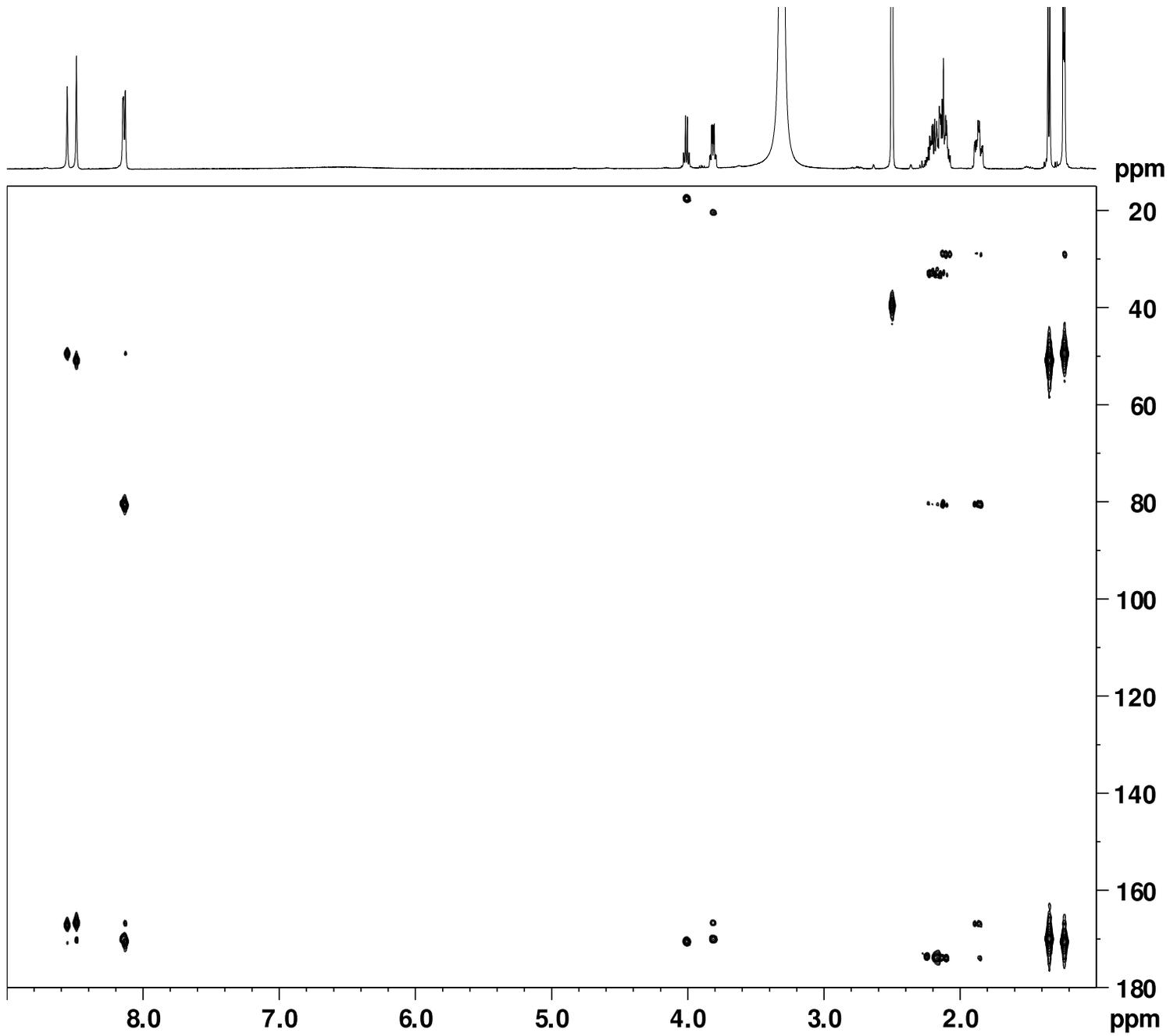
33.19  
32.36  
29.07

20.42  
17.56

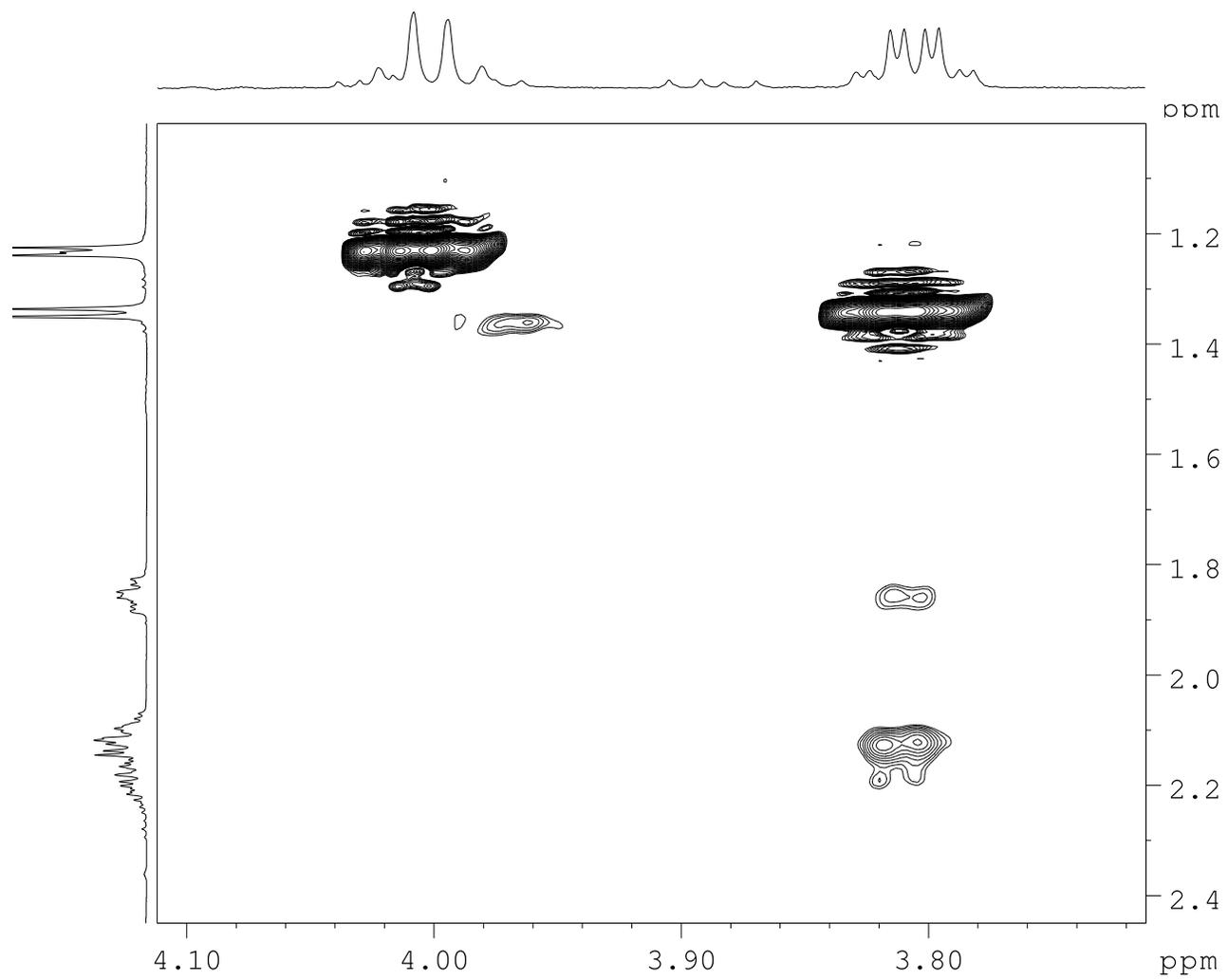


170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 ppm

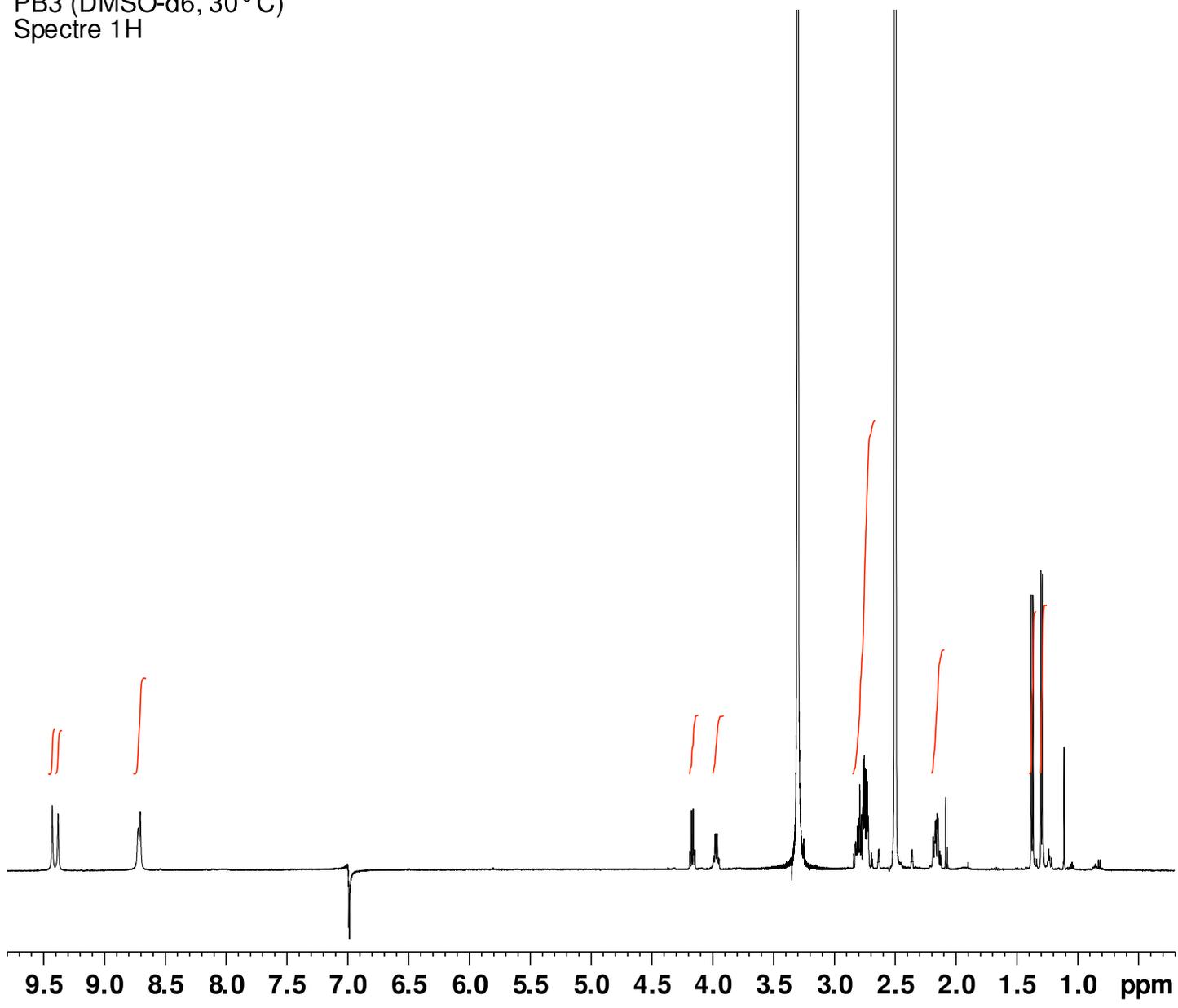
PB2 (DMSO-d6, 30 °C)  
HMBC 13C-1H



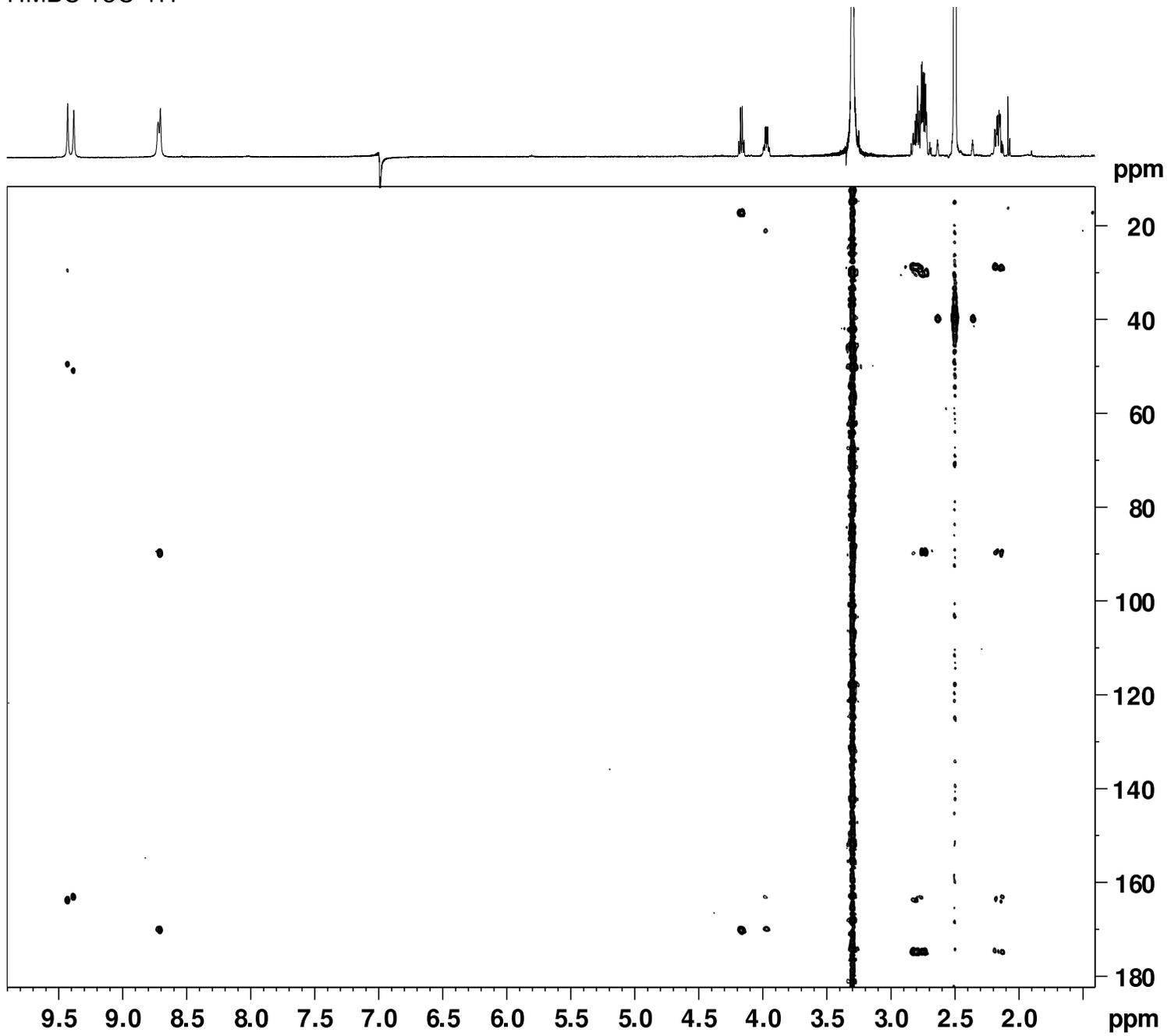
Expansion de la carte NOESY (DMSO-d6, 30°C, Temps de mélange 2s) montrant la proximité spatiale des protons Halpha Ala et Hbeta Glu du produit PB2B (S,S)



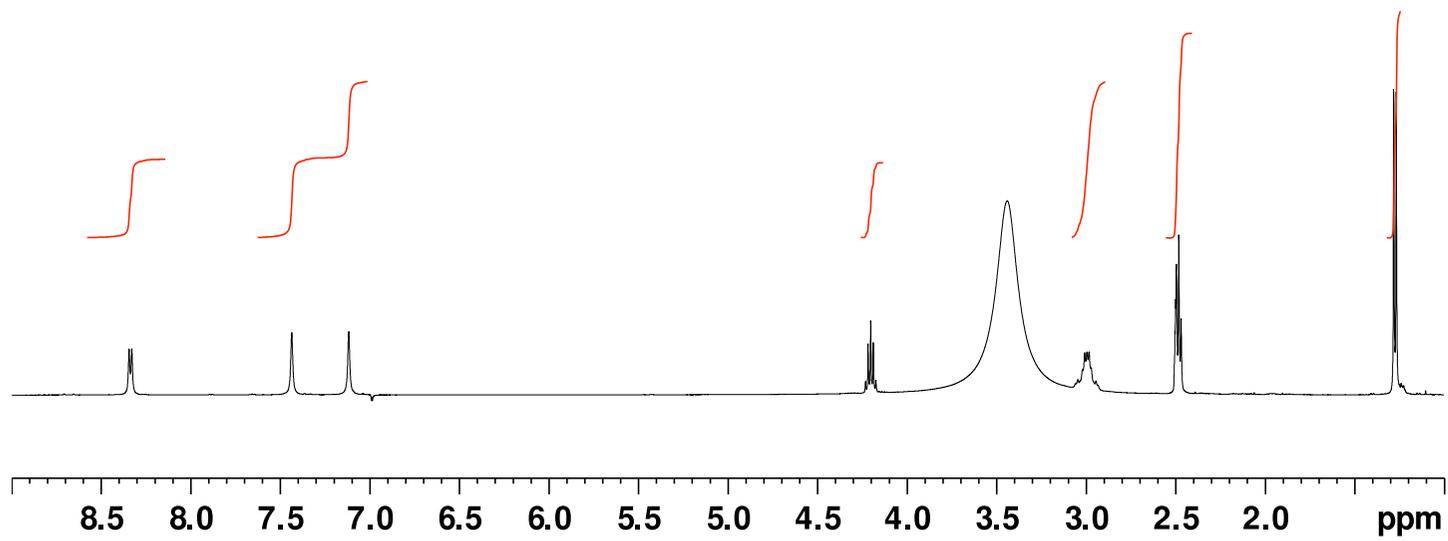
PB3 (DMSO-d6, 30°C)  
Spectre 1H



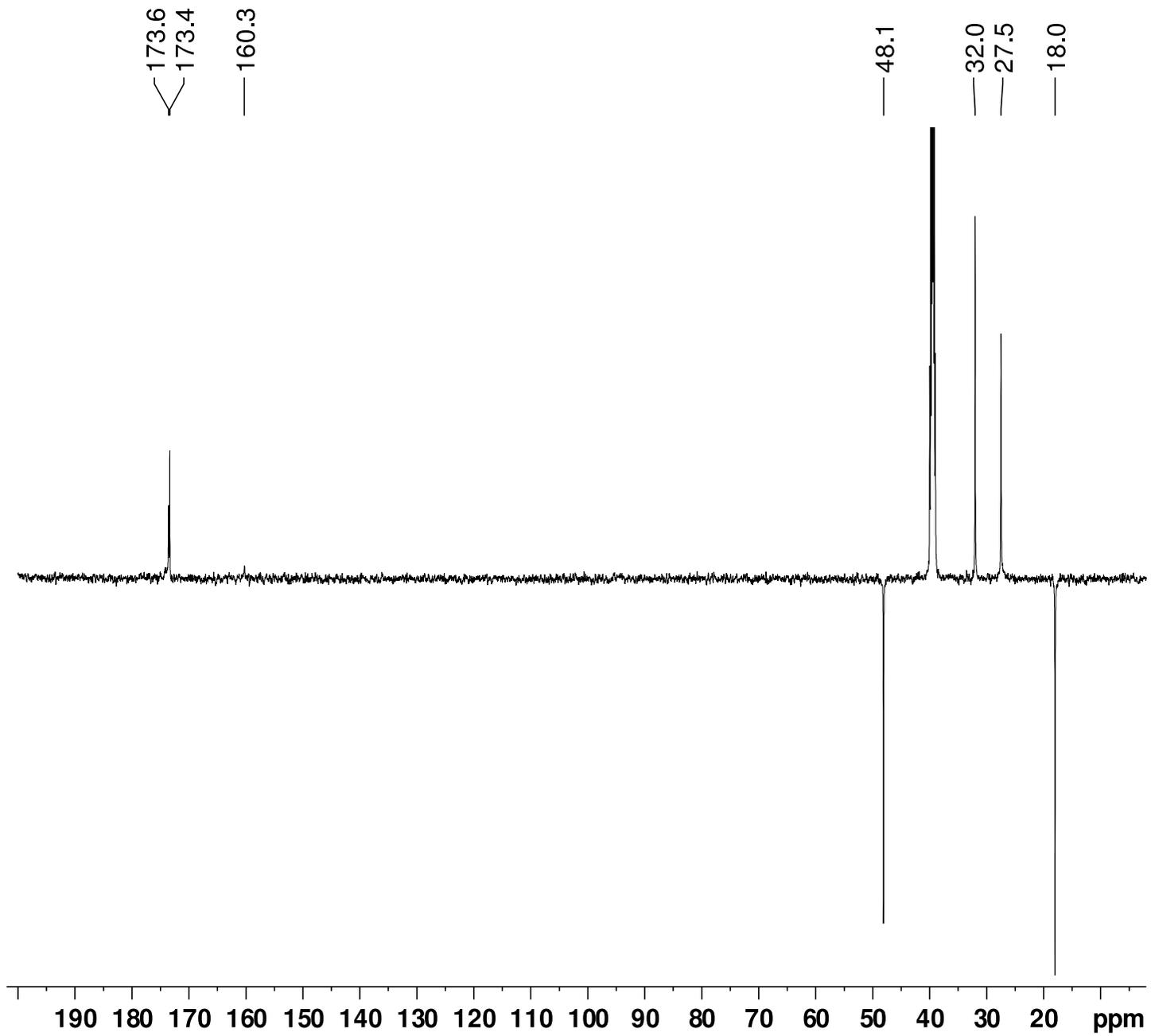
PB3 (DMSO-d6, 30 °C)  
HMBC 13C-1H



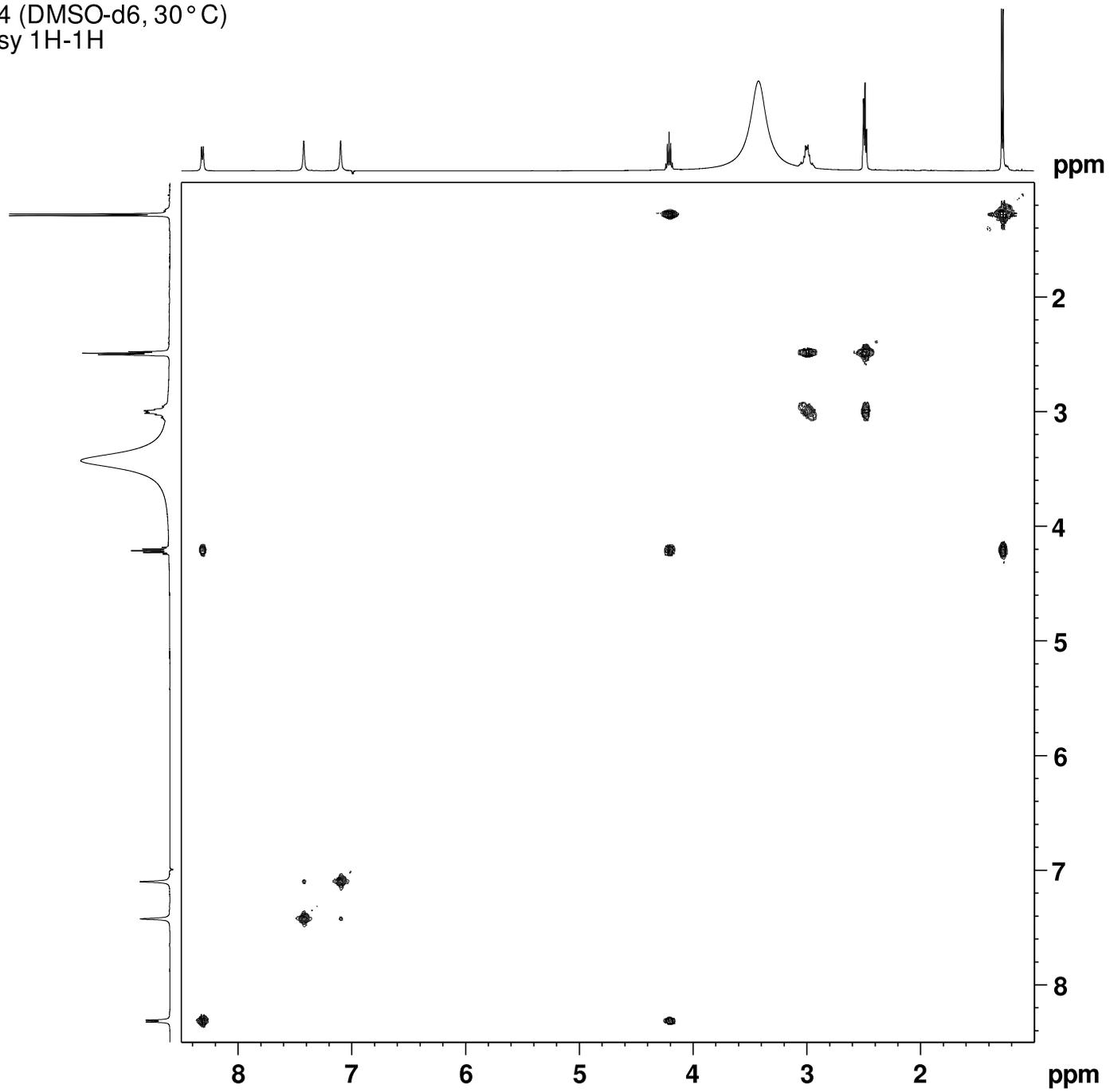
PB4 (DMSO-d6, 30° C)  
Spectre 1H



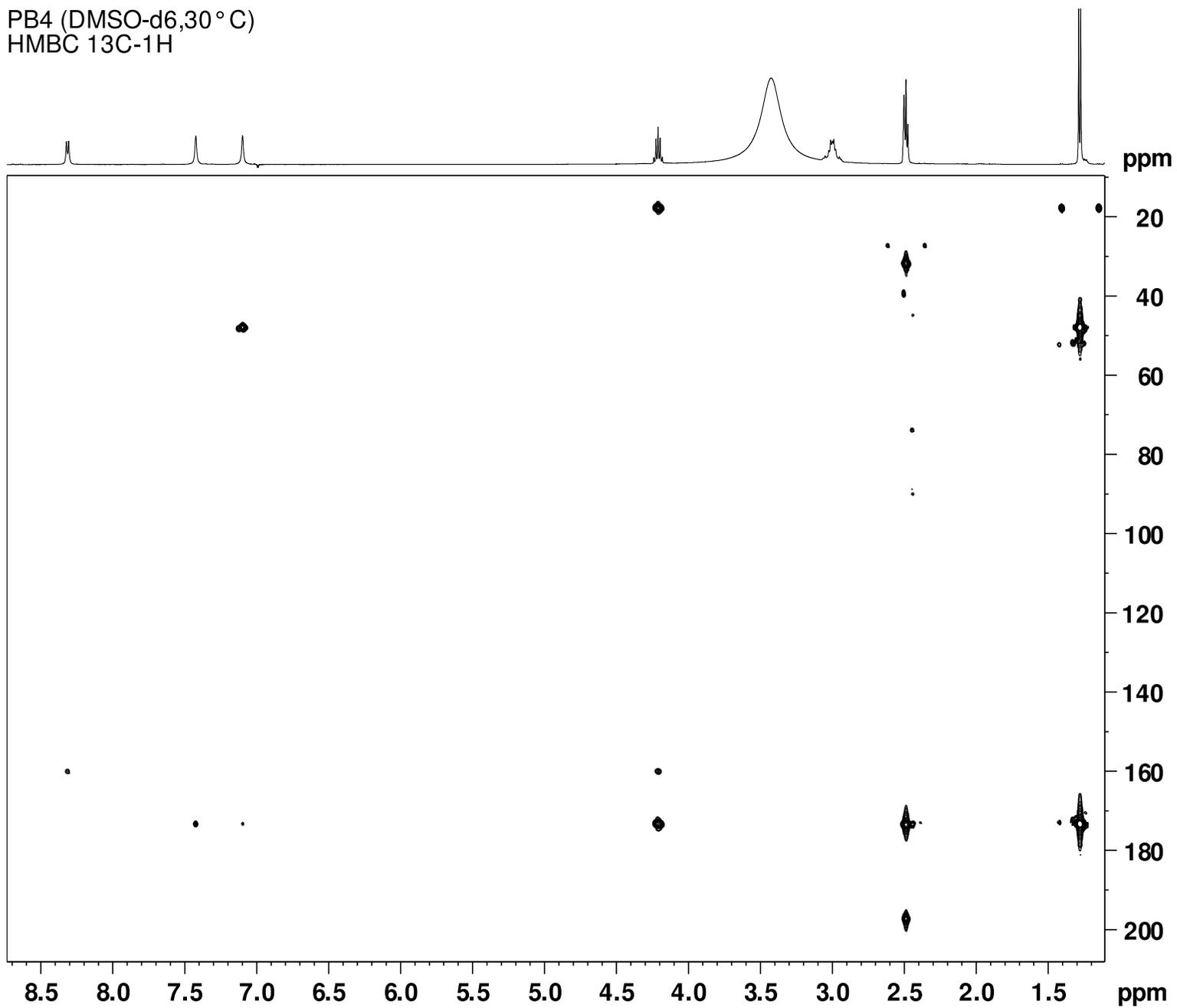
PB4 (DMSO-d6, 30°C)  
DEPTQ 13C



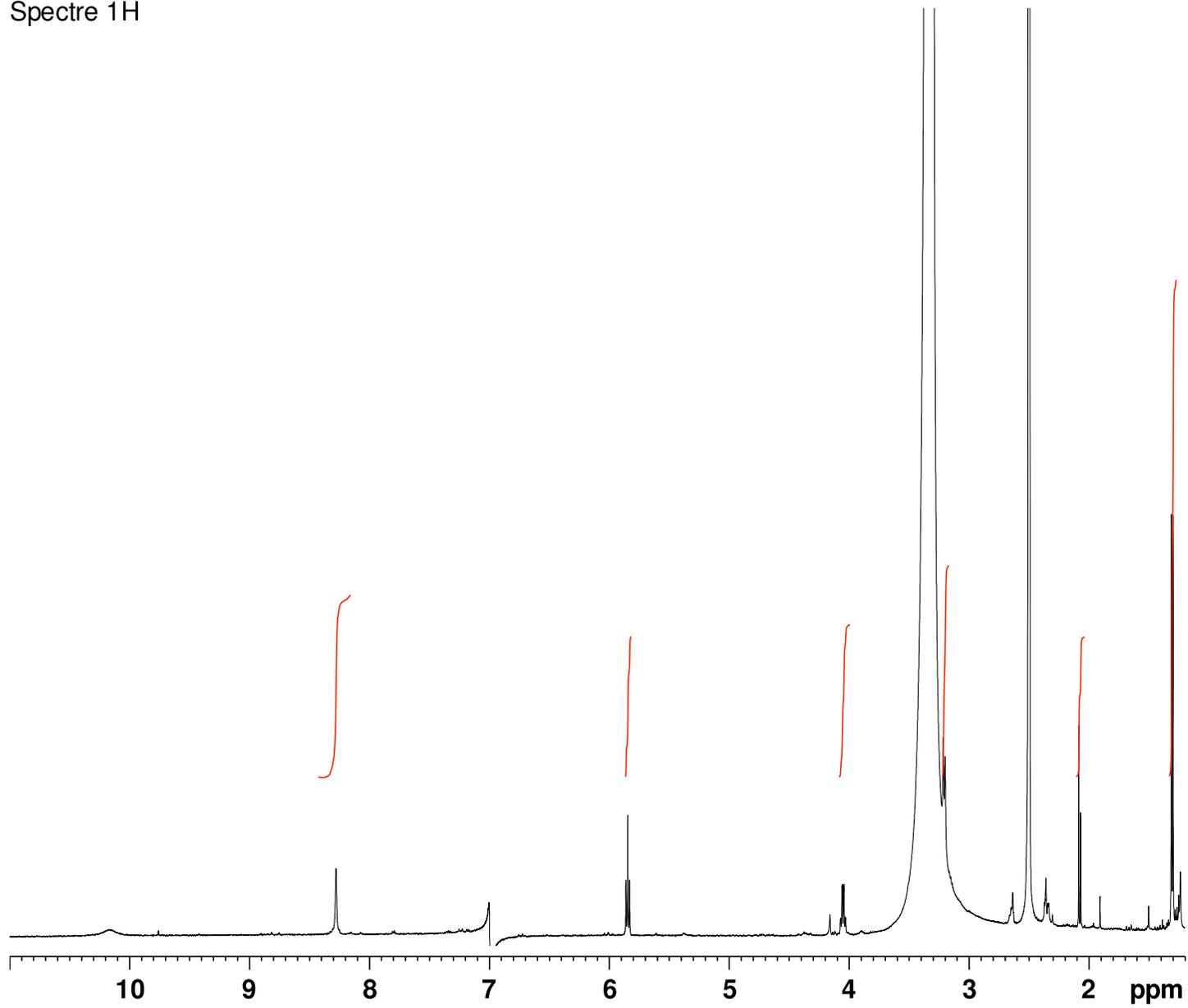
PB4 (DMSO-d6, 30 °C)  
Cosy 1H-1H



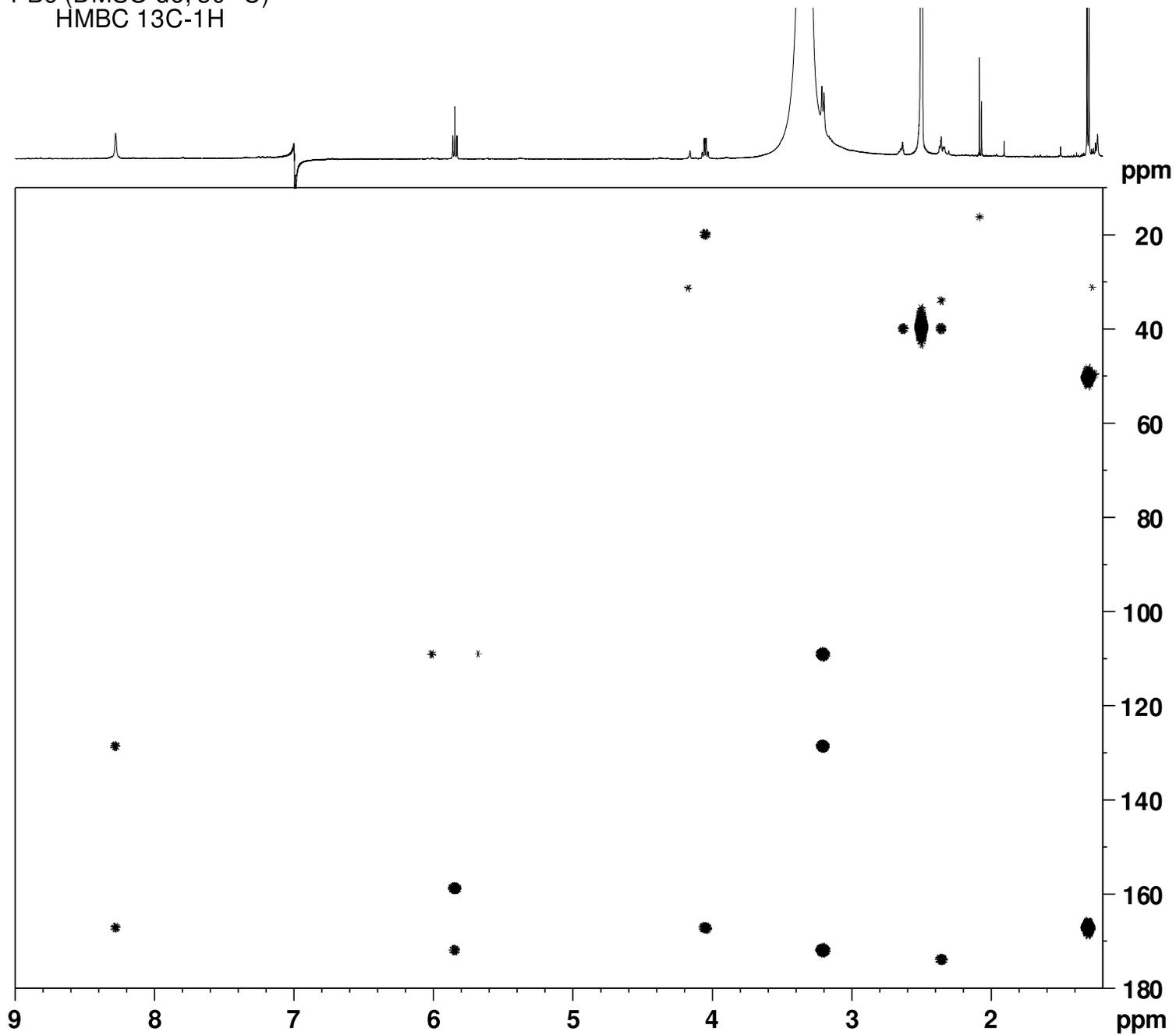
PB4 (DMSO-d<sub>6</sub>, 30 °C)  
HMBC 13C-1H



PB6 (DMSO-d6, 30 °C)  
Spectre 1H



PB6 (DMSO-d6, 30 °C)  
HMBC 13C-1H



## **RÉFÉRENCES**

1. Cragg, G. M. & Newman, D. J. Natural products: A continuing source of novel drug leads. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1830**, 3670–3695 (2013).
2. Newman, D. J. & Cragg, G. M. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *J. Nat. Prod.* **75**, 311–335 (2012).
3. Meunier, B. Does chemistry have a future in therapeutic innovations? *Angew. Chem. Int. Ed. Engl.* **51**, 8702–8706 (2012).
4. Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* **14**, 111–129 (2015).
5. Feher, M. & Schmidt, J. M. Property Distributions: Differences between Drugs, Natural Products, and Molecules from Combinatorial Chemistry. *J. Chem. Inf. Comput. Sci.* **43**, 218–227 (2003).
6. Clardy, J. & Walsh, C. Lessons from natural molecules. *Nature* **432**, 829–837 (2004).
7. Pace, N. R. A Molecular View of Microbial Diversity and the Biosphere. *Science* **276**, 734–740 (1997).
8. Walsh, C. T. & Fischbach, M. A. Natural Products Version 2.0: Connecting Genes to Molecules. *J. Am. Chem. Soc.* **132**, 2469–2493 (2010).
9. Bachmann, B. O., Lanen, S. G. V. & Baltz, R. H. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J Ind Microbiol Biotechnol* **41**, 175–184 (2014).
10. Wu, M.-C., Law, B., Wilkinson, B. & Micklefield, J. Bioengineering natural product biosynthetic pathways for therapeutic applications. *Current Opinion in Biotechnology* **23**, 931–940 (2012).
11. Challis, G. L. Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology* **154**, 1555–1569 (2008).
12. de Boer, A. L. & Schmidt-Dannert, C. Recent efforts in engineering microbial cells to produce new chemical compounds. *Curr Opin Chem Biol* **7**, 273–278 (2003).
13. Wilkinson, B. & Micklefield, J. Mining and engineering natural-product biosynthetic pathways. *Nat Chem Biol* **3**, 379–386 (2007).
14. Walsh, C. T. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303**, 1805–1810 (2004).
15. Walsh, C. T. The chemical versatility of natural-product assembly lines. *Acc. Chem. Res.* **41**, 4–10 (2008).
16. Giessen, T. W. & Marahiel, M. A. Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity. *FEBS Letters* **586**, 2065–2075 (2012).
17. Kanoh, K., Kohno, S., Katada, J., Takahashi, J. & Uno, I. (-)-Phenylahistin arrests cells in mitosis by inhibiting tubulin polymerization. *J. Antibiot.* **52**, 134–141 (1999).
18. Cui, C. B., Kakeya, H., Okada, G., Onose, R. & Osada, H. Novel mammalian cell cycle inhibitors, tryprostatins A, B and other diketopiperazines produced by *Aspergillus fumigatus*. I. Taxonomy, fermentation, isolation and biological properties. *J. Antibiot.* **49**, 527–533 (1996).
19. Kohn, H. & Widger, W. The Molecular Basis for the Mode of Action of Bicyclomycin. *Current Drug Targets - Infectious Disorders* **5**, 273–295 (2005).
20. Ström, K., Sjögren, J., Broberg, A. & Schnürer, J. *Lactobacillus plantarum* MiLAB 393 Produces the Antifungal Cyclic Dipeptides Cyclo(l-Phe-l-Pro) and Cyclo(l-Phe-trans-4-OH-l-Pro) and 3-Phenyllactic Acid. *Appl. Environ. Microbiol.* **68**, 4322–4327 (2002).
21. Waring, P., Eichner, R. D. & Müllbacher, A. The chemistry and biology of the immunomodulating agent gliotoxin and related epipolythiodioxopiperazines. *Med Res Rev* **8**, 499–524 (1988).
22. Rodriguez, P. L. & Carrasco, L. Gliotoxin: inhibitor of poliovirus RNA synthesis that blocks the viral RNA polymerase 3Dpol. *J. Virol.* **66**, 1971–1976 (1992).
23. Minelli, A. *et al.* Phosphoproteomic analysis of the effect of cyclo-[His-Pro] dipeptide on PC12 cells. *Peptides* **27**, 105–113 (2006).

24. Gondry, M. *et al.* Cyclic dipeptide oxidase from *Streptomyces noursei*. *European Journal of Biochemistry* **268**, 1712–1721 (2001).
25. Lautru, S., Gondry, M., Genet, R. & Pernodet, J.-L. The Albonoursin Gene Cluster of *S. noursei*: Biosynthesis of Diketopiperazine Metabolites Independent of Nonribosomal Peptide Synthetases. *Chemistry & Biology* **9**, 1355–1364 (2002).
26. Gondry, M. *et al.* Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. *Nature Chemical Biology* **5**, 414–420 (2009).
27. Seguin, J. *et al.* Nonribosomal Peptide Synthesis in Animals: The Cyclodipeptide Synthase of *Nematostella*. *Chemistry & Biology* **18**, 1362–1368 (2011).
28. Aravind, L., Souza, R. F. de & Iyer, L. M. Predicted class-I aminoacyl tRNA synthetase-like proteins in non-ribosomal peptide synthesis. *Biology Direct* **5**, 48 (2010).
29. Belin, P. *et al.* 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 (2009).
30. Prasad, C. Bioactive cyclic dipeptides. *Peptides* **16**, 151–164 (1995).
31. Borthwick, A. D. & Da costa, N. C. 2,5-Diketopiperazines in Food and Beverages: Taste and Bioactivity. *Critical Reviews in Food Science and Nutrition* **0**, 00–00 (2015).
32. Perry, T. L., Richardson, K. S. C., Hansen, S. & Friesen, A. J. D. Identification of the Diketopiperazine of Histidylproline in Human Urine. *J. Biol. Chem.* **240**, 4540–4542 (1965).
33. Cole, R. J. *et al.* Tremorgenic Toxin from *Penicillium verrucosum*. *Appl Microbiol* **24**, 248–250 (1972).
34. Miyoshi, T., Miyairi, N., Aoki, H., Kosaka, M. & Sakai, H. Bicyclomycin, a new antibiotic. I. Taxonomy, isolation and characterization. *J. Antibiot.* **25**, 569–575 (1972).
35. Fukushima, K., Yazawa, K. & Arai, T. Biological activities of albonoursin. *J. Antibiot.* **26**, 175–176 (1973).
36. Corey, R. B. The Crystal Structure of Diketopiperazine. *J. Am. Chem. Soc.* **60**, 1598–1604 (1938).
37. Kopple, K. D. & Ohnishi, M. Conformations of cyclic peptides. II. Side-chain conformation and ring shape in cyclic dipeptides. *J. Am. Chem. Soc.* **91**, 962–975 (1969).
38. Davies, D. B. & Khaled, M. A. Conformations of peptides in solution by nuclear magnetic resonance spectroscopy. Part II. Homoallylic coupling in cyclic dipeptides. *Journal of the Chemical Society, Perkin Transactions 2* 187 (1976). doi:10.1039/p29760000187
39. Mendham, A. P., Dines, T. J., Snowden, M. J., Chowdhry, B. Z. & Withnall, R. Vibrational spectroscopy and DFT calculations of di-amino acid cyclic peptides. Part I: cyclo(Gly-Gly), cyclo(L-Ala-L-Ala) and cyclo(L-Ala-Gly) in the solid state and in aqueous solution. *J. Raman Spectrosc.* **40**, 1478–1497 (2009).
40. Bettens, F. L., Bettens, R. P. A., Brown, R. D. & Godfrey, P. D. The Microwave Spectrum, Structure, and Ring-Puckering of the Cyclic Dipeptide Diketopiperazine. *J. Am. Chem. Soc.* **122**, 5856–5860 (2000).
41. Hirst, J. D. & Persson, B. J. *Ab Initio* Calculations of the Vibrational and Electronic Spectra of Diketopiperazine. *J. Phys. Chem. A* **102**, 7519–7524 (1998).
42. Zhu, Y., Tang, M., Shi, X. & Zhao, Y. Quantum chemical study of cyclic dipeptides. *Int. J. Quantum Chem.* **107**, 745–753 (2007).
43. MacDonald, J. C. & Whitesides, G. M. Solid-State Structures of Hydrogen-Bonded Tapes Based on Cyclic Secondary Diamides. *Chem. Rev.* **94**, 2383–2420 (1994).
44. Madison, V., Young, P. E. & Blout, E. R. Cyclic peptides. 14. Conformational energy and circular dichroism of proline-containing cyclic dipeptides. *J. Am. Chem. Soc.* **98**, 5358–5364 (1976).
45. Li, Z. & Mukamel, S. First-Principles Simulation of Amide and Aromatic Side Chain Ultraviolet Spectroscopy of a Cyclic Dipeptide. *J. Phys. Chem. A* **111**, 11579–11583 (2007).
46. Borthwick, A. D. 2,5-Diketopiperazines: Synthesis, Reactions, Medicinal Chemistry, and Bioactive Natural Products. *Chem. Rev.* **112**, 3641–3716 (2012).

47. Ando, S., Grote, A. L. & Koide, K. Diastereoselective Synthesis of Diketopiperazine Bis- $\alpha,\beta$ -epoxides. *J. Org. Chem.* **76**, 1155–1158 (2011).
48. Ressurreição, A. S. M., Delatouche, R., Gennari, C. & Piarulli, U. Bifunctional 2,5-Diketopiperazines as Rigid Three-Dimensional Scaffolds in Receptors and Peptidomimetics. *Eur. J. Org. Chem.* **2011**, 217–228 (2011).
49. Khokhlov, A. S. & Lokshin, G. B. The structure of albonoursin. *Tetrahedron Letters* **4**, 1881–1885 (1963).
50. Loria, R. Differential Production of Thaxtomins by Pathogenic *Streptomyces* Species *In Vitro*. *Phytopathology* **85**, 537 (1995).
51. Scott, P. M., Merrien, M.-A. & Polonsky, J. Roquefortine and isofumigaclavine A, metabolites from *Penicillium roqueforti*. *Experientia* **32**, 140–142 (1976).
52. Kanoh, K. *et al.* (-)-Phenylahistin: A new mammalian cell cycle inhibitor produced by *Aspergillus ustus*. *Bioorganic & Medicinal Chemistry Letters* **7**, 2847–2852 (1997).
53. Johnson, J. R., Bruce, W. F. & Dutcher, J. D. Gliotoxin, The Antibiotic Principle of *Gliocladium fimbriatum*. I. Production, Physical and Biological Properties. *J. Am. Chem. Soc.* **65**, 2005–2009 (1943).
54. Carr, G. *et al.* Plectosphaeroic Acids A, B, and C, Indoleamine 2,3-Dioxygenase Inhibitors Produced in Culture by a Marine Isolate of the Fungus *Plectosphaerella cucumerina*. *Org. Lett.* **11**, 2996–2999 (2009).
55. Huang, R.-M. *et al.* An update on 2,5-diketopiperazines from marine organisms. *Mar Drugs* **12**, 6213–6235 (2014).
56. Bellezza, I., Peirce, M. J. & Minelli, A. Cyclic dipeptides: from bugs to brain. *Trends in Molecular Medicine* doi:10.1016/j.molmed.2014.08.003
57. Birch, A. J. & Wright, J. J. Studies in relation to biosynthesis. XLII. The structural elucidation and some aspects of the biosynthesis of the brevianamides-A and -E. *Tetrahedron* **26**, 2329–2344 (1970).
58. Chen, Y.-H., Liou, S.-E. & Chen, C.-C. Two-step mass spectrometric approach for the identification of diketopiperazines in chicken essence. *Eur Food Res Technol* **218**, 589–597 (2004).
59. Furtado, N. A. J. C. *et al.* Fragmentation of diketopiperazines from *Aspergillus fumigatus* by electrospray ionization tandem mass spectrometry (ESI-MS/MS). *Journal of Mass Spectrometry* **42**, 1279–1286 (2007).
60. Xing, J., Yang, Z., Lv, B. & Xiang, L. Rapid screening for cyclo-dopa and diketopiperazine alkaloids in crude extracts of *Portulaca oleracea* L. using liquid chromatography/tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **22**, 1415–1422 (2008).
61. Lamm, A., Gozlan, I., Rotstein, A. & Avisar, D. Detection of amoxicillin-diketopiperazine-2', 5' in wastewater samples. *J Environ Sci Health A Tox Hazard Subst Environ Eng* **44**, 1512–1517 (2009).
62. Guo, Y.-C., Cao, S.-X., Zong, X.-K., Liao, X.-C. & Zhao, Y.-F. ESI-MS<sup>n</sup> study on the fragmentation of protonated cyclic-dipeptides. *Spectroscopy-an International Journal* **23**, 131–139 (2009).
63. Brack, C., Mikolasch, A. & Schauer, F. 2,5-Diketopiperazines produced by *Bacillus pumilus* during bacteriolysis of *Arthrobacter citreus*. *Mar. Biotechnol.* **16**, 385–395 (2014).
64. Szafranek, J., Palacz, Z. & Grzonka, Z. A comparison of electron impact and field ionization spectra of some 2,5-diketopiperazines. *Org. Mass Spectrom.* **11**, 920–930 (1976).
65. Guo, Y. *et al.* Fragmentation of deprotonated cyclic dipeptides by electrospray ionization mass spectrometry. *Journal of Mass Spectrometry* **44**, 1188–1194 (2009).
66. Rogalewicz, F., Hoppilliard, Y. & Ohanessian, G. Fragmentation mechanisms of  $\alpha$ -amino acids protonated under electrospray ionization: a collisional activation and *ab initio* theoretical study. *International Journal of Mass Spectrometry* **195–196**, 565–590 (2000).
67. Ginz, M. & Engelhardt, U. H. Identification of Proline-Based Diketopiperazines in Roasted Coffee. *J. Agric. Food Chem.* **48**, 3528–3532 (2000).
68. Falick, A. M., Hines, W. M., Medzihradzsky, K. F., Baldwin, M. A. & Gibson, B. W. Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass spectrometry. *Journal of the American Society for Mass Spectrometry* **4**, 882–893 (1993).

69. Papayannopoulos, I. A. The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrom. Rev.* **14**, 49–73 (1995).
70. Riedel, K. *et al.* N-Acylhomoserine-lactone-mediated communication between *Pseudomonas aeruginosa* and *Burkholderia cepacia* in mixed biofilms. *Microbiology* **147**, 3249–3262 (2001).
71. Ortiz-Castro, R. *et al.* Transkingdom signaling based on bacterial cyclodipeptides with auxin activity in plants. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7253–7258 (2011).
72. Sperandio, V., Torres, A. G., Jarvis, B., Nataro, J. P. & Kaper, J. B. Bacteria–host communication: The language of hormones. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8951–8956 (2003).
73. Hughes, D. T. & Sperandio, V. Inter-kingdom signalling: communication between bacteria and their hosts. *Nat Rev Micro* **6**, 111–120 (2008).
74. Pacheco, A. R. & Sperandio, V. Inter-kingdom signaling: chemical language between bacteria and host. *Current Opinion in Microbiology* **12**, 192–198 (2009).
75. González, J. F. & Venturi, V. A novel widespread interkingdom signaling circuit. *Trends Plant Sci.* **18**, 167–174 (2013).
76. Gray, K. M., Pearson, J. P., Downie, J. A., Boboye, B. E. & Greenberg, E. P. Cell-to-cell signaling in the symbiotic nitrogen-fixing bacterium *Rhizobium leguminosarum*: autoinduction of a stationary phase and rhizosphere-expressed genes. *J. Bacteriol.* **178**, 372–376 (1996).
77. Schripsema, J. *et al.* Bacteriocin small of *Rhizobium leguminosarum* belongs to the class of N-acyl-L-homoserine lactone molecules, known as autoinducers and as quorum sensing co-transcription factors. *J. Bacteriol.* **178**, 366–371 (1996).
78. Thorne, S. H. & Williams, H. D. Cell Density-Dependent Starvation Survival of *Rhizobium leguminosarum* bv. phaseoli: Identification of the Role of an N-Acyl Homoserine Lactone in Adaptation to Stationary-Phase Survival. *J. Bacteriol.* **181**, 981–990 (1999).
79. Chapon-Hervé, V. *et al.* Regulation of the xcp secretion pathway by multiple quorum-sensing modulons in *Pseudomonas aeruginosa*. *Molecular Microbiology* **24**, 1169–1178 (1997).
80. Piper, K. R., Beck von Bodman, S. & Farrand, S. K. Conjugation factor of *Agrobacterium tumefaciens* regulates Ti plasmid transfer by autoinduction. *Nature* **362**, 448–450 (1993).
81. Fuqua, W. C. & Winans, S. C. A LuxR-LuxI type regulatory system activates *Agrobacterium* Ti plasmid conjugal transfer in the presence of a plant tumor metabolite. *J. Bacteriol.* **176**, 2796–2806 (1994).
82. Wood, D. W. & Pierson III, L. S. The phzI gene of *Pseudomonas aureofaciens* 30–84 is responsible for the production of a diffusible signal required for phenazine antibiotic production. *Gene* **168**, 49–53 (1996).
83. Kleerebezem, M. Quorum sensing control of lantibiotic production; nisin and subtilin autoregulate their own biosynthesis. *Peptides* **25**, 1405–1414 (2004).
84. Jones, S. *et al.* The lux autoinducer regulates the production of exoenzyme virulence determinants in *Erwinia carotovora* and *Pseudomonas aeruginosa*. *EMBO J* **12**, 2477–2482 (1993).
85. Passador, L., Cook, J. M., Gambello, M. J., Rust, L. & Iglewski, B. H. Expression of *Pseudomonas aeruginosa* virulence genes requires cell-to-cell communication. *Science* **260**, 1127–1130 (1993).
86. Pearson, J. P. *et al.* Structure of the autoinducer required for expression of *Pseudomonas aeruginosa* virulence genes. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 197–201 (1994).
87. Pearson, J. P., Passador, L., Iglewski, B. H. & Greenberg, E. P. A second N-acylhomoserine lactone signal produced by *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 1490–1494 (1995).
88. Latifi, A. *et al.* Multiple homologues of LuxR and LuxI control expression of virulence determinants and secondary metabolites through quorum sensing in *Pseudomonas aeruginosa* PAO1. *Molecular Microbiology* **17**, 333–343 (1995).
89. Winson, M. K. *et al.* Multiple N-acyl-L-homoserine lactone signal molecules regulate production of virulence determinants and secondary metabolites in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9427–9431 (1995).

90. Passador, L. *et al.* Functional analysis of the *Pseudomonas aeruginosa* autoinducer PAI. *J. Bacteriol.* **178**, 5995–6000 (1996).
91. Latifi, A., Foglino, M., Tanaka, K., Williams, P. & Lazdunski, A. A hierarchical quorum-sensing cascade in *Pseudomonas aeruginosa* links the transcriptional activators LasR and RhIR (VsmR) to expression of the stationary-phase sigma factor RpoS. *Molecular Microbiology* **21**, 1137–1146 (1996).
92. Flavier, A. B., Ganova-Raeva, L. M., Schell, M. A. & Denny, T. P. Hierarchical autoinduction in *Ralstonia solanacearum*: control of acyl-homoserine lactone production by a novel autoregulatory system responsive to 3-hydroxypalmitic acid methyl ester. *J. Bacteriol.* **179**, 7089–7097 (1997).
93. Nasser, W., Bouillant, M. L., Salmond, G. & Reverchon, S. Characterization of the *Erwinia chrysanthemi* expI–expR locus directing the synthesis of two N-acyl-homoserine lactone signal molecules. *Molecular Microbiology* **29**, 1391–1405 (1998).
94. Reverchon, S., Bouillant, M. L., Salmond, G. & Nasser, W. Integration of the quorum-sensing system in the regulatory networks controlling virulence factor synthesis in *Erwinia chrysanthemi*. *Molecular Microbiology* **29**, 1407–1418 (1998).
95. Lewenza, S., Conway, B., Greenberg, E. P. & Sokol, P. A. Quorum Sensing in *Burkholderia cepacia*: Identification of the LuxRI Homologs CepRI. *J. Bacteriol.* **181**, 748–756 (1999).
96. Garg, N., Manchanda, G. & Kumar, A. Bacterial quorum sensing: circuits and applications. *Antonie van Leeuwenhoek* **105**, 289–305 (2013).
97. Swift, S., Rowe, M. C. & Kamath, M. in *Bacterial Physiology* (ed. El-Sharoud, W.) 179–232 (Springer Berlin Heidelberg, 2008).
98. Nealson, K. H., Platt, T. & Hastings, J. W. Cellular Control of the Synthesis and Activity of the Bacterial Luminescent System. *J. Bacteriol.* **104**, 313–322 (1970).
99. Nealson, K. H. & Hastings, J. W. Bacterial bioluminescence: its control and ecological significance. *Microbiol Rev* **43**, 496–518 (1979).
100. Fuqua, C., Parsek, M. R. & Greenberg, E. P. REGULATION OF GENE EXPRESSION BY CELL-TO-CELL COMMUNICATION: Acyl-Homoserine Lactone Quorum Sensing. *Annual Review of Genetics* **35**, 439–468 (2001).
101. Lerat, E. & Moran, N. A. The Evolutionary History of Quorum-Sensing Systems in Bacteria. *Mol Biol Evol* **21**, 903–913 (2004).
102. Waters, C. M. & Bassler, B. L. QUORUM SENSING: Cell-to-Cell Communication in Bacteria. *Annual Review of Cell and Developmental Biology* **21**, 319–346 (2005).
103. Schauder, S., Shokat, K., Surette, M. G. & Bassler, B. L. The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule. *Molecular Microbiology* **41**, 463–476 (2001).
104. Reading, N. C. & Sperandio, V. Quorum sensing: the many languages of bacteria. *FEMS Microbiology Letters* **254**, 1–11 (2006).
105. Wang, J.-H., Quan, C.-S., Qi, X.-H., Li, X. & Fan, S.-D. Determination of diketopiperazines of *Burkholderia cepacia* CF-66 by gas chromatography–mass spectrometry. *Anal Bioanal Chem* **396**, 1773–1779 (2010).
106. Sun, J., Daniel, R., Wagner-Döbler, I. & Zeng, A.-P. Is autoinducer-2 a universal signal for interspecies communication: a comparative genomic and phylogenetic analysis of the synthesis and signal transduction pathways. *BMC Evolutionary Biology* **4**, 36 (2004).
107. Turovskiy, Y., Kashtanov, D., Paskhover, B. & Chikindas, M. L. Quorum Sensing: Fact, Fiction, and Everything in Between. *Adv Appl Microbiol* **62**, 191–234 (2007).
108. Steindler, L. & Venturi, V. Detection of quorum-sensing N-acyl homoserine lactone signal molecules by bacterial biosensors. *FEMS Microbiology Letters* **266**, 1–9 (2007).
109. 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. *Molecular Microbiology* **33**, 1254–1266 (1999).

110. 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).
111. Teasdale, M. E., Donovan, K. A., Forscher-Dancause, S. R. & Rowley, D. C. Gram-Positive Marine Bacteria as a Potential Resource for the Discovery of Quorum Sensing Inhibitors. *Mar Biotechnol* **13**, 722–732 (2010).
112. Henke, J. M. & Bassler, B. L. Three Parallel Quorum-Sensing Systems Regulate Gene Expression in *Vibrio harveyi*. *J. Bacteriol.* **186**, 6902–6914 (2004).
113. Tommonaro, G., Abbamondi, G. R., Iodice, C., Tait, K. & Rosa, S. D. Diketopiperazines Produced by the Halophilic Archaeon, *Haloterrigena hispanica*, Activate AHL Bioreporters. *Microb Ecol* **63**, 490–495 (2011).
114. Huvenne, H., Goeminne, G., Maes, M. & Messens, E. Identification of quorum sensing signal molecules and oligolignols associated with watermark disease in willow (*Salix* sp.). *Journal of Chromatography B* **872**, 83–89 (2008).
115. Campbell, J., Lin, Q., Geske, G. D. & Blackwell, H. E. New and Unexpected Insights into the Modulation of LuxR-Type Quorum Sensing by Cyclic Dipeptides. *ACS Chem. Biol.* **4**, 1051–1059 (2009).
116. Campbell, J. & Blackwell, H. E. Efficient Construction of Diketopiperazine Macroarrays Through a Cyclative-Cleavage Strategy and Their Evaluation as Luminescence Inhibitors in the Bacterial Symbiont *Vibrio fischeri*. *J Comb Chem* **11**, 1094–1099 (2009).
117. Holden, M., Swift, S. & Williams, P. New signal molecules on the quorum-sensing block. *Trends in Microbiology* **8**, 101–103 (2000).
118. Jimenez, P. N. *et al.* The Multiple Signaling Systems Regulating Virulence in *Pseudomonas aeruginosa*. *Microbiol. Mol. Biol. Rev.* **76**, 46–65 (2012).
119. 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).
120. Park, D.-K. *et al.* Cyclo(Phe-Pro) Modulates the Expression of ompU in *Vibrio* spp. *J Bacteriol* **188**, 2214–2221 (2006).
121. Bina, X. R. & Bina, J. E. The Cyclic Dipeptide Cyclo(Phe-Pro) Inhibits Cholera Toxin and Toxin-Coregulated Pilus Production in O1 El Tor *Vibrio cholerae*. *J. Bacteriol.* **192**, 3829–3832 (2010).
122. Bina, X. R., Taylor, D. L., Vikram, A., Ante, V. M. & Bina, J. E. *Vibrio cholerae* ToxR Downregulates Virulence Factor Production in Response to Cyclo(Phe-Pro). *mBio* **4**, e00366–13 (2013).
123. Vikram, A. *et al.* Cyclo(valine-valine) inhibits *Vibrio cholerae* virulence gene expression. *Microbiology* **160**, 1054–1062 (2014).
124. Guo, X. *et al.* A case study on chemical defense based on quorum sensing: antibacterial activity of sponge-associated bacterium *Pseudoalteromonas* sp. NJ6-3-1 induced by quorum sensing mechanisms. *Ann Microbiol* **61**, 247–255 (2010).
125. Gu, Q., Fu, L., Wang, Y. & Lin, J. Identification and Characterization of Extracellular Cyclic Dipeptides As Quorum-Sensing Signal Molecules from *Shewanella baltica*, the Specific Spoilage Organism of *Pseudosciaena crocea* during 4 °C Storage. *J. Agric. Food Chem.* **61**, 11645–11652 (2013).
126. Zhu, S., Wu, H., Zeng, M., Liu, Z. & Wang, Y. The involvement of bacterial quorum sensing in the spoilage of refrigerated *Litopenaeus vannamei*. *International Journal of Food Microbiology* **192**, 26–33 (2015).
127. Loria, R. *et al.* Thaxtomin biosynthesis: the path to plant pathogenicity in the genus *Streptomyces*. *Antonie van Leeuwenhoek* **94**, 3–10 (2008).
128. Gardiner, D. M., Cozijnsen, A. J., Wilson, L. M., Pedras, M. S. C. & Howlett, B. J. The sirodesmin biosynthetic gene cluster of the plant pathogenic fungus *Leptosphaeria maculans*. *Molecular Microbiology* **53**, 1307–1318 (2004).

129. King, R. R., Lawrence, C. H. & Calhoun, L. A. Chemistry of phytotoxins associated with *Streptomyces scabies* the causal organism of potato common scab. *J. Agric. Food Chem.* **40**, 834–837 (1992).
130. Healy, F. G., Wach, M., Krasnoff, S. B., Gibson, D. M. & Loria, R. The txtAB genes of the plant pathogen *Streptomyces acidiscabies* encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity. *Molecular Microbiology* **38**, 794–804 (2000).
131. Bignell, D. R. D., Francis, I. M., Fyans, J. K. & Loria, R. Thaxtomin A Production and Virulence Are Controlled by Several bld Gene Global Regulators in *Streptomyces scabies*. *MPMI* **27**, 875–885 (2014).
132. Elliott, C. E. *et al.* Production of the toxin sirodesmin PL by *Leptosphaeria maculans* during infection of *Brassica napus*. *Molecular Plant Pathology* **8**, 791–802 (2007).
133. Elliott, C. E., Fox, E. M., Jarvis, R. S. & Howlett, B. J. The cross-pathway control system regulates production of the secondary metabolite toxin, sirodesmin PL, in the ascomycete, *Leptosphaeria maculans*. *BMC Microbiology* **11**, 169 (2011).
134. Müller, G., Barclay, S. J. & Raymond, K. N. The mechanism and specificity of iron transport in *Rhodotorula pilimanae* probed by synthetic analogs of rhodotorulic acid. *J. Biol. Chem.* **260**, 13916–13920 (1985).
135. Kluyver, A. J., van der Walt, J. P. & van Triet, A. J. Pulcherrimin, The Pigment of *Candida Pulcherrima*. *Proc. Natl. Acad. Sci. U.S.A.* **39**, 583–593 (1953).
136. Sipiczki, M. Metschnikowia Strains Isolated from Botrytized Grapes Antagonize Fungal and Bacterial Growth by Iron Depletion. *Appl. Environ. Microbiol.* **72**, 6716–6724 (2006).
137. Boiko, S. S., Gudasheva, T. A., Vichuzhanin, M. V., Zherdev, V. P. & Seredenin, S. B. Regional and Subcellular Localization of Cycloprolylglycine in Rat Brain. *Bull Exp Biol Med* **149**, 709–711 (2010).
138. Fujimoto, H., Fujimaki, T., Okuyama, E. & Yamazaki, M. Immunomodulatory Constituents from an Ascomycete, *Microascus tardifaciens*. *Chemical & Pharmaceutical Bulletin* **47**, 1426–1432 (1999).
139. Ravikanth, V. *et al.* An immunosuppressive tryptophan-derived alkaloid from *Lepidagathis cristata*. *Phytochemistry* **58**, 1263–1266 (2001).
140. Hayashi, H., Takiuchi, K., Murao, S. & Arai, M. Structure and Insecticidal Activity of New Indole Alkaloids, Okaramines A and B, from *Penicillium simplicissimum* AK-40. *Agricultural and Biological Chemistry* **53**, 461–469 (1989).
141. Paterson, R. R. M., Simmonds, M. J. S., Kimmelmeier, C. & Blaney, W. M. Effects of brevianamide A, its photolysis product brevianamide D, and ochratoxin A from two *Penicillium* strains on the insect pests *Spodoptera frugiperda* and *Heliothis virescens*. *Mycological Research* **94**, 538–542 (1990).
142. Hayashi, H., Fujiwara, T., Murao, S. & Arai, M. Okaramine C, a New Insecticidal Indole Alkaloid from *Penicillium simplicissimum*. *Agricultural and Biological Chemistry* **55**, 3143–3145 (1991).
143. SHIONO, Y., AKIYAMA, K. & HAYASHI, H. Okaramines N, O, P, Q and R, New Okaramine Congeners, from *Penicillium simplicissimum* ATCC 90288. *Bioscience, Biotechnology, and Biochemistry* **64**, 103–110 (2000).
144. Walchshofer, D. N. *et al.* Anthelmintic activity of 3,6-dibenzyl-2,5-dioxopiperazine, cyclo(L-Phe-L-Phe). *Amino Acids* **12**, 41–47 (1997).
145. Sprogøe, K., Manniche, S., Larsen, T. O. & Christophersen, C. Janoxepin and brevicompanine B: antiplasmodial metabolites from the fungus *Aspergillus janus*. *Tetrahedron* **61**, 8718–8721 (2005).
146. Watts, K. R. *et al.* Assessing the trypanocidal potential of natural and semi-synthetic diketopiperazines from two deep water marine-derived fungi. *Bioorganic & Medicinal Chemistry* **18**, 2566–2574 (2010).
147. Jamie, H., Kilian, G., Dyason, K. & Milne, P. J. The effect of the isomers of cyclo(Trp-Pro) on heart and ion-channel activity. *Journal of Pharmacy and Pharmacology* **54**, 1659–1665 (2002).
148. Kilian, G., Jamie, H., Brauns, S. C. A., Dyason, K. & Milne, P. J. Biological activity of selected tyrosine-containing 2,5-diketopiperazines. *Pharmazie* **60**, 305–309 (2005).
149. Takase, S., Kawai, Y., Uchida, I., Tanaka, H. & Aoki, H. Structure of amauromine, a new hypotensive vasodilator produced by *amauroascus* sp. *Tetrahedron* **41**, 3037–3048 (1985).

150. Hof, H. & Kupfahl, C. Gliotoxin in *Aspergillus fumigatus*: an example that mycotoxins are potential virulence factors. *Mycotox Res* **25**, 123–131 (2009).
151. Yamazaki, M., Suzuki, S. & Miyaki, K. Tremorgenic toxins from *Aspergillus fumigatus* Fres. *Chem. Pharm. Bull.* **19**, 1739–1740 (1971).
152. Arnold, D. L., Scott, P. M., McGuire, P. F., Harwig, J. & Nera, E. A. Acute toxicity studies on roquefortine and PR toxin, metabolites of *Penicillium roqueforti*, in the mouse. *Food and Cosmetics Toxicology* **16**, 369–371 (1978).
153. Aninat, C., Hayashi, Y., André, F. & Delaforge, M. Molecular Requirements for Inhibition of Cytochrome P450 Activities by Roquefortine. *Chem. Res. Toxicol.* **14**, 1259–1265 (2001).
154. Sabater-Vilar, M., Nijmeijer, S. & Fink-Gremmels, J. Genotoxicity Assessment of Five Tremorgenic Mycotoxins (Fumitremorgen B, Paxilline, Penitrem A, Verruculogen, and Verrucosidin) Produced by Molds Isolated from Fermented Meats. *Journal of Food Protection* **66**, 2123–2129 (2003).
155. Milne, P. J., Hunt, A. L., Rostoll, K., Van Der Walt, J. J. & Graz, C. J. M. Medicinal Chemistry: The Biological Activity of Selected Cyclic Dipeptides. *Journal of Pharmacy and Pharmacology* **50**, 1331–1337 (1998).
156. Wang, W. *et al.* Cerebrosides of the Halotolerant Fungus *Alternaria raphani* Isolated from a Sea Salt Field. *J. Nat. Prod.* **72**, 1695–1698 (2009).
157. Tompsett, R., McDermott, W. & Kidd, J. G. Tuberculostatic activity of blood and urine from animals given gliotoxin. *J. Immunol.* **65**, 59–63 (1950).
158. Sugie, Y. *et al.* A new antibiotic CJ-17,665 from *Aspergillus ochraceus*. *J. Antibiot.* **54**, 911–916 (2001).
159. Isaka, M., Palasarn, S., Rachtawee, P., Vimuttipong, S. & Kongsaree, P. Unique Diketopiperazine Dimers from the Insect Pathogenic Fungus *Verticillium hemipterigenum* BCC 1449. *Org. Lett.* **7**, 2257–2260 (2005).
160. Shimi, I. R., Abedallah, N. & Fathy, S. Cairomycin B, a New Antibiotic. *Antimicrob Agents Chemother* **11**, 373–375 (1977).
161. Hauser, D., Weber, H. P. & Sigg, H. P. Isolierung und Strukturaufklärung von Chaetocin. *HCA* **53**, 1061–1073 (1970).
162. Qi, S.-H., Xu, Y., Gao, J., Qian, P.-Y. & Zhang, S. Antibacterial and antilarval compounds from marine bacterium *Pseudomonas rhizosphaerae*. *Annals of Microbiology* **59**, 229–233 (2009).
163. Rhee, K.-H., Choi, K.-H., Kim, C.-J. & Kim, C.-H. Identification of *Streptomyces* sp. AMLK-335 producing antibiotic substance inhibitory to vancomycin-resistant enterococci. *Journal of microbiology and biotechnology* **11**, 469–474 (2001).
164. Fdhila, F., Vázquez, V., Sánchez, J. L. & Riguera, R. dd-Diketopiperazines: Antibiotics Active against *Vibrio anguillarum* Isolated from Marine Bacteria Associated with Cultures of *Pecten maximus*. *J. Nat. Prod.* **66**, 1299–1301 (2003).
165. Mehdi, R. B. A. *et al.* Five naturally bioactive molecules including two rhamnopyranoside derivatives isolated from the *Streptomyces* sp. strain TN58. *Natural Product Research* **23**, 1095–1107 (2009).
166. Miyamura, S., Ogasawara, N., Otsuka, H., Niwayama, S. & Tanaka, H. Antibiotic 5879 produced by *Streptomyces aizunensis*, identical with bicyclomycin. *J. Antibiot.* **26**, 479–484 (1973).
167. Rhee, K.-H. Isolation and characterization of *Streptomyces* sp KH-614 producing anti-VRE (vancomycin-resistant enterococci) antibiotics. *J. Gen. Appl. Microbiol.* **48**, 321–327 (2002).
168. Kritzinger, A. L. The medicinal chemistry of cyclo (Ser-Ser) and cyclo (Ser-Tyr). (2008).
169. McClelland, K. *et al.* An investigation into the biological activity of the selected histidine-containing diketopiperazines cyclo(His-Phe) and cyclo(His-Tyr). *Journal of Pharmacy and Pharmacology* **56**, 1143–1153 (2004).
170. Huberman, L. *et al.* Antibacterial substances of low molecular weight isolated from the blowfly, *Lucilia sericata*. *Medical and Veterinary Entomology* **21**, 127–131 (2007).

171. Michalský, J., Čtvrtník, J., Horáková, Z. & Bydžovský, V. Über die tuberkulostatische Aktivität von 2,5-Bis-(aminooxymethyl)-3,6-diketopiperazin, eines Umwandlungsproduktes des Cycloserins. *Experientia* **18**, 217–218 (1962).
172. McMahon, T. C., Stanley, S., Kazyanskaya, E., Hung, D. & Wood, J. L. A scaleable formal total synthesis of dehydrogliotoxin. *Tetrahedron Letters* **52**, 2262–2264 (2011).
173. Li, X. *et al.* A New Antibacterial Dioxopiperazine Alkaloid Related to Gliotoxin from a Marine Isolate of the Fungus *Pseudallescheria*. *J. Antibiot* **59**, 248–250 (2006).
174. Zhang, Y., Liu, S., Che, Y. & Liu, X. Epicoccins A-D, epipolythiodioxopiperazines from a Cordyceps-colonizing isolate of *Epicoccum nigrum*. *J. Nat. Prod.* **70**, 1522–1525 (2007).
175. Tsumagari, N. *et al.* MPC1001, a new antitumor antibiotic produced by *Cladorrhinum* sp. *J. Antibiot.* **57**, 532–534 (2004).
176. Kossuga, M. H. *et al.* Antibacterial modified diketopiperazines from two ascidians of the genus *Didemnum*. *Journal of the Brazilian Chemical Society* **20**, 704–711 (2009).
177. Kopp-Holtwiesche, B. & Rehm, H. J. Antimicrobial action of roquefortine. *J. Environ. Pathol. Toxicol. Oncol.* **10**, 41–44 (1990).
178. Kamiya, T., Maeno, S., Hashimoto, M. & Mine, Y. Bicyclomycin, a new antibiotic. II. Structural elucidation and acyl derivatives. *J. Antibiot.* **25**, 576–581 (1972).
179. Miyamura, S., Ogasawara, N., Otsuka, H., Niwayama, S. & Tanaka, H. Antibiotic no. 5879, a new water-soluble antibiotic against gram-negative bacteria. *J. Antibiot.* **25**, 610–612 (1972).
180. Zwiefka, A., Kohn, H. & Widger, W. R. Transcription termination factor rho: the site of bicyclomycin inhibition in *Escherichia coli*. *Biochemistry* **32**, 3564–3570 (1993).
181. Park, H. *et al.* Bicyclomycin and Dihydrobicyclomycin Inhibition Kinetics of *Escherichia coli* rho-Dependent Transcription Termination Factor ATPase Activity. *Archives of Biochemistry and Biophysics* **323**, 447–454 (1995).
182. Magyar, A., Zhang, X., Kohn, H. & Widger, W. R. The Antibiotic Bicyclomycin Affects the Secondary RNA Binding Site of *Escherichia coli* Transcription Termination Factor Rho. *J. Biol. Chem.* **271**, 25369–25374 (1996).
183. Skordalakes, E., Brogan, A. P., Park, B. S., Kohn, H. & Berger, J. M. Structural Mechanism of Inhibition of the Rho Transcription Termination Factor by the Antibiotic Bicyclomycin. *Structure* **13**, 99–109 (2005).
184. Park Hg, H., Zhang, X., Widger, W. R. & Kohn, H. Role of the C(1) Triol Group in Bicyclomycin: Synthesis and Biochemical and Biological Properties. *J. Org. Chem.* **61**, 7750–7755 (1996).
185. Park Hg, H., Zhang, Z., Zhang, X., Widger, W. R. & Kohn, H. Role of the C(5)-C(5a) Exomethylene Group in Bicyclomycin: Synthesis, Structure, and Biochemical and Biological Properties. *J. Org. Chem.* **61**, 7764–7776 (1996).
186. Vincent, F., Srinivasan, J., Santillán, Alejandro, Widger, W. R. & Kohn, H. C(5)–C(5a)-Modified Bicyclomycins: Synthesis, Structure, and Biochemical and Biological Properties. *J. Org. Chem.* **66**, 2251–2264 (2001).
187. Brogan, A. P. *et al.* Development of a Technique to Determine Bicyclomycin-Rho Binding and Stoichiometry by Isothermal Titration Calorimetry and Mass Spectrometry. *J. Am. Chem. Soc.* **127**, 2741–2751 (2005).
188. Vincent, F., Widger, W. R., Openshaw, M., Gaskell, S. J. & Kohn, H. 5a-Formylbicyclomycin: Studies on the Bicyclomycin–Rho Interaction†. *Biochemistry* **39**, 9067–9076 (2000).
189. Park, B.-S., Widger, W. & Kohn, H. Fluorine-substituted dihydrobicyclomycins: Synthesis and biochemical and biological properties. *Bioorganic & Medicinal Chemistry* **14**, 41–61 (2006).
190. Usui, T., Kondoh, M., Cui, C. B., Mayumi, T. & Osada, H. Tryprostatin A, a specific and novel inhibitor of microtubule assembly. *Biochem. J.* **333 ( Pt 3)**, 543–548 (1998).
191. Jain, H. D. *et al.* Synthesis and structure–activity relationship studies on tryprostatin A, an inhibitor of breast cancer resistance protein. *Bioorganic & Medicinal Chemistry* **16**, 4626–4651 (2008).

192. Depew, K. M., Danishefsky, S. J., Rosen, N. & Sepp Lorenzino, L. Total synthesis of tryprostatin B: Generation of a nucleophilic prenylating species from a prenylstannane. *J. Am. Chem. Soc.* **118**, 12463–12464 (1996).
193. Zhao, S. *et al.* Biological Activity of the Tryprostatins and Their Diastereomers on Human Carcinoma Cell Lines. *J. Med. Chem.* **45**, 1559–1562 (2002).
194. Woehlecke, H., Osada, H., Herrmann, A. & Lage, H. Reversal of breast cancer resistance protein-mediated drug resistance by tryprostatin A. *Int. J. Cancer* **107**, 721–728 (2003).
195. He, H. Y., Rabindran, S. G., Greenberger, L. M. & Carter, G. T. Fumitremorgin C analogs that reverse mitoxantrone resistance in human colon carcinoma cells. *Med. Chem. Res.* **9**, 424–437 (1999).
196. van Loevezijn, A., Allen, J. D., Schinkel, A. H. & Koomen, G.-J. Inhibition of BCRP-mediated drug efflux by fumitremorgin-type indolyl diketopiperazines. *Bioorganic & Medicinal Chemistry Letters* **11**, 29–32 (2001).
197. Allen, J. D. *et al.* Potent and Specific Inhibition of the Breast Cancer Resistance Protein Multidrug Transporter *in Vitro* and in Mouse Intestine by a Novel Analogue of Fumitremorgin C. *Mol Cancer Ther* **1**, 417–425 (2002).
198. Gomez-Monterrey, I. *et al.* Design, Synthesis, and Cytotoxic Evaluation of a New Series of 3-Substituted Spiro[(dihydropyrazine-2,5-dione)-6,3'-(2',3'-dihydrothieno[2,3-b]naphtho-4',9'-dione)] Derivatives. *J. Med. Chem.* **50**, 1787–1798 (2007).
199. Gomez-Monterrey, I. *et al.* Spiro[(dihydropyrazin-2,5-dione)-6,3'-(2',3'-dihydrothieno[2,3-b]naphtho-4',9'-dione)]-Based Cytotoxic Agents: Structure–Activity Relationship Studies on the Substituent at N4-Position of the Diketopiperazine Domain. *J. Med. Chem.* **51**, 2924–2932 (2008).
200. Minelli, A., Bellezza, I., Grottelli, S. & Galli, F. Focus on cyclo(His-Pro): history and perspectives as antioxidant peptide. *Amino Acids* **35**, 283–289 (2007).
201. Bellezza, I. *et al.* Neuroinflammation and endoplasmic reticulum stress are coregulated by cyclo(His-Pro) to prevent LPS neurotoxicity. *The International Journal of Biochemistry & Cell Biology* **51**, 159–169 (2014).
202. Minelli, A. *et al.* Cyclo(His-Pro) promotes cytoprotection by activating Nrf2-mediated up-regulation of antioxidant defence. *Journal of Cellular and Molecular Medicine* **13**, 1149–1161 (2009).
203. Taubert, D., Grimberg, G., Stenzel, W. & Schömig, E. Identification of the Endogenous Key Substrates of the Human Organic Cation Transporter OCT2 and Their Implication in Function of Dopaminergic Neurons. *PLoS ONE* **2**, e385 (2007).
204. Minelli, A. *et al.* Cyclo(His-Pro) up-regulates heme oxygenase 1 via activation of Nrf2-ARE signalling. *Journal of Neurochemistry* **111**, 956–966 (2009).
205. Teixidó, M., Zurita, E., Malakoutikhah, M., Tarragó, T. & Giralt, E. Diketopiperazines as a Tool for the Study of Transport across the Blood–Brain Barrier (BBB) and Their Potential Use as BBB-Shuttles. *J. Am. Chem. Soc.* **129**, 11802–11813 (2007).
206. Prakash, K. R. C. *et al.* Synthesis and Biological Activity of Novel Neuroprotective Diketopiperazines. *Bioorganic & Medicinal Chemistry* **10**, 3043–3048 (2002).
207. Faden, A. I. *et al.* Novel diketopiperazine enhances motor and cognitive recovery after traumatic brain injury in rats and shows neuroprotection *in vitro* and *in vivo*. *Journal of Cerebral Blood Flow and Metabolism* **23**, 342–354 (2003).
208. Faden, A. I. *et al.* Neuroprotective and nootropic actions of a novel cyclized dipeptide after controlled cortical impact injury in mice. *Journal of Cerebral Blood Flow and Metabolism* **23**, 355–363 (2003).
209. Faden, A. I., Movsesyan, V. A., Knoblach, S. M., Ahmed, F. & Cernak, I. Neuroprotective effects of novel small peptides *in vitro* and after brain injury. *Neuropharmacology* **49**, 410–424 (2005).
210. Faden, A. I., Movsesyan, V. A., Fang, X. L. & Wang, S. M. Identification of novel neuroprotective agents using pharmacophore modeling. *Chemistry & Biodiversity* **2**, 1564–1570 (2005).
211. Cornacchia, C. *et al.* 2,5-Diketopiperazines as Neuroprotective Agents. *Mini-Reviews in Medicinal Chemistry* **12**, 2–12 (2012).

212. Suzuki, K., Sasaki, Y., Endo, N. & Mihara, Y. Acetic Acid-Catalyzed Diketopiperazine Synthesis. *Chemical & Pharmaceutical Bulletin* **29**, 233–237 (1981).
213. Depew, K. M. *et al.* Total Synthesis of 5-N-Acetylardeemin and Amauromine: Practical Routes to Potential MDR Reversal Agents. *J. Am. Chem. Soc.* **121**, 11953–11963 (1999).
214. Tullberg, M., Grötli, M. & Luthman, K. Efficient synthesis of 2,5-diketopiperazines using microwave assisted heating. *Tetrahedron* **62**, 7484–7491 (2006).
215. Tullberg, M., Grötli, M. & Luthman, K. Synthesis of Functionalized, Unsymmetrical 1,3,4,6-Tetrasubstituted 2,5-Diketopiperazines. *J. Org. Chem.* **72**, 195–199 (2007).
216. Pérez-Picaso, L., Escalante, J., Olivo, H. F. & Rios, M. Y. Efficient Microwave Assisted Syntheses of 2,5-Diketopiperazines in Aqueous Media. *Molecules* **14**, 2836–2849 (2009).
217. Cabon, G., Gaucher, B., Gegout, A., Heulle, S. & Masquelin, T. Novel solution- and solid-phase syntheses of heterocyclic systems. *Chimia* **57**, 248–254 (2003).
218. Kennedy, A. L., Fryer, A. M. & Josey, J. A. A New Resin-Bound Universal Isonitrile for the Ugi 4CC Reaction: Preparation and Applications to the Synthesis of 2,5-Diketopiperazines and 1,4-Benzodiazepine-2,5-diones. *Org. Lett.* **4**, 1167–1170 (2002).
219. Couladouros, E. A. & Magos, A. D. Solid-phase total synthesis of (-)-Phenylhistine and (-)-Aurantiamine. Synthesis of a diverse dehydro-2,5-diketopiperazine library. Part II. *Mol. Divers.* **9**, 111–121 (2005).
220. Domling, A. Recent developments in isocyanide based multicomponent reactions in applied chemistry. *Chem. Rev.* **106**, 17–89 (2006).
221. Hulme, C., Morrisette, M. M., Volz, F. A. & Burns, C. J. The solution phase synthesis of diketopiperazine libraries via the Ugi reaction: Novel application of Armstrong's convertible isonitrile. *Tetrahedron Lett.* **39**, 1113–1116 (1998).
222. Rhoden, C. R. B. *et al.* Rapid Access to N-Substituted Diketopiperazines by One-Pot Ugi-4CR/Deprotection plus Activation/Cyclization (UDAC). *J. Comb. Chem.* **11**, 1078–1082 (2009).
223. Hulme, C. *et al.* Novel safety-catch linker and its application with a Ugi/De-BOC/Cyclization (UDC) strategy to access carboxylic acids, 1,4-benzodiazepines, diketopiperazines, ketopiperazines and dihydroquinoxalinones. *Tetrahedron Letters* **39**, 7227–7230 (1998).
224. Chen, J. J., Golebiowski, A., Klopfenstein, S. R. & West, L. The universal Rink-isonitrile resin: applications in Ugi reactions. *Tetrahedron Letters* **43**, 4083–4085 (2002).
225. Falorni, M., Giacomelli, G., Porcheddu, A. & Taddei, M. Solution-phase synthesis of mixed amide libraries by simultaneous addition of functionalities (SPSAF) to a diketopiperazine tetracarboxylic acid scaffold monitored by GC analysis of isobutyl alcohol. *Eur. J. Org. Chem.* 1669–1675 (2000).
226. Gonzalez, A., Vorobeve, S. & Linares, A. Synthesis and Structural Study of Piperazine-2,5-Diones Derived from (r)-Cysteine. *Tetrahedron: Asymmetry* **6**, 1357–1366 (1995).
227. Giraud, M., Bernad, N., Martinez, J. & Cavelier, F. New general strategy of dimerization of bioactive molecules. *Tetrahedron Letters* **42**, 1895–1897 (2001).
228. Santagada, V. *et al.* A convenient strategy of dimerization by microwave heating and using 2,5-diketopiperazine as scaffold. *Tetrahedron Lett.* **44**, 1145–1148 (2003).
229. Jainta, M., Nieger, M. & Braese, S. Microwave-Assisted Stereoselective One-Pot Synthesis of Symmetrical and Unsymmetrical 2,5-Diketopiperazines from Unprotected Amino Acids. *Eur. J. Org. Chem.* 5418–5424 (2008).
230. Majumdar, K. C., Ray, K. & Ganai, S. Intramolecular Aza-Wittig Reaction: A New Efficient Tool for the Construction of Piperazine 2,5-Dione Derivatives. *Synlett* 2122–2124 (2010).
231. Hazra, A. *et al.* Synthesis of symmetrically 1,4-disubstituted piperazine-2,5-diones: a new class of antileishmanial agents. *J. Chem. Res* 381–383 (2007).
232. Cho, S. D. *et al.* One-pot synthesis of symmetrical 1,4-disubstituted piperazine-2,5-diones. *Bull. Korean Chem. Soc.* **25**, 415–416 (2004).

233. Marcaccini, S., Pepino, R. & Pozo, M. C. A facile synthesis of 2,5-diketopiperazines based on isocyanide chemistry. *Tetrahedron Lett.* **42**, 2727–2728 (2001).
234. Santra, S. & Andreana, P. R. A one-pot, microwave-influenced synthesis of diverse small molecules by multicomponent reaction cascades. *Org. Lett.* **9**, 5035–5038 (2007).
235. Sheradsky, T. & Silcoff, E. R. Synthesis of (2R,5R)-2-amino-5-hydroxyhexanoic acid by intramolecular cycloaddition. *Molecules* **3**, 80–87 (1998).
236. Peng, J. & Clive, D. L. J. Asymmetric Synthesis of the ABC-Ring System of the Antitumor Antibiotic MPC1001. *J. Org. Chem.* **74**, 513–519 (2009).
237. Herzon, S. B. & Myers, A. G. Enantioselective synthesis of stephacidin B. *J. Am. Chem. Soc.* **127**, 5342–5344 (2005).
238. Siwicka, A. *et al.* Diastereodivergent synthesis of 2,5-diketopiperazine derivatives of beta-carboline and isoquinoline from L-amino acids. *Tetrahedron: Asymmetry* **16**, 975–993 (2005).
239. Maw, G. N., Allerton, C. M. N., Gbektor, E. & Million, W. A. Design, synthesis and biological activity of beta-carboline-based type-5 phosphodiesterase inhibitors. *Bioorg. Med. Chem. Lett.* **13**, 1425–1428 (2003).
240. Bull, S. D., Davies, S. G., Epstein, S. W. & Ouzman, J. V. A. Deracemisation of  $\alpha$ -amino acids—(R)- and (S)-phenylalanine from the same enantiomer of a homochiral auxiliary. *Tetrahedron: Asymmetry* **9**, 2795–2798 (1998).
241. Balducci, D., Lazzari, I., Monari, M., Piccinelli, F. & Porzi, G. (S)- $\alpha$ -methyl, $\alpha$ -amino acids: a new stereocontrolled synthesis. *Amino Acids* **38**, 829–837 (2010).
242. Balducci, D., Contaldi, S., Lazzari, I. & Porzi, G. A highly efficient stereocontrolled synthesis of (S)-2',6'-dimethyltyrosine [(S)-DMT]. *Tetrahedron: Asymmetry* **20**, 1398–1401 (2009).
243. Chai, C. L. L., Elix, J. A. & Huleatt, P. B. The synthetic versatility of alkoxy carbonyl- and hydroxymethyl-piperazine-2,5-diones. *Tetrahedron* **61**, 8722–8739 (2005).
244. Dubey, R. & Olenyuk, B. Direct organocatalytic coupling of carboxylated piperazine-2,5-diones with indoles through conjugate addition of carbon nucleophiles to indolenine intermediates. *Tetrahedron Letters* **51**, 609–612 (2010).
245. Dubey, R., Polaske, N. W., Nichol, G. S. & Olenyuk, B. Efficient organocatalytic  $\alpha$ -sulfenylation of substituted piperazine-2,5-diones. *Tetrahedron Letters* **50**, 4310–4313 (2009).
246. Badran, T., Chai, C., Easton, C., Harper, J. & Page, D. Directing Bromination of Piperazine-2,5-diones. *Aust. J. Chem.* **48**, 1379–1384 (1995).
247. Williams, R. M., Armstrong, R. W., Maruyama, L. K., Dung, J. S. & Anderson, O. P. Divergent, generalized synthesis of unsymmetrically substituted 2,5-piperazinediones. *J. Am. Chem. Soc.* **107**, 3246–3253 (1985).
248. Marcuccio, S. M. & Elix, J. A. A structural revision of picroroccellin. *Tetrahedron Letters* **24**, 1445–1448 (1983).
249. Park, D., Lee, J., Kwon, T., Chung, D. & Jung, H. Synthesis of Symmetrically Alpha-N-Functionalized Piperazine-2,5-Diones. *Bull. Korean Chem. Soc.* **15**, 332–333 (1994).
250. Bull, S. D. *et al.* 2-Halo-diketopiperazines as chiral glycine cation equivalents. *Tetrahedron: Asymmetry* **15**, 3989–4001 (2004).
251. Whitlock, C. R. & Cava, M. P. A total synthesis of dragmacidin B. *Tetrahedron Letters* **35**, 371–374 (1994).
252. Kienle, M., Argyrakis, W., Baro, A. & Laschat, S. Diketopiperazine-derived hydroperoxide for chemoselective oxidations of sulfides and enantioselective Weitz-Scheffer epoxidations. *Tetrahedron Lett.* **49**, 1971–1974 (2008).
253. Baran, P. S. & Corey, E. J. A Short Synthetic Route to (+)-Austamide, (+)-Deoxyisoaustamide, and (+)-Hydratoaustamide from a Common Precursor by a Novel Palladium-Mediated Indole  $\rightarrow$  Dihydroindolozocine Cyclization. *J. Am. Chem. Soc.* **124**, 7904–7905 (2002).

254. Liebscher, J. & Jin, S. 3-Ylidenepiperazine-2,5-diones as versatile organic substrates. *Chem. Soc. Rev.* **28**, 251–259 (1999).
255. Yoshimura, J., Yamaura, M., Suzuki, T. & Hashimoto, H. Oxidative Removal of N-(para-Methoxybenzyl) Group on Diketopiperazine Skeleton with Ceric Ammonium-Nitrate. *Chem. Lett.* 1001–1002 (1983). doi:10.1246/cl.1983.1001
256. Ohta, A. *et al.* Catalytic-Hydrogenation of 2,5-Dialkylpyrazines and 3,6-Dialkyl-2-Hydroxypyrazines. *Heterocycles* **26**, 2691–2701 (1987).
257. Chai, C. L. L., Elix, J. A. & Huleatt, P. B. Multi-purpose functionality for the structural elaboration of the piperazine-2,5-dione motif. *Tetrahedron Lett.* **44**, 263–265 (2003).
258. Bianco, A., Sonksen, C. P., Roepstorff, P. & Briand, J. P. Solid-phase synthesis and structural characterization of highly substituted hydroxyproline-based 2,5-diketopiperazines. *J. Org. Chem.* **65**, 2179–2187 (2000).
259. Caballero, E., Avendaño, C. & Menéndez, J. C. Stereochemical issues related to the synthesis and reactivity of pyrazino[2',1'-5,1]pyrrolo[2,3-b]indole-1,4-diones. *Tetrahedron: Asymmetry* **9**, 967–981 (1998).
260. Kim, J., Ashenhurst, J. A. & Movassaghi, M. Total Synthesis of (+)-11,11'-Dideoxyverticillin A. *Science* **324**, 238–241 (2009).
261. Movassaghi, M., Schmidt, M. A. & Ashenhurst, J. A. Concise Total Synthesis of (+)-WIN 64821 and (-)-Ditryptophenaline. *Angewandte Chemie International Edition* **47**, 1485–1487 (2008).
262. Schkeryantz, J. M., Woo, J. C. G., Silphaivanh, P., Depew, K. M. & Danishefsky, S. J. Total Synthesis of Gypsetin, Deoxybrevianamide E, Brevianamide E, and Tryprostatin B: Novel Constructions of 2,3-Disubstituted Indoles. *J. Am. Chem. Soc.* **121**, 11964–11975 (1999).
263. Baran, P. S., Guerrero, C. A. & Corey, E. J. Short, enantioselective total synthesis of okaramine N. *J. Am. Chem. Soc.* **125**, 5628–5629 (2003).
264. Avendano, C. & la Cuesta, E. Synthetic Chemistry with N-Acyliminium Ions Derived from Piperazine-2,5-Diones and Related Compounds. *Current Organic Synthesis* **6**, 143–168 (2009).
265. Falorni, M., Porcheddu, A. & Giacomelli, G. Synthesis of (2R, 5S)- and (2S, 5S)-2-carboxy-1,4-diaza-[4.3.0]bicyclononane as building blocks for the synthesis of new potential HIV protease inhibitors. *Tetrahedron: Asymmetry* **7**, 1999–2005 (1996).
266. Nakamura, D., Kakiuchi, K., Koga, K. & Shirai, R. Design and synthesis of novel C-2-symmetric chiral piperazines and an application to asymmetric acylation of sigma-symmetric 1,2-diols. *Org. Lett.* **8**, 6139–6142 (2006).
267. Soukara, S. & Wünsch, B. Regioselective thionation of bicyclic piperazinediones. *Tetrahedron* **57**, 4359–4363 (2001).
268. Carlsson, A.-C. *et al.* Microwave-assisted synthesis of the Schöllkopf chiral auxiliaries: (3S)- and (3R)-3,6-dihydro-2,5-diethoxy-3-isopropyl-pyrazine. *Tetrahedron Letters* **47**, 5199–5201 (2006).
269. Chaignaud, M., Gillaizeau, I., Ouhamou, N. & Coudert, G. New highlights in the synthesis and reactivity of 1,4-dihydropyrazine derivatives. *Tetrahedron* **64**, 8059–8066 (2008).
270. Hernández, F., Lumetzberger, A., Avendaño, C. & Söllhuber, M. Regioselective N-Acylation of 3-Arylmethylpiperazine-2,5-diones: Short Synthesis of (-)-Glyantrypine and (-)-Fumiquinazoline F. *Synlett* **2001**, 1387–1390 (2001).
271. Veerman, J. J. N. *et al.* Synthesis of 2,6-Bridged Piperazine-3-ones by N-Acyliminium Ion Chemistry. *J. Org. Chem.* **68**, 4486–4494 (2003).
272. Robertz, B., Keul, H. & Hocker, H. Synthesis of cyclo(amide-ester)s by ring-expansion of N-(acyl)-lactams. *Macromol. Chem. Phys.* **200**, 1034–1040 (1999).
273. Meier, J. L. & Burkart, M. D. The chemical biology of modular biosynthetic enzymes. *Chem. Soc. Rev.* **38**, 2012–2045 (2009).

274. Mootz, H. D. & Marahiel, M. A. The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179**, 6843–6850 (1997).
275. Schwarzer, D., Finking, R. & Marahiel, M. A. Nonribosomal peptides: from genes to products. *Natural Product Reports* **20**, 275 (2003).
276. Stein, T. *et al.* The Multiple Carrier Model of Nonribosomal Peptide Biosynthesis at Modular Multienzymatic Templates. *J. Biol. Chem.* **271**, 15428–15435 (1996).
277. Keating, T. A. & Walsh, C. T. Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Current Opinion in Chemical Biology* **3**, 598–606 (1999).
278. Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *The EMBO Journal* **16**, 4174–4183 (1997).
279. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology* **6**, 493–505 (1999).
280. Eppelmann, K., Stachelhaus, T. & Marahiel, M. A. Exploitation of the Selectivity-Confering Code of Nonribosomal Peptide Synthetases for the Rational Design of Novel Peptide Antibiotics†. *Biochemistry* **41**, 9718–9726 (2002).
281. Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W. & Huson, D. H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucl. Acids Res.* **33**, 5799–5808 (2005).
282. Röttig, M. *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucl. Acids Res.* **39**, W362–W367 (2011).
283. Bachmann, B. O. & Ravel, J. Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data. *Meth. Enzymol.* **458**, 181–217 (2009).
284. Stachelhaus, T., Hüser, A. & Marahiel, M. A. Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. *Chemistry & Biology* **3**, 913–921 (1996).
285. Ehmann, D. E., Shaw-Reid, C. A., Losey, H. C. & Walsh, C. T. The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: Sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2509–2514 (2000).
286. Linne, U. & Marahiel, M. A. Control of Directionality in Nonribosomal Peptide Synthesis: Role of the Condensation Domain in Preventing Misinitiation and Timing of Epimerization†. *Biochemistry* **39**, 10439–10447 (2000).
287. Stachelhaus, T., Mootz, H. D., Bergendahl, V. & Marahiel, M. A. Peptide Bond Formation in Nonribosomal Peptide Biosynthesis CATALYTIC ROLE OF THE CONDENSATION DOMAIN. *J. Biol. Chem.* **273**, 22773–22781 (1998).
288. Bergendahl, V., Linne, U. & Marahiel, M. A. Mutational analysis of the C-domain in nonribosomal peptide synthesis. *European Journal of Biochemistry* **269**, 620–629 (2002).
289. Felnagle, E. A. *et al.* Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products. *Mol. Pharmaceutics* **5**, 191–211 (2008).
290. Miller, D. A., Walsh, C. T. & Luo, L. C-methyltransferase and cyclization domain activity at the intraprotein PK/NRP switch point of yersiniabactin synthetase. *J. Am. Chem. Soc.* **123**, 8434–8435 (2001).
291. Keating, T. A., Marshall, C. G. & Walsh, C. T. Reconstitution and characterization of the *Vibrio cholerae* vibriobactin synthetase from VibB, VibE, VibF, and VibH. *Biochemistry* **39**, 15522–15530 (2000).
292. Silakowski, B. *et al.* New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J. Biol. Chem.* **274**, 37391–37399 (1999).

293. Tang, L. *et al.* Cloning and heterologous expression of the epothilone gene cluster. *Science* **287**, 640–642 (2000).
294. Du, L., Chen, M., Sánchez, C. & Shen, B. An oxidation domain in the BlmIII non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in *Streptomyces verticillus* ATCC15003. *FEMS Microbiol. Lett.* **189**, 171–175 (2000).
295. Weber, G., Schörgendorfer, K., Schneider-Scherzer, E. & Leitner, E. The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame. *Curr. Genet.* **26**, 120–125 (1994).
296. Haese, A., Schubert, M., Herrmann, M. & Zocher, R. Molecular characterization of the enniatin synthetase gene encoding a multifunctional enzyme catalysing N-methyldepsipeptide formation in *Fusarium scirpi*. *Molecular Microbiology* **7**, 905–914 (1993).
297. Haese, A., Pieper, R., Ostrowski, T. von & Zocher, R. Bacterial Expression of Catalytically Active Fragments of the Multifunctional Enzyme Enniatin Synthetase. *Journal of Molecular Biology* **243**, 116–122 (1994).
298. Burmester, J., Haese, A. & Zocher, R. Highly conserved N-methyltransferases as an integral part of peptide synthetases. *Biochem. Mol. Biol. Int.* **37**, 201–207 (1995).
299. Schauwecker, F., Pfennig, F., Grammel, N. & Keller, U. Construction and *in vitro* analysis of a new bimodular polypeptide synthetase for synthesis of N-methylated acyl peptides. *Chemistry & Biology* **7**, 287–297 (2000).
300. Stachelhaus, T. & Walsh, C. T. Mutational Analysis of the Epimerization Domain in the Initiation Module PheATE of Gramicidin S Synthetase†. *Biochemistry* **39**, 5775–5787 (2000).
301. Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**, 486–489 (1999).
302. Cosmina, P. *et al.* Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis*. *Molecular Microbiology* **8**, 821–831 (1993).
303. Steller, S. *et al.* Structural and functional organization of the fengycin synthetase multienzyme system from *Bacillus subtilis* b213 and A1/3. *Chemistry & Biology* **6**, 31–41 (1999).
304. Silakowski, B. *et al.* The myxochelin iron transport regulon of the myxobacterium *Stigmatella aurantiaca* Sg a15. *European Journal of Biochemistry* **267**, 6476–6485 (2000).
305. Schwarzer, D. & Marahiel, M. A. Multimodular biocatalysts for natural product assembly. *Naturwissenschaften* **88**, 93–101 (2001).
306. Du, L., Sánchez, C. & Shen, B. Hybrid Peptide–Polyketide Natural Products: Biosynthesis and Prospects toward Engineering Novel Molecules. *Metabolic Engineering* **3**, 78–95 (2001).
307. Du, L. & Shen, B. Biosynthesis of hybrid peptide-polyketide natural products. *Curr Opin Drug Discov Devel* **4**, 215–228 (2001).
308. Du, L., Sánchez, C., Chen, M., Edwards, D. J. & Shen, B. The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chemistry & Biology* **7**, 623–642 (2000).
309. Molnár, I. *et al.* The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from *Sorangium cellulosum* So ce90. *Chem. Biol.* **7**, 97–109 (2000).
310. Walsh, C. T. *et al.* Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Current Opinion in Chemical Biology* **5**, 525–534 (2001).
311. Balibar, C. J. & Walsh, C. T. GliP, a Multimodular Nonribosomal Peptide Synthetase in *Aspergillus fumigatus*, Makes the Diketopiperazine Scaffold of Gliotoxin†. *Biochemistry* **45**, 15029–15038 (2006).
312. Maiya, S., Grundmann, A., Li, S.-M. & Turner, G. The Fumitremorgin Gene Cluster of *Aspergillus fumigatus*: Identification of a Gene Encoding Brevianamide F Synthetase. *ChemBioChem* **7**, 1062–1069 (2006).

313. Lazos, O. *et al.* Biosynthesis of the Putative Siderophore Erythrochelin Requires Unprecedented Crosstalk between Separate Nonribosomal Peptide Gene Clusters. *Chemistry & Biology* **17**, 160–173 (2010).
314. Correia, T., Grammel, N., Ortel, I., Keller, U. & Tudzynski, P. Molecular Cloning and Analysis of the Ergopeptide Assembly System in the Ergot Fungus *Claviceps purpurea*. *Chemistry & Biology* **10**, 1281–1292 (2003).
315. García-Estrada, C. *et al.* A Single Cluster of Coregulated Genes Encodes the Biosynthesis of the Mycotoxins Roquefortine C and Meleagrins in *Penicillium chrysogenum*. *Chemistry & Biology* **18**, 1499–1512 (2011).
316. Yin, W.-B., Grundmann, A., Cheng, J. & Li, S.-M. Acetylszonalenin Biosynthesis in *Neosartorya fischeri* IDENTIFICATION OF THE BIOSYNTHETIC GENE CLUSTER BY GENOMIC MINING AND FUNCTIONAL PROOF OF THE GENES BY BIOCHEMICAL INVESTIGATION. *J. Biol. Chem.* **284**, 100–109 (2009).
317. Oide, S. *et al.* NPS6, Encoding a Nonribosomal Peptide Synthetase Involved in Siderophore-Mediated Iron Metabolism, Is a Conserved Virulence Determinant of Plant Pathogenic Ascomycetes. *Plant Cell* **18**, 2836–2853 (2006).
318. Wilhite, S. E., Lumsden, R. D. & Straney, D. C. Peptide Synthetase Gene in *Trichoderma virens*. *Appl. Environ. Microbiol.* **67**, 5055–5062 (2001).
319. Lee, S. G. & Lipmann, F. Isolation of amino acid activating subunit-pantetheine protein complexes: Their role in chain elongation in tyrocidine synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2343–2347 (1977).
320. 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).
321. Scharf, D. H. *et al.* Biosynthesis and function of gliotoxin in *Aspergillus fumigatus*. *Applied Microbiology and Biotechnology* **93**, 467–472 (2012).
322. Gardiner, D. M. & Howlett, B. J. Bioinformatic and expression analysis of the putative gliotoxin biosynthetic gene cluster of *Aspergillus fumigatus*. *FEMS Microbiology Letters* **248**, 241–248 (2005).
323. Schrettl, M. *et al.* Self-Protection against Gliotoxin—A Component of the Gliotoxin Biosynthetic Cluster, GliT, Completely Protects *Aspergillus fumigatus* Against Exogenous Gliotoxin. *PLoS Pathog* **6**, e1000952 (2010).
324. Gu, B., He, S., Yan, X. & Zhang, L. Tentative biosynthetic pathways of some microbial diketopiperazines. *Appl Microbiol Biotechnol* **97**, 8439–8453 (2013).
325. Moreau, C. Les mycotoxines neurotropes de l'*Aspergillus fumigatus*. *Bull. Soc. Myc. Fr.* **98**, 261–273 (1982).
326. Mundt, K., Wollinsky, B., Ruan, H.-L., Zhu, T. & Li, S.-M. Identification of the Verruculogen Prenyltransferase FtmPT3 by a Combination of Chemical, Bioinformatic and Biochemical Approaches. *ChemBioChem* **13**, 2583–2592 (2012).
327. Li, S.-M. Genome mining and biosynthesis of fumitremorgin-type alkaloids in ascomycetes. *J. Antibiot* **64**, 45–49 (2011).
328. Kato, N. *et al.* Identification of Cytochrome P450s Required for Fumitremorgin Biosynthesis in *Aspergillus fumigatus*. *ChemBioChem* **10**, 920–928 (2009).
329. Grundmann, A. & Li, S.-M. Overproduction, purification and characterization of FtmPT1, a brevianamide F prenyltransferase from *Aspergillus fumigatus*. *Microbiology* **151**, 2199–2207 (2005).
330. Grundmann, A., Kuznetsova, T., Afiyatullo, S. S. & Li, S.-M. FtmPT2, an N-Prenyltransferase from *Aspergillus fumigatus*, Catalyses the Last Step in the Biosynthesis of Fumitremorgin B. *ChemBioChem* **9**, 2059–2063 (2008).
331. Steffan, N., Grundmann, A., Afiyatullo, S., Ruan, H. & Li, S.-M. FtmOx1, a non-heme Fe(II) and  $\alpha$ -ketoglutarate-dependent dioxygenase, catalyses the endoperoxide formation of verruculogen in *Aspergillus fumigatus*. *Org. Biomol. Chem.* **7**, 4082–4087 (2009).

332. King, R. R., Lawrence, C. H., Clark, M. C. & Calhoun, L. A. Isolation and characterization of phytotoxins associated with *Streptomyces scabies*. *J. Chem. Soc., Chem. Commun.* 849–850 (1989). doi:10.1039/C39890000849
333. Kers, J. A. *et al.* A large, mobile pathogenicity island confers plant pathogenicity on *Streptomyces* species. *Molecular Microbiology* **55**, 1025–1033 (2005).
334. Kers, J. A. *et al.* Nitration of a peptide phytotoxin by bacterial nitric oxide synthase. *Nature* **429**, 79–82 (2004).
335. Barry, S. M. *et al.* Cytochrome P450-catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. *Nat Chem Biol* **8**, 814–816 (2012).
336. Johnson, E. G. *et al.* 4-Nitrotryptophan is a substrate for the non-ribosomal peptide synthetase TxtB in the thaxtomin A biosynthetic pathway. *Molecular Microbiology* **73**, 409–418 (2009).
337. Healy, F. G., Krasnoff, S. B., Wach, M., Gibson, D. M. & Loria, R. Involvement of a Cytochrome P450 Monooxygenase in Thaxtomin A Biosynthesis by *Streptomyces acidiscabies*. *J. Bacteriol.* **184**, 2019–2029 (2002).
338. King, R. R. & Calhoun, L. A. The thaxtomin phytotoxins: Sources, synthesis, biosynthesis, biotransformation and biological activity. *Phytochemistry* **70**, 833–841 (2009).
339. Walzel, B., Riederer, B. & Keller, U. Mechanism of alkaloid cyclopeptide synthesis in the ergot fungus *Claviceps purpurea*. *Chemistry & Biology* **4**, 223–230 (1997).
340. Haarmann, T., Lorenz, N. & Tudzynski, P. Use of a nonhomologous end joining deficient strain ( $\Delta ku70$ ) of the ergot fungus *Claviceps purpurea* for identification of a nonribosomal peptide synthetase gene involved in ergotamine biosynthesis. *Fungal Genetics and Biology* **45**, 35–44 (2008).
341. Wallwey, C. & Li, S.-M. Ergot alkaloids: structure diversity, biosynthetic gene clusters and functional proof of biosynthetic genes. *Nat. Prod. Rep.* **28**, 496–510 (2011).
342. Quigley, F. R. & Floss, H. G. Mechanism of amino acid  $\alpha$ -hydroxylation and formation of the lysergyl moiety in ergotamine biosynthesis. *J. Org. Chem.* **46**, 464–466 (1981).
343. Haarmann, T. *et al.* The ergot alkaloid gene cluster in *Claviceps purpurea*: Extension of the cluster sequence and intra species evolution. *Phytochemistry* **66**, 1312–1320 (2005).
344. Schwarzer, D., Mootz, H. D., Linne, U. & Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14083–14088 (2002).
345. Belin, P. *et al.* The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways. *Nat. Prod. Rep.* **29**, 961–979 (2012).
346. Giessen, T. W., Tesmar, A. M. von & Marahiel, M. A. Insights into the Generation of Structural Diversity in a tRNA-Dependent Pathway for Highly Modified Bioactive Cyclic Dipeptides. *Chemistry & Biology* **20**, 828–838 (2013).
347. Li, Y., Lai, Y.-M., Lu, Y., Yang, Y.-L. & Chen, S. Analysis of the biosynthesis of antibacterial cyclic dipeptides in *Nocardiopsis alba*. *Arch Microbiol* **196**, 765–774 (2014).
348. Giessen, T. W., Tesmar, A. M. von & Marahiel, M. A. A tRNA-Dependent Two-Enzyme Pathway for the Generation of Singly and Doubly Methylated Dityryptophan 2,5-Diketopiperazines. *Biochemistry* **52**, 4274–4283 (2013).
349. Döhren, H. von. Charged tRNAs charge into secondary metabolism. *Nat Chem Biol* **5**, 374–375 (2009).
350. RajBhandary, U. L. & Söll, D. Aminoacyl-tRNAs, the bacterial cell envelope, and antibiotics. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5285–5286 (2008).
351. Shepherd, J. & Ibba, M. Direction of aminoacylated transfer RNAs into antibiotic synthesis and peptidoglycan-mediated antibiotic resistance. *FEBS Lett* **587**, 2895–2904 (2013).
352. Jahn, D., Verkamp, E. & Söll, D. Glutamyl-transfer RNA: a precursor of heme and chlorophyll biosynthesis. *Trends in Biochemical Sciences* **17**, 215–218 (1992).
353. Roy, H. & Ibba, M. RNA-dependent lipid remodeling by bacterial multiple peptide resistance factors. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4667–4672 (2008).

354. Garg, R. P., Qian, X. L., Alemany, L. B., Moran, S. & Parry, R. J. Investigations of valanimycin biosynthesis: Elucidation of the role of seryl-tRNA. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6543–6547 (2008).
355. Mainardi, J.-L., Villet, R., Bugg, T. D., Mayer, C. & Arthur, M. Evolution of peptidoglycan biosynthesis under the selective pressure of antibiotics in Gram-positive bacteria. *FEMS Microbiology Reviews* **32**, 386–408 (2008).
356. Watanabe, K. *et al.* Protein-based peptide-bond formation by aminoacyl-tRNA protein transferase. *Nature* **449**, 867–871 (2007).
357. Zhang, W., Ntai, I., Kelleher, N. L. & Walsh, C. T. tRNA-dependent peptide bond formation by the transferase PacB in biosynthesis of the pacidamycin group of pentapeptidyl nucleoside antibiotics. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12249–12253 (2011).
358. Benson, T. E. *et al.* X-Ray Crystal Structure of *Staphylococcus aureus* FemA. *Structure* **10**, 1107–1115 (2002).
359. Biarrotte-Sorin, S. *et al.* Crystal Structures of *Weissella viridescens* FemX and Its Complex with UDP-MurNAc-Pentapeptide: Insights into FemABX Family Substrates Recognition. *Structure* **12**, 257–267 (2004).
360. Fonvielle, M. *et al.* The Structure of FemXWv in Complex with a Peptidyl-RNA Conjugate: Mechanism of Aminoacyl Transfer from Ala-tRNAAla to Peptidoglycan Precursors. *Angew. Chem. Int. Ed.* **52**, 7278–7281 (2013).
361. Sauguet, L. *et al.* Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis. *Nucl. Acids Res.* **39**, 4475–4489 (2011).
362. Vetting, M. W., Hegde, S. S. & Blanchard, J. S. The Structure and Mechanism of the *Mycobacterium tuberculosis* Cyclodityrosine Synthetase. *Nat Chem Biol* **6**, 797–799 (2010).
363. Bonnefond, L. *et al.* Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3912–3917 (2011).
364. Moutiez, M. *et al.* Unravelling the mechanism of non-ribosomal peptide synthesis by cyclodipeptide synthases. *Nat Commun* **5**, (2014).
365. Moutiez, M. *et al.* Specificity determinants for the two tRNA substrates of the cyclodipeptide synthase AlbC from *Streptomyces noursei*. *Nucleic Acids Res* **42**, 7247–7258 (2014).
366. Agarwal, V., Pierce, E., McIntosh, J., Schmidt, E. W. & Nair, S. K. Structures of Cyanobactin Maturation Enzymes Define a Family of Transamidating Proteases. *Chemistry & Biology* **19**, 1411–1422 (2012).
367. Cryle, M. J., Bell, S. G. & Schlichting, I. Structural and biochemical characterization of the cytochrome P450 CypX (CYP134A1) from *Bacillus subtilis*: a cyclo-L-leucyl-L-leucyl dipeptide oxidase. *Biochemistry* **49**, 7282–7296 (2010).
368. Kanzaki, H., Imura, D., Nitoda, T. & Kawazu, K. Enzymatic dehydrogenation of cyclo(l-Phe-l-Leu) to a bioactive derivative, albonoursin1. *Journal of Molecular Catalysis B: Enzymatic* **6**, 265–270 (1999).
369. Kanzaki, H., Imura, D., Sashida, R., Kobayashi, A. & Kawazu, K. Effective production of dehydro cyclic dipeptide albonoursin exhibiting pronuclear fusion inhibitory activity. I. Taxonomy and fermentation. *J. Antibiot.* **52**, 1017–1022 (1999).
370. Kanzaki, H., Yanagisawa, S. & Nitoda, T. Biosynthetic intermediates of the tetrahydro cyclic dipeptide albonoursin produced by *Streptomyces albulus* KO-23. *J. Antibiot.* **53**, 1257–1264 (2000).
371. Shin, C., Hayakawa, M., Mikami, K. & Yoshimura, J. Synthesis and configurational assignments of albonoursin and its three geometric isomers. *Tetrahedron Letters* **18**, 863–866 (1977).
372. Arunrattiyakorn, P., Nitoda, T. & Kanzaki, H. Enzymatic conversion-based method for screening cyclic dipeptide-producing microbes. *Peptides* **27**, 633–639 (2006).
373. Kanzaki, H., Yanagisawa, S. & Nitoda, T. Enzymatic Synthesis of Dehydro Cyclo(His&ndash;Phe)s, Analogs of the Potent Cell Cycle Inhibitor, Dehydrophenylahistin, and Their Inhibitory Activities toward Cell Division. *Bioscience, Biotechnology, and Biochemistry* **68**, 2341–2345 (2004).

374. Tang, M. R., Sternberg, D., Behr, R. K., Sloma, A. & Berka, R. M. Use of transcriptional profiling & bioinformatics to solve production problems: Eliminating red pigment production in a *Bacillus subtilis* strain producing hyaluronic acid. *Industrial Biotechnology* **2**, 66–74 (2006).
375. Guengerich, F. P. Common and Uncommon Cytochrome P450 Reactions Related to Metabolism and Chemical Toxicity. *Chem. Res. Toxicol.* **14**, 611–650 (2001).
376. Isin, E. M. & Guengerich, F. P. Complex reactions catalyzed by cytochrome P450 enzymes. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1770**, 314–329 (2007).
377. Hannemann, F., Bichet, A., Ewen, K. M. & Bernhardt, R. Cytochrome P450 systems—biological variations of electron transport chains. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1770**, 330–344 (2007).
378. Gao, Q. *et al.* Gene expression diversity among *Mycobacterium tuberculosis* clinical isolates. *Microbiology* **151**, 5–14 (2005).
379. McLean, K. J. *et al.* Characterization of active site structure in CYP121. A cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. *J. Biol. Chem.* **283**, 33406–33416 (2008).
380. Leys, D. *et al.* Atomic Structure of *Mycobacterium tuberculosis* CYP121 to 1.06 Å Reveals Novel Features of Cytochrome P450. *J. Biol. Chem.* **278**, 5141–5147 (2003).
381. Seward, H. E., Roujeinikova, A., McLean, K. J., Munro, A. W. & Leys, D. Crystal Structure of the *Mycobacterium tuberculosis* P450 CYP121-Fluconazole Complex Reveals New Azole Drug-P450 Binding Mode. *J. Biol. Chem.* **281**, 39437–39443 (2006).
382. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* **32**, 1792–1797 (2004).
383. Gouy, M., Guindon, S. & Gascuel, O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol Biol Evol* **27**, 221–224 (2010).
384. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol* **59**, 307–321 (2010).
385. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
386. Letunic, I. & Bork, P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucl. Acids Res.* **39**, W475–W478 (2011).
387. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2012).
388. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990).
389. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
390. Sambrook, J. & Russell, D. W. *Molecular cloning: a laboratory manual*. (Cold Spring Harbor Laboratory Press, 2001).
391. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23–28 (1990).
392. Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).
393. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41**, 207–234 (2005).
394. Jeedigunta, S., Krenisky, J. M. & Kerr, R. G. Diketopiperazines as Advanced Intermediates in the Biosynthesis of Ecteinascidins. *Tetrahedron* **56**, 3303–3307 (2000).
395. Braud, S. *et al.* Dual expression system suitable for high-throughput fluorescence-based screening and production of soluble proteins. *J. Proteome Res.* **4**, 2137–2147 (2005).

396. Mazmouz, R., Chapuis-Hugon, F., Pichon, V., Méjean, A. & Ploux, O. The Last Step of the Biosynthesis of the Cyanotoxins Cylindrospermopsin and 7-epi-Cylindrospermopsin is Catalysed by CyrI, a 2-Oxoglutarate-Dependent Iron Oxygenase. *ChemBioChem* **12**, 858–862 (2011).
397. Watanabe, K. & Oikawa, H. Robust platform for de novo production of heterologous polyketides and nonribosomal peptides in *Escherichia coli*. *Org. Biomol. Chem.* **5**, 593–602 (2007).
398. McDowell, D. G., Burns, N. A. & Parkes, H. C. Localised sequence regions possessing high melting temperatures prevent the amplification of a DNA mimic in competitive PCR. *Nucl. Acids Res.* **26**, 3340–3347 (1998).
399. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
400. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402 (1997).
401. Challis, G. L., Ravel, J. & Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**, 211–224 (2000).
402. Cheeseman, P., Toms-Wood, A. & Wolfe, R. S. Isolation and properties of a fluorescent compound, factor 420, from *Methanobacterium* strain M.o.H. *J. Bacteriol.* **112**, 527–531 (1972).
403. Isabelle, D., Simpson, D. R. & Daniels, L. Large-scale production of coenzyme F420-5,6 by using *Mycobacterium smegmatis*. *Appl. Environ. Microbiol.* **68**, 5750–5755 (2002).
404. Jenkins, C. M. & Waterman, M. R. NADPH–Flavodoxin Reductase and Flavodoxin from *Escherichia coli*: Characteristics as a Soluble Microsomal P450 Reductase. *Biochemistry* **37**, 6106–6113 (1998).
405. Persson, B., Hedlund, J. & Jörnvall, H. Medium- and short-chain dehydrogenase/reductase gene and protein families. *Cell. Mol. Life Sci.* **65**, 3879–3894 (2008).
406. Vasiliou, V., Pappa, A. & Petersen, D. R. Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. *Chemico-Biological Interactions* **129**, 1–19 (2000).
407. Liu, Z.-J. *et al.* The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold. *Nat Struct Mol Biol* **4**, 317–326 (1997).
408. Vasiliou, V., Thompson, D. C., Smith, C., Fujita, M. & Chen, Y. Aldehyde dehydrogenases: From eye crystallins to metabolic disease and cancer stem cells. *Chemico-Biological Interactions* **202**, 2–10 (2013).
409. Hausinger, R. P. Fe(II)/ $\alpha$ -Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Critical Reviews in Biochemistry and Molecular Biology* **39**, 21–68 (2004).
410. Aik, W., McDonough, M. A., Thalhammer, A., Chowdhury, R. & Schofield, C. J. Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Current Opinion in Structural Biology* **22**, 691–700 (2012).
411. Hangasky, J. A., Taabazuing, C. Y., Valliere, M. A. & Knapp, M. J. Imposing function down a (cupin)-barrel: secondary structure and metal stereochemistry in the  $\alpha$ KG-dependent oxygenases. *Metallomics* **5**, 287–301 (2013).
412. Hibi, M. & Ogawa, J. Characteristics and biotechnology applications of aliphatic amino acid hydroxylases belonging to the Fe(II)/ $\alpha$ -ketoglutarate-dependent dioxygenase superfamily. *Appl Microbiol Biotechnol* **98**, 3869–3876 (2014).
413. Macheroux, P., Kappes, B. & Ealick, S. E. Flavogenomics – a genomic and structural view of flavin-dependent proteins. *FEBS Journal* **278**, 2625–2634 (2011).
414. Huijbers, M. M. E., Montersino, S., Westphal, A. H., Tischler, D. & van Berkel, W. J. H. Flavin dependent monooxygenases. *Archives of Biochemistry and Biophysics* **544**, 2–17 (2014).
415. Struck, A.-W., Thompson, M. L., Wong, L. S. & Micklefield, J. S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications. *ChemBioChem* **13**, 2642–2655 (2012).
416. Schubert, H. L., Blumenthal, R. M. & Cheng, X. Many paths to methyltransfer: a chronicle of convergence. *Trends in Biochemical Sciences* **28**, 329–335 (2003).

417. Zhang, Q., van der Donk, W. A. & Liu, W. Radical-Mediated Enzymatic Methylation: A Tale of Two SAMS. *Acc. Chem. Res.* **45**, 555–564 (2012).
418. Barreiro, E. J., Kümmerle, A. E. & Fraga, C. A. M. The Methylation Effect in Medicinal Chemistry. *Chem. Rev.* **111**, 5215–5246 (2011).
419. Leung, C. S., Leung, S. S. F., Tirado-Rives, J. & Jorgensen, W. L. Methyl Effects on Protein–Ligand Binding. *J. Med. Chem.* **55**, 4489–4500 (2012).
420. Kodera, T. *et al.* A novel l-isoleucine hydroxylating enzyme, l-isoleucine dioxygenase from *Bacillus thuringiensis*, produces (2S,3R,4S)-4-hydroxyisoleucine. *Biochemical and Biophysical Research Communications* **390**, 506–510 (2009).
421. Salinas, A. E. & Wong, M. G. Glutathione S-transferases--a review. *Curr. Med. Chem.* **6**, 279–309 (1999).
422. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**, 675–679 (1991).
423. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **348**, 176–178 (1990).
424. Cochrane, J. R., White, J. M., Wille, U. & Hutton, C. A. Total Synthesis of Mycrocyclosin. *Org. Lett.* **14**, 2402–2405 (2012).
425. Luo, L. *et al.* An assay for Fe(II)/2-oxoglutarate-dependent dioxygenases by enzyme-coupled detection of succinate formation. *Analytical Biochemistry* **353**, 69–74 (2006).
426. Schäberle, T. F., Siba, C., Höver, T. & König, G. M. An easy-to-perform photometric assay for methyltransferase activity measurements. *Analytical Biochemistry* **432**, 38–40 (2013).



**Résumé de la thèse :**

Les cyclodipeptides, et leurs dérivés, les 2,5-dicétopipérazines (DKP), sont une classe de molécules naturelles particulièrement représentée chez les micro-organismes. L'objectif de ces travaux a été d'explorer la diversité moléculaire encodée par les voies de biosynthèse de DKP dépendant des synthèses de cyclodipeptides (CDPS).

**Mots-clés :**

Produits naturels, Métabolites secondaires, 2,5-dicétopipérazines (DKP) et dérivés, Synthèse peptidique non ribosomique, Synthase de cyclodipeptides (CDPS), Enzyme utilisant des ARNt chargés.

**Laboratoire de rattachement :**

Equipe Enzymologie et Biosynthèse Peptidique Non Ribosomale  
CEA/iBiTec-S, I2BC/UMR 9198  
SBIGeM, Bâtiment 144  
91191 Gif-sur-Yvette Cedex  
FRANCE

PÔLE : INGENIERIE DES PROTEINES ET CIBLES THERAPEUTIQUES

UNIVERSITÉ PARIS-SUD 11  
UFR «FACULTÉ DE PHARMACIE DE CHATENAY-MALABRY »  
5, rue Jean Baptiste Clément  
92296 CHÂTENAY-MALABRY Cedex