



HAL
open science

Caractérisation de voies de biosynthèse de dicétopipérazines chez les actinobactéries

Jerzy Witwinowski

► **To cite this version:**

Jerzy Witwinowski. Caractérisation de voies de biosynthèse de dicétopipérazines chez les actinobactéries. Biochimie [q-bio.BM]. Université Paris-Saclay, 2017. Français. NNT : 2017SACLS273 . tel-01875027

HAL Id: tel-01875027

<https://theses.hal.science/tel-01875027>

Submitted on 16 Sep 2018

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.

NNT : 2017SACLS273

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
L'UNIVERSITE PARIS-SUD

ECOLE DOCTORALE N° 577
Structure et Dynamique des Systèmes Vivants (SDSV)
Spécialité de doctorat : sciences de la vie et de la santé

Par

Jerzy WITWINOWSKI

Caractérisation de voies de biosynthèse de dicétopipérazines chez les actinobactéries

Thèse présentée et soutenue à Orsay, le 15/09/2017 :

Composition du Jury :

Pr. Nicolas BAYAN	Professeur, Université Paris-Sud	Président du jury
Pr. Annick MEJEAN	Professeure, Université Paris-Diderot	Rapporteuse
Pr. Christophe CORRE	Professeur, University of Warwick	Rapporteur
Dr Yanyan LI	Chargée de recherche, MNHN	Examinatrice
Pr. Pierre LEBLOND	Professeur, Université de Lorraine	Examineur
Dr Jean-Luc PERNODET	Directeur de recherche, CNRS	Directeur de thèse
Dr Emmanuelle DARBON	MdC, Université Paris-Sud	Co-encadrante de thèse

Remerciements

Il y a probablement trop de personnes qui soit m'ont aidé directement dans mon travail, soit par leur soutien moral ont contribué à ce que ce travail de thèse arrive à sa fin, soit ont tout simplement créé un environnement professionnel agréable qui me motivait à me lever chaque matin et à faire cent trente kilomètres de route pour aller travailler, pour que j'arrive à les remercier toutes sans oublier personne... Donc si vous ne trouvez pas votre nom par là, je suis désolé et je vous demande de m'excuser. Ce n'est pas que je ne sois pas reconnaissant pour votre apport, ce sont peut-être juste les signes d'un début de la maladie d'Alzheimer.

Je voudrais commencer en remerciant mon directeur de thèse, **Jean-Luc Pernodet**. Non seulement c'est grâce à lui que je peux maintenant soumettre ce manuscrit en vue d'obtention du grade de docteur. Mais surtout au fil de ces années il a été le directeur de thèse idéal. Une personne qui trouve toujours du temps pour discuter, malgré le fait qu'il soit en permanence surchargé (les casquettes de directeur d'équipe, directeur du département, responsable du site d'Orsay, directeur adjoint de l'école doctorale, et ainsi de suite). Une personne pleine d'idées créatives dès que les obstacles arrivent. Une personne dont les connaissances sont tout simplement impressionnantes et qui a une encyclopédie entière dans la tête. Mais aussi un directeur de thèse qui ne met jamais de pression, qui ne met jamais mal à l'aise, qui est toujours ouvert aux suggestions et qui sait laisser beaucoup d'autonomie. Un directeur de thèse qui ne rejette jamais une idée venant de son doctorant et le laisse faire des expériences qu'il souhaite faire même en n'étant pas convaincu que ce soit la meilleure voie pour obtenir les résultats. Et puis il s'agit de quelqu'un de très sympa humainement parlant,

toujours prêt à rigoler, à s'amuser, à boire un coup. Un directeur de thèse parfait, quoi. Un grand merci, Jean-Luc !

Je voudrais remercier au même titre ma co-encadrante, **Emmanuelle Darbon**. D'une efficacité redoutable, elle aussi a toujours été disponible malgré ses nombreuses responsabilités (enseignement, comité magasin, et autres) et j'ai grandement profité de son expertise scientifique. J'ai également toujours apprécié sa franchise ; quand elle a des choses à redire, elle n'y va pas par quatre chemins. Toujours bienveillante, toujours prête à aider, probablement aussi capable de recourber l'espace-temps (elle remplissait son rôle d'encadrante avec dévouement même quand elle manquait cruellement de temps). Cela a vraiment été un plaisir de travailler avec elle. Mais c'est surtout sur le plan personnel que je tiens à la remercier. Depuis le début elle m'a offert un soutien moral exceptionnel, et au fil des années elle est devenue la plus proche amie que j'aie, jusqu'à redéfinir mon concept de l'amitié... Probablement il n'y a pas de mots dans la langue française qui me permettraient d'exprimer pleinement l'immensité de ma gratitude et reconnaissance pour son dévouement en tant qu'amie. Un grand merci, Manue !

Je tiens également à remercier **Mireille Moutiez** dont la curiosité scientifique, l'intelligence, la passion pour la science, la persévérance et la bienveillance n'ont cessé de m'émerveiller. C'est grâce à ses compétences extraordinaires que l'on a pu sortir une hypothèse de la voie de biosynthèse de la bicyclomycine qui, à quelques détails près, s'est avérée correcte. Mais surtout j'ai adoré chaque minute que l'on a passée à travailler ensemble. Merci beaucoup Mireille, dommage que tu ne prennes pas de postdoc dans l'avenir proche...

Un gros merci à **Matthieu Coupet**, le stagiaire de mes rêves que j'ai eu le plaisir d'encadrer ; intelligent, compétent, sympa, persévérant, curieux, autonome, et qui en plus a réussi à débloquent mon projet scientifique là où ça coïnçait... De gros remerciements aussi pour mon autre stagiaire, **Jordan Duarte**, qui m'a beaucoup aidé avec la construction de certains plasmides et qui s'est coltiné aussi quelques manips qui n'ont abouti à aucun résultat

intéressant, qui ne sont même pas mentionnées dans cette thèse, mais qu'il fallait pourtant tenter...

Je tiens à remercier les autres personnes qui ont été impliquées dans mon projet de thèse, **Laëtitia Philippe**, qui m'a beaucoup aidé vers la fin quand le temps pressait et que je n'arrivais pas à être dans trois endroits à la fois, **Corinne Saulnier**, qui m'a donné un coup de main par ci et par là, qui a également rendu mes pauses clope agréables et qui a participé à des batailles d'eau avec autant d'énergie, **Antonio Ruzzini**, qui a déblayé le chemin pour mon travail et dont la présence a été le catalyseur pour que l'on fasse la fête au labo tous les vendredis, **Emmanuel Favry**, le rigolo du CEA qui m'a encadré (avec beaucoup de patience) pendant ma partie des purifications de protéines, **Pascal Belin**, le fétichiste des cytochrome P450 monooxygénases, qui m'a encadré (avec beaucoup moins de patience) pendant ma partie des réactions *in vitro*, **Jérôme Seguin**, qui, quand le temps commençait à presser, m'a beaucoup aidé dans la purification des protéines, **Muriel Gondry**, la big boss au CEA, toujours prête à tout interrompre pour donner un conseil, pour discuter, pour aider, **Nicolas Canu**, qui était toujours extrêmement arrangeant pour le partage de la HPLC et qui m'a spontanément proposé de l'aide et par conséquent donné un coup de main à de nombreuses reprises, **Robert Thai**, le spécialiste de la spectrométrie de masse avec un grand sourire accueillant toujours présent sur le visage... Je voudrais également remercier nos collaborateurs de l'UPMC, **Olivier Lequin**, **Isabelle Correia** et **Lucrèce Matheron** pour leur aide avec les analyses RMN et HRMS.

Merci à mes tuteurs, **Olivier Namy** et **Yanyan Li** pour leur regard extérieur et frais sur mon travail.

Un grand merci à **Cécile Martel**, la seule autre amatrice du métal que j'ai rencontrée sur mon chemin professionnel, qui m'a appris beaucoup de techniques de microbiologie particulières à *Streptomyces*, et à **Drago Haas**, le misanthrope asocial, probablement une des personnes les plus intelligentes que j'aie jamais rencontrées, et qui lui aussi, m'a appris pas mal de techniques et surtout m'a fait une véritable formation sur l'utilisation de la HPLC.

Merci également à **Sylvie Lautru** et **Isabelle Jacques** pour les discussions scientifiques et quelques coups de main.

Et finalement un grand merci à **Sylvain Pendino**, un adjoint technique avec la culture générale et curiosité dignes d'un professeur, pour tous les milieux de culture, parfois très ch***** à préparer, et pour toutes nos discussions fascinantes pendant les pauses clope.

Ensuite, je tiens à remercier toutes les personnes qui ne sont pas directement intervenues dans mon travail, mais qui, grâce à leur présence, caractère et personnalité, ont rendu le lieu de travail agréable. Un grand merci à **Britany Marta**, probablement la personne la plus gentille que j'aie jamais rencontrée, quelqu'un qui pourrait partager son énergie positive et son optimisme avec tout le monde tellement elle en a et une véritable maîtresse de la communication tactile (et oui, parfois un câlin vaut plus que mille mots de consolation). Merci également à **Soumaya Najah** pour toutes nos discussions interminables, nos blagues folles et nos conneries partagées. Merci à **Céline Aubry** et à **Clara Lejeune** pour toutes nos pauses café – le moment tant attendu de chaque journée (je remercie Céline aussi pour la bouteille de Ronmiel de Canarias partagée dans le RER et pour son aide avec la HPLC). Un grand merci à **Agathe Marcelot** ; je n'oublierai jamais nos discussions passionnantes (même si par leur fréquence et durée elles ont sûrement diminué ma productivité scientifique pendant les mois qu'elle a passés dans notre laboratoire). Merci à un misanthrope encore plus grand que Drago, **Marc Boutant**, le spécialiste des heures de travail impossibles. Merci également à **Armel Guyonvarch**, dont les connaissances et la culture générale sont tout simplement époustouflantes, et avec qui j'ai adoré discuter. Puis merci à **Morgan Babin**, la fana du moyen âge, à **Laura Giudici**, la sculpteuse dont le départ a signifié la fin de la belle époque pour moi, à **Ikram Medjaoui**, une personne absolument adorable, aux inséparables supportrices du PSG, **Audrey Vingadassalon** et **Florence Lorieux**, au gros blagueur **Thierry Dulermo**, à la dame nature (et occasionnellement mon chauffeur privé) **Catherine Esnault**, à l'expert des cas de PCR désespérés **Alain Raynal**, à la folle stakhanoviste **Mélanie Miranda**, à **Victor Kreiss**, le dépressif qui avait toujours des histoires

impossibles à raconter, à **Jennifer Perrin** alias Gueunnifer qui m'a tellement rappelé la manière d'être d'Audrey, à **Karine Tuphile**, la rigolote, et à tous ceux avec qui je n'ai pas eu l'occasion de beaucoup interagir, mais dont la présence a été vraiment agréable, notamment **Antoine Mottier, Laura Marín, Luísa Santos, Chloë Danioux, Quentin Masson, Moran Célestin, Søren Helmark, Inès Boukhris, Anthony Degournay, Esma Bentchikou, Michelle David, Stéphanie Bury-Moné, Christiane Elie, Marie-Joëlle Virolle et al.** Sans vous cela n'aurait pas été pareil. Finalement un grand merci pour le club de discussions autour de la clope, surtout à **Stéphane Demais**, l'amateur des théories du complot, des faits alternatifs, et des séries B (C, D, voire Z) et à **Philippe Le Flèche**, le passionné de la nature.

Je tiens également à remercier deux personnes qui ont eu la confiance en moi au moment où moi je l'avais complètement perdue, qui ont vu du potentiel là où moi je ne voyais que l'échec et dont l'encouragement et le soutien ont été vraiment décisifs pour ma réussite. Il s'agit de **Françoise Praz**, ma maître de stage en master 1, et de **Guennadi Sezonov**, le responsable du master 2 que j'ai suivi par la suite. Un grand merci, sans vous je ne serais jamais arrivé là où je suis. A cet endroit je tiens également à remercier **Pierre Netter**, qui lui aussi a joué un rôle non négligeable dans ces événements et à qui je suis aussi immensément reconnaissant.

Finalement, je tiens à remercier mon cercle d'amies proches pour leur soutien moral quand cela a été difficile pour moi, et pour m'avoir tiré par les oreilles de la m**** à plusieurs reprises... Je remercie donc, en plus d'Emmanuelle Darbon, **Joanna Pocij, Aleksandra Włodek** et **Małgorzata Kobus**. Un gros merci également à mes amies un peu moins proches (mais pas moins importantes pour autant), **Magdalena Potępa, Agata Andrzejewska** et **Anne Chevallereau**. Et, last but not least, un énorme merci à ma merveilleuse femme que j'aime tellement, **Małgorzata Kaszuba**.

Sommaire

Remerciements	1
Abréviations.....	13
CHAPITRE I : INTRODUCTION	17
1. Métabolisme secondaire	19
1.1. Généralités.....	19
1.2. Classification chimique des métabolites secondaires.....	20
1.2.1. Alcaloïdes	20
1.2.2. Terpénoïdes	21
1.2.3. Glycosides.....	22
1.2.4. Polycétides	23
1.2.5. Peptides	24
2. Dicétopipérazines.....	26
2.1. Généralités.....	26
2.2. Diverses fonctions des dicétopipérazines naturelles	27
2.2.1. Introduction.....	27
2.2.2. Communication	28
2.2.3. Défense contre les concurrents et les prédateurs.....	29
2.2.4. Virulence	31
2.2.5. Sidérophores.....	32
2.3. Activités biologiques des dicétopipérazines artificielles obtenues par synthèse chimique	33
3. Biosynthèse des dicétopipérazines.....	34
3.1. Biosynthèse du noyau dicétopipérazine	34
3.1.1. Introduction.....	34

3.1.2. NRPS	35
3.1.3. CDPS	39
3.2. Voies de biosynthèse de dicétopipérazines	42
3.2.1 Introduction.....	42
3.2.2. Voies de biosynthèse dépendantes des NRPS.....	42
3.2.3. Voies de biosynthèse dépendantes des CDPS	45
3.3 Mise en évidence et caractérisation des voies de biosynthèse de dicétopipérazines.....	48
4. <i>Streptomyces</i>	50
4.1. Ecologie	50
4.2. Cycle de vie	52
4.3. Génome.....	54
4.4. Métabolisme secondaire	56
5. Objectifs du travail de thèse	57
CHAPITRE II : ETUDE DE LA DIVERSITE DES CYCLODIPEPTIDE SYNTHASES (CDPS)	59
Introduction.....	61
Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity	64
Abstract.....	65
Introduction.....	65
Results	67
Search for active CDPSs identified 41 new members	67
Different catalytic residues for the two CDPS subfamilies	72
An enlarged diversity of cyclodipeptides	75
Toward prediction of CDPS specificity	79
Discussion	84

Acknowledgements.....	86
Author contributions.....	87
Online methods	87
Chemicals, bacterial strains and plasmids	87
Bioinformatics	88
Identification of the new putative CDPSs.....	88
Tree calculation.	89
Tree of sequences available at the time of writing (September 2014)	90
Selection of a set of 49 sequences.	90
Selection of a set of 12 additional sequences.....	91
HHPred Prediction.	91
Topology diagrams.....	91
Sequence Logos.	91
CDPSs genes.....	92
Expression of CDPSs in medium throughput format	93
Cyclodipeptide identification	95
Chemical synthesis of cFA, cFV, cFT, cFN and cFE.	97
Competing financial interests statement.....	99
CHAPITRE III : ETUDE D'UN CLUSTER CYCLODITYROSYL SYNTHASE – CYTOCHROME P450 CYP121 CHEZ <i>STREPTOMYCES VENEZUELAE</i>	101
Introduction.....	103
Matériel et méthodes.....	105
Souches bactériennes, conditions de culture et vecteurs	105
Manipulation de l'ADN et de l'ARN	107
Traitement bioinformatique.....	110

Obtention des vecteurs intégratifs permettant la surexpression du <i>cluster Sven_1864-5</i> de <i>S. venezuelae</i>	110
Obtention des vecteurs intégratifs permettant l'expression hétérologue du <i>cluster Rv2275-2276</i> de <i>M. tuberculosis</i>	111
Obtention des mutants de délétion des gènes <i>Sven_1864</i> et <i>Sven_1865</i>	111
Etude biochimique du cytochrome <i>SvCYP121</i>	112
Analyse des surnageants de culture de la souche sauvage de <i>S. venezuelae</i>	114
Surexpression du <i>cluster Sven_1864-5</i> chez <i>S. venezuelae</i>	114
Expression hétérologue des <i>clusters Sven_1864-5</i> et <i>Rv2275-6</i> chez <i>S. coelicolor</i> M1154.....	114
Complémentation des mutants	115
Analyses LC-MS/MS.....	115
Analyse phénotypique des mutants.....	116
Résultats	116
Organisation génétique du <i>cluster</i> codant la <i>SvCDPS</i> et le <i>SvCYP121</i> chez <i>S. venezuelae</i>	116
<i>SvCYP121</i> est capable de transformer <i>in vitro</i> le cYY en mycocyclosine	117
Le cYY est détecté dans les surnageants de culture de <i>S. venezuelae</i> mais pas la mycocyclosine.....	120
Dans les conditions de production du cYY, les gènes du <i>cluster</i> sont exprimés et forment un opéron.....	122
La surexpression du <i>cluster Sven_1864-5</i> ne permet pas d'observer la mycocyclosine	123
Expression hétérologue du <i>cluster CDPS-CYP121</i> de <i>S. venezuelae</i> et <i>M. tuberculosis</i>	123

Aucun phénotype n'a pu être associé à la délétion ou surexpression des gènes <i>Sven_1864-5</i>	124
Discussion	125
CHAPITRE IV : ETUDE DE LA VOIE DE BIOSYNTHESE DE BICYCLOMYCINE.....	129
Introduction.....	131
Bicyclomycin biosynthesis: multiple modifications of a cyclodipeptide synthesized in a tRNA-dependent CDPS manner.....	133
Introduction.....	133
Material and methods	135
Bacterial strains, culture conditions, strain manipulation and plasmids	135
DNA manipulations.....	136
Bioinformatic analyses.....	140
<i>bcmA</i> mutant construction.....	140
Construction of the plasmid expressing <i>bcmA</i>	141
Deletion of the whole <i>bcm</i> cluster.....	141
Construction of a plasmid containing the <i>bcm</i> cluster	142
Deletion of <i>bcm</i> genes on the pJWe14.1 plasmid	142
Construction of plasmids expressing <i>bcmH</i> or <i>bcmA</i>	143
Construction of expression vector of the <i>bcmA</i> gene under its native promoter	143
Cloning of <i>bcm</i> genes into an <i>E. coli</i> expression vector	143
Protein production and purification.....	144
Chemical synthesis of cIL.....	145
Enzymatic assays with purified proteins	145
Production assays in <i>S. aizunensis</i> strains	147

HPLC and LC-MS analysis of the <i>S. aizunensis</i> culture supernatants	147
HPLC and LC-MS analysis of the enzymatic assays	148
Preparative HPLC of <i>in vitro</i> enzymatic reaction products	148
NMR Analysis	149
HRMS analysis	150
Results	150
Mining the <i>S. aizunensis</i> genome for a candidate bicyclomycin biosynthetic gene cluster	150
.....	151
The putative CDPS gene is involved in bicyclomycin biosynthesis..	152
Bioinformatic analysis of the bicyclomycin biosynthetic gene cluster.	152
All biosynthetic enzymes within the <i>bcm</i> cluster are required for bicyclomycin biosynthesis	155
BcmA is a functional CDPS that synthesizes cIL.....	158
Functional characterization of the bicyclomycin BGC	158
BcmF catalyzes the transformation of dihydrobicyclomycin into bicyclomycin, the last step of the biosynthetic pathway	160
All other tailoring enzymes act in a specific order	160
<i>In vitro</i> characterization of the decoration enzymes.....	162
BcmH is dispensible for bicyclomycin export in <i>S. aizunensis</i>	166
The proposed structures of the bicyclomycin biosynthetic pathway intermediates are confirmed by NMR.....	167
Stereochemical analysis of compounds 2 and 12	170
HMRS analysis of compound 19 and bicyclomycin.....	172
Discussion	172
Acknowledgements.....	176

Conclusion générale.....	177
Bibliographie	181

Abréviations

aaRS – Aminoacyl-tRNA Synthetase

aa-tRNA – Aminoacyl-tRNA

ADN – Acide Désoxyribonucléique

AMP – Adenosine Monophosphate

ANR – Agence Nationale de la Recherche

AP – Alkaline Phosphatase

ARN – Acide Ribonucléique

ARNt – ARN de Transfert

ATP – Adénosine Triphosphate

BCM - Bicyclomycin

BGC – Biosynthetic Gene Cluster

Blast – Basic Local Alignment Search Tool

BOC - Butyloxycarbonyl

CAS – Clavaminic Acid Synthase

CDO – Cyclodipeptide Oxidase

CDPS – Cyclodipeptide Synthase

CEA – Commissariat à l'Energie Atomique et aux Energies Alternatives

CNRS – Centre National de la Recherche Scientifique

COSY – Correlation Spectroscopy

CP1 – Connector Peptide 1

cXX – cyclo-(L-X-L-X) (où X représente un acide aminé en code à une lettre)

DEPTQ - Distorsionless Enhancement by Polarization Transfer Including the
Detection of Quaternary Nuclei

DKP - Diketopiperazine

DMSO – Dimethyl Sulfoxide

DNA – Deoxyribonucleic Acid

DO – Densité Optique
DOPA – 3,4-dihydroxyphenylalanine
DTT - Dithiothreitol
EDTA – Ethylenediaminetetraacetatic Acid
EIC – Extracted Ionic Current
ELSD – Evaporative Light Scattering Detector
EMBL – European Molecular Biology Laboratory
ESI – Electrospray Ionisation
ESI-MS – Electrospray Ionisation – Mass Spectrometry
FA – Formic Acid
FPLC – Fast Protein Liquid Chromatography
HAD – Haloacid Dehydrogenase
HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMBC – Heteronuclear Multiple Bond Correlation
HPLC – High Performance Liquid Chromatography
HRMS – High Resolution Mass Spectrometry
HSQC – Heteronuclear Single Quantum Coherence
I2BC – Institute for Integrative Biology of the Cell
IPAP-HSQMBNC – In Phase / Anti Phase – Heteronuclear Single Quantum
Multiple Bond Correlation
IPTG - Isopropyl β -D-1-thiogalactopyranoside
LC-MS – Liquid Chromatography – Mass Spectrometry
LC-MS/MS – Liquid Chromatography – Tandem Mass Spectrometry
MALDI-TOF – Matrix-Assisted Laser Desorption-Ionization – Time of Flight
MFS – Major Facilitator Superfamily
MS – Mass Spectrometry
NADPH - Nicotinamide Adenine Dinucleotide Phosphate (reduced form)
NCBI – National Center for Biotechnology Information
NMR – Nuclear Magnetic Resonance
NOE – Nuclear Overhauser Effect
NRPS – Non Ribosomal Peptide Synthetase
OD – Optical Density
PCP – Peptidyl Carrier Protein

PCR – Polymerase Chain Reaction
PKS – Polyketide Synthase
PMSF – Phenylmethylsulfonyl Fluoride
PSI-Blast – Position Specific Iterative Blast
RBS – Ribosome Binding Site
RNA – Ribonucleic Acid
rpm – Rotations Per Minute
RP-HPLC – Reverse Phase High Performance Liquid Chromatography
RT-PCR – Reverse Transcription – Polymerase Chain Reaction
SARP – Streptomyces Antibiotic Regulatory Protein
SDS-PAGE – Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis
SIDA – Sndrome d’Immunodeficience Acquise
SIMPL - Signalling Molecule that Associates with Mouse Pelle-like kinase
TAP – Telomere-Associated Protein
TFA – Trifluoroacetic Acid
TOCSY – Total Correlation Spectroscopy
TP – Terminal Protein
tRNA – Transfert RNA
UMR – Unité Mixte de Recherche
UV – Ultraviolet
VIH – Virus de l’Immunodeficience Humaine
V/V – Volume/Volume
WHO – World Health Organisation
XRE – Xenobiotic Response Element

Chapitre I

Introduction

1. Métabolisme secondaire

1.1. Généralités

L'homme a utilisé les propriétés médicinales des plantes depuis l'aube du temps. Il s'est par exemple vite aperçu que l'extrait du pavot à opium (*Papaver somniferum*) calmait la douleur alors que l'extrait de la digitale (*Digitalis spp.*) administré à faible dose régularisait le rythme des battements cardiaques. L'avancement de la recherche a permis de découvrir que ces propriétés étaient dues à la présence de petites molécules bioactives, notamment la morphine dans le cas de *P. somniferum* et la digoxine dans le cas de *Digitalis spp.* (figure I.1) (Courtwright 2009, Withering 1785).

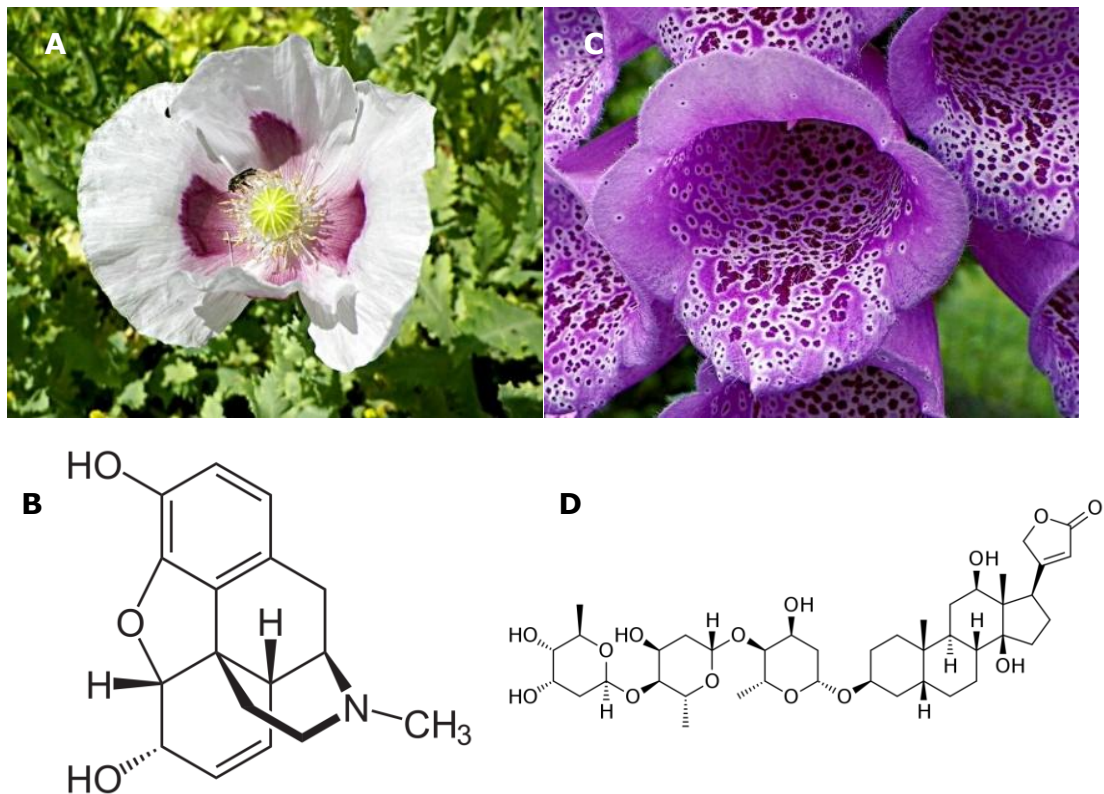


Figure I.1. Exemple de plantes médicinales et de leur molécule active: *P. somniferum* (A) et la morphine (B); *Digitalis purpurea* (C) et la digoxine (D).

La morphine et la digoxine font partie de ce que l'on appelle aujourd'hui des métabolites secondaires (ou métabolites spécialisés). Il s'agit de petites molécules synthétisées par des organismes vivants, qui sont caractérisées surtout par ce qu'elles ne sont pas. Elles ne sont pas synthétisées par un

grand nombre d'espèces, elles ne sont pas produites à toutes les étapes du développement de l'organisme producteur, et elles ne sont pas essentielles pour la croissance, la survie et la reproduction de l'organisme producteur dans les conditions de culture en laboratoire, même si elles lui procurent un avantage sélectif dans son environnement naturel (Davies 2013, Davies & Ryan 2011).

Les métabolites secondaires sont très répandus dans le monde du vivant. Ils sont notamment produits par des plantes (comme c'est le cas pour les exemples de la morphine et de la digoxine) où ils jouent souvent un rôle important dans la protection contre les prédateurs, mais on trouve également une grande abondance de métabolites secondaires chez les champignons et les bactéries où ils sont souvent impliqués dans la compétition entre les espèces et dans la signalisation intra- et interspécifique.

1.2. Classification chimique des métabolites secondaires

Les métabolites secondaires peuvent être classifiés de différentes manières. Un moyen de le faire est de se baser sur leur structure chimique, comme dans la présentation des classes majeures de métabolites secondaires qui va suivre. Cette introduction ne se veut en aucun cas exhaustive et omet quelques classes mineures de métabolites secondaires, comme les phénazines ou les phénols. Il faut aussi souligner que d'un côté la frontière entre les différentes classes est parfois floue, de l'autre les molécules hybrides ne sont pas rares, et par conséquent une molécule peut parfois appartenir à plusieurs classes en même temps.

1.2.1. *Alcaloïdes*

Les alcaloïdes constituent probablement la classe de métabolites secondaires qui comprend le plus de membres caractérisés et qui est, en même temps, la moins bien définie. Leur point commun chimiquement parlant est la présence d'un atome d'azote et le fait qu'ils soient des dérivés d'acides aminés (Waller & Nowacki 1978). On a longtemps cru qu'ils étaient produits uniquement par les plantes (avec les exemples phares de la

morphine (figure I.1B) qui fut le premier alcaloïde purifié, la quinine (figure I.2A) ou la cocaïne (figure I.2B)), chez lesquelles ils jouent souvent un rôle dans la protection contre les prédateurs grâce à leur toxicité. Néanmoins, on les retrouve aussi produits dans d'autres parties de l'arbre du vivant (Waller & Nowacki 1978) : la tétrodotoxine est par exemple trouvée chez le poisson fugu (*Takifugu spp.*) (figure I.2C) et la psilocybine est produite par le champignon *Psilocybe semilanceolata* (figure I.2D) (Guzmán 2005 ; Narahashi *et al.* 1964). De nombreux alcaloïdes ont trouvé un usage médical (comme la morphine en traitement de la douleur ou la quinine en traitement du paludisme) ou récréatif (comme la cocaïne qui est un stupéfiant ou la psilocybine qui est un hallucinogène puissant).

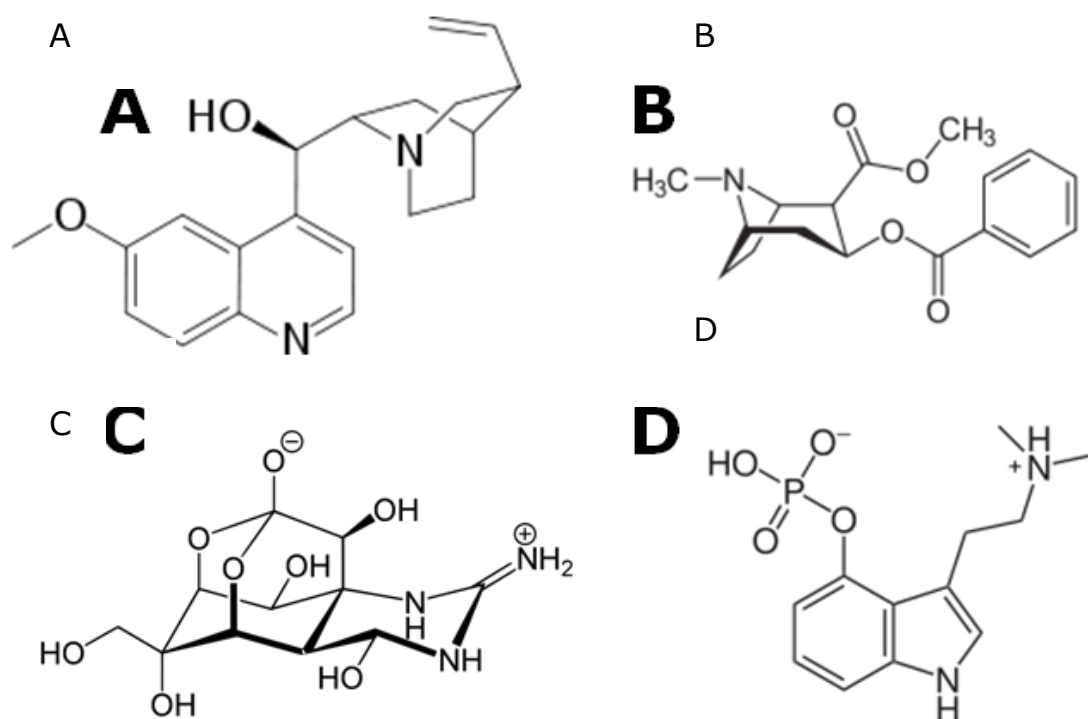


Figure I.2. Exemples d'alcaloïdes. A – quinine (produit naturel des plantes du genre *Chinchona*), B – cocaïne (produit naturel des plantes du genre *Erythroxylum*), C – tétrodotoxine (produit naturel des poissons du genre *Takifugu*), D – psilocybine (produit naturel des champignons du genre *Psilocybe*).

1.2.2. Terpénoïdes

Les terpénoïdes sont des molécules dérivées d'unités d'isoprène assemblées de différente manière et souvent fortement modifiées. C'est une classe extrêmement riche et diversifiée. On peut ainsi citer des exemples trouvés chez les plantes, comme le menthol (figure I.3A), apprécié pour son

arôme, ou l'artémisinine (figure I.3B), qui est utilisée en médecine pour ses propriétés antipaludéennes (Tholl 2015), mais aussi des métabolites secondaires de champignons, comme l'aphidicoline (figure I.3C), ou encore d'actinobactéries, comme la terpentécine (figure I.3D), les deux ayant des propriétés antibiotiques (Toyomasu *et al.* 2014, Hamano *et al.* 2002).

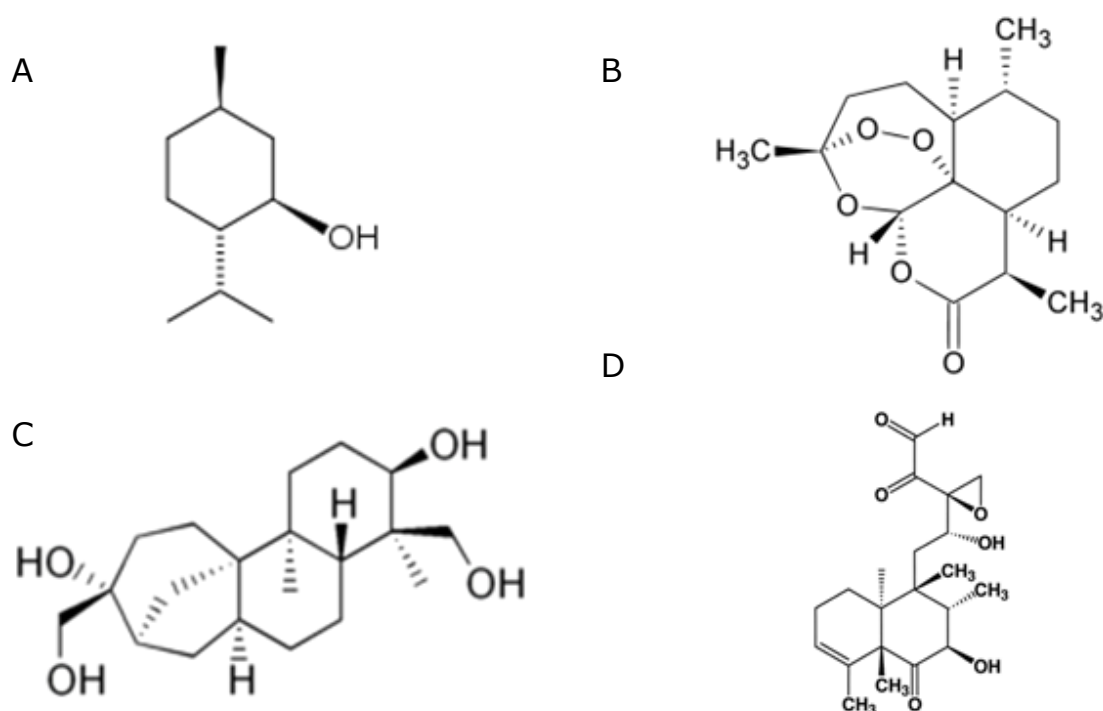


Figure I.3. Exemples de terpénoïdes. A – menthol (produit naturel des plantes du genre *Mentha*), B – artémisinine (produit naturel de la plante *Artemisia annua*), C – aphidicoline (produit naturel du champignon *Nigrospora oryzae*), D – terpentécine (produit naturel de la bactérie *Streptomyces griseolosporeus*)

1.2.3. Glycosides

Le terme « glycoside » décrit toute molécule contenant des sucres liés par une liaison glycosidique, pouvant ainsi englober de nombreuses molécules du métabolisme primaire comme secondaire. Les glycosides diffèrent par les sucres contenus ainsi que par la partie de la molécule ne contenant pas de sucre que l'on appelle aglycone. Les molécules appartenant à cette famille peuvent être trouvées dans tout l'arbre du vivant, mais la plupart des exemples viennent des plantes et des bactéries. Le premier métabolite spécialisé appartenant à cette classe qui ait été découvert est

l'amygdaline (Robiquet & Boutron-Charland 1830), responsable de l'arôme des amandes amères, molécule très toxique en grande quantité (figure I.4A). La digoxine mentionnée au début de ce chapitre fait aussi partie de cette catégorie. Un autre exemple de molécule qui protège la plante qui la produit contre la prédation, mais appréciée par l'homme pour son goût, est la sinigrine, dont le produit de dégradation donne son arôme à la moutarde (figure I.4B) (Erickson & Feeney 1974). Enfin, les bactéries du genre *Streptomyces* produisent aussi des aminoglycosides. Ceux-ci constituent une des classes majeures d'antibiotiques utilisés en médecine avec les exemples de la streptomycine (figure I.4C), le premier antibiotique utilisé dans la thérapie de la tuberculose, ou de la kanamycine (figure I.4D) (Yanai & Murakami 2004, Sing & Mitchison 1954).

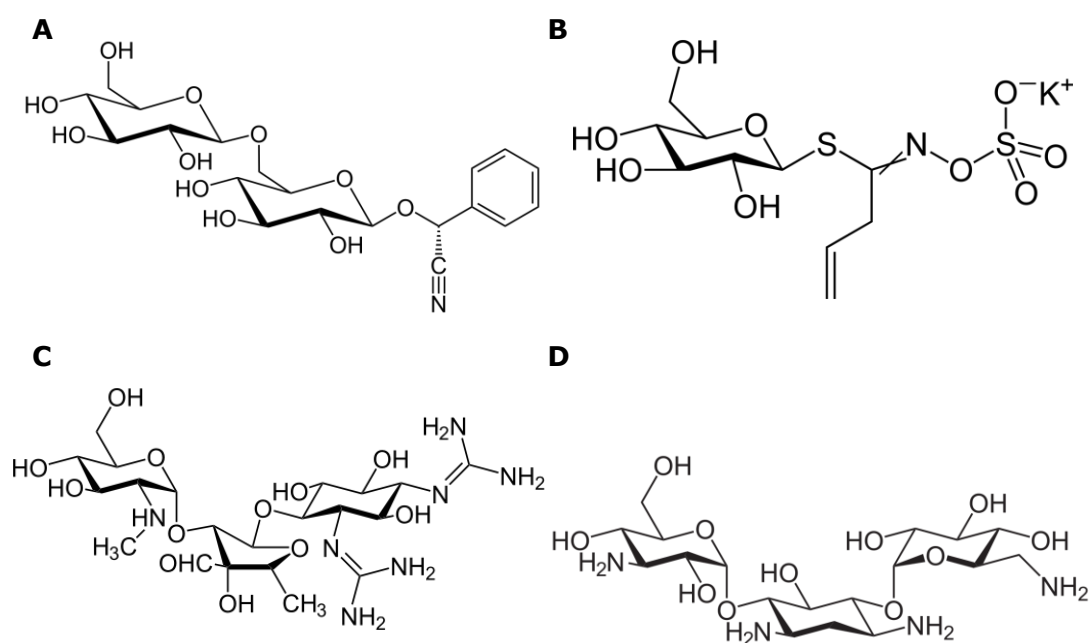


Figure I.4. Exemples de glycosides. A – amygdaline (produit naturel des plantes du genre *Prunus*), B – sinigrine (produit naturel des plantes du genre *Brassica*), C – streptomycine (produit naturel de la bactérie *Streptomyces griseus*), D – kanamycine (produit naturel de la bactérie *Streptomyces kanamyceticus*).

1.2.4. Polycétides

Les polycétides sont des produits issus de la condensation itérative d'unités acyles catalysée par des enzymes appelées polycétide synthases (PKS – Polyketide Synthase). Les PKS présentent un certain nombre de

similitudes avec les synthèses d'acides gras. Cette classe de molécules est très représentée dans les produits naturels des bactéries, mais aussi des champignons. Leur fonction primaire est la compétition entre les espèces et la protection contre les prédateurs (Staunton & Weissman 2001). De nombreux composés utilisés en médecine appartiennent à cette classe. On peut par exemple citer les antibiotiques synthétisés par les actinobactéries comme l'érythromycine (figure I.5A) ou la chlortétracycline (figure I.5B), l'antifongique nystatine (figure I.5C), mais aussi la toxine cancérigène d'origine fongique, l'aflatoxine (figure I.5D) (Staunton & Weissman 2001).

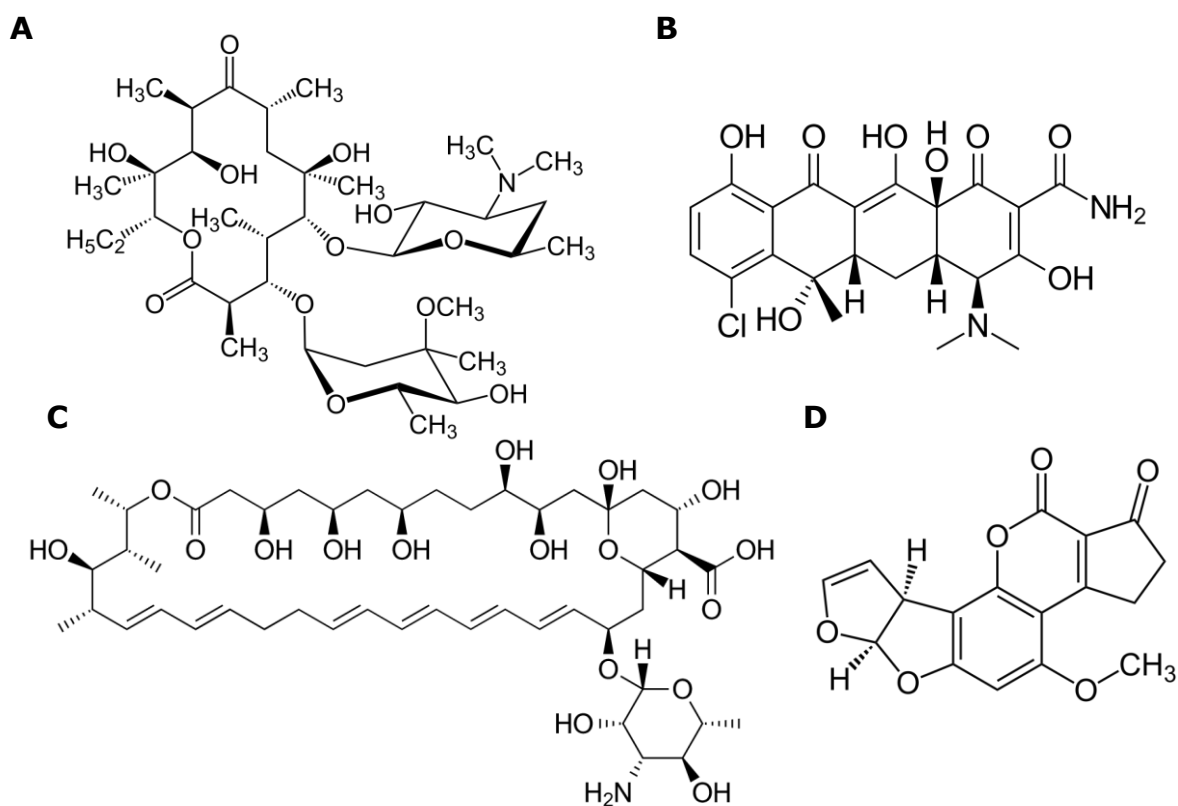


Figure I.5. Exemples de polycétides. A - érythromycine (produit naturel de la bactérie *Saccharopolyspora erythraea*), B - chlortétracycline (produit naturel de la bactérie *Streptomyces aureofaciens*), C - nystatine (produit naturel de la bactérie *Streptomyces noursei*), D - aflatoxine (produit naturel du champignon *Aspergillus flavus*).

1.2.5. Peptides

Les peptides, comme leur nom l'indique, se caractérisent par la présence d'une ou plusieurs liaisons peptidiques (amides) dans leur structure chimique. Comme pour les glycosides, il s'agit avant tout de

molécules du métabolisme primaire ; néanmoins de nombreux métabolites secondaires appartiennent aussi à cette classe. Ils peuvent être synthétisés aussi bien par la voie ribosomale (comme les protéines) puis éventuellement fortement modifiés après la traduction, que par des voies indépendantes du ribosome, notamment grâce à des synthétases de peptides non ribosomaux (NRPS – Non Ribosomal Peptide Synthetase) ou à des cyclodipeptide synthases (CDPS) (cf. la troisième partie de cette introduction). Ils peuvent être synthétisés par tous les domaines du vivant et leurs fonctions peuvent être très diverses, allant de la compétition et protection contre les prédateurs (comme dans le cas des antibiotiques produits par des actinobactéries comme le thiostrepton (figure I.6A) ou la vancomycine (figure I.6B)), en

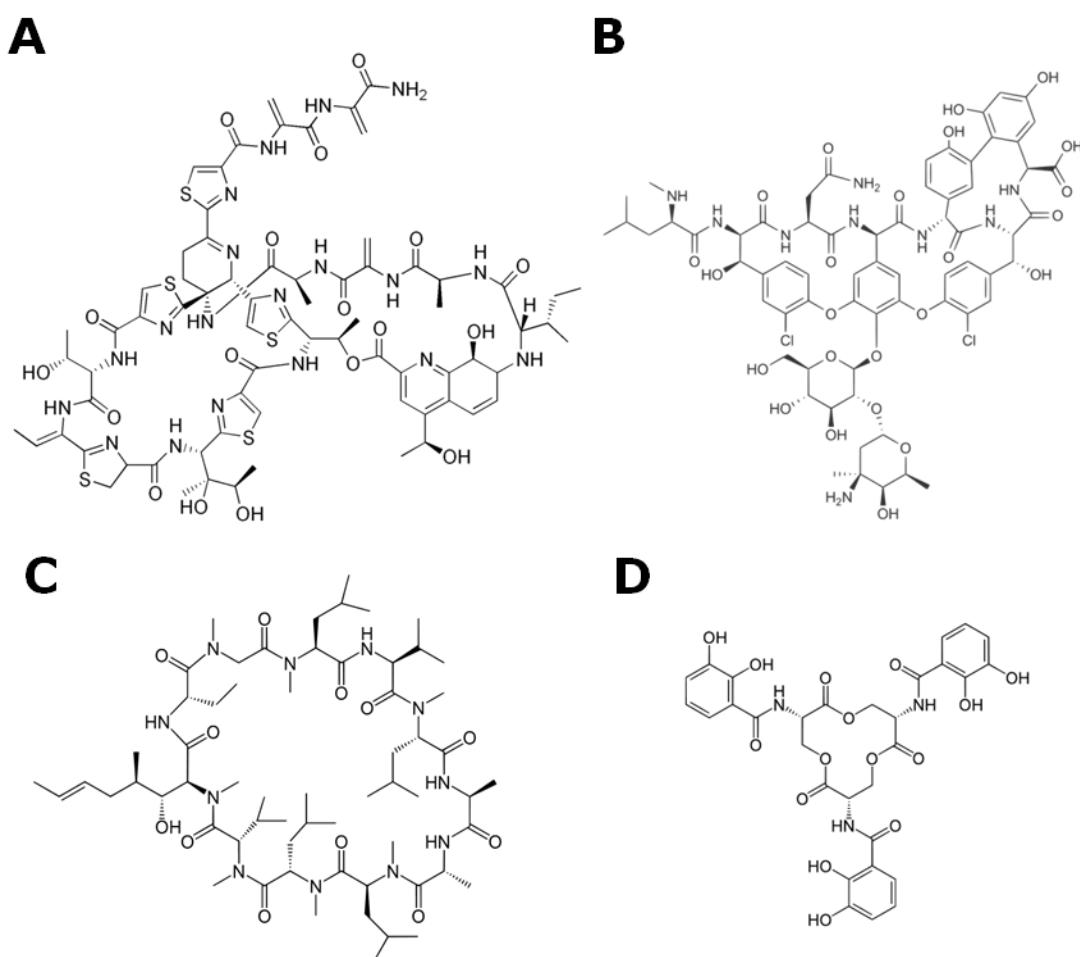


Figure I.6. Exemples de peptides. A – thiostrepton (produit naturel de la bactérie *Streptomyces azureus*), B – vancomycine (produit naturel de la bactérie *Amycolatopsis orientalis*), C – cyclosporine (produit naturel du champignon *Tolypocladium inflatum*), D – entérobactine (produit naturel de la bactérie *Salmonella typhimurium*).

passant par l'implication dans la virulence (comme dans le cas des molécules ayant des propriétés immunosuppressives et utilisées par l'homme en thérapie après les greffes, comme la cyclosporine (figure I.6C)) ou encore dans la capture du fer (les sidérophores comme l'entérobactine (figure I.6D)) (Svarstad *et al.* 2000, Pollack & Neilands 1970, Donovanick *et al.* 1955, McCormick *et al.* 1955). Les dicétopipérazines, auxquelles sera consacrée la partie suivante de l'introduction, constituent un exemple particulier de peptides.

2. Dicétopipérazines

2.1. Généralités

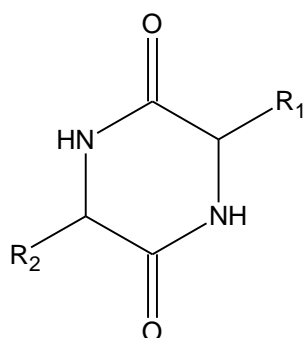


Figure I.7. Noyau pipérazine-2,5-dione. Ce noyau est à la base de toutes les dicétopipérazines. R_1 et R_2 sont les chaînes latérales des acides aminés alpha qui forment le noyau.

Les dicétopipérazines (figure I.7) sont des molécules se caractérisant par la présence d'un noyau pipérazine-2,5-dione, obtenu par formation d'une double liaison peptidique entre deux acides aminés alpha. Par conséquent les noms « dicétopipérazine » et « cyclodipeptide » sont interchangeables dans la majorité des cas. De nombreux produits naturels de tous les domaines du vivant appartiennent à cette classe de molécules. Le noyau dicétopipérazine constitue un squelette chimique avec une conformation contrainte, où de la diversité peut être introduite dans les six positions et la stéréochimie contrôlée à quatre positions. Les dicétopipérazines sont résistantes à la protéolyse, contrairement aux peptides linéaires (Borthwick 2012). Les fonctions biologiques des dicétopipérazines chez l'organisme producteur ainsi que leurs activités biologiques exploitables par l'homme peuvent être très diverses et seront passées en revue de manière non exhaustive ci-dessous.

2.2. Diverses fonctions des dicétopipérazines naturelles

2.2.1. Introduction

Avant de commencer une revue des dicétopipérazines classifiées par rapport à leur fonction, il faut définir deux concepts souvent confondus : la fonction biologique (nommée aussi rôle biologique) et l'activité biologique. La notion de fonction est définie par rapport à l'organisme producteur du métabolite en question. Elle permet de répondre à la question « en quoi la molécule est utile pour le producteur ? ». L'activité biologique en revanche est un effet mesurable et observable de la molécule qui peut être mis en évidence par des méthodes scientifiques. Souvent les deux sont liées : une molécule donnée a une fonction biologique donnée grâce à son activité biologique. Néanmoins, en faisant les recherches systématiques on peut trouver des activités biologiques qui n'ont pas d'utilité visible pour l'organisme producteur, mais que l'homme peut exploiter. Pour clarifier, prenons l'exemple du champignon *T. inflatum* mentionné dans le chapitre précédent. Pendant le stade végétatif, c'est un saprophyte vivant dans le sol. Mais durant le stade sexuel, c'est un parasite des coléoptères. Ce parasitisme est facilité par la production d'un métabolite secondaire, la cyclosporine, dont la fonction biologique est donc facteur de virulence. La cyclosporine agit en supprimant les défenses immunitaires de l'hôte (ce qui facilite l'infection), son activité biologique est donc immunosuppressive. Des études ultérieures ont montré que l'activité de la cyclosporine n'est pas limitée aux insectes, mais qu'elle est capable aussi d'inhiber l'activité des lymphocytes T humains (activité biologique accessoire, en soi inutile à l'organisme producteur). Cette dernière activité a été utilisée par l'homme qui emploie la cyclosporine comme médicament immunosuppresseur, permettant la généralisation des greffes (Matsuda et Koyasu 2000, Bissett 1983). Par conséquent, même si dans les paragraphes suivants les molécules sont classées par leur fonction biologique, il est également souvent mention de leur activité biologique, tant naturelle, par laquelle cette fonction peut être réalisée, qu'accessoire, ainsi que de leur utilisation par l'homme.

2.2.2. Communication

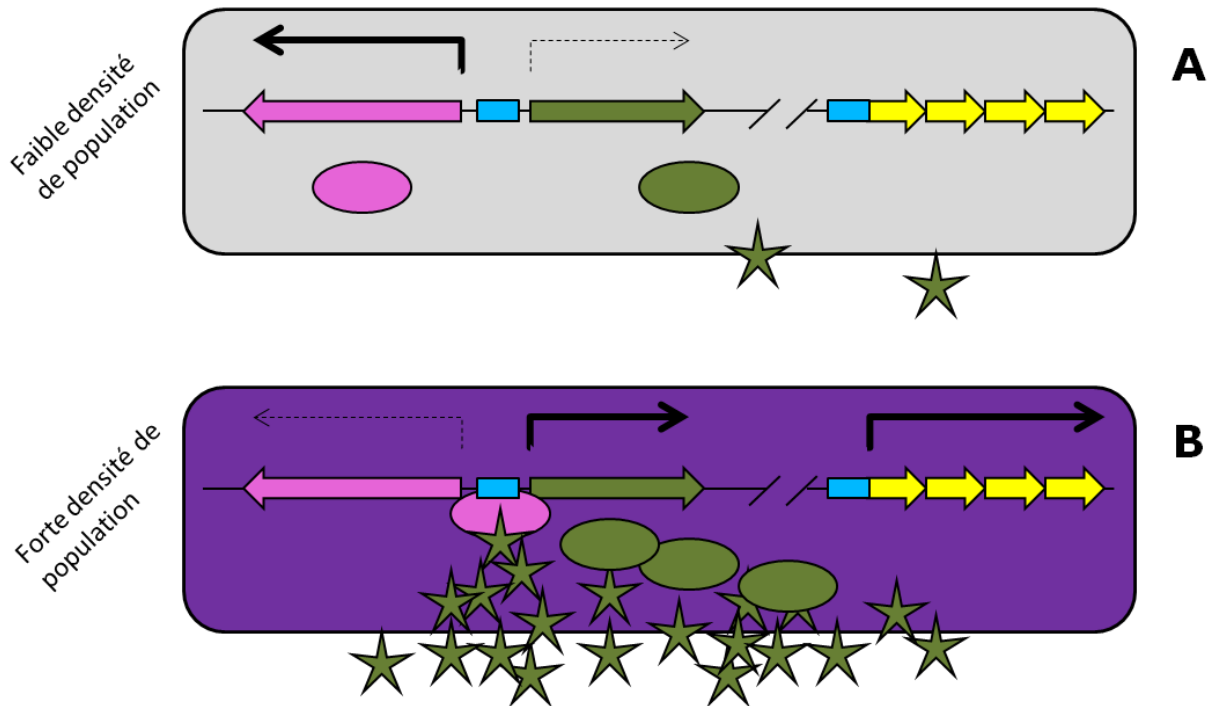


Figure I.8. Quorum sensing. En rose: le gène codant le récepteur (flèche) ainsi que le récepteur (ellipse). En vert le(s) gène(s) codant la (les) protéine(s) synthétisant le signal (flèche), la (les) protéine(s) (ellipse) et la molécule signal (étoile). En jaune les gènes codant les protéines effectrices. En bleu la région opératrice, une région d'ADN fixant le récepteur activé. Quand la densité de population est faible (**A**), la molécule signal diffuse dans le milieu et n'a pas d'activité biologique. Le promoteur du gène codant le récepteur est très actif, en revanche transcription du ou des gènes de biosynthèse de la molécule signal est faible. Quand la densité de population devient forte (**B**), la concentration de la molécule signal dépasse une valeur seuil, cette molécule se lie à son récepteur, qui a son tour se lie à son opérateur (séquence reconnue sur ADN) en permettant l'activation de la transcription des gènes effecteurs qui sont responsables d'une réponse biologique (ici marquée schématiquement par le changement de couleur de la cellule). La transcription du ou des gènes codant les protéines responsables de la biosynthèse du signal devient forte, dans un mécanisme de rétrocontrôle positif, alors que la transcription du gène codant le récepteur devient faible (rétrocontrôle négatif).

Un des processus biologiques dans lequel les dicétopipérazines jouent un rôle est le *quorum sensing*, un phénomène de communication permettant aux bactéries d'estimer leur densité de population et de coordonner l'expression de certains gènes en fonction de cette densité (figure I.8). L'exemple le mieux connu a été étudié chez plusieurs espèces de *Vibrio*, notamment *Vibrio vulnificus*, *Vibrio cholerae*, *Vibrio parahaemolyticus* et *Vibrio harveyi* qui produisent du cyclo-(L-phénylalaninyl-L-prolyl) (cFP, figure I.9A) comme signal naturel de *quorum sensing*. L'expression de gènes de

virulence (notamment des gènes *ctxA* et *ctxB* codant la toxine cholérique chez *V. cholerae*) est contrôlée par la concentration de cFP. Cette activation transcriptionnelle dépend de la présence du senseur transmembranaire ToxR (Park *et al.* 2006).

L'implication des dicétopipérazines dans la communication entre les cellules n'est pas observée uniquement chez les bactéries. En effet chez l'homme, le cyclo-(L-histidyl-L-prolyl) (cHP, figure I.9B), un métabolite issu de la dégradation de l'hormone thyroïdienne, fait partie d'un système complexe qui régule la signalisation dopaminergique (Borthwick 2012, Prasad 1995).

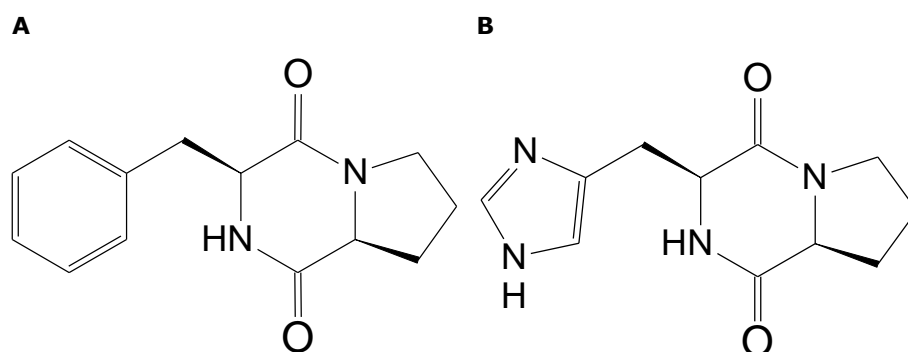


Figure I.9. Exemples de dicétopipérazines impliquées dans la communication entre les cellules. A – cFP, signal de *quorum sensing* chez *Vibrio spp.* **B** – cHP, assurant des fonctions neurohormonales chez *Homo sapiens*.

2.2.3. Défense contre les concurrents et les prédateurs

Deux autres processus majeurs auxquels les dicétopipérazines participent sont la concurrence entre les microorganismes et la défense contre les prédateurs (surtout dans le cas des champignons développant des structures macroscopiques). Ces molécules peuvent ainsi être intéressantes pour le développement de médicaments, notamment des antibiotiques, des antifongiques, des anticancéreux, mais aussi des inhibiteurs d'enzymes.

Un des exemples est la bicyclomycine (figure I.10A), un antibiotique utilisé en médecine, actif contre un large spectre de bactéries à Gram négatif. La bicyclomycine est un inhibiteur du facteur de terminaison Rho (Zwiefka *et al.* 1993). Il s'agit d'un produit naturel des bactéries *Streptomyces aizunensis* et *Streptomyces cinnamoneus* (connu aussi sous le nom synonyme de *Streptomyces sapporonensis*) (Bentley *et al.* 1993, Miyamura *et*

al. 1973, Miyoshi *et al.* 1972) leur permettant de se défendre contre les compétiteurs.

Un autre exemple d'une molécule ayant des propriétés antibiotiques est l'albonoursine (figure I.10B) synthétisée par *Streptomyces noursei* et *Streptomyces albus* (Borthwick 2012, Fukushima *et al.* 1972). Là aussi la défense contre les compétiteurs peut être envisagée comme fonction, bien que l'on n'en ait pas la certitude.

Le gliocladrine (figure I.10C) isolé d'un champignon marin du genre *Gliocladium* est cytotoxique envers les cellules eucaryotes, ce qui permet à son producteur de se défendre contre les compétiteurs et les prédateurs. Accessoirement, il a été démontré que le gliocladrine était capable de tuer les cellules d'une lignée de mélanome (Borthwick 2012, Yao *et al.* 2009).

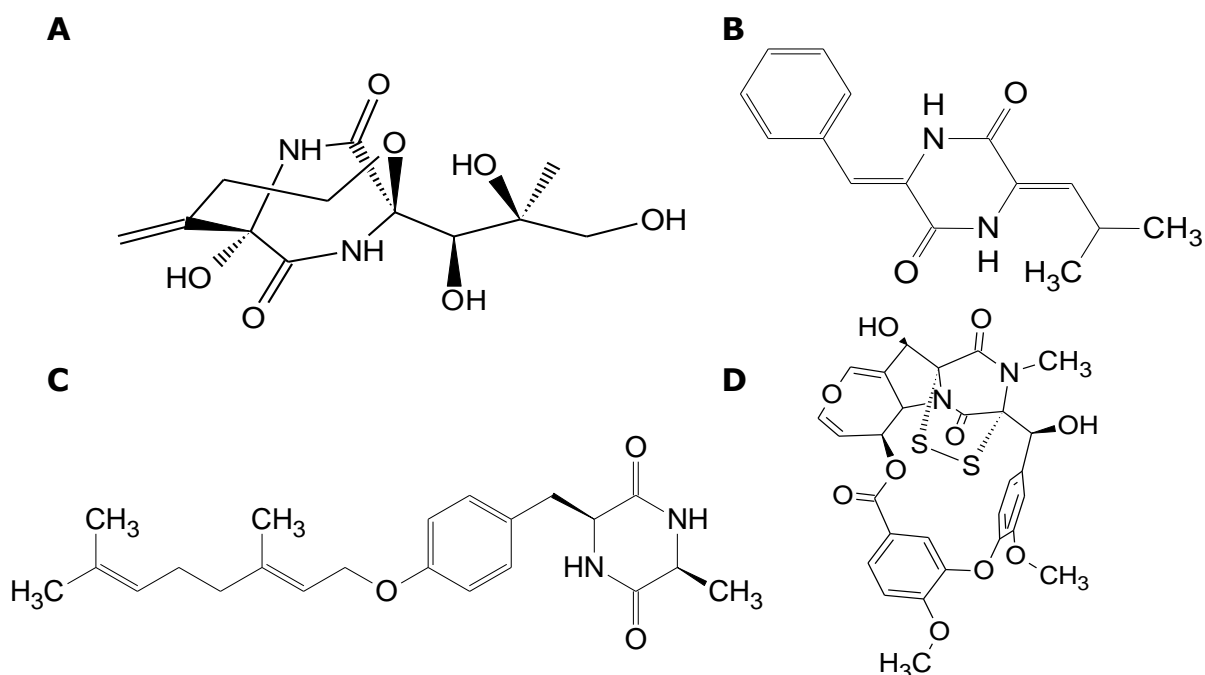


Figure I.10. Exemples de dicétopipérazines impliquées dans la concurrence entre les microorganismes. A – bicyclomycine, antibiotique utilisé en médecine, produit par *S. aizunensis* et *S. cinnamoneus*. B – albonoursine, molécule ayant des propriétés antibactériennes, produite par *S. noursei* et *S. albus*. C – gliocladrine, molécule cytotoxique produite par *Gliocladium sp.* D – MPC1001, molécule antifongique produite par *Cladorrhinum sp.* KY4922.

Le MPC1001 (figure I.10D) est une molécule intéressante tant du point de vue chimique, étant un hybride entre polycétide et dicétopipérazine, que du point de vue biologique. Ce métabolite secondaire du champignon *Cladorrhinum sp.* KY4922 possède de fortes propriétés antifongiques,

antimitotiques, mais aussi bactéricides, ce qui permet à son producteur de se défendre contre une grande variété de compétiteurs et prédateurs potentiels (Borthwick 2012, Onodera *et al.* 2004).

2.2.4. Virulence

Une partie des dicétopipérazines produites par des pathogènes obligatoires ou opportunistes facilite l'infection de leur hôte.

Par exemple, la cFP (figure I.9A), citée précédemment comme signal naturel du *quorum sensing* de plusieurs espèces de *Vibrio*, semble posséder une autre propriété : elle interfère avec la voie de signalisation NF- κ B (en inhibant la kinase I κ B) et bloque ainsi la réponse immunitaire innée de l'hôte, favorisant le développement de la maladie (Kim *et al.* 2015).

La thaxtomine A (figure I.11A) est une dicétopipérazine qui inhibe la biosynthèse de cellulose des plantes. Elle est le facteur de virulence indispensable pour le pouvoir pathogène de *Streptomyces acidiscabies* et *Streptomyces scabies*, des bactéries responsables de la galle commune des pommes de terre. Il suffit d'éliminer les gènes de biosynthèse de la thaxtomine pour que la souche devienne avirulente (Healy *et al.* 2000).

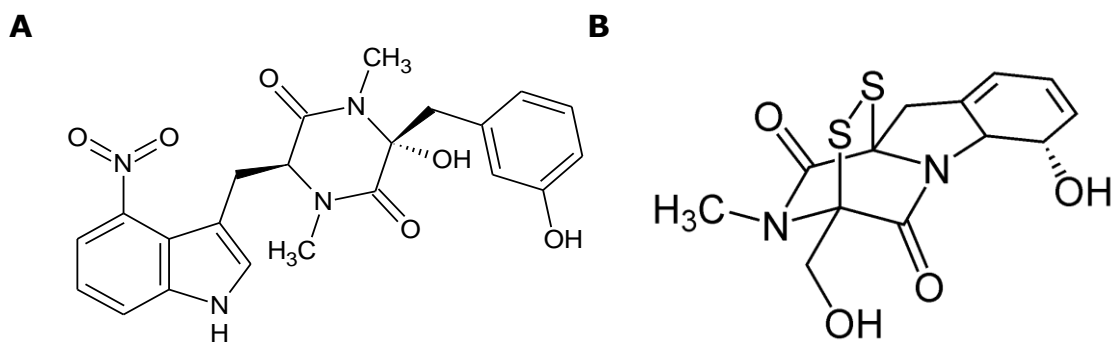


Figure I.11. Exemples de dicétopipérazines impliquées dans la virulence. A – thaxtomine A, facteur de virulence de *S. acidiscabies*. B – gliotoxine, facteur de virulence de *A. fumigatus*.

La gliotoxine (figure I.11B), une dicétopipérazine fortement toxique pour les macrophages et les lymphocytes T, est un facteur de virulence produit par *Aspergillus fumigatus*, un pathogène opportuniste s'attaquant aux humains immunodéprimés. (Gu *et al.* 2013, Borthwick 2012).

2.2.5. Sidérophores

Les sidérophores sont des molécules capables de chélater le fer, permettant ainsi aux bactéries de capturer ce métal dans les environnements où sa biodisponibilité est faible. Le complexe sidérophore-fer est par la suite importé par la bactérie à l'aide d'un transporteur spécifique. Plusieurs dicétopipérazines jouent ce rôle. C'est le cas notamment de la foroxymithine (figure I.12A) synthétisée par *Streptomyces nitrosporeus* ainsi que de l'erythrochéline (figure I.12B), produite par *Saccharopolyspora erythraea* (Lazos *et al.* 2010, Umezawa *et al.* 1985). L'acide pulcherriminique (figure I.12C) produit par *Bacillus subtilis* et les espèces apparentées joue un rôle un peu différent. Bien qu'il soit capable de capturer le fer efficacement, il n'est pas réimporté par la cellule bactérienne. Le fait que sa production soit inhibée en condition de déplétion en fer suggère qu'il ne joue pas le rôle de sidérophore. Il pourrait plutôt jouer un rôle dans la compétition, en appauvrissant l'environnement en fer disponible pour d'autres bactéries, plutôt qu'en aidant l'organisme producteur à le capturer (Randazzo *et al.* 2016, Tang *et al.* 2006).

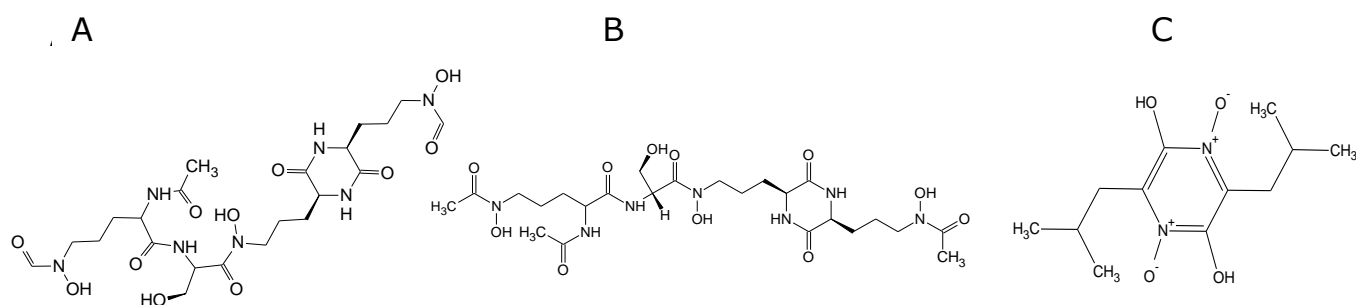


Figure I.12. Exemples de dicétopipérazines jouant le rôle de sidérophores ou chélateurs de fer. A – foroxymithine de *S. nitrosporus*. B – erythrochéline de *S. erythraea*. C – acide pulcherriminique de *B. subtilis* et *Bacillus licheniformis*

2.3. Activités biologiques des dicétopipérazines artificielles obtenues par synthèse chimique

La recherche pharmaceutique, en plus d'explorer les dicétopipérazines naturellement produites, a cherché à obtenir des produits biologiquement actifs appartenant à cette famille par synthèse chimique ou par optimisation chimique de produits naturels.

Parmi les dicétopipérazines qui ne sont pas des produits naturels, mais qui sont très importantes en médecine, on peut notamment citer le tadalafil (figure I.13A), mieux connu sous son nom commercial Cialis, un médicament utilisé pour soigner la dysfonction érectile (Borthwick 2012, Brock *et al.* 2002).

L'epelsiban (figure I.13B) est une dicétopipérazine développée par GlaxoSmithKline pour traiter l'éjaculation précoce (Shinghal *et al.* 2013, Borthwick 2012).

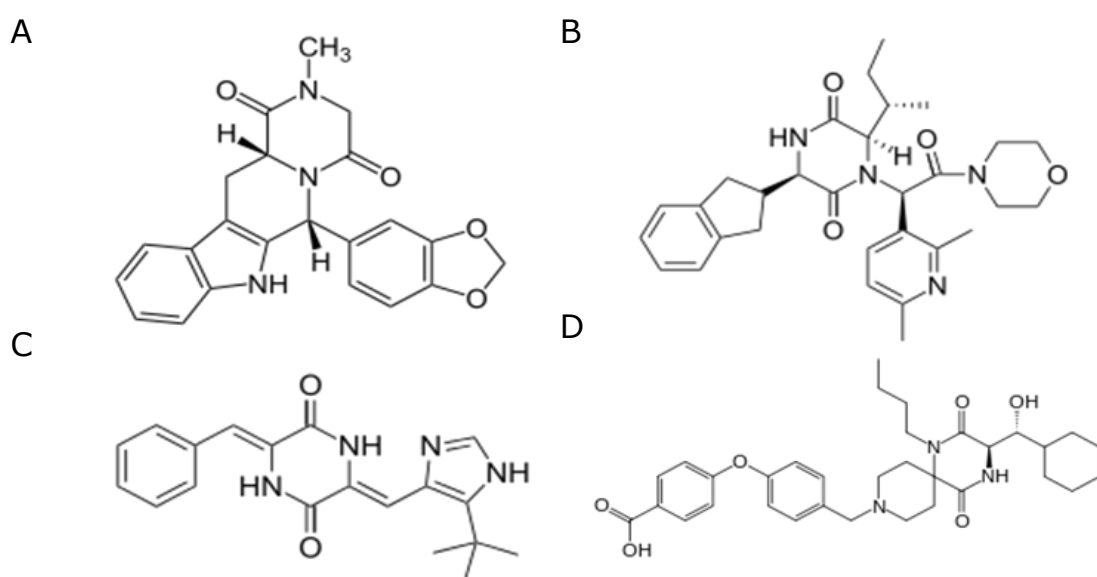


Figure I.13. Exemples de dicétopipérazines artificielles et semi-artificielles utilisées en médecine ou étudiées dans le cadre d'un usage médical potentiel. A – tadalafil (Cialis) utilisé pour traiter le dysfonctionnement érectile. B – Epelsiban, en développement par GlaxoSmithKline pour traiter l'éjaculation précoce. C – plinabuline, sous essai clinique de phase III pour le traitement du cancer. D – aplaviroc, molécule étudiée dans le cadre de la thérapie contre le SIDA.

La plinabuline (figure I.13C) est une molécule qui pourrait être employée dans la thérapie du cancer. Son développement a commencé avec la phénylahistine, un métabolite secondaire du champignon *Aspergillus ustus* qui a été soumis à l'action de la cyclodipeptide oxidase, une enzyme de la voie de biosynthèse de l'albonoursine (cf. la partie 3.2.3 de cette introduction et la figure I.22). Finalement un groupement tert-butyl y a été ajouté par des moyens chimiques, aboutissant à la plinabuline, une molécule ayant une activité antitumorale extrêmement forte, y compris sur les tumeurs de type MDR (MultiDrug Resistant). En se liant aux microtubules, la plinabuline provoque une diminution massive et sélective de la vascularisation de la tumeur et par conséquent sa nécrose. Cette molécule est en essai clinique de phase III (Borthwick 2012, Bertelsen *et al.* 2011).

L'aplaviroc (figure I.13D) a été une molécule très prometteuse inhibant l'entrée du VIH dans les cellules humaines. Malheureusement les recherches ont été interrompues à cause d'une forte hépatotoxicité (Borthwick 2012, Nichols *et al.* 2008).

Tous ces exemples montrent le potentiel des dicétopipérazines et justifient l'intérêt que leur porte la recherche scientifique.

3. Biosynthèse des dicétopipérazines

3.1. Biosynthèse du noyau dicétopipérazine

3.1.1. Introduction

Le noyau pipérazine-2,5-dione peut être synthétisé soit de façon spécifique par des enzymes dédiées, soit de façon accessoire, par libération d'un produit tronqué d'une voie de biosynthèse d'une autre molécule. Concernant le premier cas de figure, il existe deux types d'enzymes qui peuvent catalyser la formation du noyau dicétopipérazine. Les premières s'appellent synthétases de peptides non ribosomiaux (NRPS), les deuxièmes cyclodipeptide synthases (CDPS). Concernant le deuxième cas de figure, le

plus souvent il s'agit d'une voie de biosynthèse dépendant d'une NRPS normalement dédiée à la biosynthèse d'une molécule de grande taille où un produit tronqué se libère de l'enzyme en se cyclisant. Finalement il est également possible que des dicétopipérazines se forment par dégradation des peptides, bien que cette dernière voie de synthèse n'ait pas été étudiée en détail (c'est le cas notamment du cHP chez l'homme issu de la dégradation de l'hormone thyroïdienne) (Gu *et al.* 2013, Belin *et al.* 2012).

3.1.2. NRPS

Les synthétases de peptides non ribosomiaux sont des enzymes ou complexes enzymatiques avec une structure modulaire. En effet, une NRPS peut être divisée en plusieurs modules, chacun responsable de l'incorporation d'un acide aminé dans la molécule synthétisée. Chaque module peut être subdivisé en domaines ayant des fonctions catalytiques spécifiques (figure I.14). Chez les champignons, une NRPS est constituée par une seule protéine comportant plusieurs modules en général. Chez les bactéries, une NRPS est souvent un complexe hétéromultimérique où les modules et parfois les domaines se trouvent sur des protéines séparées.

Chaque module d'une NRPS est composé typiquement au minimum de trois domaines : le domaine d'adénylation, le domaine de thiolation et le domaine de condensation (ce dernier étant absent dans le module d'initiation), même si des exceptions peuvent exister. A ceux-ci peuvent s'ajouter d'autres domaines supplémentaires (comme domaines de méthylation, d'épimérisation, etc.) qui ne sont pas présents dans toutes les NRPS. L'incorporation des substrats se fait le plus souvent de façon linéaire et séquentielle, même si des exceptions existent aussi. Bien sûr, cette catégorie d'enzymes ne produit pas que des dicétopipérazines, mais aussi d'autres types de peptides non ribosomiaux (Gulick 2016, Finking & Marahiel 2004).

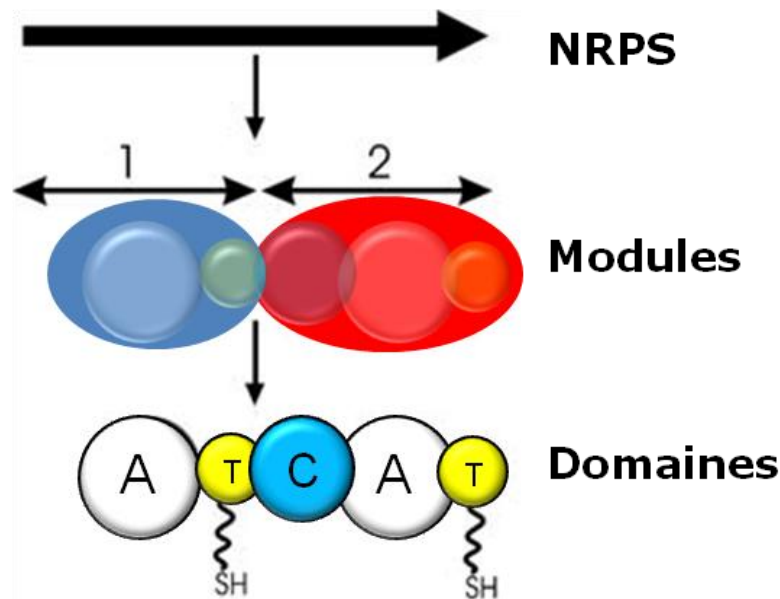


Figure I.14. Organisation modulaire des NRPS. Une NRPS se décompose en plusieurs modules, chacun incorporant un substrat dans la chaîne peptidique synthétisée. Les modules se divisent en domaines catalytiques. 1 – module d’initiation. 2 – module d’élargissement. A – domaine d’adénylation. T – domaine de thiolation. C – domaine de condensation. Adaptée de Finking & Marahiel 2004.

Le domaine d’adénylation (figure I.15A) composé d’environ 550 résidus est responsable du choix de l’acide aminé (ou acide carboxylique) qui va être incorporé. Il l’active en aminoacyl-adénylate en consommant de l’ATP.

Le domaine de thiolation (figure I.15B) appelé également le domaine PCP (pour Peptidyl Carrier Protein) est de petite taille (80-100 résidus) et comporte un cofacteur 4’-phosphopantéthéinyl fixé covalamment. Ce cofacteur est rajouté à la protéine par une phosphopantéthéinyl transférase. Il agit en tant que bras flexible permettant au substrat d’être transporté entre les différents sites catalytiques de la NRPS. Le substrat est récupéré à partir du domaine d’adénylation, fixé sur le bras phosphopantéthéinyl et à la fin transféré au domaine de condensation.

Le domaine de condensation (figure I.15C) composé d’environ 450 résidus est responsable de la catalyse de la formation du lien peptidique entre les deux substrats fixés sur deux domaines de thiolation adjacents. Il

présente aussi une certaine spécificité envers le substrat fixé au domaine de thiolation en aval qui joue le rôle de nucléophile dans la réaction.

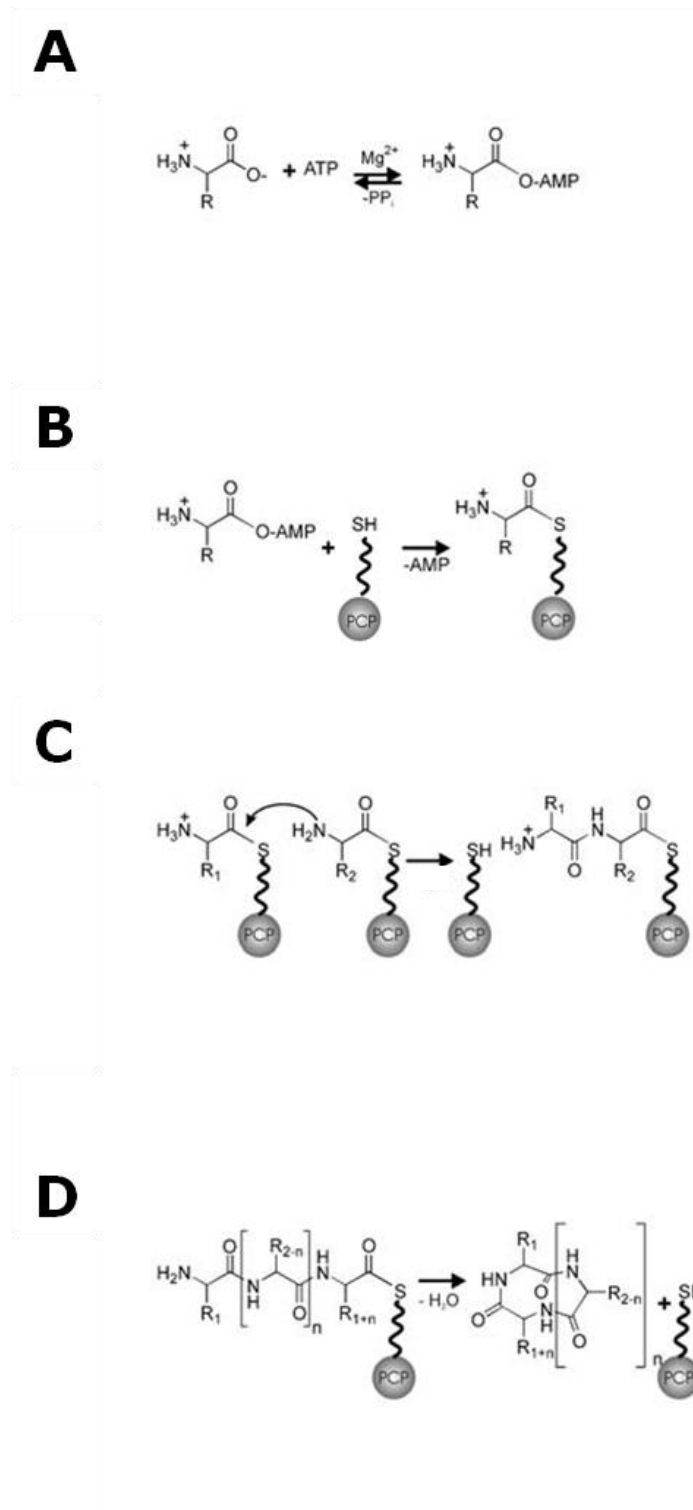


Figure I.15. Réactions catalysées par les domaines de NRPS. A – domaine d’adénylation. **B** – domaine de thiolation (domaine PCP), **C** – domaine de condensation, **D** – domaine thioestérase. Adaptée de Finking & Marahiel 2004.

Une NRPS classique se termine par un domaine thioestérase (figure I.15D) d'environ 250 résidus responsable de la libération du produit fini du domaine de thiolation du dernier module. Cette libération se fait par formation d'un intermédiaire covalent peptidyl-thioestérase, un complexe qui soit est hydrolysé pour libérer un peptide linéaire, soit subit une attaque nucléophile de la part d'un des acides aminés de la molécule synthétisée et est libéré sous forme de macrocycle. Il faut cependant mentionner que la majorité des NRPS spécialisées dans la biosynthèse de dicétopipérazines ne possède pas de domaine thioestérase, ayant à la place un domaine de condensation supplémentaire permettant de fermer le noyau pipérazine-2,5-dione.

En plus des domaines typiques décrits précédemment, un certain nombre de NRPS contient des domaines supplémentaires modifiant le squelette de la molécule synthétisée. Notamment, les domaines d'épimérisation peuvent transformer les acides aminés L en D et inversement, les domaines de cyclisation peuvent former des hétérocycles entre des résidus contenant des groupements hydroxyle ou thiol dans leurs chaînes latérales. Les domaines de méthylation peuvent quant à eux méthyler le substrat et finalement les domaines réductases et oxygénases peuvent réaliser des réactions d'oxydoréduction (Finking & Marahiel 2004).

Pour résumer, une NRPS, ayant pour fonction la synthèse d'un cyclodipeptide simple, doit avoir environ 2200 résidus. Elle consomme de l'ATP et peut utiliser une grande diversité d'acides aminés tant protéogéniques que non protéogéniques.

3.1.3. CDPS

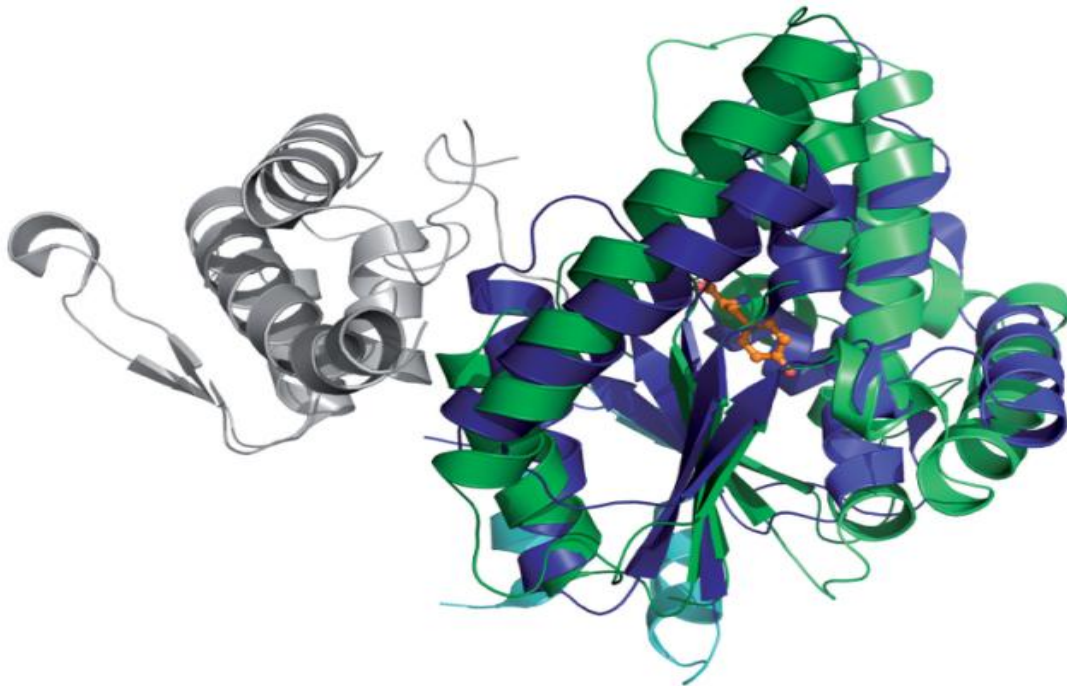


Figure I.16. Superposition d'une CDPS (AlbC de *S. noursei*) et d'une aminoacyl-ARNt synthétase (TyrRS de *Methanococcus jannaschii*). Le substrat (Y) est montré en orange. Le motif Rossmann et le CP1 de l'aa-ARNt synthétase sont colorés respectivement en bleu foncé et en bleu clair, tandis que les domaines équivalents de la CDPS sont colorés respectivement en vert foncé et en vert clair. Le domaine de reconnaissance de l'anticodon de l'aa-ARNt synthétase est coloré en gris. La région N-terminale de l'aa-ARNt synthétase est colorée en bleu très clair. D'après Sauguet et *al.* (2011).

Les CDPS forment une classe de petites enzymes de 200-300 résidus utilisant comme substrats les ARNt chargés et ne consommant pas directement d'ATP (Gondry *et al.* 2009, Lautru *et al.* 2002). Même si certaines enzymes sont connues pour utiliser les aminoacyl-ARNt (par exemple les glutamyl réductases, les aminoacyl-phosphatidylglycérol synthétases ou encore les Fem aminoacyl-transférases), leur structure est très différente de celle des CDPS et aucune ne catalyse la réaction de cyclisation de deux acides aminés (Moutiez *et al.* 2017, Belin *et al.* 2012). La structure des CDPS présente des similarités avec celle des aminoacyl-ARNt synthétases de classe I (figure I.16). En effet ces deux classes d'enzymes partagent l'architecture du domaine catalytique, notamment la présence d'un motif Rossmann rattaché à un sous-domaine de polypeptide connecteur

(CP1) au niveau de la poche catalytique. En revanche, le motif HIGH de fixation de l'ATP qui sert à l'activation des acides aminés ainsi que le domaine de reconnaissance de la région anticodon de l'ARNt ne sont pas conservés au sein des CDPS. La reconnaissance du premier aminoacyl-ARNt se fait principalement grâce à la partie acide aminé. La reconnaissance du deuxième substrat se fait à la fois grâce à son acide aminé et à son ARNt. En effet, il a été démontré que lorsque plusieurs isoformes d'ARNt coexistent, toutes ne sont pas reconnues en tant que substrat. Finalement les aminoacyl-ARNt synthétases de classe I fonctionnent toujours sous forme de dimère, tandis que parmi les CDPS caractérisées jusqu'à présent, seule la CDPS de *Mycobacterium tuberculosis* pourrait être dimérique. Cependant le mode de dimérisation de cette CDPS est différent de celui des aminoacyl-ARNt synthétases et certains auteurs considèrent que cette dimérisation ne serait qu'un artefact de cristallisation (Moutiez *et al.* 2014a, Belin *et al.* 2012, Bonnefond *et al.* 2011, Sauguet *et al.* 2011).

Le mode d'action catalytique des CDPS implique un mécanisme de type ping-pong. Il a récemment été étudié en détail pour la protéine AlbC de *S. noursei* (figure I.17). Le premier aminoacyl-ARNt se lie à l'enzyme et l'acide aminé est accommodé au niveau de la poche P1. La Tyr178, conservée dans la majorité des CDPS décrites, polarise le groupement ester du substrat ce qui facilite l'attaque nucléophile de la Ser37, un résidu catalytique strictement conservé. L'activation du résidu Ser37 résulte probablement d'un brassage des protons impliquant les groupements hydroxyles adjacents du nucléotide A76 de l'ARNt. Un intermédiaire aminoacyl-enzyme est formé, l'acide aminé étant fixé de façon covalente sur la Ser37 et l'ARNt est libéré. Le deuxième substrat se lie ensuite à l'enzyme, l'acide aminé étant accommodé au niveau de la poche P2. Le groupe aminé du premier substrat est déprotoné grâce au résidu Glu182, très majoritairement conservé, et réalise l'attaque nucléophile sur le groupement carbonyle du deuxième substrat. Un intermédiaire dipeptidyl-enzyme est alors formé et l'ARNt est libéré. Finalement le résidu Tyr202, un deuxième résidu catalytique strictement conservé dans toutes les CDPS, déprotone le groupe aminé du deuxième substrat qui attaque la liaison ester entre l'enzyme et le dipeptidyle

et permet la fermeture du noyau dicétopipérazine (Moutiez *et al.* 2017, Moutiez *et al.* 2014b, Bonnefond *et al.* 2011, Sauguet *et al.* 2011, Vetting *et al.* 2011).

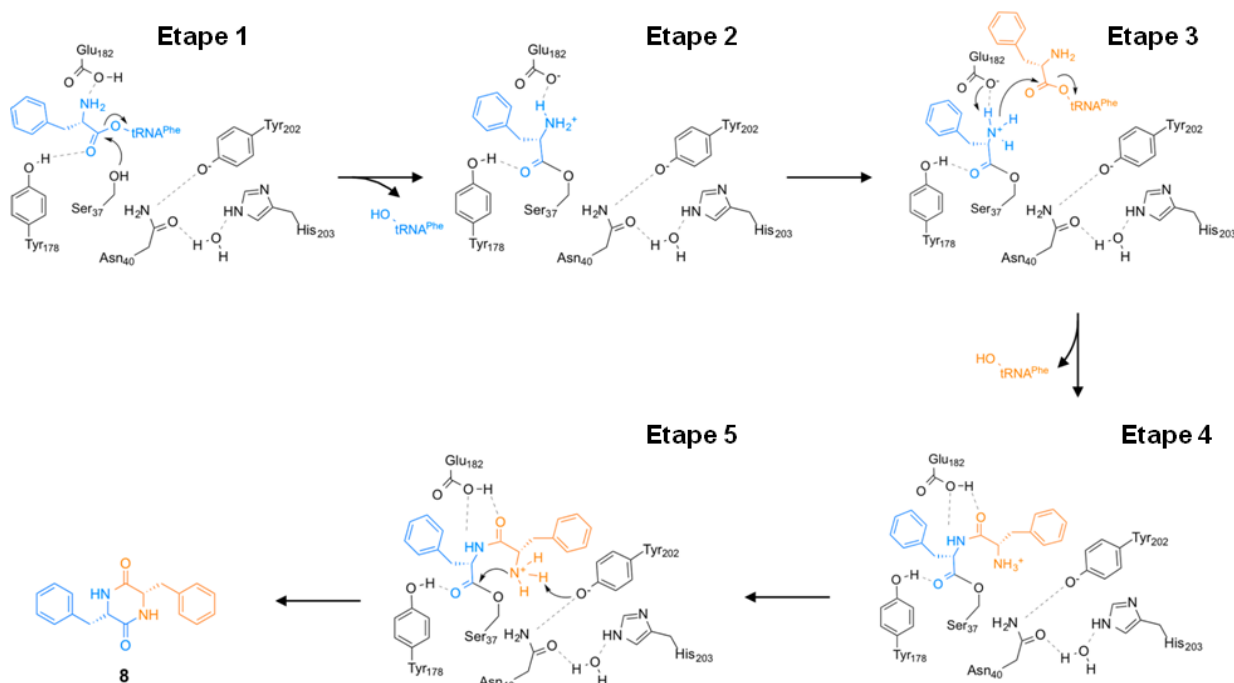


Figure I.17. Mécanisme catalytique des CDPS sur l'exemple d'AlbC de *S. noursei*. Les deux substrats consécutifs sont marqués respectivement en bleu et en jaune. Pour la description détaillée cf. le texte. Adaptée de Moutiez *et al.* 2014b.

Avant le début de cette thèse, la majorité des CDPS caractérisées avait une spécificité de substrat relaxée ce qui veut dire qu'elles catalysaient la formation d'un cyclodipeptide majeur et de plusieurs dicétopipérazines secondaires (Belin *et al.* 2012). Le travail faisant partie de cette thèse a prouvé que presque la moitié des CDPS étudiées étaient très spécifiques.

Pour résumer et comparer avec les NRPS, on peut dire que les CDPS sont des enzymes de petite taille (200-300 résidus), ne consommant pas directement de l'ATP et pouvant incorporer dans le noyau dicétopipérazine uniquement des acides aminés protéogéniques, puisqu'elles utilisent des aminoacyl-ARNt comme substrats.

3.2. Voies de biosynthèse de dicétopipérazines

3.2.1 *Introduction*

La formation du noyau pipérazine-2,5-dione est très souvent juste une des étapes d'une voie de biosynthèse plus complexe faisant intervenir une ou plusieurs enzymes de modification, également appelées enzymes de décoration. Chez les bactéries et chez les champignons, en général tous les gènes codant les enzymes nécessaires à la biosynthèse d'une molécule sont groupés dans le génome au sein de ce que l'on appelle un *cluster* de gènes de biosynthèse. Néanmoins cette règle, comme la majorité des règles en biologie, connaît des exceptions. Enfin on peut évoquer une autre règle, à savoir que dans les voies de biosynthèse dépendantes des NRPS, les modifications peuvent être faites avant, pendant et après la synthèse du noyau dicétopipérazine, tandis que dans les voies de biosynthèse dépendantes des CDPS, la formation du noyau pipérazine-2,5-dione constitue toujours la première étape (Gu *et al.* 2012).

3.2.2. *Voies de biosynthèse dépendantes des NRPS*

Le premier exemple présenté est la voie de biosynthèse de la thaxtomine A (figure I.11A), un inhibiteur de la cellulose synthase et facteur majeur de virulence de *S. acidiscabies* et *S. scabies*, pathogènes des plantes, représentée sur la figure I.18. Brièvement : la NO-synthase TxtD produit le monoxyde d'azote à partir de l'arginine. Cette molécule est prise en charge par le cytochrome P450 TxtE qui l'utilise pour nitrosyler le tryptophane. Le nitrotryptophane et une phénylalanine sont ensuite chargés sur la NRPS composée de deux protéines TxtA et TxtB. Les deux sous-unités de la NRPS comportent un domaine de méthylation qui méthyle les acides aminés chargés. La NRPS catalyse la formation du cyclo-(méthylnitrotryptophanyl-méthylphénylanyl) qui subit ensuite deux hydroxylations par un deuxième cytochrome P450 du *cluster*, TxtC, afin de devenir le produit final, la thaxtomine A (Gu *et al.* 2013, Barry *et al.* 2012, Healy *et al.* 2000).

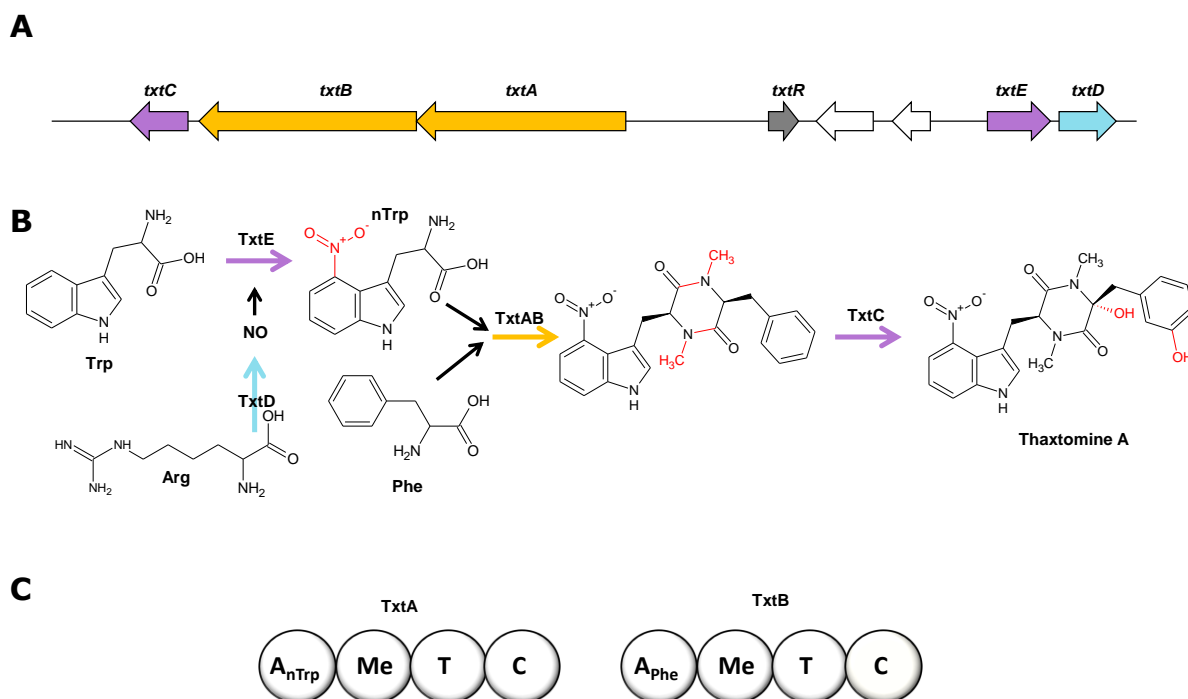


Figure I.18. Cluster (A) et voie (B) de biosynthèse de la thaxtomine A de *S. acidiscabies*. Concernant les gènes et les protéines qu'ils codent: La NRPS est marquée en orange, les cytochromes en violet, la NO-synthase en bleu clair, l'activateur transcriptionnel en gris et les transposases en blanc. **C – représentation schématique de l'organisation modulaire des deux sous-unités de la NRPS** : A – domaine d'adénylation, Me – domaine de méthylation, T – domaine de thiolation et C – domaine de condensation ; pour les domaines d'adénylation, leur spécificité est indiquée en indice.

Un autre exemple est la voie de biosynthèse de l'érythrochéline (figure I.12B), un sidérophore produit par *S. erythraea*. Cette voie de biosynthèse, représentée sur la figure I.19, est atypique parce qu'elle fait appel à des enzymes codés par des gènes qui appartiennent à deux *clusters* séparés de 2 Mbp sur le génome. Tout commence avec la synthèse de l'hydroxyornithine à partir de l'ornithine par l'ornithine monooxygénase ErcB. Une fraction de l'hydroxyornithine produite est alors acétylée par l'acétyltransférase Mcd qui est codée par un gène se trouvant dans un autre *cluster* de biosynthèse. Ce deuxième *cluster* contient également un gène codant une NRPS non fonctionnelle du fait de la modification du cadre de lecture. Finalement la NRPS ErcD synthétise l'érythrochéline en utilisant un acétyl-CoA, deux résidus acétylhydroxyornithine, un résidu hydroxyornithine et un résidu sérine (Gu *et al.* 2013, Lazos *et al.* 2010).

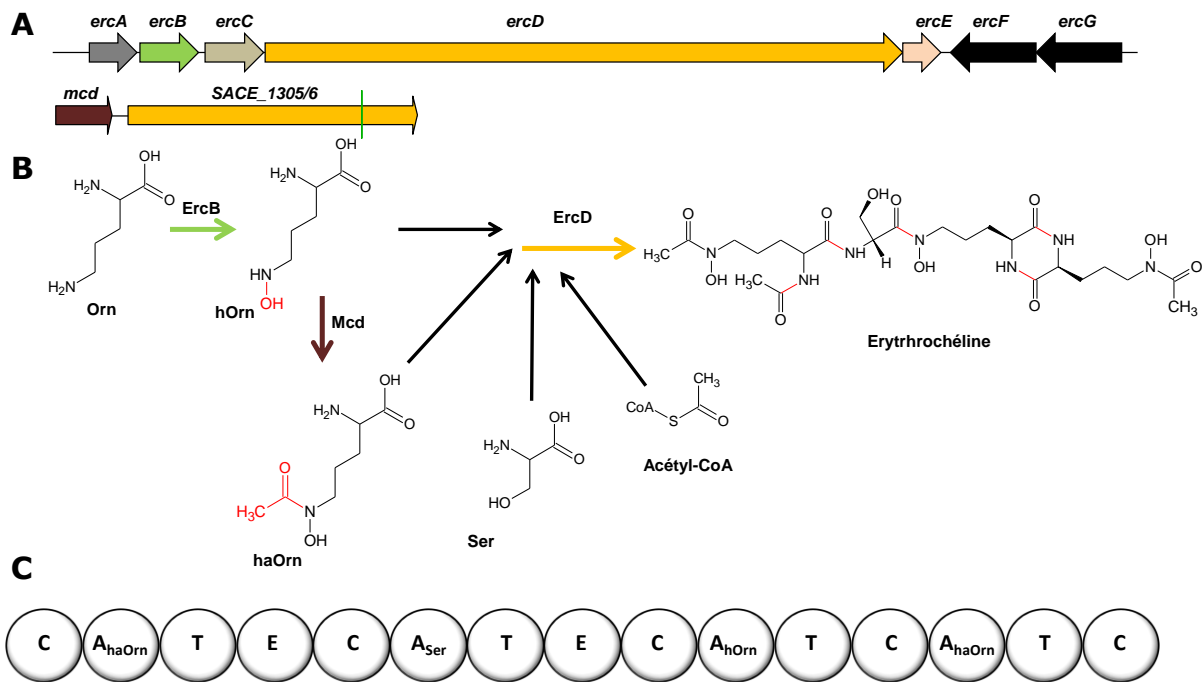


Figure I.19. Clusters (A) et voie (B) de biosynthèse de l'érythrochéline de *S. erythraea*. Les fonctions des protéines codées par les gènes du *cluster* sont les suivantes: ErcA - régulateur transcriptionnel, ErcB - ornithine monooxygénase, ErcC - récepteur à sidérophore, ErcD - NRPS, ErcE - protéine MbtH, ErcFG - transporteur de type ABC, Mcd - acétyltransférase. Le gène SACE_1305/6 codant la NRPS non fonctionnelle est également représenté, la position de la mutation du cadre de lecture est signalisée par un trait vert. **C - représentation schématique de l'organisation modulaire de la NRPS ErcD** : A - domaine d'adénylation, E - domaine d'épimérisation, T - domaine de thiolation et C - domaine de condensation ; pour les domaines d'adénylation, leur spécificité est indiquée en indice.

Enfin, le dernier exemple concerne une voie de biosynthèse impliquant la libération d'un produit tronqué. Chez *Salinispora arenicola*, le *cluster cym* permet la synthèse d'un peptide cyclique ayant des propriétés anti-inflammatoires, appelé cyclomarine A. Le *cluster* est aussi à l'origine de la synthèse d'une dicétopipérazine appelée cyclomarine A ayant des propriétés antibactériennes. Cette voie de biosynthèse est représentée sur la figure I.20. Le cytochrome P450 CymS synthétise l'hydroxyleucine à partir de la leucine. La prényltransférase CymD prényle le tryptophane. Ce prényltryptophane peut ensuite être hydroxylé par la dioxygénase CymW ou pas. Le prényltryptophane et l'hydroxyleucine sont alors incorporés sur les deux premiers modules de la NRPS CymA. Si le prényltryptophane a été hydroxylé, la voie de biosynthèse complète est engagée et la cyclomarine A est synthétisée. Cependant s'il n'est pas hydroxylé, la synthèse s'arrête au niveau d'un dipeptide, méthylé sur l'hydroxyleucine par le domaine de

méthylation de la NRPS. Le produit subit alors une cyclisation et est libéré de la NRPS, probablement avec l'aide de la thioestérase de type II CymQ. Ainsi, la cyclomarazine A est obtenue (Schulz *et al.* 2008).

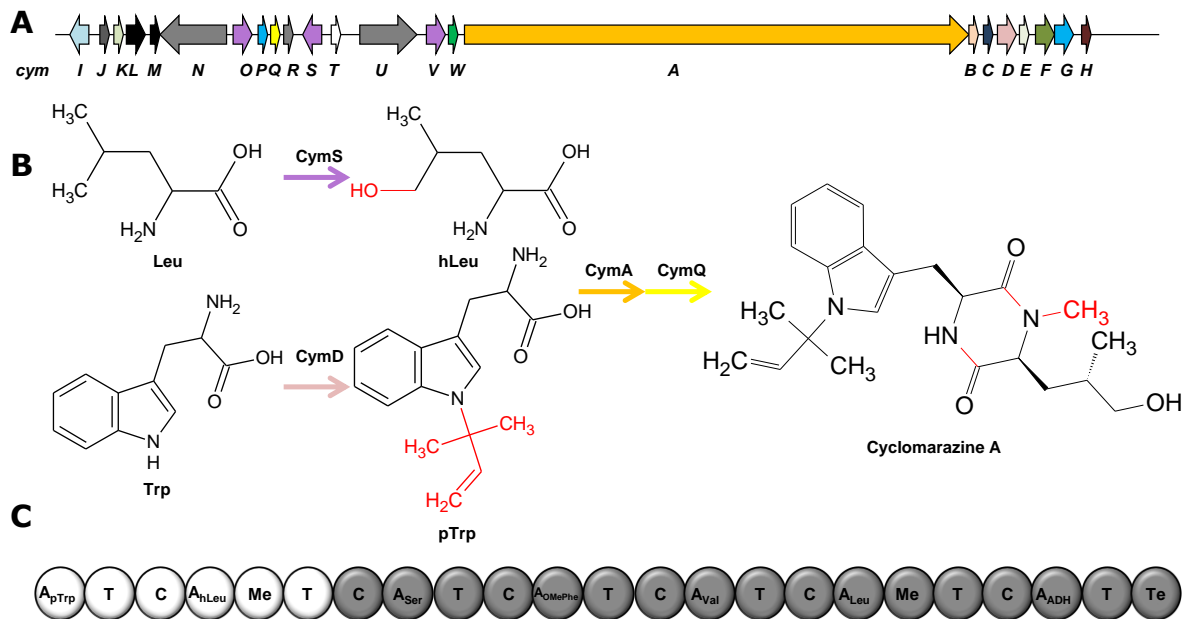


Figure I.20. Cluster (A) et voie (B) de biosynthèse de la cyclomarazine A de *S. arenicola*. Les fonctions des protéines codées par les gènes du *cluster* sont les suivantes: CymI – phosphoénolpyruvate synthase, CymJ et CymR – protéines hypothétiques, CymK – fructose-2,6-bisphosphatase, CymL et CymM – transporteurs, CymN et CymU – régulateurs transcriptionnels, CymO, CymS et CymV – cytochromes P450, CymP et CymG – méthyltransférases, CymQ – thioestérase de type II, CymT – transposase, CymW – dioxygénase, CymA – NRPS, CymB – protéine de type Mbth, CymC – isporényl diphosphate isomérase, CymD – prényltransférase, CymE – acétaldéhyde déshydrogénase, CymF – 4-hydroxy-2-cétovalérate aldolase, CymH – 2-céto-4-pentoate hydratase. **C – représentation schématique de l'organisation modulaire de la NRPS** : A – domaine d'adénylation, Me – domaine de méthylation, T – domaine de thiolation, C – domaine de condensation et Te – domaine thioestérase; pour les domaines d'adénylation leur spécificité est indiquée en indice. La partie en gris ne participe pas à la biosynthèse de cyclomarazine. En revanche la NRPS en entier peut synthétiser la cyclomarazine A.

3.2.3. Voies de biosynthèse dépendantes des CDPS

Le premier exemple que nous présentons est la voie de biosynthèse de l'albonoursine (figure I.10B), une molécule aux propriétés antibiotiques produite par *S. noursei* et *S. albulus*. Cette voie est représentée sur la figure I.21.

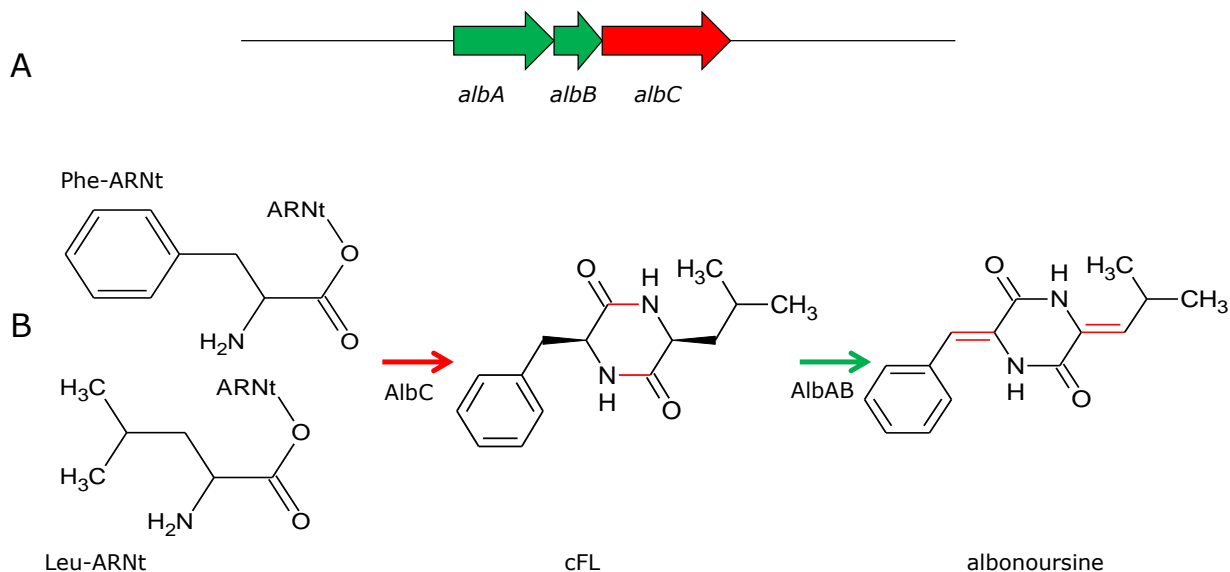


Figure I.21. Cluster (A) et voie (B) de biosynthèse de l'albonoursine de *S. noursei*. En rouge sont indiqués la CDPS et le gène la codant, en vert la cyclodipeptide oxydase et les gènes codant des deux sous-unités.

La CDPS AlbC synthétise principalement du cyclo-(L-phénylalanyl-L-leucyl) (cFL), qui est ensuite oxydé par la cyclodipeptide oxydase AlbAB pour donner l'albonoursine. Il faut rajouter que c'est la première voie de biosynthèse de dicétopipérazines impliquant une CDPS qui ait été élucidée (Gondry *et al.* 2001, Lautru *et al.* 2002).

Une voie de biosynthèse très similaire est utilisée par *Nocardioopsis dassonvillei* pour la biosynthèse des molécules de la famille des nocazines ayant des propriétés cytotoxiques et présentant *in vitro* une activité anticancéreuse. La voie de biosynthèse est représentée sur la figure I.22. La CDPS Ndas_1148 peut catalyser la formation de cYF, cYY et cFF à partir de phénylalanyl-ARNt et de tyrosyl-ARNt. Ces molécules peuvent subir une déshydrogénation par la cyclodipeptide oxydase Ndas_1146-7 au niveau d'une ou de deux chaînes latérales. Finalement la N-méthyltransférase Ndas_1145 peut introduire des groupements méthyle au niveau du noyau dicétopipérazine alors que la O-méthyltransférase Ndas_1149 peut méthyler le résidu tyrosine (Giessen *et al.* 2013a).

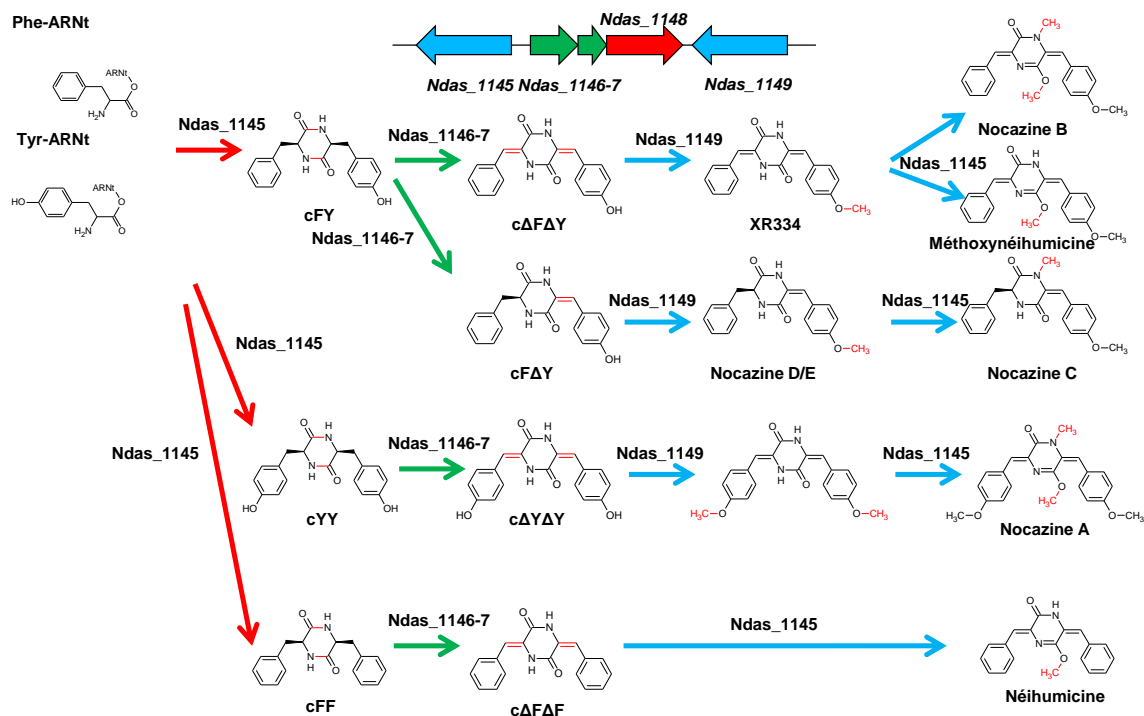


Figure I.22. Cluster et voie de biosynthèse des nocazines de *N. dassonvillei*. En rouge sont indiqués la CDPS et le gène la codant, en vert la CDO et les gènes la codant et en bleu les méthyltransférases et les gènes les codant.

Chez plusieurs espèces de *Bacillus*, notamment *B. subtilis* et *B. licheniformis*, le cLL synthétisé par la CDPS YvmC est ensuite oxydé par le cytochrome P450 CypX pour former l'acide pulcherriminique, un composé chélateur qui après avoir fixé du fer se transforme en pulcherrimine (figure I.23), un pigment rouge (Randazzo *et al.* 2016, Bonnefond *et al.* 2011, Cryle *et al.* 2010, Gondry *et al.* 2009, Tang *et al.* 2006).

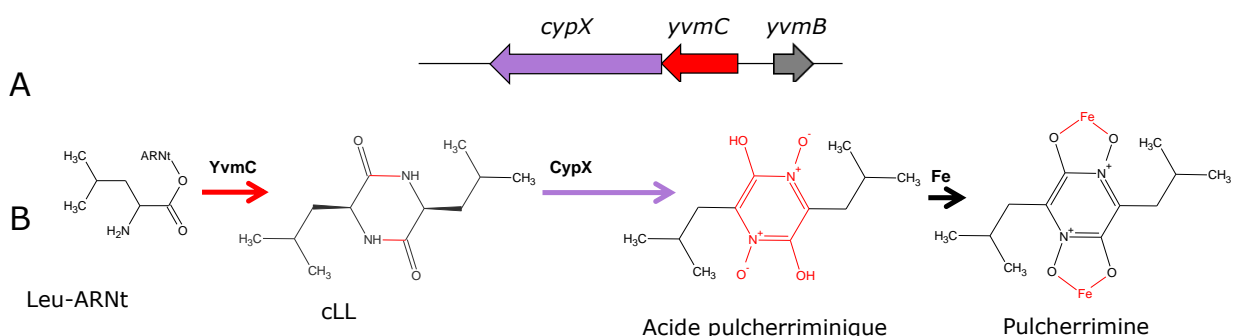


Figure I.23. Cluster (A) et voie (B) de biosynthèse de l'acide pulcherriminique de *B. subtilis*. En rouge sont indiqués la CDPS et le gène la codant, en violet le cytochrome P450 et le gène le codant, en gris le gène régulateur.

3.3 Mise en évidence et caractérisation des voies de biosynthèse de dicétopipérazines

Le sous-chapitre précédent a eu pour but de présenter quelques voies de biosynthèse de dicétopipérazines connues à ce jour. Il ne répond pourtant pas à la question : comment ces voies ont-elles été mises en évidence ? En effet le chemin entre la molécule détectée dans les surnageants de culture et à laquelle on peut attribuer une activité biologique et l'élucidation de sa voie de biosynthèse peut être long. C'était surtout le cas avant les développements technologiques qui ont permis le séquençage massif des génomes et la création d'outils bioinformatiques pour l'analyse des données ainsi générées. Il s'agissait de longues années de travail fastidieux de construction de banques d'ADN génomique, de mutagenèse, et de techniques de génétique classique.

La découverte de la voie de biosynthèse de l'albonoursine (figure I.21) en est un bon exemple. Dans une première étape, les chercheurs testèrent l'hypothèse que le cFL était un précurseur qui subissait ensuite deux oxydations. Pour cela ils ajoutèrent du cFL tritié à la culture et observèrent la production de l'albonoursine tritiée. La cinétique de cette transformation fut aussi réalisée afin de trouver le moment où l'enzyme catalysant cette réaction était présente en abondance. Des extraits protéiques totaux furent alors réalisés afin de purifier l'enzyme par chromatographie. Trois purifications successives furent nécessaires et à chaque étape il fallut vérifier l'activité de chacune des fractions. Par la suite, des fragments tryptiques de l'enzyme furent séquencés par la technique de dégradation d'Edman. La séquence de ces fragments servit pour générer des amorces dégénérées afin d'amplifier une région interne du gène *albA* par RT-PCR, car les tentatives de PCR sur l'ADN génomique avaient échoué. Ce fragment servit pour sonder une banque d'ADN génomique de *S. noursei*. Le cosmide ainsi retrouvé fut séquencé. Les auteurs de l'étude démontrèrent par expression hétérologue que les gènes *albA*, *albB* et *albC* étaient nécessaires et suffisants pour synthétiser l'albonoursine et que la protéine AlbC était suffisante pour synthétiser le précurseur cFL, ce qui fut une grande

surprise. En effet, ils s'attendaient à trouver une NRPS, cependant ils trouvèrent une enzyme dix fois plus petite qui de plus ne présentait aucune similitude de séquence avec des protéines caractérisées. Cette protéine nouvellement découverte fut baptisée cyclodipeptide synthase (Lautru *et al.* 2002, Gondry *et al.* 2001).

L'avènement de l'ère post-génomique a énormément facilité le travail des chercheurs. Désormais la voie la plus rapide pour trouver le *cluster* de biosynthèse d'une molécule donnée passe par le séquençage du génome d'un organisme producteur, ce qui est dorénavant relativement rapide et peu coûteux, et ensuite par l'analyse bioinformatique des séquences obtenues. La conjonction de l'accumulation des connaissances sur le mode d'action catalytique des principales enzymes de biosynthèse des métabolites secondaires, et des développements bioinformatiques, permettent maintenant de trouver plus facilement les *clusters* de biosynthèse, et parfois même de prédire le squelette chimique de la molécule qui va être synthétisée. Le travail en aval ne diffère pas tant que ça de ce que l'on faisait auparavant. On étudie la fonction de chaque gène soit par l'expression hétérologue, soit en créant des mutants dans la souche productrice et en analysant la production de la molécule en question ou de ses précurseurs, mais tout le travail en amont est largement simplifié. Des outils comme AntiSMASH (pour ANTIbiotics and Secondary Metabolites Analysis SHell) accessibles en ligne avec une interface simple et intuitive intègrent aujourd'hui de nombreux algorithmes de recherche des *clusters* de métabolisme secondaire ainsi que de prédiction de structure de la molécule synthétisée (Blin *et al.* 2017, Weber *et al.* 2015, Weber 2014).

La découverte de la voie de biosynthèse de l'érythrochéline (figure I.19) constitue un bon exemple de cette nouvelle manière de procéder. Le génome de *S. erythraea* fut séquencé et les recherches bioinformatiques permirent de mettre en évidence un *cluster* codant une NRPS de grande taille (figure I.19C) où les domaines d'adénylation des modules 2 et 3 étaient prédits pour incorporer respectivement la sérine et l'hydroxyornithine. Les prédictions pour les deux autres domaines d'adénylation étaient plus vagues, mais il était clair qu'ils incorporaient le même substrat. Or d'après la structure de

l'érythrochéline deux molécules d'hydroxyacétylornithine étaient nécessaires. De plus, la NRPS contenait un domaine de condensation dans le module d'initiation qui selon les prédictions pouvait incorporer un groupement acétyl directement à partir de l'acétyl-CoA ; or un des résidus hydroxyacétylornithine est acétylé. La NRPS se terminait par un domaine de condensation atypique, soupçonné de catalyser la formation du noyau dicétopipérazine. Le *cluster* contenait également un gène codant une ornithine monooxygénase, nécessaire pour la formation d'un des précurseurs. En revanche il manquait un gène codant une acétyltransférase nécessaire pour la formation de l'hydroxyacétylornithine. De nouveau, les moyens bioinformatiques s'avérèrent utiles, puisqu'une recherche d'homologues de gènes d'acétyltransférase permit de localiser un gène candidat codant une acétyltransférase dans un autre *cluster* éloigné de 2Mbp. Finalement l'obtention de mutants de la NRPS et de l'acétyltransférase permit de confirmer qu'elles étaient impliquées dans la biosynthèse de l'érythrochéline (Lazos *et al.* 2010).

4. Streptomyces

4.1. Ecologie

Streptomyces est un genre de bactéries à Gram positif au sein de la classe des actinobactéries (*Actinobacteria*), et de l'ordre des actinomycètes (*Actinomycetales*). Leur style de vie ressemble fortement au style de vie des champignons filamenteux : les deux poussent sous forme d'hyphes branchées formant un mycélium végétatif et se dispersent grâce aux spores qui se développent sur les hyphes aériennes émergeant de la colonie vers l'air. Ces similarités sont probablement dues à l'occupation de niches écologiques identiques, même si leur origine évolutive est indépendante (Flårdh & Buttner 2009, Kieser *et al.* 2000).

La grande majorité des espèces décrites jusqu'à présent sont des saprophytes vivant dans le sol où ils dégradent les matières organiques (Flårdh & Buttner 2009, Kieser *et al.* 2000). Ceci est possible grâce à un

énorme réservoir d'enzymes hydrolytiques secrétées dans le milieu par les souches de *Streptomyces* afin de dégrader les nutriments. Le génome de l'espèce modèle *Streptomyces coelicolor* (en réalité *Streptomyces violaceoruber*, mais appelé *coelicolor* pour des raisons historiques) contient par exemple 819 gènes codant des protéines prédites comme étant secrétées, dont au moins une soixantaine de protéases, treize chitinases, huit cellulases, trois amylases et deux pectate lyases (Chater *et al.* 2010). De nombreuses espèces de *Streptomyces* jouent le même rôle dans les sédiments marins. D'autres colonisent des végétaux et autres environnements (Flärdh & Buttner 2009, Kieser *et al.* 2000). La figure I.24 présente l'aspect des colonies de deux espèces saprophytiques du sol.

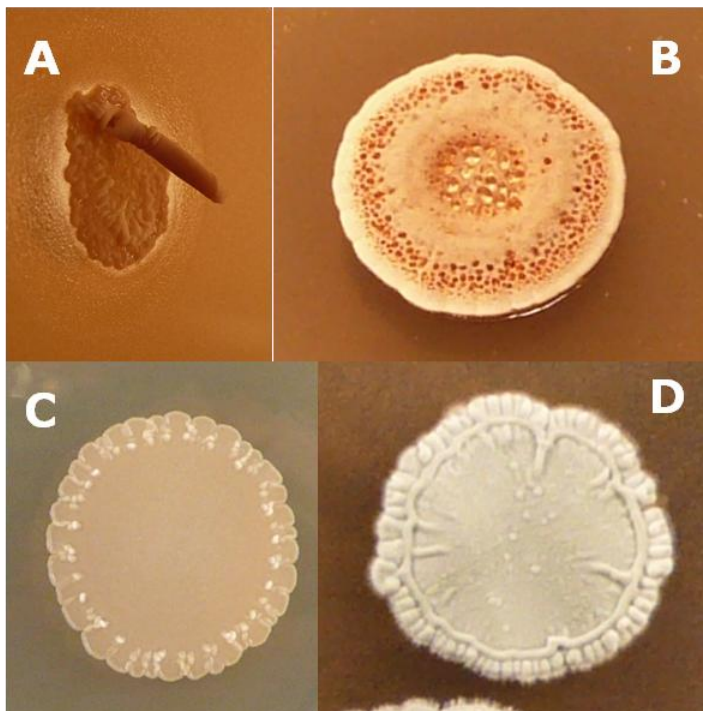


Figure I.24. Formes de colonies de *Streptomyces*.
A et B - *S. aizunensis* (A - non sporulée, B - sporulée).
C et D - *Streptomyces venezuelae* (C - non sporulée, D - sporulée).

Les espèces pathogènes sont assez rares chez *Streptomyces*. *Streptomyces somaliensis* peut par exemple provoquer une actinomycétose sévère chez les patients immunodéprimés (Kieser *et al.* 2000). Quelques autres espèces sont pathogènes de plantes, comme *S. acidiscabies* évoqué précédemment ou encore *S. scabies*, qui sont les agents causaux de la galle commune de la pomme de terre (figure I.25A) (Healy *et al.* 2000).



Figure I.25. Exemples des interactions de *Streptomyces* avec d'autres organismes. A – galle commune de la pomme de terre, maladie provoquée par *S. scabies*. B – Ouvrière de la fourmi champignonnière *Acromyrmex* sp. Le dépôt farineux sur la cuticule correspond aux actinomycètes symbiotiques aidant la fourmi à lutter contre le champignon parasite du genre *Escovopsis*.

D'autres espèces ont développé des relations symbiotiques. Par exemple *Streptomyces* sp. ZC2 fait partie du microbiote intestinal des termites de l'espèce *Zootermopsis angusticollis* où il participe à la dégradation de la cellulose et d'autres résidus d'origine végétale (Schäffer *et al.* 1996). Chez les fourmis champignonnières de la tribu *Attini*, on trouve de nombreux actinomycètes colonisant les replis de la cuticule et aidant leurs hôtes à protéger leurs cultures du champignon du genre *Leucoagaricus* contre le champignon pathogène du genre *Escovopsis* (figure I.25B). Par exemple *Streptomyces* sp. Ao10 isolé de la fourmi *Acromyrmex octospinosus* produit la candidicine D, un polycétide qui présente une activité antifongique sélective contre *Escovopsis* mais pas contre *Leucoagaricus* (Haeder *et al.* 2009).

4.2. Cycle de vie

Le cycle de vie des *Streptomyces* (figure I.26) commence avec une spore d'où émergent un ou deux tubes de germination. Les tubes poussent exclusivement par croissance apicale, ce qui différencie *Streptomyces* et autres actinomycètes des autres bactéries, mais encore une fois constitue une ressemblance frappante avec les champignons filamenteux. Des branchements se produisent parfois dans le mycélium végétatif (appelé aussi mycélium basal) et celui-ci pénètre profondément dans le substrat d'où il

puise ses nutriments. De nouveaux branchements sont alors produits dans les parties âgées du mycélium, loin de parties apicales. C'est dans ces parties âgées du mycélium que la formation des cloisons prend place, résultant en mycélium semi-syncytial (Flärdh & Buttner 2009, Kieser *et al.* 2000).

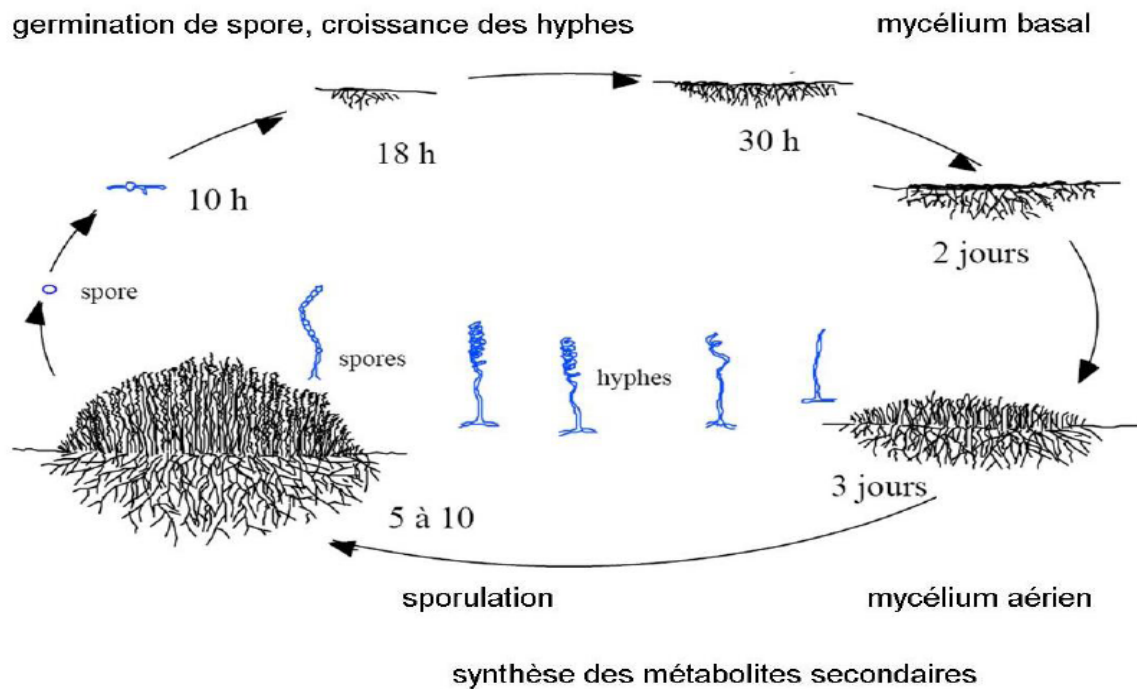


Figure I.26. Cycle de vie de *Streptomyces* sur milieu solide.

Quand les colonies grandissent et que le mycélium le plus âgé devient compact, un changement physiologique prend place. Ce changement est induit par des stress, notamment des carences alimentaires, mais aussi des phénomènes de signalisation cellulaire du type *quorum sensing*. Il est caractérisé par cinq phénomènes : (1) augmentation de la production de certaines protéines extracellulaires (majoritairement des hydrolases) ; (2) lyse de certains compartiments du mycélium basal (qui fournira des nutriments au mycélium aérien) ; (3) accumulation des nutriments dans les parties supérieures du mycélium basal ; (4) formation du mycélium aérien ; et (5) activation du métabolisme secondaire (notamment la production des antibiotiques, antifongiques et autres molécules importantes pour la défense

des nutriments accumulés et partiellement libérés par la lyse partielle du mycélium basal contre les concurrents et prédateurs potentiels). La lyse du mycélium basal évoque un événement de mort cellulaire programmé. C'est à partir des hyphes qui n'ont pas subi la mort cellulaire programmée que se développe le mycélium aérien qui se décolle de la surface humide du sol. Pour ce faire, ce mycélium a besoin de vaincre la tension de surface ainsi que de devenir hydrophobe. Ceci est obtenu par l'incorporation de protéines hydrophobes, les chaplines, à la surface extérieure de la paroi bactérienne, et par la production de surfactants (Flärdh & Buttner 2009, Manteca *et al.* 2007, Kieser *et al.* 2000).

A l'extrémité des hyphes aériennes se développe un long compartiment non cloisonné, appelé la cellule sporogénique. La réplication de l'ADN y est particulièrement active et une cinquantaine (voire plus) de copies du chromosome bactérien y est produite. La croissance et la réplication s'arrêtent et la cellule sporogénique commence à former des cloisons. Finalement des spores sont formées. Chaque spore contient une seule copie du chromosome bactérien et est entourée d'une paroi épaisse et hydrophobe saturée avec un polycétide connu sous le nom de pigment gris de spores. Parfois des molécules ressemblant à la mélanine sont aussi présentes. La spore est un stade de vie dormant, permettant aux *Streptomyces* d'une part d'être dispersés, et d'autre part, de résister aux conditions de vie défavorables, notamment à la dessiccation (Flärdh & Buttner 2009, Kieser *et al.* 2000).

4.3. Génome

Le génome de *Streptomyces* est de grande taille pour une bactérie, entre 6 et 12 Mbp avec la moyenne autour de 8 Mbp. Il a un contenu en GC élevé, d'environ 70% (Kirby 2011). Les chromosomes de *Streptomyces* ont la particularité d'être linéaires avec une origine de réplication centrale. Afin d'assurer la réplication du brin retardé aux extrémités du chromosome, la solution retenue par l'évolution consiste à utiliser deux protéines qui permettent d'initier la synthèse d'ADN. Une d'entre elles, la protéine terminale (TP, codée par le gène *tpg*) est liée à l'extrémité du chromosome de

façon covalente. L'autre, la protéine associée aux télomères (TAP) se fixe aux séquences palindromes de l'extrémité du chromosome ainsi qu'à la TP. Les deux, lors de la réplication de l'ADN, initient la synthèse sur le brin retardé, en synthétisant un petit fragment d'ADN de treize nucléotides complémentaire à l'extrémité du chromosome. L'ADN polymérase prend le relais et continue la synthèse afin d'éviter le raccourcissement des télomères à chaque cycle de réplication (Yang *et al.* 2015).

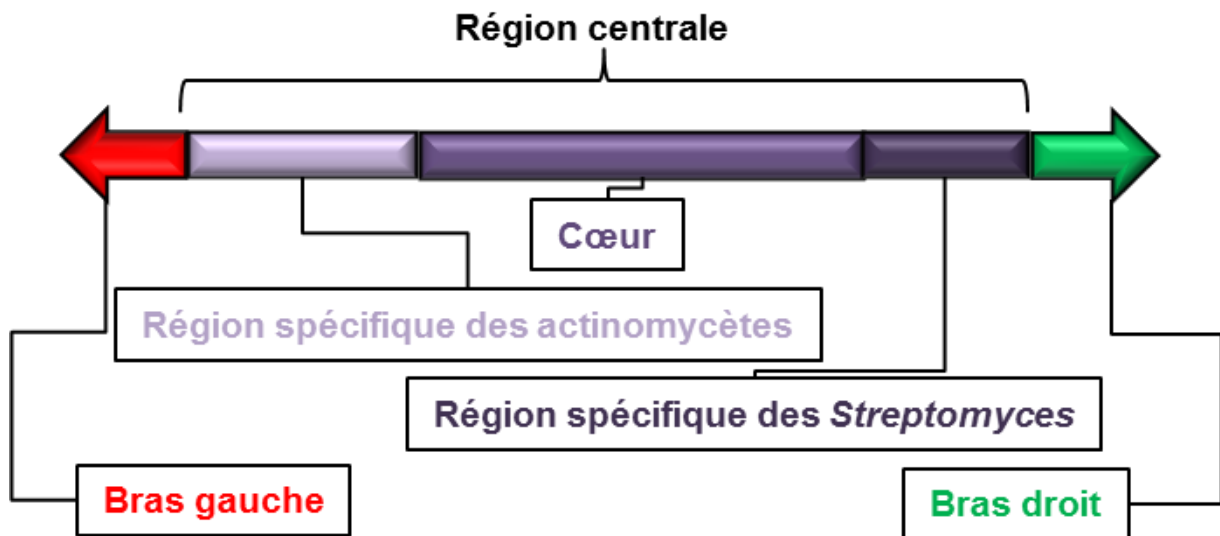


Figure I.27. Structure du chromosome linéaire de *Streptomyces* .

La figure I.27 représente l'organisation du chromosome de *Streptomyces*. Cinq régions différentes peuvent être observées : la région centrale qui contient des gènes conservés chez les *Streptomyces*, et qui peut être subdivisée en région du cœur avec les gènes de métabolisme primaire, bordée d'un côté par une région spécifique d'actinomycètes et contenant des gènes communs à l'ordre *Actinomycetales*, et de l'autre par une spécifique des *Streptomyces* (Kirby 2011). Les parties les plus excentriques du chromosome sont appelées les bras. Les *clusters* de gènes liés au métabolisme secondaire, en général une trentaine par espèce, sont localisés préférentiellement dans les bras. Ces gènes sont très variables non seulement entre espèces, mais même entre souches au sein d'une même

espèce. Dans les bras on observe généralement une absence complète de gènes essentiels. Cette organisation présente des avantages. Notamment, puisqu'aucun gène essentiel ne se trouve au niveau des bras, ce sont donc des sites privilégiés pour d'un côté servir de sites d'intégration pour un matériel génétique acquis par des échanges génétiques horizontaux, de l'autre pour une évolution plus classique par de nombreux réarrangements chromosomiques et autres types de mutations (Choulet *et al* 2006, Fisher *et al* 1998). En effet des réarrangements chromosomiques spontanés s'observent au laboratoire avec la fréquence par cycle de vie complet de 0,1 – 1% selon l'espèce. (Kirby 2011, Kieser *et al.* 2000).

4.4. Métabolisme secondaire

Les sous-chapitres précédents ont déjà donné un aperçu de la richesse du métabolisme secondaire des *Streptomyces* avec de nombreuses molécules ayant des propriétés antibiotiques (comme la terpentécine (figure I.3D), la streptomycine (figure I.4C), la kanamycine (figure I.4D), la chlortétracycline (figure I.5B), le thiostrepton (figure I.6A) ou encore la bicyclomycine (figure I.10A)), fongicides (comme la nystatine (figure I.5D)), des facteurs de virulence (comme la thaxtomine (figure I.11A)) ou encore des sidérophores (comme la foroxymithine (figure I.12A)). Avant le développement des techniques de séquençage massif du génome, beaucoup d'espèces étaient connues pour produire une seule molécule bioactive, ou, dans quelques cas, pour en produire un nombre limité, comme *S. venezuelae* connu pour la production du chloramphénicol, de la jadomycine, de la pikromycine, de la méthymycine et les dérivés de ces deux dernières molécules (narbomycine et néométhymycine, respectivement). Cependant l'analyse des génomes a permis de se rendre compte que chaque espèce possède en moyenne une trentaine de *clusters* de biosynthèse de métabolites secondaires, qui sont en grande majorité silencieux en conditions de laboratoire. Enfin, le fait qu'environ 60% des antibiotiques utilisés en médecine sont des molécules ou des dérivés des molécules découvertes pour la première fois en tant que produits naturels de *Streptomyces*, explique bien l'intérêt que ce genre bactérien suscite (Chater *et al.* 2010, Challis 2008, Kieser *et al.* 2000).

Les gènes de biosynthèse des métabolites secondaires sont en général organisés en *clusters*, c'est-à-dire qu'ils sont regroupés sur une même région du génome. Ces *clusters* se trouvent le plus souvent dans les bras du chromosome. Une grande partie d'entre eux est spécifique d'une espèce, voire d'une souche donnée, ce qui assure une énorme diversité attendant toujours d'être découverte (Gu *et al.* 2013, Kirby 2011, Lazos *et al.* 2010, Kieser *et al.* 2000).

5. Objectifs du travail de thèse

Comme de très nombreuses molécules bioactives à usage médicinal sont des dicétopipérazines et comme nous venons de le voir, les *Streptomyces* constituent une source inépuisable de métabolites secondaires, ils est donc intéressant de se focaliser sur la biosynthèse des dicétopipérazines chez *Streptomyces*.

Les voies de biosynthèse de dicétopipérazines dépendantes des CDPS suscitent un intérêt particulier, notamment parce qu'il s'agit d'une famille d'enzymes récemment découverte et toujours assez peu étudiée surtout par rapport aux NRPS (Belin *et al.* 2012, Lautru *et al.* 2002), mais aussi parce que le séquençage massif des génomes fait littéralement exploser le nombre d'enzymes appartenant à cette famille. Au moment de la découverte de la première CDPS chez *S. noursei*, seuls deux homologues étaient présents dans les bases de données, l'un chez *M. tuberculosis* et l'autre chez *B. subtilis* (Lautru *et al.* 2002). Dix ans plus tard, une cinquantaine de CDPS putatives pouvaient être dénombrées (Belin *et al.* 2012). A l'heure actuelle, une analyse par psi-blast permet d'en détecter plusieurs centaines et si la progression continue, ce nombre dépassera mille séquences non redondantes avant 2018. Avant le début de ce travail, uniquement neuf CDPS étaient caractérisées, toutes utilisant comme substrats des acides aminés hydrophobes et aromatiques. Un des objectifs de ma thèse a donc consisté à explorer la diversité des CDPS et des produits qu'elles sont capables de synthétiser. Ce travail est présenté dans le chapitre II.

Le rôle biologique de certaines dicétopipérazines chez l'organisme producteur étant complètement élusif, il était donc intéressant d'essayer de l'appréhender. Pour cela, au cours de ma thèse, je me suis focalisé sur un *cluster* de deux gènes, codant une CDPS et un cytochrome P450, chez *S. venezuelae*. Ce *cluster* m'a particulièrement intéressé parce qu'il est homologue au *cluster* de biosynthèse de la mycocyclosine de *Mycobacterium tuberculosis*, chez laquelle il a été suggéré que le cytochrome était essentiel (McLean *et al.* 2008). Les résultats de ce travail sont présentés dans le chapitre III.

La voie de biosynthèse de certaines dicétopipérazines complexes reste inconnue, mais certains éléments suggèrent qu'elles sont synthétisées par des voies dépendantes de CDPS. Il est donc intéressant d'élucider complètement les chemins métaboliques qui conduisent à la synthèse de ces molécules. Au cours de ma thèse, je me suis focalisé sur l'étude de la voie de biosynthèse de la bicyclomycine qui est intéressante pour plusieurs raisons. D'abord il s'agit d'un antibiotique utilisé en médecine qui possède un mode d'action original car c'est le seul inhibiteur spécifique connu du facteur de terminaison Rho (Zwiefka *et al.* 1993). De plus la bicyclomycine suscite un nouvel intérêt suite aux travaux qui ont démontré une forte synergie avec d'autres antibiotiques (Malik *et al.* 2014). Enfin la bicyclomycine est intéressante du point de vue chimique, vu le nombre de modifications du noyau DKP et des chaînes latérales. L'élucidation de la voie de biosynthèse de la bicyclomycine est présentée dans le chapitre IV.

Chapitre II

Etude de la diversité des cyclodipeptide synthases (CDPS)

Introduction

Comme mentionné dans l'introduction générale de la thèse, avant le début de ce travail, seules neuf CDPS étaient caractérisées expérimentalement. Elles avaient toutes été exprimées chez *Escherichia coli* en tant que protéines recombinantes et les cyclodipeptides présents dans les surnageants de culture avaient été caractérisés. L'état des connaissances concernant l'activité de ces neuf CDPS caractérisées avant le début de cette étude est présenté dans le tableau II.1 (Belin *et al.* 2012).

CDPS	Organisme producteur	Cyclodipeptides produits ^a
AlbC	<i>Streptomyces noursei</i>	cFL (54), cFY (20), cFF (19), cFM (10), cYM (6), cYL (5), cLM (4), cLL (2), cMM (1), cYA (<1), cVL (<1) et cYY (<1)
Rv2275	<i>Mycobacterium tuberculosis</i>	cYY (29), cYF (2), cYL (<0.4), cYA (<0.3), cYM (0.1) et cYW (0.1)
YvmC	<i>Bacillus licheniformis</i>	cLL (71), cLM (6), cLF (4) et cXX (1)
YvmC	<i>Bacillus subtilis</i>	cLL (97), cLF (7), cLM (5) et cXX (0.9)
YvmC	<i>Bacillus thuringiensis</i>	cLL (60), cLF (8), cLM (3), cFF (1) et cXX (0.2)
pSHaeC06	<i>Staphylococcus haemolyticus</i>	cLL (118), cLM (10), cLF (8) et cXX (0.5)
Plu0297	<i>Photobacterium luminescens</i>	cLL (17), cLF (<1) et cLM (<1)
JK0923	<i>Corynebacterium jeikeium</i>	cLL (0.5) et cLM (0.2)
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)

Tableau II.1. Activité des CDPS caractérisées expérimentalement avant le début de cette étude. Les CDPS ont été étudiées en expression hétérologue chez *E. coli*. ^a-les cyclodipeptides sont indiqués en tant que cXX où X correspond à un acide aminé selon le code à une lettre. Entre parenthèses sont indiquées les quantités produites en mg/l. Adapté de Belin *et al.* 2012.

CDPS	Organisme producteur	Cyclodipeptides produits	Référence
Amir_4627	<i>Actinosynnema mirum</i>	cWW	Giessen <i>et al.</i> (2013b) <i>Biochemistry</i> 52 : 4274-4283
Ndas_1148	<i>Nocardiopsis dassonvillei</i>	cFY, cFF, cFL, cYL, cYY, cFA, cFM	Giessen <i>et al.</i> (2013a) <i>Chemistry & Biology</i> 20 : 828-838
B005_4463	<i>Nocardiopsis alba</i>	cFL, cYL	Li <i>et al.</i> (2014) <i>Archives of Microbiology</i> . 196 :765
KALB_4567	<i>Kutzneria albida</i>	cFL, cFM, cFY, cFF	Rebets <i>et al.</i> (2014) <i>BMC Genomics</i> 15 :885
NozA	<i>Nocardiopsis sp.</i> CMB-M0232	cWW	Alqahtani <i>et al.</i> (2015) <i>Organic & Biomolecular Chemistry</i> 13 :7177-7192

Tableau II.2. Activité des CDPS caractérisées expérimentalement par les équipes concurrentes pendant la durée de cette étude.

On voit clairement que les CDPS caractérisées incorporaient principalement des acides aminés hydrophobes et apolaires et que la diversité des cyclodipeptides produits était relativement faible. Un certain nombre de CDPS a été caractérisé par les équipes concurrentes pendant la durée de notre étude, elles sont reprises dans le tableau II.2.

En ne prenant en compte que les séquences des CDPS putatives disponibles dans les bases de données au début de 2012, il est clair que toutes les CDPS caractérisées étaient groupées dans une même branche de l'arbre phylogénétique déduit de ces séquences (Figure II.1). Il était donc essentiel de poursuivre l'étude d'autres CDPS putatives, d'autant plus que certaines d'entre elles présentaient des pourcentages d'identité de séquence protéique très faibles (inférieurs à 10%) avec les CDPS déjà caractérisées (Belin *et al.* 2012). Il était donc justifié de se demander si ces protéines possédaient effectivement une activité de cyclodipeptide synthase.

Pour ces différentes raisons, une exploration plus poussée de la diversité des CDPS a été entreprise dans le cadre du projet ANR CyDiPepS coordonné par Muriel Gondry. J'ai été associé à ces travaux dont les résultats sont présentés dans les pages qui suivent sous forme de la version acceptée d'un manuscrit qui a été publié dans *Nature Chemical Biology* 11:721-727 (2015). Pour faciliter la lecture, les figures et les figures supplémentaires ont été incorporées dans le texte et leur numérotation a été modifiée pour suivre l'ordre logique de la thèse. Les figures et les tableaux supplémentaires trop grands pour être incorporés directement (pour question de lisibilité) sont accessibles à l'adresse <https://www.nature.com/nchembio/journal/v11/n9/full/nchembio.1868.html#supplementary-information>.

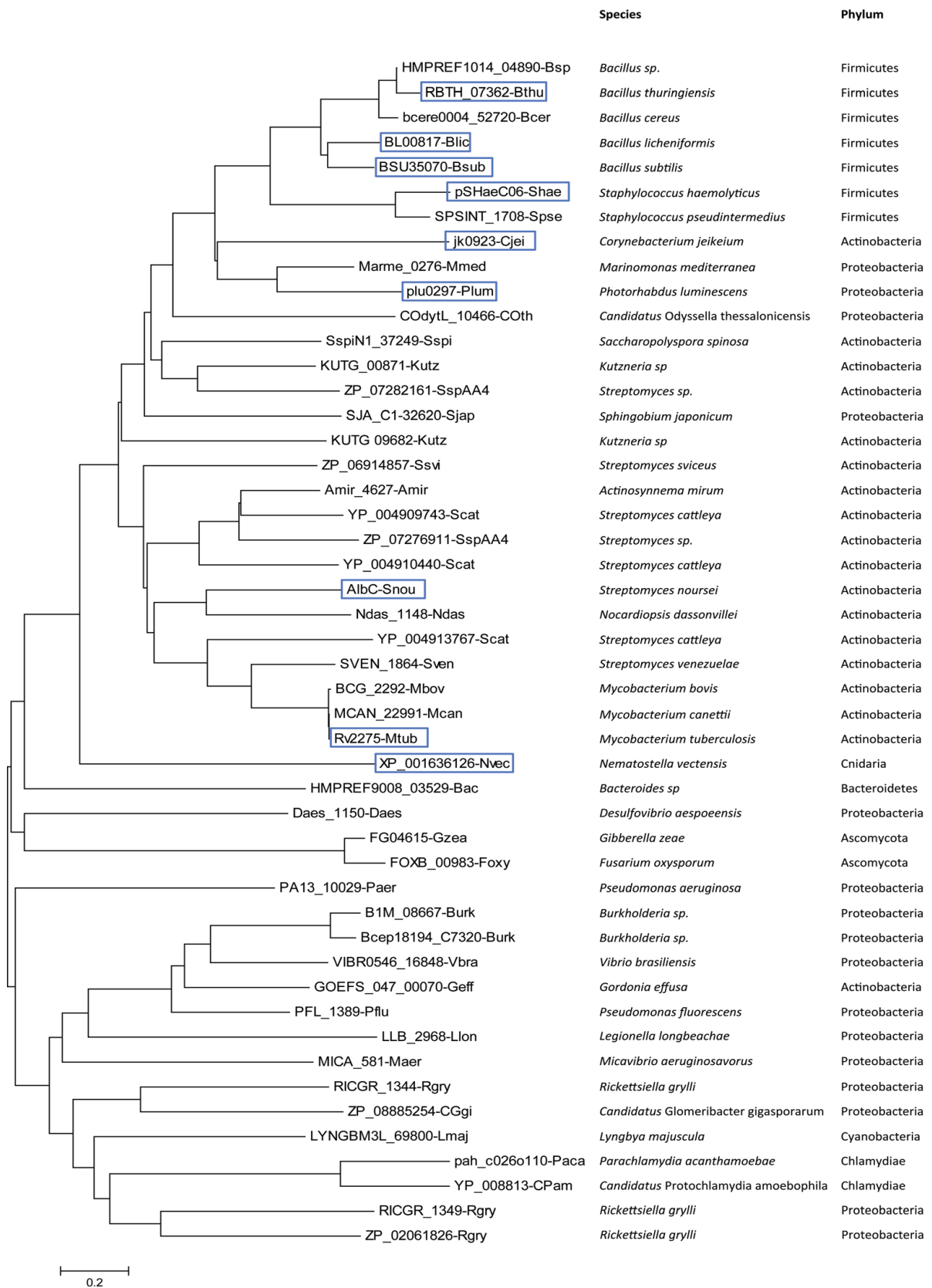


Figure II.1. Arbre phylogénétique des CDPS putatives et caractérisées expérimentalement avant le début de cette étude. Les CDPS caractérisées sont entourées en bleu. Adapté de Belin *et al.* 2012.

Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity

Isabelle B. JACQUES^{1,4}, Mireille MOUTIEZ^{1,3,4}, Jerzy WITWINOWSKI^{2,3}, Emmanuelle DARBON², Cécile MARTEL², Jérôme SEGUIN^{1,4}, Emmanuel FAVRY^{1,4}, Robert THAI¹, Alain LECOQ¹, Steven DUBOIS¹, Jean-Luc PERNODET², Muriel GONDRY^{1,4*}, and Pascal BELIN^{1,4*}

¹Commissariat à l'énergie atomique et aux énergies alternatives (CEA), Institut de Biologie et de Technologies de Saclay (iBiTec-S), Service d'Ingénierie Moléculaire des Protéines, Gif-sur-Yvette, France.

²Université Paris-Sud, Centre National de la Recherche Scientifique (CNRS), Unité Mixte de Recherche (UMR) 8621, Institut de Génétique et Microbiologie, Orsay, France

³These authors contributed equally to this work

⁴Present address: CEA, iBiTec-S, Service de Biologie Intégrative et Génétique Moléculaire, I2BC, UMR9198, F-91191 Gif-sur-Yvette, France

*To whom correspondence should be addressed. Tel: +33 169082576; Fax: +33 169089071; Email: pascal.belin@cea.fr; Correspondence may also be addressed to Muriel Gondry. Tel: +33 169087647; Fax: +33 169089071; Email: muriel.gondry@cea.fr

Abstract

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 XYP. All 11 previously characterized CDPSs belong to the NYH subfamily, suggesting that further special features may yet to be discovered in the other subfamily. CDPSs synthesize a large diversity of cyclodipeptides made 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, are a first step towards being able to predict the cyclodipeptides synthesized by newly discovered CDPSs. The determination of the activity of 10 further CDPSs with predicted function constitutes a first experimental validation of this predictive approach.

Introduction

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 (Gondry *et al.* 2009). 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) (Bonfond *et al.* 2011, Sauguet *et al.* 2011, Vetting *et al.* 2010). 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 (Bonfond *et al.* 2011, Sauguet *et al.* 2011, Seguin *et al.* 2011, Vetting *et al.* 2010). 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 (Moutiez *et al.* 2014b). Two recent studies on the CDPS AlbC give clues on how CDPSs and their two substrates interact (Moutiez *et al.* 2014a,b). 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, 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 N¹-N⁷² base pair of the acceptor stem. Only ten bacterial CDPSs and one eukaryotic CDPS have been biochemically characterized (Giessen *et al.* 2013a,b, Seguin *et al.* 2011, Gondry *et al.* 2009). 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 (Giessen et Marahiel 2014, Belin *et al.* 2012, Aravind *et al.* 2010) 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 show the reliability of prediction for few selected groups by characterizing 10 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) protein database. We retrieved about 80 new sequences corresponding to putative CDPSs (May 2013) and used these to construct a phylogenetic tree (Fig. II.2). We selected 49 sequences representative of the whole sequence set (designated CDPS 1–49; Supplementary Data Set 1, Fig. II.2) 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 the recovery of cyclodipeptides in culture supernatants upon CDPS expression in this host (Gondry *et al.* 2009). 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 (Fig. II.3). We found that 41 of the 49 recombinant CDPS candidates tested had cyclodipeptide-synthesizing activity (Fig. II.2; Supplementary Data Set 2, Table II.3).

We found eight of the putative CDPSs inactive although they were effectively produced (Figure II.4). Sequences analysis revealed several possible explanations for their 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 (Sauguet *et al.* 2011). 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.

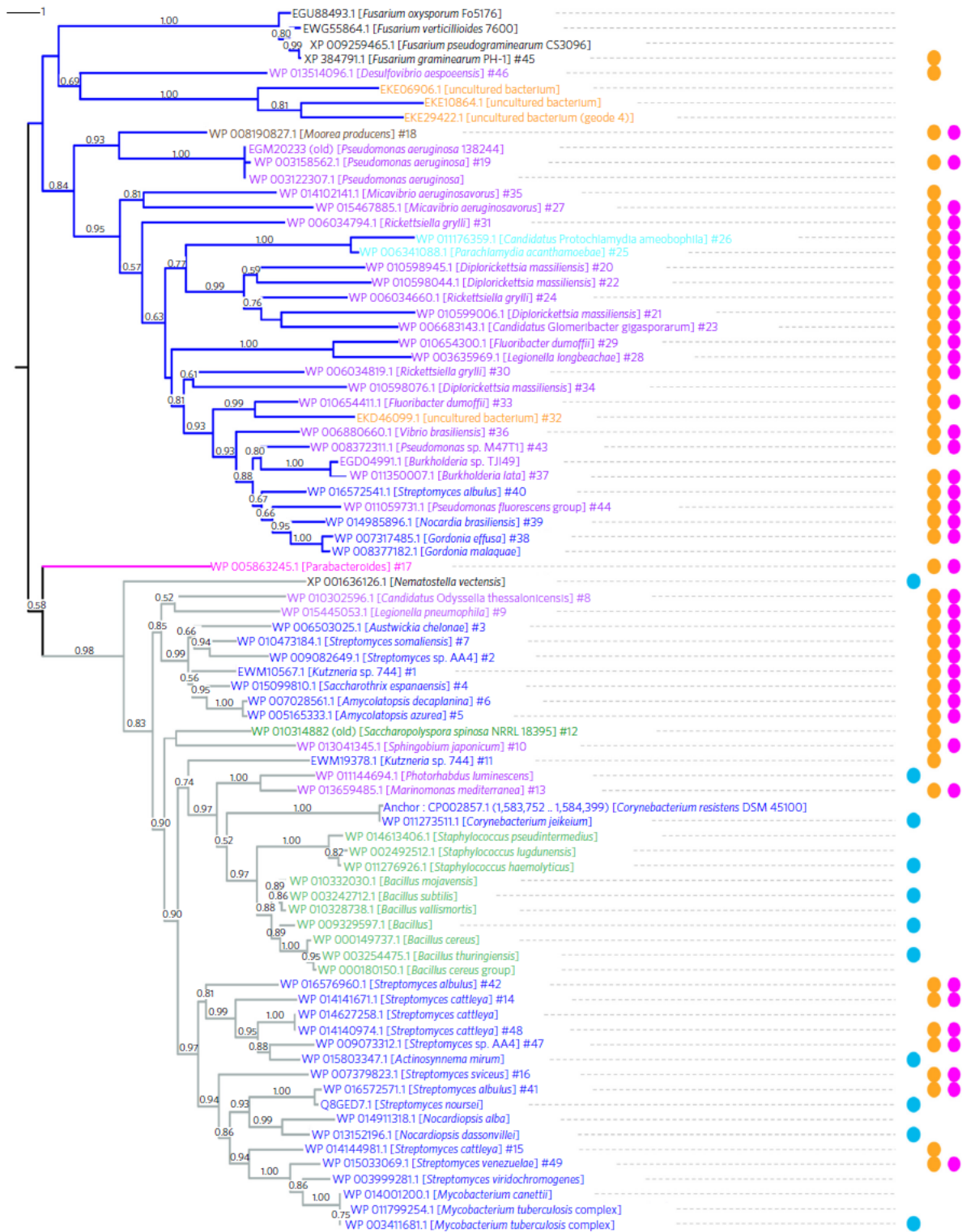


Figure II.2 Phylogenetic tree of known CDPs and putative CDPs retrieved from bioinformatics searches in databases (May 2013). The set of sequences was curated: seven partial sequences were removed and the start of eight sequences manually corrected (see **Figs. II.11** and **II.12**). CDPs are named by their protein accession numbers in NCBI and their host organisms; the numbers given to the CDPs 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*])). CDPs 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 CDPs, the 49 putative CDPs studied herein, and the 41 new active CDPs are labeled with cyan, orange, and magenta dots, respectively.

CDPS	Species	CDPS subfamily	In vivo activity ^a
1	<i>Kutzneria</i> sp. 744	NYH	cCC (100%)
2	<i>Streptomyces</i> sp. AA4	NYH	cCC (100%)
3	<i>Austwickia cheloneae</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 vicius</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%)
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 aespoensis</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%)

Table II.3. The Major cyclodipeptides produced by the newly identified CDPS.

^a-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.

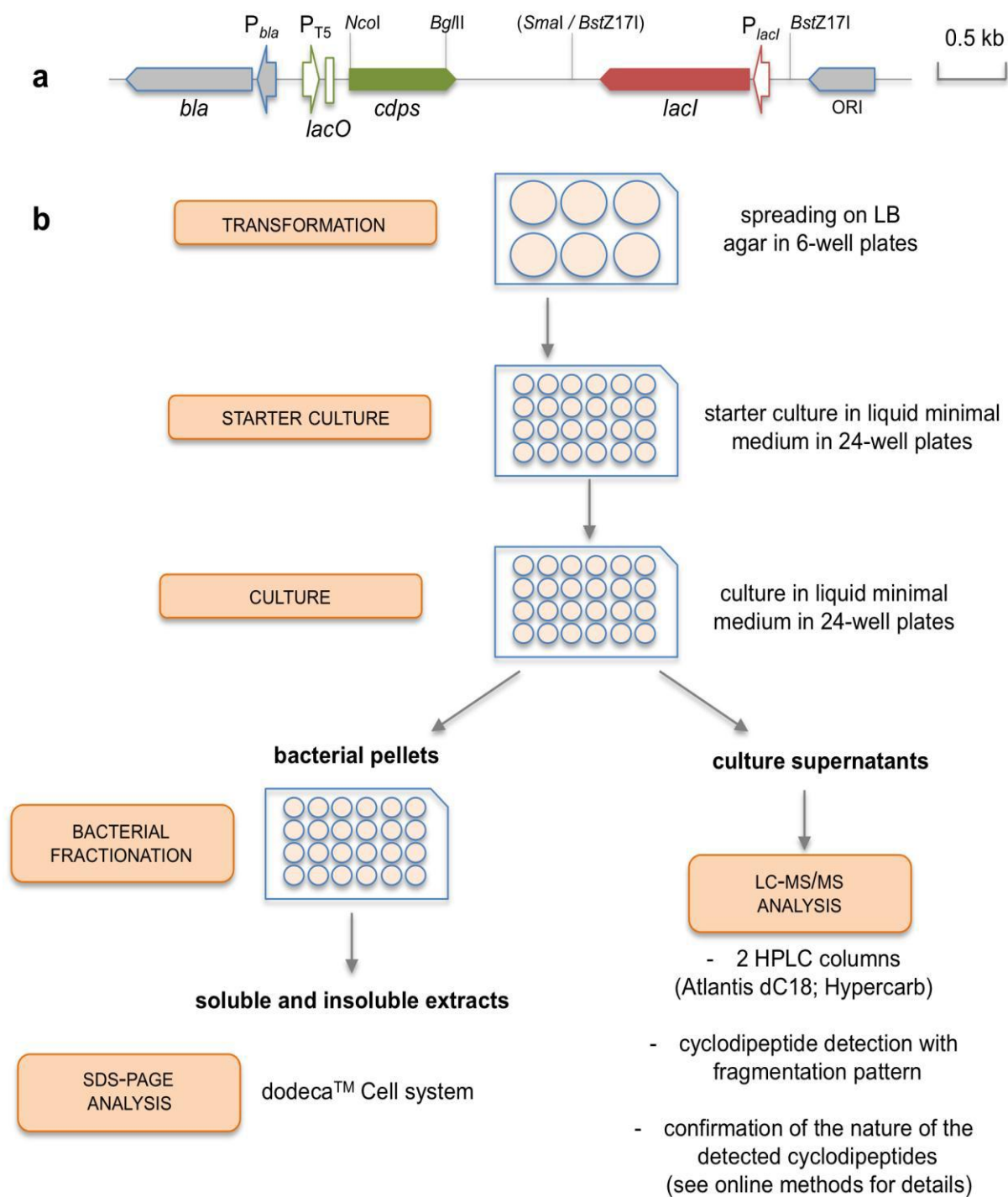


Figure II.3. Screening of CDPS activities in *Escherichia coli*. (a) Schematic representation of a pIJ196-derived plasmid encoding a CDPS. The CDPS gene (colored in green) is under the control of the P_{T5} promoter and two *lacO* sites (only one is represented). *lacI* encodes the repressor protein LacI and *bla* encodes resistance to ampicillin. ORI is the origin of replication. The cloning sites used for inserting the CDPS gene (*NcoI* and *BglII*) and the *lacI*-containing cassette (*SmaI* and *BstZ171*) are indicated, those in brackets being lost after cloning. (b) Work flow chart for screening for in vivo CDPS activity and production. The various steps following the transformation of *E. coli* bacteria with the pIJ196-derived plasmids are indicated and detailed in the Methods section.

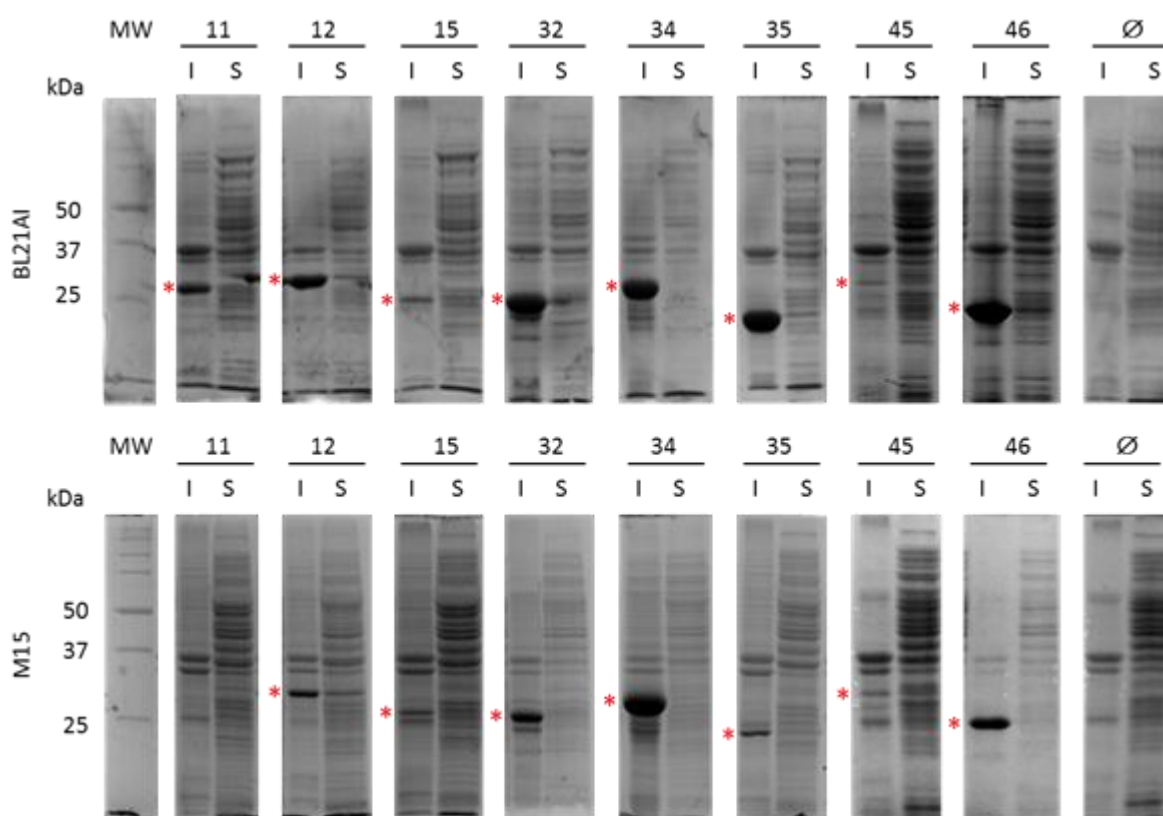


Figure II.4. SDS-PAGE analysis of the expression of putative CDPSs with no detected cyclodipeptide-synthesizing activities. BL21AI, BL21AI-pREP4 (for 45-46) or M15-pREP4 bacteria bearing pIJ196-derived plasmids encoding a putative CDPS or not (pIJ196, marked as \emptyset) were grown in minimal medium and induced for CDPS production. After overnight growth, bacteria were collected and fractionated into soluble (S) and insoluble (I) fractions. These fractions were analyzed by 12% SDS-PAGE and staining with Coomassie blue. Overproduced proteins are indicated by red stars. Only the lanes, extracted from different gels, for 11, 12, 15, 32, 34, 45 and 46 for which no CDPS activity was detected are represented.

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 reexamination of the features of this family. Most known CDPSs are found in three bacterial phyla, Actinobacteria, Firmicutes and Proteobacteria, and a few CDPSs are found in Bacteroidetes (1), Chlamydiae (2) and Cyanobacteria (1). Only one active CDPS has been found in an eukaryotic phylum (Cnidaria) (Seguin *et al.* 2011) and no CDPS has been found in Archaea (Fig. II.2). CDPSs are typically 200 to 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, E182, H203), indeed even strictly conserved (S37, Y202) (AlbC numbering, used throughout the paper) (Moutiez *et al.* 2014b, Sauguet *et al.* 2011, Gondry *et al.* 2009), allowing a functional sequence signature for the CDPS enzymes to be defined (Fig. II.5). We previously showed that the strictly conserved catalytic residue Y202 participates in a hydrogen bond network, also involving residues N40 and H203 that are essential for the accurate positioning of the two loops bearing the catalytic residues in AlbC (Moutiez *et al.* 2014b). We observed that about half of the CDPSs do indeed have the “N40, H203” pair whereas others have a “X40, P203” pair (X being a non-conserved 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. II.2). 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 functional subsequence signature (Fig. II.5b)—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.

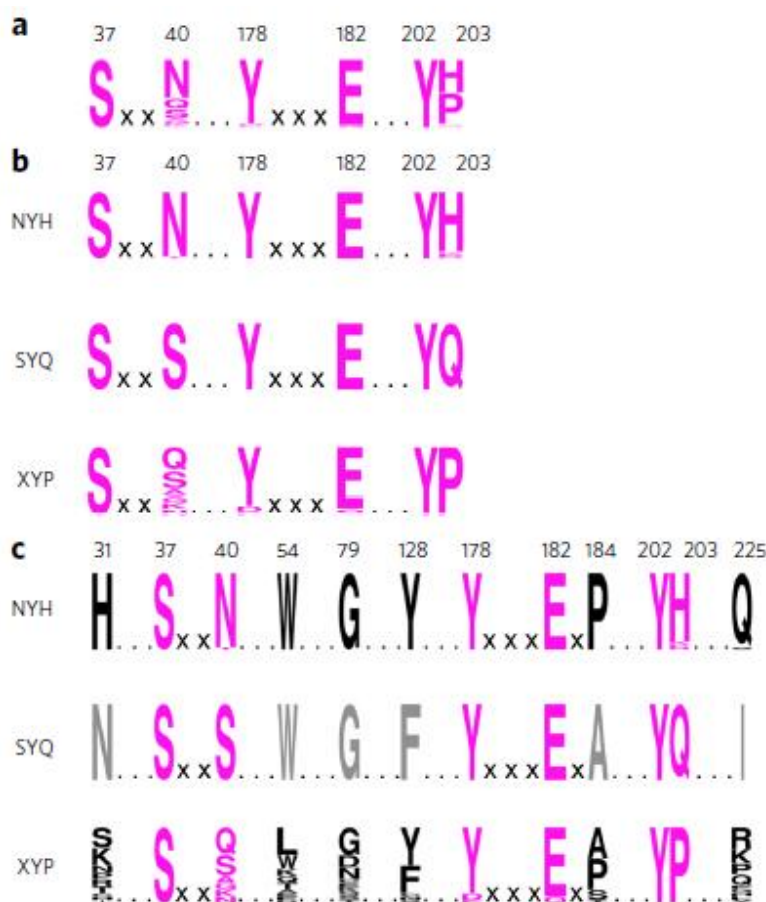


Figure II.5 CDPS subfamilies. (a) Functional sequence signature for CDPSs shown with sequence logos corresponding to the frequency plot of the catalytic amino acids at positions 37, 40, 178, 182, 202 and 203 (AlbC numbering). (b) Functional subsequence signatures for the CDPS subfamilies. Subsequence 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.

The NYH subfamily has six conserved residues, in addition to the six catalytic residues, (Supplementary Data Set 5; Fig. II.6), including three residues identified in AlbC as being essential for enzyme activity (H31 and Y128) or structural integrity (W54) (Sauguet *et al.* 2011). The XYP subfamily

has a more degenerate subsequence, with only five catalytic residues conserved and no other conserved residues. The single SYQ sequence has half of the non-catalytic conserved residues of the NYH subfamily (Fig. II.5c). The basic patch involved in the binding of the tRNA moiety of the first substrate (Moutiez *et al.* 2014a) is found in all members of the NYH subfamily and contains five to 11 basic residues. This patch is much smaller in the XYP class with only two to six basic residues. The XYP subfamily is generally far more diverse than the NYH subfamily; variants of even the highly conserved catalytic sequence YxxxE are found: CDPSs 25 and 26 have the sequence DxxxQ and not YxxxE (Supplementary Data Set 5). The role of the E residue that acts as a catalytic base, essential for generating the dipeptidyl-AlbC intermediate, and that of the Y residue for the positioning of the aminoacyl moiety of the first substrate in AlbC (Moutiez *et al.* 2014b, Sauguet *et al.* 2011) 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 presenting sequence insertions up to 30-residues long between predicted secondary structures, and deletions of a few residues, mostly in predicted α 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 (Fig. II.7).

An enlarged diversity of cyclodipeptides

The 19 cyclodipeptides synthesized by the 11 NYH enzymes, known before this study, are mostly hydrophobic, with aromatic and aliphatic lateral chains (Seguin *et al.* 2011, Giessen *et al.* 2013a,b, Gondry *et al.* 2009). Here, we identify 35 cyclodipeptides that have not previously been described as CDPS products. Many of these cyclodipeptides 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. II.8a). Consequently, CDPSs now incorporate 17 of 20 proteinogenic amino acids into cyclodipeptides.

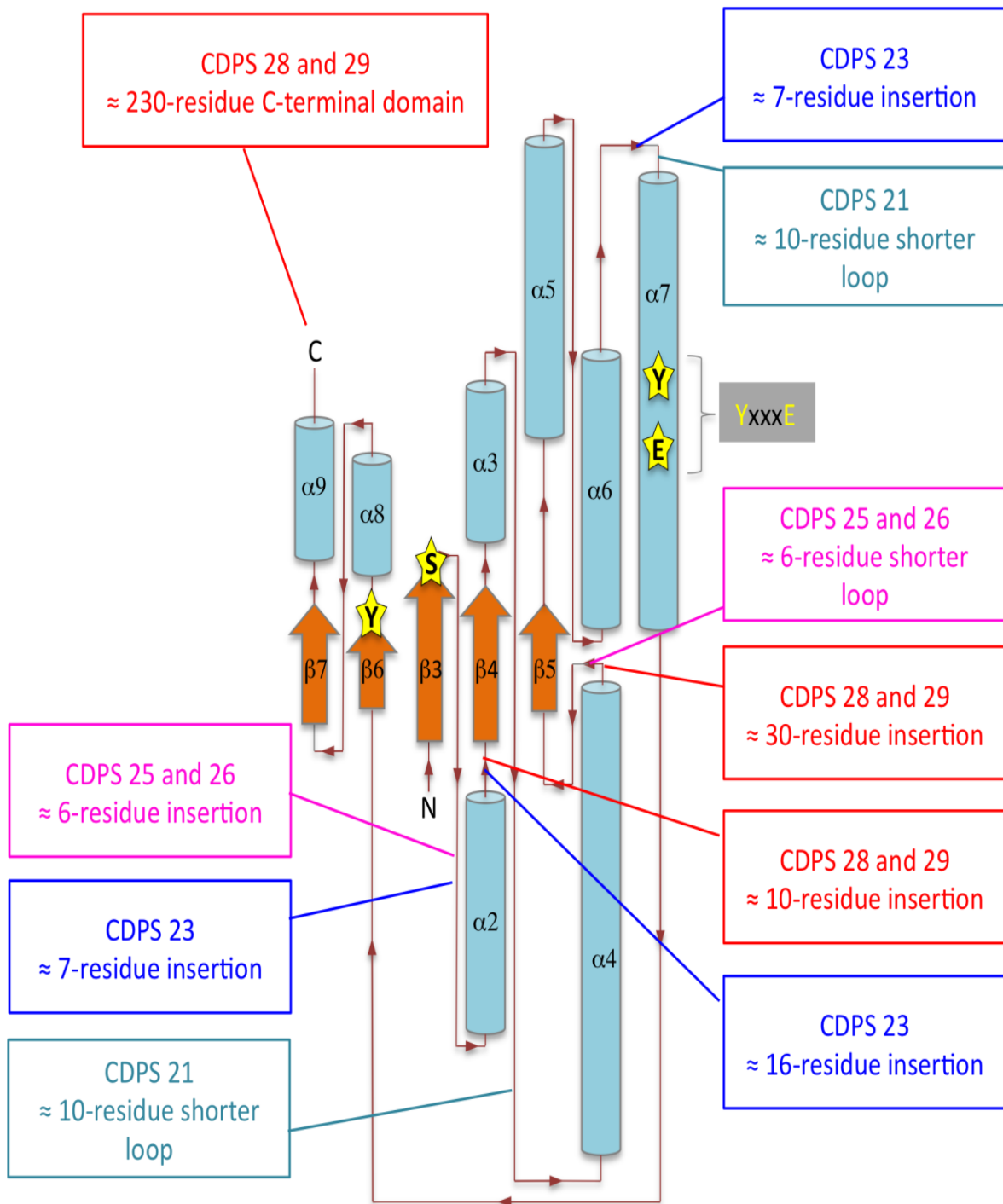


Figure II.7. Two-dimensional topology outline of insertions and deletions in the CDPS fold. The N- and C-terminal extremities are indicated by N and C, respectively. Secondary structure elements are annotated as previously described (Gondry *et al.* 2009). Catalytic residues are indicated as yellow stars, and the YxxxE motif in helix $\alpha 7$ is highlighted. The most significant insertions and deletions observed in the CDPS fold of the newly characterized XYP CDPSs are indicated in rectangles, according to the multiple sequence alignment manually adjusted with HHPRED results (Supplementary Data Set 4). A color code was used for each CDPS: turquoise, CDPS 21; dark blue, CDPS 23; fuchsia, CDPSs 25 and 26; red, CDPSs 28 and 29.

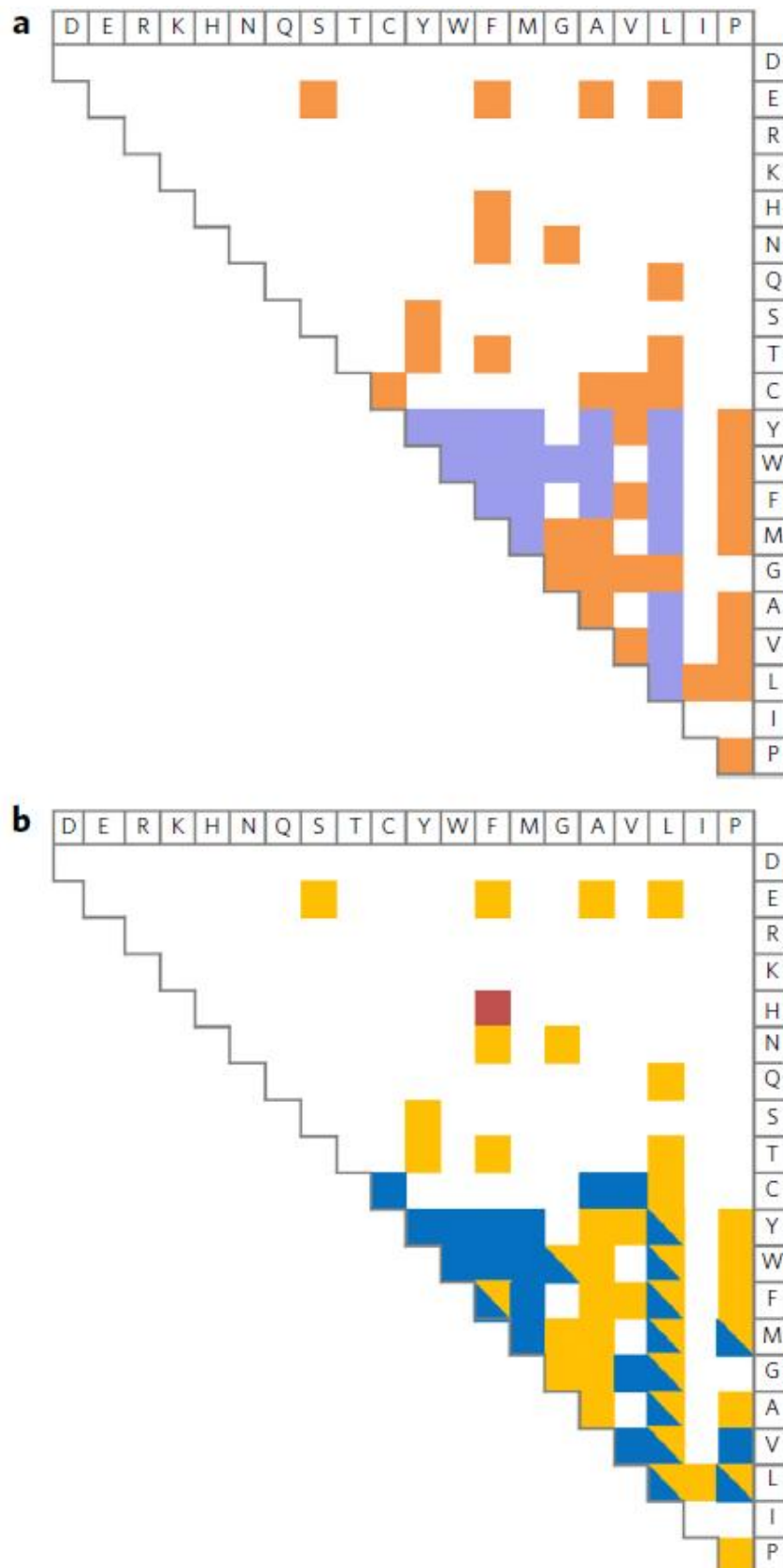


Figure II.8 Diversity of cyclodipeptides produced by the CDPS subfamilies. (a) Cyclodipeptides produced by the 11 previously characterized CDPSs (in violet) and the newly identified CDPSs (in orange). The two amino acids constituting the cyclodipeptides are indicated in the one-letter code at the top 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 dark red) members. The presence and nature of cyclodipeptides are indicated as described in a.

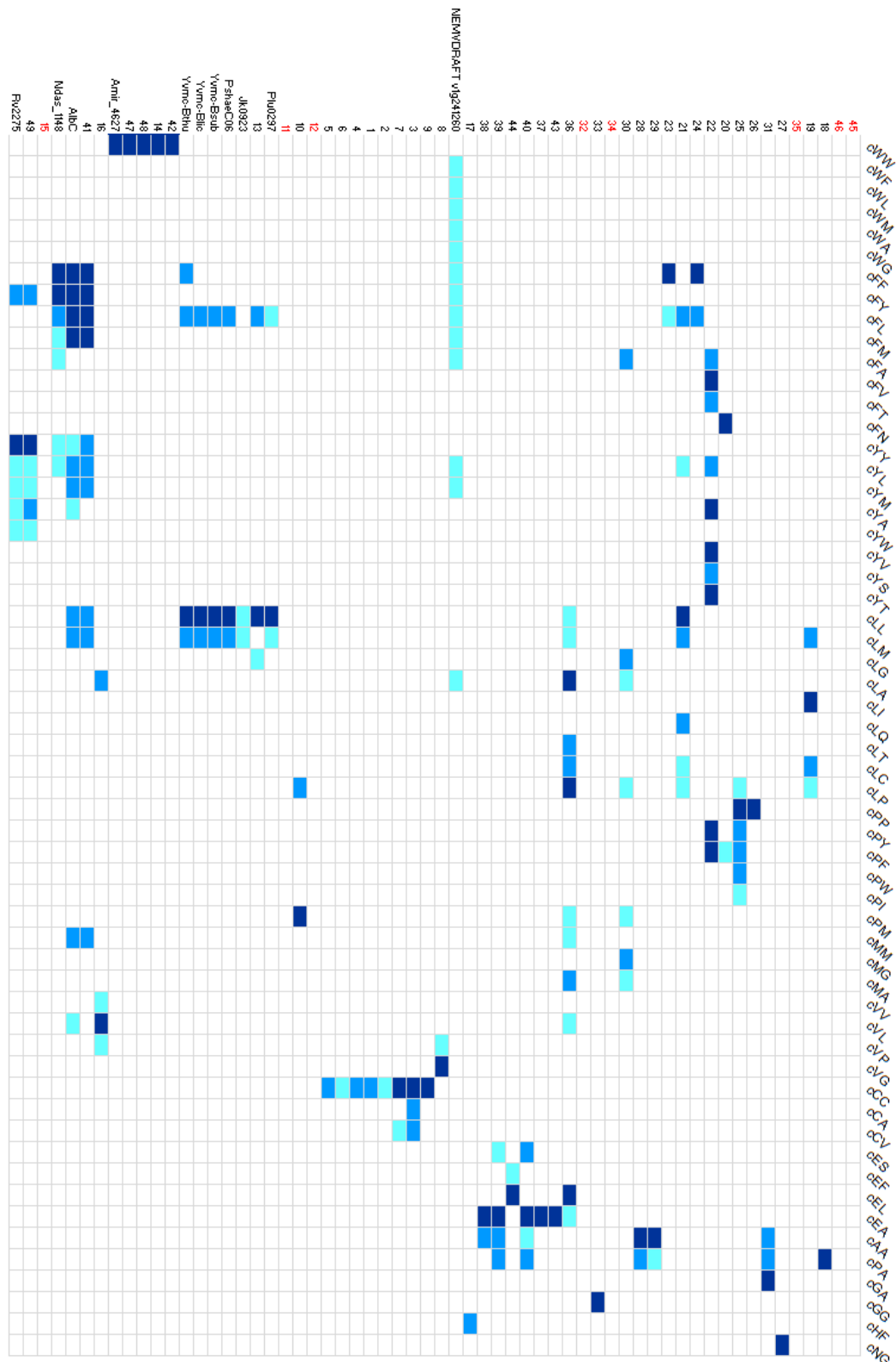


Figure II.9. Activity screening of the CDPS family. The presence of cyclodipeptides (upper line) in culture supernatants of recombinant bacteria each expressing one CDPS (left line) is reported according to Supplementary Data Set 2. CDPSs were expressed in one, two or three bacterial strains and results from one expression experiment are reported. Blue rectangles indicate detection of the corresponding cyclodipeptide according to peak area (A) on chromatograms recorded at 214 nm for the CDPSs studied herein (dark blue, > 1000; medium blue, 100<A≤1000; light blue, 1≤A≤100) or according to the quantity (q) of cyclodipeptide synthesized for already published CDPSs2-5 (dark blue, q>10 mg.L-1 of culture; medium blue, 1<q≤10; light blue, 0,01≤A≤1). CDPSs are ranked according to the tree presented in **Figure II.2**. CDPSs are indicated by their number or the name used in previous publications. Putative CDPSs found in this study to be inactive are labeled in red.

We looked at how cyclodipeptide diversity is distributed among CDPS subfamilies. NYH members incorporate ten different amino acids into cyclodipeptides, 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. II.8b). 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 (Fig. II.6).

The previously characterized NYH members show a degree of substrate promiscuity, Amir_4628 being the only one synthesizing specifically one cyclodipeptide: cWW (Giessen *et al.* 2013a). About a third of the newly characterized CDPSs show high substrate specificity (Table II.3; Fig. II.9). 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 and reflect interaction with either a particular tRNA sequence—the N¹-N⁷² 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 6)—or an aminoacyl moiety with specific physical properties or chemical properties (Moutiez *et al.* 2014a). The other newly characterized CDPSs demonstrated promiscuity toward different amino acids in our experimental conditions (Table II.3; Supplementary Data Set 2; Fig. II.9), 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

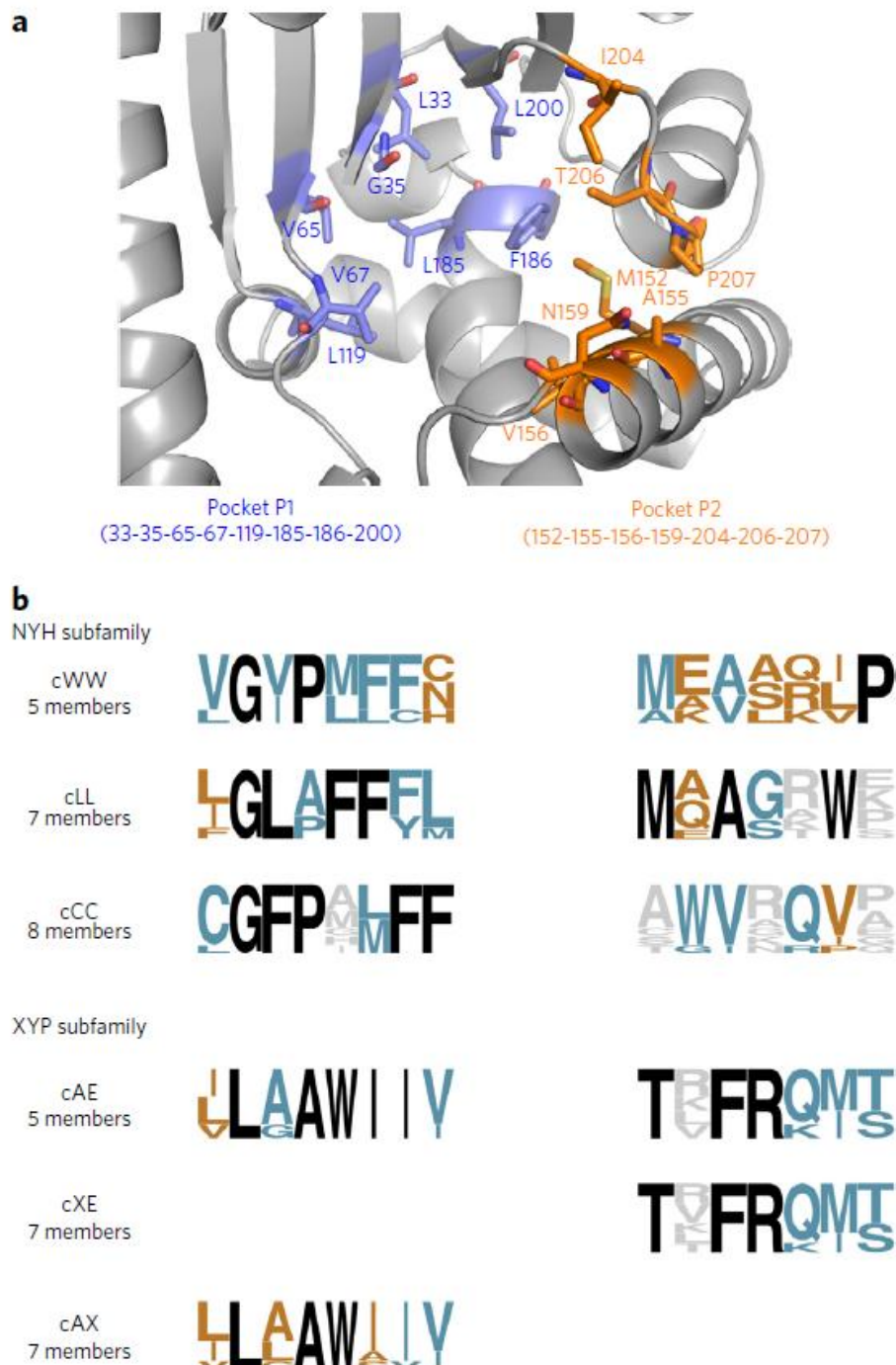


Figure II.10 Sequence logos of the amino acids constituting the two binding pockets, P1 and P2, for groups of CDPs synthesizing the same cyclodipeptides. (a) Three-dimensional view of the binding pockets P1 and P2 accommodating the aminoacyl moieties of the first and second aa-tRNA in AlbC6. The eight residues constituting P1 are in violet and the seven constituting P2 are in orange. (b) Residues suspected to delineate P1 and P2 were determined for each CDPS (Supplementary Data Set 7). Logo sequences corresponding to the frequency plot of amino acids at positions in P1 and P2 were generated for groups of CDPs 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, YvmC_BSU, YvmC_BLIC and YvmC_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, teal blue, tan or gray, respectively.

synthetases (Challis *et al.* 2000, Stachelhaus *et al.* 1999). 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 (Moutiez *et al.* 2014b) (Fig. II.10). 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 7). We predicted for all active CDPSs secondary structures similar to those observed in the crystal structures of AlbC, YvmC 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 whereas there are significant differences between subfamilies, indicating different combinations of amino acids to synthesize the same product, in agreement with separate evolution. Thus, CDPSs of the same subfamily and having the same activity appear to constitute a specific group (Supplementary Data Set 7).

For groups that were sufficiently populated, *i.e.* 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. II.10b). As P1 and P2 pockets are independent (Moutiez *et al.* 2014b), we considered combining the groups accommodating the same residue in one of the pockets, regardless the residue accommodated in the other pocket. However, the 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 promiscuous like AlbC and produce several cyclodipeptides of general formulae c(AA₁-X), AA₁ being the preferred amino acyl and X being

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, 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 (7 members) to create a sequence logo for P1. Similarly, we combined cAE- and cLE-synthesizing groups to give the cXE-synthesizing group (7 members) and create a sequence logo of P2 (Fig. II.10b). 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 for 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 7), 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.

CDPS ^a	Species	Pocket residues ^b		Prediction	Sequence identity ^c	Cyclodipeptides synthesized ^d
		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	CGFPSLFF	AWVRQVWF	cCC	32–57%	cCC (100%)
53	<i>Staphylococcus pseudintermedius</i>	IGLAFFEM	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	VGVPMFFC	MDVAQLP	cWW	39–60%	cWW (100%)
58	<i>Nocardioopsis halophila</i>	VGVPMFFN	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%)

Table II.4. Prediction and validation of the cyclodipeptide-synthesizing activities of CDPS

^a-CDPSs are grouped according to the prediction of activity: 50–52, cCC; 53–55, cLL; 56–58, cWW; 59–61, cAE. ^bThe predicted residues of the P1 and P2 pockets are indicated for each CDPS. cClustal Omega at the EMBL website was used to determine sequence identities between each CDPS and members of the predicted group (as defined in **Fig. II.10b**). ^d-, no cyclodipeptide detected.

Sequence motifs allow discrimination between CDPSs clustered on the tree but having 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. II.2; Table II.3). However, its P1 and P2 residues differ substantially from those in the sequence motif of the cCC-synthesizing CDPSs (Supplementary Data Set 7).

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 8). All new putative CDPSs are distributed into the three subfamilies and fulfill the functional subsequence signatures described above. The residues constituting P1 and P2 of all the new putative CDPSs were compared to the sequence motifs described above (Fig. II.10b). 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 active and the main cyclodipeptides produced in all cases were as predicted (Table II.4; 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 II.4). Among the 200 newly identified putative CDPSs, about 30% have P1- and P2-constituting residues very similar to those of poorly populated groups (less than five members). These putative CDPSs are listed with the appropriate groups in Supplementary Data Set 9. 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 biochemically characterized—can be classified into two main phylogenetically distinct subfamilies. The NYH and XYP 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 8), 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 towards 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 non-canonical pair X40-X203. Availability of new CDPS sequences is likely to reveal if the atypical CDPSs form relevant subgroups in the different subfamilies and to what extent their study may reveal novel features of CDPSs.

Current understanding of CDPSs has mostly come from the crystallographic structures of three NYH members determined so far (Moutiez *et al.* 2014b, Bonnefond *et al.* 2011, Seguin *et al.* 2011, Vetting *et al.* 2010), and related biochemical studies (Moutiez *et al.* 2014a). The identification of XYP 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 compared to NYH members, and affect both the positioning of catalytic residues and composition of amino acids delineating the binding pocket P2 accommodating the aminoacyl moiety of the second aa-tRNA substrate (Moutiez *et al.* 2014b). 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 8), 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 N¹-N⁷² base pair of the tRNA moiety participate in the recognition of the second substrate (Moutiez *et al.* 2014a,b). As (i) the sequence of the tRNA moiety does not matter for the binding of the first substrate, (ii) the N¹-N⁷² base pair is strongly conserved in prokaryotic sequences (Supplementary Data Set 6), 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 spp.* (Gondry *et al.* 2009, Tang *et al.* 2006), cFL for AlbC from *Streptomyces noursei* (Lautru *et al.* 2002, Fukushima *et al.* 1973), and cFY for Ndas_1148 from *Nocardioopsis dassonvillei* (Giessen *et al.* 2013b). 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 (Moutiez *et al.* 2014b, Giessen *et al.* 2013a,b, Gondry *et al.* 2009). However, the experimental conditions used in our study probably do not reflect those, 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 (Gu *et al.* 2013, Belin *et al.* 2012). 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 (Giessen *et al.* 2013a, 2013b, Cryle *et al.* 2010, Belin *et al.* 2009, Lautru *et al.* 2002, Gondry *et al.* 2001). The identification of the cyclodipeptide-synthesizing activities of many CDPSs provides valuable data for unraveling a large number of CDPS-dependent pathways and to elucidate and describe 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.

Acknowledgements

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 Vincent Dive, head of the Service d'Ingénierie Moléculaire des Protéines, for his continuous support and encouragement throughout this work. We thank Olivier Lespinet for advice about the building of phylogenetic trees, David Vallenet, Mark Stam and Maria Sorokina for helpful discussion on bioinformatics, and Adèle Ponties for technical assistance in cloning experiments. We are indebted to Ludovic Beuvier for skillful assistance in mass data analysis. We thank François Fenaille for kindly performing the

experiments using the Orbitrap mass spectrophotometer. We thank Pascal Kessler and Olivier Lequin for skillful assistance in NMR experiments.

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.

Online methods

Chemicals, bacterial strains and plasmids

All chemicals were from Sigma-Aldrich unless otherwise stated. We used *Escherichia coli* DH5 α 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-

*Sma*I fragment carrying the *lacI* gene and its promoter was excised from pREP4 and ligated into the *Bst*Z17I restriction site of pQE60 such that replication from ColE1 and transcription from *lacI* promoter are in the same direction on the plasmid (Fig. II.3). 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 (Sambrook *et al.* 2001).

Bioinformatics

Identification of the new putative CDPSs.

Sequence databases at NCBI and Broad Institute were screened for genes encoding putative CDPSs: PSI-Blast searches (Schäffer *et al.* 2001, Altschul *et al.* 1997) 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 (Fig. II.11), 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 Pro and Ala residues (Fig. II.12). 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.

A

```
AlbC -----MLAGLVPA PDHGMREEILGDRSRLIRQRGEHALIGISAG----- 39
C791_7200 -----
SsomD4_010100015357 -----MCIGVSPF----- 8
SJA_C1-32620 -----MNIGDYGHALLIAVSLF----- 16
Dmas2_010100010809 -----
K530_11977 -----
pc1814 -----METEKPVIILVSTHSSFHEKISG 23
WP_003122307 -----
AlbC ----NSYFSQKNTVMLLQWAGQRFERTDVVYVDTH-IDEMLIADGRSAQEAERSVKRTLK 94
C791_7200 -----MSAFTGFHFFVDPGP-SAFTLEALGYDPIRAAQKQRQGN 39
SsomD4_010100015357 ----NSYFSTPRLRRLADWALTRFGVHFFVPDAA-AAYTLEALGYDPARARHKAQRQGN 63
SJA_C1-32620 ----NSYYSTDRITTLVRWAKHTFERSHIPVFDLP-HAYTILSARKGHSPGSVRRARKEGR 71
Dmas2_010100010809 -----MTDSLQEHYLG LNP LVP T NIKRLAHKKGT 30
K530_11977 -----MLLQWAGQHFERIDVVYVDTH-IDDMLMADGRSAQEAERSVKRTLK 15
pc1814 DLKMNAFVSTIRN--HV----KG--KITVLLSDRAHINTMSL----- 57
WP_003122307 -----MPEEDARREAVRLG 14
```

B

```
AlbC -----MLAGLVPA PDHGMREEILGDRSRLIRQRGEHALIGISAG----- 39
C791_7200-EXT -----MLRPLPLETEHCADPLARAEHVCIGISPF----- 28
SsomD4_010100015357-EXT -----MFEIEPLTERCRGLLPEAVHVCIGVSPF----- 28
SJA_C1-32620-EXT -----MDVTVLSHRALT VTSIREPDRMNIGDYGHALLIAVSLF----- 37
Dmas2_010100010809-EXT -----MELHA-----KINGLE---KHL--KQHPTKRDFSIVNVIS---MDQPSQR 37
K530_11977-EXT -----MLSGLVPA LDHSMREEILGNRGRKIRQRGEHALIGISAG----- 39
pc1814-EXT MNKYTWRKNFTAILSNQKEIKGKGL---AFKTEETHWLETEKPVIIIVSTHSSFHEKISG 57
WP_003122307-EXT ----MNLIAEDFSTRSLLAGERYKAKIAFVSPHTRRNSFEDEPKCFLGVSLE----- 48
AlbC ----NSYFSQKNTVMLLQWAGQRFERTDVVYVDTH-IDEMLIADGRSAQEAERSVKRTLK 94
C791_7200-EXT ----NSYFTVDRIADLARWAMS AFTGFHFFVDPGP-SAFTLEALGYDPIRAAQKQRQGN 83
SsomD4_010100015357-EXT ----NSYFSTPRLRRLADWALTRFGVHFFVPDAA-AAYTLEALGYDPARARHKAQRQGN 83
SJA_C1-32620-EXT ----NSYYSTDRITTLVRWAKHTFERSHIPVFDLP-HAYTILSARKGHSPGSVRRARKEGR 92
Dmas2_010100010809-EXT GEKFEAFTRRINLLHKK----RGVRKLVIIITDSLQEHYLG LNP LVP T NIKRLAHKKGT 93
K530_11977-EXT ----NSYFSQKNTVMLLQWAGQHFERIDVVYVDTH-IDDMLMADGRSAQEAERSVKRTLK 94
pc1814-EXT DLKMNAFVSTIRN--HV----KG--KITVLLSDRAHINTMSL----- 91
WP_003122307-EXT ----NRNFTPNRFHSMVEWAARRFEKCSILIGDHI-HRITLQSTRDMPEEDARREAVRLG 103
```

Figure II.11. Multiple alignment with AlbC of putative CDPSs lacking their N-terminal part. The putative CDPSs identified as lacking a N-terminal part were aligned with AlbC using Clustal O at the EMBL-EBI website (panel A). The name of the coding sequence locus was used as the entry for each putative CDPS. The extended versions (suffix -EXT) obtained after examination of the 5' DNA sequences were also aligned with AlbC (panel B). The catalytic residue S37 of AlbC is in red. The NCBI entry WP_003122307 (locus PABE171_2006) was selected for the initial bioinformatics analysis (**Figure II.2**), but it was not selected for the biochemical characterization.

```
AlbC -----MLAGLVPA PDHGMREE 16
SCAT_0901 MHDNGHRPTAAGVRPQT VQARPAAGTASLTAPAPAPAPAPEFAPATAPVPAMRSGDFNCR 60
SCAT_0901_SHORT -----MRS GDFNCR 9

AlbC ILGDRSRLIRQRGEHALIGISAGNSYFSQKNTVMLLQWAGQRFERTDVVYVDTHIDEMLI 76
SCAT_0901 PLTPNCARVVDAGEHAVVG VSLGNSYFRGDILRALLRWLSGRFAQLD VVIPDSSYRDNLL 120
SCAT_0901_SHORT PLTPNCARVVDAGEHAVVG VSLGNSYFRGDILRALLRWLSGRFAQLD VVIPDSSYRDNLL 69
```

Figure II.12. Alignment of the putative CDPS encoded by SCAT_0901 with AlbC. The putative CDPS identified with a N-terminal extension was aligned with AlbC using Clustal O at the EMBL-EBI website. The shortened version (suffix -SHORT) was also aligned. The catalytic residue S37 of AlbC is in red.

Tree calculation.

Tree of sequences available in May 2013 (Figure II.2): After PSI-Blast searches and a set of relevant sequences had been obtained as described above, we generated a multiple sequence alignment using Muscle (Edgar

2004) integrated into Seaview (Gouy *et al.* 2010). 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) (Gouy *et al.* 2010) 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 (Letunic and Bork 2011, 2007).

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 (Edgar 2004) integrated into Seaview (Gouy *et al.* 2010) 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. II.10b), 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^{-3} ; alignment mode, local) (Remmert *et al.* 2012).

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 (Schneider and Stephens 1990). They were obtained at the WebLogo website (Crooks *et al.* 2004). 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 μ l of a phenol/chloroform/isoamyl alcohol mixture (25:24:1 v/v/v) and 400 μ 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 μ l of propan-2-ol and 120 μ l of 3M sodium acetate. The pellet was centrifuged (16 000 g, 15 min), washed with 70% ethanol and resuspended in TE buffer containing RNase at 50 μ 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 Table II.5. The PCR cycle was as follows: 98°C for 30s; 30 cycles of 98°C for 10s, 72°C for 60s; 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 α bacteria and verified by DNA sequencing of the promoter and CDPS-encoding regions (Eurofins-MWG).

CDPS	Amplified locus	Sequences
47	SSMG_00951	forward primer: 5'-CCATGGCGATCAGCAACGGCTTCCGC-3' reverse primer: 5'-AGATCTGGCTGGGCGGATGGTGGCGT-3'
48	SCAT_0198	forward primer: 5'-CCATGGCGCTGCATCGAACGTCCTTCTC-3' reverse primer: 5'-AGATCTCGGCTCGGCGGGCAGTTGCT-3'
49	SVEN_1864	forward primer: 5'-CCATGGCGACTCAGGCGGCTTTCGCC-3' reverse primer: 5'-AGATCTTCGCGGGCTCCTCGGTCGG-3'

Table II.5. Sequences of the forward and reverse primers used for PCR amplification.

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 (Gondry *et al.* 2009), 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 (Studier 2005). M15 bacteria were grown at 37°C until the OD₆₀₀ 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 24h at 20°C. BL21AI bacteria were grown in an autoinduced medium (Studier 2005) 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 hours, and transferred to 20°C for 20.5 hours. 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 lysosyme, 10 µM phosphoramidon, 1 mM phenylmethylsulfonylfluoride (PMSF), 100 mM Tris HCl, pH 8.0) (Braud *et al.* 2005). After 1.5 h on ice, MgCl₂ (10 mM final concentration) and 5 units of benzonase (Sigma-Aldrich) were added and the samples incubated for 1h 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 BioRad.

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 H₂O, solvent B: 0.1% (v/v) formic acid in acetonitrile/H₂O (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 (Gondry *et al.* 2009). 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 (Bajad *et al.* 2006). 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 (Belin *et al.* 2009, Gondry *et al.* 2009).

Cyclodipeptides were detected by both their *m/z* value and their daughter ion spectra (Armirotti *et al.* 2007, Stark and Hofmann 2005, Chet

et al. 2004, Papayannopoulos 1995 Falick *et al.* 1993, Johnson *et al.* 1987, Roepstorf and Fohlman 1984). 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 (Chen *et al.* 2004); cWW (Giessen *et al.* 2013b); cLL, cLM, cFL, cMM, cFM, cYM, cFF, cYF, cYA, cYL and cYY (Gondry *et al.* 2009)) or by comparison with authentic standards purchased from Bachem (cGG, cGA, cAA, cLG, cVV, cEA, cPL, cPM, cFP, cYS, cYP, cHF), 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 (0.1% TFA in water), buffer B (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 above. 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 (Gondry *et al.* 2009) (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⁺] calculated for C₁₄H₁₉N₂O₃ 263.1396, found 263.1393; cGN, [MH⁺] calculated for C₆H₁₀N₃O₃ 172.0722, found 172.0713; cGV, [MH⁺] calculated for

$C_7H_{13}N_2O_2$ 157.0977, found 157.0974; cAP, $[MH^+]$ calculated for $C_8H_{13}N_2O_2$ 169.0977, found 169.0966; cLE, $[MH^+]$ calculated for $C_{11}H_{19}N_2O_4$ 243.1345, found 243.1324; cyclocystine, $[MH^+]$ calculated for $C_6H_9N_2O_2S_2$ 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 (Jeedigunta *et al.* 2000). They were purified by semi preparative RP-HPLC (Purospher® STAR RP-18 endcapped LiChroCART, 10 mm × 250 mm, 5 μm, 120 Å, Merk) 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 described¹ and NMR spectroscopy. The NMR experiments were recorded on a Bruker Avance 250 spectrometer. The lyophilized samples were dissolved in

DMSO-*d*₆ (Eurisotop) and spectra were recorded at 25°C. The data from NMR experiments were processed and analyzed with Bruker TOPSPIN 2.0 program. cFA. ¹H NMR (250 MHz, DMSO): δ 8.13 (s, 1H, H^N F or A), 8.02 (s, 1H, H^N F or A), 7.32–7.13 (m, 5H, H^{δ,ε,ζ} F), 4.17 (m, 1H, H^α F), 3.61 (qt, J=7, 1.5 Hz, 1H, H^α A), 3.13 (dd, J=13.5, 3.75 Hz, 1H, H^β F), 2.85 (dd, J=13.5, 5 Hz, 1H, H^{β'} F), 0.45 (d, J=7Hz, 3H, H^β A); ¹³C NMR (62.8 MHz, DMSO): δ 167.7 (C^F or A), 165.8 (C^F or A), 136.1 (C^γ F), 130.4 (C^δ F), 128.1 (C^ε F), 126.7 (C^ζ F), 55.4 (C^α F), 49.7 (C^α A), 38.3 (C^β F), 19.7 (C^β A). HRMS (*m/z*): [MH⁺] calculated for C₁₂H₁₅N₂O₂, 219.1133, found 219.1143;

cFV. ¹H NMR (250 MHz, DMSO): δ 8.13 (s, 1H, H^N F or V), 7.92 (s, 1H, H^N F or V), 7.27–7.14 (m, 5H, H^{δ,ε,ζ} F), 4.21 (m, 1H, H^α F), 3.53 (m, 1H, H^αV), 3.15 (dd, J=13.5, 4.25 Hz, 1H, H^β F), 2.86 (dd, J=13.5, 5 Hz, 1H, H^{β'} F), 1.72 (m, 1H, H^β V), 0.64 (d, J=7.1 Hz, 3H, H^γ V), 0.24 (d, J=6.8 Hz, 3H, H^{γ'} V) ; ¹³C NMR (62.8 MHz, DMSO): δ 166.6 (C^F or V), 166.4 (C^F or V), 136.3 (C^γ F), 130.4 (C^δ F), 128.0 (C^ε F), 126.5 (C^ζ F), 59.2 (C^α V), 55.0 (C^α F), 37.8 (C^β F), 31.0 (C^β V), 18.3 (C^γ V), 16.2 (C^{γ'} V). HRMS (*m/z*): [MH⁺] calculated for C₁₄H₁₉N₂O₂, 247.1446, found 247.1444;

cFT. ¹H NMR (250 MHz, DMSO): δ 8.05 (d, J=2.75 Hz, 1H, H^N F or T), 7.89 (d, J=2.75 Hz, 1H, H^N F or T), 7.32–7.16 (m, 5H, H^{δ,ε,ζ} F), 5.00 (s, 1H, H^{OH} T), 3.93 (m, 1H, H^α F), 3.69 (m, 1H, H^βT), 3.52 (t , J= ~3 Hz, 1H, H^α T), 3.16–3.01 (ABX, 2H, H^β F, H^{β'} F), 0.96 (d, J=6.5 Hz, 3H, H^γ T); ¹³C NMR (62.8 MHz, DMSO): δ 167.4 (C^F or T), 166.7 (C^F or T), 137.4 (C^γ F), 129.8 (C^δ F), 128.2 (C^ε F), 126.4 (C^ζ F), 67.2 (C^β T), 60.3 (C^α T), 55.9 (C^α F), 40.8 (C^β F), 19.7 (C^γ T). HRMS (*m/z*): [MH⁺] calculated for C₁₃H₁₇N₂O₂, 249.1239, found 249.1236;

cFN. ¹H NMR (250 MHz, DMSO): δ 8.10 (s, 1H, H^N F or N), 7.62 (s, 1H, H^N F or N), 7.32–7.17 (m, 6H, H^{δ,ε,ζ} F, H^{NH2} N), 6.94 (s, 1H, H^{NH2}), 4.19 (m 1H, H^α F), 4.02 (m, 1H, H^α N), 3.10 (dd, J=13.75, 4.75 Hz, 1H, H^β F), 2.92 (dd, J=13.75, 5.25 Hz, 1H, H^{β'} F), 2.26 (dd, J=16, 4 Hz, 1H, H^β N), 1.386 (dd, J=16, 8.75 Hz, 1H, H^{β'} N); ¹³C NMR (62.8 MHz, DMSO): δ 171.6 (C^γ N), 166.9 (C^F or N), 166.4 (C^F or N), 136.4 (C^γ F), 130.1 (C^δ F), 128.1 (C^ε F), 126.7 (C^ζ

F), 55.2 (C^α F), 51.1 (C^α N), 38.1 (C^β F or N), 38.1 (C^β F or N). HRMS (*m/z*): [MH⁺] calculated for C₁₃H₁₆N₃O₃, 262.1192, found 262.1186;

cFE. ¹H NMR (250 MHz, DMSO): δ 11.91 (s, 1H, H^{COOH} E), 8.21 (s, 1H, H^N F or E), 8.10 (s, 1H, H^N F or E), 7.28–7.14 (m, 5H, H^{δ-ε-ζ} F), 4.19 (s, 1H, H^α F), 3.69 (m, 1H, H^α E), 3.14 (dd, J=13.25, 3.5 Hz, 1H, H^β F), 2.84 (dd, J=13.25, 4.75 Hz, 1H, H^{β'} F), 1.66 (m, 2H, H^γ E), 1.32–0.97 (m, 2H, H^β E); ¹³C NMR (62.8 MHz, DMSO): δ 173.9 (C^δ E), 166.7 (C^γ F or E), 166.4 (C^γ F or E), 136.1 (C^γ F), 130.3 (C^δ F), 128.1 (C^ε F), 126.8 (C^ζ F), 55.32 (C^α F), 53.0 (C^α E), 38.1 (C^β F), 28.6 (C^β or ^γ E), 28.5 (C^β or ^γ E). HRMS (*m/z*): [MH⁺] calculated for C₁₄H₁₇N₂O₄, 277.1188, found 277.1187.

Competing financial interests statement

The authors declare no competing financial interests.

Chapitre III

Etude d'un cluster cyclodityrosyl synthase – cytochrome P450 CYP121 chez *Streptomyces venezuelae*

Introduction

La tuberculose, provoquée par l'actinobactérie *Mycobacterium tuberculosis* est un problème majeur de santé publique. C'est une maladie qui affecte neuf millions de personnes par an et est responsable de 1,5 million de décès par an. Des traitements sont disponibles ; néanmoins ils sont de longue durée (six mois minimum), contraignants, avec beaucoup d'effets secondaires. De plus, l'émergence de souches multirésistantes aux antibiotiques rend le traitement de la maladie de plus en plus difficile (WHO, 2016). Il est donc urgent de développer de nouvelles molécules efficaces contre les cibles thérapeutiques connues chez *M. tuberculosis*, ou alors de trouver de nouvelles cibles potentielles, de les caractériser et ensuite d'identifier des molécules interagissant avec ces cibles inhibant ainsi la croissance de *M. tuberculosis* ou conduisant à sa mort.

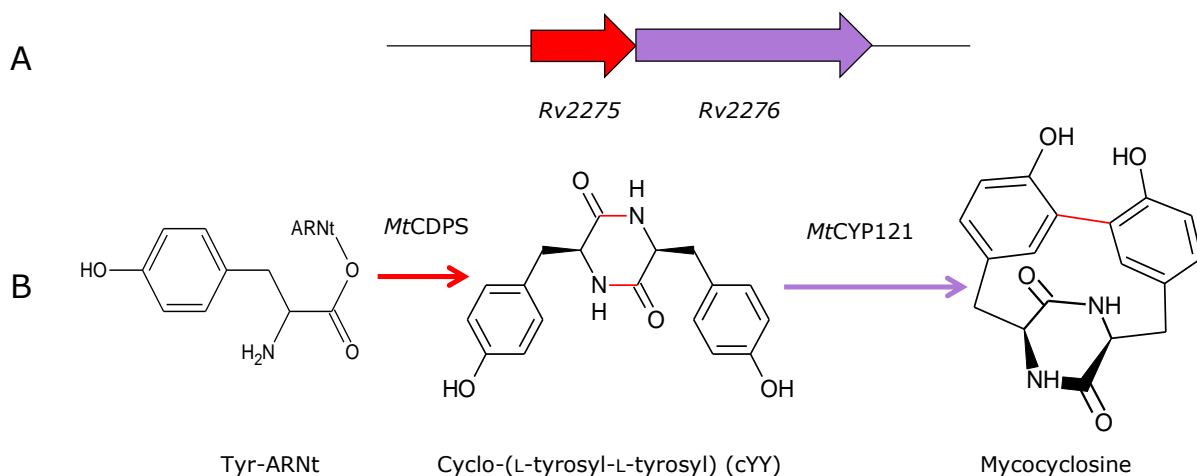


Figure III.1. Cluster (A) et voie (B) de biosynthèse putatifs de la mycocyclosine de *M. tuberculosis*. En rouge est indiquée la cyclodityrosyl synthase *MtCDPS* (une enzyme de la famille des cyclodipeptide synthases), en violet la mycocyclosine synthase *MtCYP121* (une enzyme de la famille des cytochromes P450). La voie est inférée des études en expression hétérologue (*MtCDPS*) et *in vitro* (*MtCYP121*).

Parmi les cibles thérapeutiques potentielles chez ce microorganisme, il y a le cytochrome P450 CYP121, appelé aussi mycocyclosine synthase, codé par le gène *Rv2276* (MacLean *et al* 2008), que l'on appellera par la suite *MtCYP121*. MacLean et ses collaborateurs (2008) ont établi qu'il s'agissait d'un gène essentiel chez *M. tuberculosis* et ont obtenu des résultats

préliminaires suggérant que l'inhibition du *MtCYP121* par des composés azolés aurait un effet bactéricide sur *M. tuberculosis*. Le gène codant ce cytochrome est associé au gène *Rv2275* codant une cyclodipeptide synthase (CDPS), appelée également cyclodityrosyl synthase, que l'on appellera par la suite *MtCDPS*. *MtCDPS* et *MtCYP121* sont censés synthétiser le composé nommé mycocyclosine, selon la voie représentée sur la figure III.1 (Gondry *et al.* 2009, Belin *et al.* 2009). Des études structurales et mécanistiques très poussées ont été effectuées tant sur la *MtCDPS* que sur le *MtCYP121* (Dumas *et al.* 2014, Fonvielle *et al.* 2013, Vetting *et al.* 2011, Belin *et al.* 2009, MacLean *et al.* 2008). Cependant plusieurs questions restent encore sans réponse à ce jour. En effet l'activité de la *MtCDPS* a été étudiée en expression hétérologue chez *Escherichia coli* et l'activité du *MtCYP121* a été étudiée *in vitro* à partir d'une protéine recombinante purifiée (Gondry *et al.* 2009, Belin *et al.* 2009). La production de la mycocyclosine n'a jamais été décrite *in vivo* chez *M. tuberculosis*. De plus, même si MacLean et ses collaborateurs (2008) n'ayant pas réussi à inactiver le gène *Rv2276* à moins de fournir une copie fonctionnelle du gène en *trans* en ont conclu qu'il s'agissait d'un gène essentiel, cette essentialité est remise en question par le fait qu'il existe des isolats cliniques avec de grandes délétions dans cette partie du génome incluant le *cluster MtCDPS – MtCYP121* (Belin *et al.* 2012, MacLean *et al.* 2008). Une des hypothèses envisageables pour expliquer cet état des choses serait que l'essentialité du gène *Rv2276* serait liée à la présence du gène *Rv2275*, parce que la cyclodityrosine non modifiée présenterait une toxicité pour la cellule bactérienne. Enfin, aucun rôle biologique n'a été jusqu'à présent attribué à la mycocyclosine (si tant est que cette molécule soit bel et bien produite *in vivo*). Il serait donc intéressant d'approfondir les études du *cluster*. Mais *M. tuberculosis* est un organisme à croissance très lente (temps de génération d'environ 20 h) et dont la manipulation exige des contraintes de sécurité poussées (niveau de sécurité biologique 3). *Mycobacterium smegmatis*, une mycobactérie non pathogène à croissance rapide, souvent utilisée comme modèle pour les études sur les mycobactéries, ne peut quant à elle pas être utilisée, car elle est dépourvue des homologues des gènes *Rv2275* et *Rv2276* (Belin *et al.* 2012).

Cependant des homologues de *Rv2275* et de *Rv2276* de *M. tuberculosis* ont été trouvés chez *Streptomyces venezuelae*.

S. venezuelae est une bactérie du sol, connue et étudiée depuis la moitié du XX^e siècle, notamment en tant que producteur naturel du chloramphénicol ainsi que de la pikromycine, le premier antibiotique de la famille des macrolides à avoir été découvert. L'homologue du gène *Rv2275* est le gène *Sven_1864* codant la CDPS que l'on appellera par la suite *SvCDPS* (59% d'identité de séquence protéique avec *MtCDPS*). L'homologue du gène *Rv2276* est le gène *Sven_1865* codant le cytochrome CYP121 que l'on appellera par la suite *SvCYP121* (60% d'identité de séquence protéique avec *MtCYP121*) (Belin *et al.* 2012). *S. venezuelae* n'est pas pathogène, se cultive facilement et croît rapidement (temps de génération d'environ 1 h) ; la séquence de son génome est connue et les techniques pour manipuler génétiquement cette souche sont décrites.

De plus, le travail que j'ai effectué décrit dans le chapitre précédent prouve que la *SvCDPS* en expression hétérologue produit majoritairement du cYY, comme son homologue *MtCDPS* (Jacques *et al.* 2015). Compte tenu de ce résultat et de la forte similitude de séquence entre *SvCYP121* et *MtCYP121*, on pouvait émettre l'hypothèse que les gènes de *S. venezuelae* dirigeraient la biosynthèse de la mycocyclosine et que *S. venezuelae* constituerait un bon modèle pour étudier la biosynthèse de la mycocyclosine et aborder la question de son rôle biologique.

Matériel et méthodes

Souches bactériennes et conditions de culture

Les souches bactériennes utilisées dans cette étude sont répertoriées dans le tableau III.1 et les milieux utilisés dans le tableau III.2. Les souches d'*E. coli* ont été manipulées selon les protocoles standards (Green et Sambrook 2012). Si nécessaire, les antibiotiques ont été ajoutés à la concentration suivante : ampicilline : 100 mg/l, apramycine: 50 (milieux liquides) ou 75 (milieux solides) mg/l, chloramphénicol: 30 mg/l,

kanamycine: 25 mg/l, hygromycine: 50 (milieux liquides) ou 100 (milieux solides) mg/l. De plus les cultures d'*E. coli* contenant de l'hygromycine ont été supplémentées avec 1 mM d'IPTG. Les souches de *Streptomyces* ont été manipulées selon les procédures standards (Kieser *et al.* 2000). Les spores de *S. venezuelae* ont été préparées sur le milieu GYM, celles de *Streptomyces coelicolor* sur le milieu Gruau. Les cellules d'*E. coli* électrocompétentes ont été préparées selon le protocole établi par Robey et ses collaborateurs (1996). Les conjugaisons entre *E. coli* ET12567pUZ8002 et *Streptomyces spp.* ont été réalisées selon le protocole de Kieser *et al.* (2000) modifié, utilisant le milieu SFM, complété avec 10 mM MgCl₂, et l'application de la sélection antibiotique dans 3 ml de milieu SNA, contenant de la fosfomycine à la place de l'acide nalidixique. Pour *S. venezuelae*, l'incubation entre l'étalement de la conjugaison et la mise en place de la sélection antibiotique a été effectuée à la température ambiante. Les exconjugants ont été repiqués sur milieu SFM contenant de la fosfomycine et les antibiotiques de sélection. Les concentrations d'antibiotiques utilisés sont les suivantes : apramycine : 30 mg/l, fosfomycine : 50 mg/l, hygromycine: 50 mg/l, thiostrepton: 30 mg/l.

Souche	Génotype	Plasmide	Source
<i>Escherichia coli</i> BL21-AI	F ⁻ ompT hsdS _B (r _B ⁻ m _B ⁻) gal dcm araB::T7RNAP-tetA	-	Thermo Fisher Scientific
<i>Escherichia coli</i> DH5α	F ⁻ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(r _K ⁻ m _K ⁺), λ ⁻	-	Promega
<i>Escherichia coli</i> ET12567pUZ8002	F ⁻ dam13::Tn9 dcm-6 hsdM hsdR zjj-202::Tn10 recF143 galK2 galT22 ara-14 lacY1 xyl-5 leuB6 thi-1 tonA31 rpsL136 hisG4 tsx-78 mtl-1 glnV44	pUZ8002	Flett <i>et al.</i> (1997)
<i>Micrococcus luteus</i> DSM1790	souche sauvage	-	DSM
<i>Streptomyces coelicolor</i> M1154	Δact Δred Δcpk Δcda rpoB[C1298T] rpsL[A262G]	-	Gomez-Escribano et Bibb (2010)
<i>Streptomyces venezuelae</i> ATCC 10712 = <i>S. venezuelae</i> WT	souche sauvage	-	ATTC
<i>Streptomyces venezuelae</i> ΔSven_1864	ΔSven_1864	-	cette étude
<i>Streptomyces venezuelae</i> ΔSven_1865	ΔSven_1865	-	cette étude
<i>Streptomyces venezuelae</i> ΔSven_1864-5	ΔSven_1864-5	-	cette étude

Tableau III.1. Souches utilisées dans cette étude

Milieu	Composition ou référence bibliographique
5006	Saccharose 3g/l, dextrine 15,0 g/l, extrait de viande 1,0 g/l, extrait de levure 2,0 g/l, tryptone de soja (Oxoid) 5,0 g/l, NaCl 0,5 g/l, K ₂ HPO ₄ 0,5 g/l, MgSO ₄ • 7H ₂ O 0,5 g/l, FeSO ₄ • 7H ₂ O 0,01 g/l
BHI	Brain Heart Infusion (Difco) 37 g/l
DAM5	Grove et Randall (1955)
GPYE	Shirling et Gottlieb (1966)
Gruau	farine de gruau 30 g/l, glucose 5 g/l yeast extract 5 g/l, pH ajusté à 7,2 avec NaOH
GYM	glucose : 4 g/l, yeast extract : 4 g/l, malt extract : 10 g/l, CaCO ₃ : 2 g/l, pH ajusté à 7,2 avec NaOH
HT	Kieser <i>et al.</i> (2000)
ISP5	Shirling et Gottlieb (1966)
LB	Green et Sambrook (2012)
MP5	Pernodet <i>et al.</i> (1993)
R2-S	R2YE (Kieser <i>et al.</i> 2000) sans saccharose ni yeast extract
R2M	R2-S avec glucose remplacé par la même quantité de maltose
R2YE-S	R2YE (Kieser <i>et al.</i> 2000) sans saccharose
SAF	Sánchez et Braña (1996)
SFM	Kieser <i>et al.</i> (2000)
SNA	Kieser <i>et al.</i> (2000)
SSG	Gottardi <i>et al.</i> (2011)
TB	Green et Sambrook (2012)
TSB	Tryptone Soya Broth (Bacto) 30 g/l

Tableau III.2. Milieux de culture utilisés dans cette étude. Pour les milieux solides, 20 g/l d'agar (Bacto) a été ajouté.

Vecteurs, manipulation de l'ADN et de l'ARN

Tous les plasmides utilisés dans cette étude sont répertoriés dans le tableau III.3. Les réactions PCR pour les clonages ont été effectuées avec la polymérase à haute fidélité Phusion (Thermo Fischer Scientific) en utilisant le protocole du fournisseur spécifique des ADN à haut contenu en GC. Les mélanges réactionnels ont été supplémentés avec la bêtaïne à 1 M final. Les réactions d'amplification dans le cas de vérification de constructions et pendant la RT-PCR ont été effectuées avec la polymérase Dream Taq (Thermo Fischer Scientific), le mélange réactionnel ayant été supplémenté avec la Q-Solution (Qiagen) à 1x final.

Les amorces sont répertoriées dans le tableau III.4. Elles ont été commandées auprès d'Integrated DNA Technologies.

Les produits PCR ont été purifiés avec le kit NucleoSpin Gel and PCR Clean-up (Macherey-Nagel).

Les plasmides ont été extraits avec le kit NucleoSpin Plasmid (Macherey-Nagel).

Plasmide	Description	Source
pET15b	Vecteur d'expression <i>E. coli</i> ajoutant une étiquette 6HisTag à l'extrémité N-terminale de la protéine exprimée	Novagen / Merck-Millipore
pGEM-T Easy	Vecteur de clonage <i>E. coli</i>	Promega
pIJ196	Vecteur d'expression <i>E. coli</i> , ajoutant une étiquette 6HisTag à l'extrémité C-terminale de la protéine exprimée, dérivé de pQE60 exprimant constitutivement <i>lacI</i>	Jacques <i>et al</i> (2015)
pJWc05	pGEM-T Easy portant comme insert le fragment amont du gène <i>Sven_1864</i>	cette étude
pJWc06	pGEM-T Easy portant comme insert le fragment amont du gène <i>Sven_1865</i>	cette étude
pJWc07	pGEM-T Easy portant comme insert le fragment aval du gène <i>Sven_1864</i>	cette étude
pJWc08	pGEM-T Easy portant comme insert le fragment aval du gène <i>Sven_1865</i>	cette étude
pJWe05	pOSV666 portant le gène <i>Sven_1864</i>	cette étude
pJWe06	pOSV666 portant le gène <i>Sven_1865</i>	cette étude
pJWe07	pOSV666 portant les gènes <i>Sven_1864-5</i>	cette étude
pJWe08	pET-15b portant le gène <i>Sven_1865</i>	cette étude
pJWe19	pOSV666 portant les gènes <i>Rv_2275-6</i>	cette étude
pJWm01	Dérivé du plasmide pOSV400 portant la construction pour remplacer le gène <i>Sven_1864</i> par la cassette $\Omega att1aac$	cette étude
pJWm02	Dérivé du plasmide pOSV400 portant la construction pour remplacer le gène <i>Sven_1865</i> par la cassette $\Omega att1aac$	cette étude
pJWm03	Dérivé du plasmide pOSV400 portant la construction pour remplacer les gènes <i>Sven_1864-5</i> par la cassette $\Omega att1aac$	cette étude
pOSV215	Source de la cassette excisable $\Omega att1aac$	Raynal <i>et al</i> (2006)
pOSV236	Vecteur permettant l'excision de la cassette $\Omega att1aac$	Raynal <i>et al</i> (2006)
pOSV400	Vecteur suicide pour la délétion des gènes chez <i>Streptomyces</i> par recombinaison homologue	Vingadassalon <i>et al</i> , unpublished data
pOSV666	pRT801 avec <i>ermE*</i> _p , TipA RBS, et sites de restriction <i>NdeI</i> , <i>PacI</i> et <i>HindIII</i> dérivé de pRT801	Witwinowski <i>et al.</i> (2017), en préparation (cf. chapitre IV)

Tableau III.3. Plasmides utilisés dans cette étude

Les enzymes de restriction et de modification d'ADN ont été fournies par Thermo Fischer Scientific. L'ADN des vecteurs a systématiquement été déphosphorylé avec la FastAP (Thermo Fischer Scientific). Tous les produits de digestion ont été purifiés sur gel d'agarose à l'aide du kit NucleoSpin Gel and PCR Clean-up (Macherey-Nagel). Les ligations ont été effectuées avec l'ADN ligase du phage T4 (Thermo Fischer Scientific). L'ADN génomique de *S. venezuelae* et *S. coelicolor* a été extrait selon le protocole décrit dans Jacques *et al.* (2015). Le séquençage d'inserts de plasmide ou de produits PCR a été réalisé par la société Beckman Coulter Genomics.

Nom	Séquence	Description
JW-RT1	CCCTGCAGAGACGCTCCGAC	Amorce FWD pour la RT-PCR de <i>Sven_1864</i>
JW-RT2	GTCACAGGTCTCGCGCACCA	Amorce REV pour la RT-PCR de <i>Sven_1864</i>
JW-RT3	GCCTACGCCGACCCGTA	Amorce FWD pour la RT-PCR de <i>Sven_1865</i>
JW-RT4	AGCAGTTCGTCCATCGCCCC	Amorce REV pour la RT-PCR de <i>Sven_1865</i>
JW-RTc1	TCGTTCCGAGAGGTTGTTCG	Amorce FWD pour la RT-PCR de <i>Sven_5499 (rpoD)</i>
JW-RTc2	CTCGGTCTCTTCTCCTCGC	Amorce REV pour la RT-PCR de <i>Sven_5499 (rpoD)</i>
JWm1	AAGCTTGCACTTCACGATCCCGGCGA	Amorce FWD pour amplifier le fragment amont de <i>Sven_1864</i> ajoutant le site <i>HindIII</i>
JWm2	GATATCCTCTGCAGGGGCGAAAGCCG	Amorce REV pour amplifier le fragment amont de <i>Sven_1864</i> ajoutant le site <i>EcoRV (Eco32I)</i>
JWm3	GATATCAACCAGGGCCACGGAATCCT	Amorce FWD pour amplifier le fragment aval de <i>Sven_1864</i> ajoutant le site <i>EcoRV (Eco32I)</i>
JWm4	ACTAGTCAACCGCCGTCTGAAGGCCA	Amorce REV pour amplifier le fragment aval de <i>Sven_1864</i> ajoutant le site <i>SpeI (BcuI)</i>
JWm5	AAGCTTTGTTCTACGACGAGACGGGC	Amorce FWD pour amplifier le fragment amont de <i>Sven_1865</i> ajoutant le site <i>HindIII</i>
JWm6	GATATCGGCCGGCAGGGTCTCGACGG	Amorce REV pour amplifier le fragment amont de <i>Sven_1865</i> ajoutant le site <i>EcoRV (Eco32I)</i>
JWm7	GATATCCTCCCGAGCGGCTGCCGGT	Amorce FWD pour amplifier le fragment aval de <i>Sven_1865</i> ajoutant le site <i>EcoRV (Eco32I)</i>
JWm8	ACTAGTGCGCCACGCACCGGGTGTAC	Amorce REV pour amplifier le fragment aval de <i>Sven_1865</i> ajoutant le site <i>SpeI</i>
JWseq1	CGGTGGGCGTGACGCTGTTC	Amorce pour séquencer le fragment aval <i>Sven_1865</i>
JWseq4	CTGGATCGACTACAAGCGCA	Amorce pour séquencer la cicatrice Δ <i>Sven_1864-1865</i>
JWseq5	CAAGGCGACCGGGCAG	Amorce pour séquencer la cicatrice Δ <i>Sven_1864-1865</i>
JWseq6	CCCTTTTTGGCCTTGAAT	Amorce FWD pour séquencer l'insert dans pOSV666
JWseq7	GAATGTGAGCGGATAACAA	Amorce REV pour séquencer l'insert dans pET15b
JWseq42	CGCTGACTGTTCACCCGAA	Amorce interne au fragment <i>Rv2275-6</i> pour séquençage
JWseq43	CGCTTTCATCACCGCCTTGC	Amorce interne au fragment <i>Rv2275-6</i> pour séquençage
JWseq44	CCAGCGGGTCGGCGAAGTCA	Amorce interne au fragment <i>Rv2275-6</i> pour séquençage
JWc3	GGGAATTCATATGACTCAGGCGGCTTTCGCCC	Amorce FWD pour cloner <i>Sven_1864</i> dans pOSV666 ajoutant le site <i>NdeI</i>
JWc4	CCCAAGCTTTCATCGCGCGGCTCCTCGGT	Amorce REV pour cloner <i>Sven_1864</i> dans pOSV666 ajoutant le site <i>HindIII</i>
JWc5	GGGAATTCATATGAGCGTGC	Amorce FWD pour cloner <i>Sven_1865</i> dans pOSV666 ajoutant le site <i>NdeI</i>
JWc6	CCCAAGCTTCTACCAGGCCACCGGCAGCC	Amorce REV pour cloner <i>Sven_1865</i> dans pOSV666 ajoutant le site <i>HindIII</i>
JWc6b	CCGCTCGAGTACCAGGCCACCGGCAGCC	Amorce REV pour cloner <i>Sven_1865</i> dans pET-15b ajoutant le site <i>XhoI</i>
JWc27	GGGAATTCATATGTCATACGTGGCTGCCGA	Amorce FWD pour cloner <i>RV2275-6</i> avec le terminateur naturel ajoutant le site <i>NdeI</i>
JWc28	CCCAAGCTTCGTCGCTCGCTGT	Amorce REV pour cloner <i>RV2275-6</i> avec le terminateur naturel ajoutant le site <i>HindIII</i>
pQE60-F	CCCGAAAAGTGCCACCTG	Amorce FWD pour séquencer l'insert dans pIJ196
pQE60-R	GTTCTGAGGTCATTACTGG	Amorce FWD pour séquencer l'insert dans pIJ196
pSET152-F	CGGGCCTCTTCGCTATTAC	Amorce forward pour séquencer l'insert dans pGEM-T
pSET152-R	TTATGCTTCCGGCTCGTATG	Amorce reverse pour séquencer l'insert dans pGEM-T
AG-aac3	CGACGTGGCAGGATCGAACG	Vérification de la cassette <i>Ωatt1aac4</i> (Amorce sortant, début)
AG-aac4	GTCAACTGGGCCGAGATCCG	Vérification de la cassette <i>Ωatt1aac4</i> (Amorce sortant, fin)
AG-hyg1	TCGGATGATTCCTACGCGAG	Vérification de la cassette <i>hph</i> (Amorce interne, FWD)
AG-hyg2	TCCTCGAACACCTCGAAGTC	Vérification de la cassette <i>hph</i> (Amorce interne, REV)

Tableau III.4. Amorces utilisées dans cette étude

L'ARN a été isolé avec les kits NucleoSpin RNA et NucleoSpin RNA/DNA Buffer Set (Macherey-Nagel). Les extraits d'ARN ont été traités avec la RNase-free DNase (Ambion) et purifiés avec le kit NucleoSpin RNA Clean-up (Macherey-Nagel). L'absence de contamination par de l'ADN génomique a été vérifiée par une PCR de 30 cycles avec les amorces JW-RTc1 et JW-RTc2 utilisant l'extrait d'ARN en tant que matrice. La transcription inverse a été effectuée avec le kit RevertAid First Strand cDNA Synthesis (Thermo Fischer Scientific) en utilisant le Random Hexamer Primer (amorce d'hexamères aléatoires), selon le protocole du fournisseur pour les matrices à haut contenu en GC. La PCR réalisée sur l'ADNc avait 28 cycles avec 30 s de dénaturation à 95°C, 30 s d'hybridation à 60°C et 30 s d'élongation à 72°C.

Traitement bioinformatique

Les alignements multiples de séquences ont été réalisés à l'aide du logiciel ClustalW (Combet *et al.* 2000). Pour l'analyse des séquences, la recherche des sites de restriction, et la planification des clonages, le logiciel Clone Manager Professional 9 (Educational & Scientific Software) a été utilisé. La conception des amorces a été réalisée soit avec le logiciel Clone Manager Professional 9 soit avec le logiciel en ligne Primer Blast (NCBI, <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Toutes les séquences ont été obtenues du NCBI.

Obtention des vecteurs intégratifs permettant la surexpression du cluster *Sven_1864-5* de *S. venezuelae*

Les gènes *Sven_1864*, *Sven_1865* ainsi que le *cluster* entier ont été amplifiés par PCR à partir de l'ADN génomique de la souche sauvage de *S. venezuelae* en utilisant respectivement les couples d'amorces JWc3/JWc4, JWc5/JWc6 et JWc3/JWc6. Les produits PCR ont été clonés dans le vecteur pOSV666, entre les sites *HindIII* et *NdeI*, donnant respectivement les plasmides pJWe05, pJWe06 et pJWe07. L'intégrité de chaque insert a été vérifiée par séquençage.

Obtention des vecteurs intégratifs permettant l'expression hétérologue du cluster Rv2275-2276 de *M. tuberculosis*

Le cluster Rv2275-6 a été amplifié à partir de l'ADN génomique de la souche H37Rv de *M. tuberculosis* (don de Guislaine Refrégier, I2BC, Orsay) à l'aide des amorces JWc27/JWc28. Il a été cloné dans le vecteur pOSV666 entre les sites *Hind*III/*Nde*I, donnant le vecteur pJWe19. L'intégrité de l'insert a été vérifiée par séquençage.

Obtention des mutants de délétion des gènes *Sven_1864* et *Sven_1865*

La stratégie utilisée pour inactiver les gènes *Sven_1864* et/ou *Sven_1865* consiste à remplacer par recombinaison homologue dans le génome la majeure partie de la séquence codante du gène cible par la cassette *Ωatt1aac* conférant la résistance à l'apramycine.

La cassette utilisée présente l'avantage de pouvoir être excisée chez *Streptomyces* par un événement de recombinaison spécifique de site afin d'éviter tout éventuel effet polaire sur l'expression des gènes situés en aval si ces gènes sont co-transcrits. Après excision subsistent 30 paires de bases au début du gène et 60 (*Sven_1864*) ou 30 (*Sven_1865*) paires de bases à la fin du gène.

Les fragments amont et aval des gènes à déléter ont été amplifiés par PCR à partir de l'ADN génomique de la souche sauvage de *S. venezuelae* à l'aide des couples d'amorces suivants : JWm1/JWm2 (amont *Sven_1864*), JWm3/JWm4 (aval *Sven_1864*), JWm5/JWm6 (amont *Sven_1865*) et JWm7/JWm8 (aval *Sven_1865*). Après purification, les extrémités 3' ont alors été traitées avec la polymérase Taq (Qiagen) sans amorces en présence de dNTP à 72°C pendant 10 min afin de rajouter une adénine. Les quatre fragments ont été clonés dans le vecteur pGEM-T Easy, donnant respectivement, les plasmides pJWc05, pJWc06, pJWc07 et pJWc08. L'intégrité de leur insert a été vérifiée par séquençage. Les fragments amont ont ensuite été excisés des vecteurs pJWc05 et pJWc07 par digestion *Hind*III/*Eco*32I. Les fragments aval ont quant à eux été excisés des vecteurs pJWc06 et pJWc08 par digestion *Eco*32I/*Bcu*I. La cassette *Ωatt1aac* a été

excisée du vecteur pOSV215 par digestion *Eco32I*. Le vecteur suicide pOSV400 a été digéré par les enzymes *HindIII/BclI*. Trois ligations avec quatre fragments (vecteur suicide + fragment amont + cassette *Qatt1aac* + fragment aval) ont été réalisées. La ligation utilisant les fragments amont et aval de *Sven_1864* a donné le plasmide pJWm01. La ligation utilisant les fragments amont et aval de *Sven_1865* a donné le plasmide pJWm02. La ligation utilisant le fragment amont de *Sven_1864* et le fragment aval de *Sven_1865* a donné le plasmide pJWm03. L'intégrité des constructions a été vérifiée par profil de digestion. Les trois constructions ont ensuite été transférées dans *S. venezuelae* par conjugaison. Les exconjugants ont été sélectionnés pour la résistance à l'apramycine. Les clones résultant potentiellement d'un événement de double recombinaison ont été criblés pour leur sensibilité à l'hygromycine. L'ADN génomique de ces clones sensibles à l'hygromycine a été extrait et le remplacement des gènes d'intérêt par la cassette *Qatt1aac* ainsi que l'absence de la cassette *hph* conférant la résistance à l'hygromycine ont été vérifiés par PCR. Le plasmide pOSV236, permettant d'exciser la cassette de résistance, a ensuite été introduit dans les trois mutants obtenus par conjugaison entre *E. coli* et *Streptomyces*. Les exconjugants ont été sélectionnés pour la résistance au thiostrepton. Après un à deux repiquages sur milieu contenant du thiostrepton, ils ont ensuite été criblés pour la sensibilité à l'apramycine (excision de cassette réussie). Le plasmide pOSV236 a enfin été curé par repiquages successifs sur un milieu sans sélection (les clones ayant perdu le plasmide étant criblés pour leur sensibilité au thiostrepton). L'ADN génomique a été extrait et l'excision a été confirmée par séquençage d'un produit PCR.

Etude biochimique du cytochrome *Su*CYP121

Tous les produits chimiques ont été fournis par Sigma Aldrich. Le gène *Sven_1865* a été amplifié à partir de l'ADN génomique de la souche sauvage de *S. venezuelae* avec les amorces JWc5/JWc6b. Il a été cloné dans le vecteur pET15b au niveau des sites de restriction *NdeI/XhoI*, donnant le vecteur pJWe08. L'intégrité de l'insert a été vérifiée par séquençage. Le plasmide pJWe08 a ensuite été transféré par transformation dans *E. coli*

BL21-AI. Des précultures de la souche résultante BL21-AI/pJWe08 ont été réalisées sur la nuit dans 100 ml de milieu TB additionné d'ampicilline. Une culture de 2 l de milieu TB additionné d'ampicilline et de 100 mg/l d'acide aminolévulinique a étéensemencée à une DO initiale de 0,03 à l'aide de la préculture de nuit puis répartie dans 6 erlenmeyers de 1 l et mise à incuber à 37°C sous agitation (200 rpm) jusqu'à la DO = 0,5. Les 6 cultures ont alors été refroidies à 20°C pendant 15 min, puis l'expression du cytochrome *SvCYP121* a été induite par l'ajout d'IPTG (0,5 mM) et d'arabinose (0,02%). Après une nuit d'incubation, les bactéries ont été récupérées par centrifugation, suspendues dans le tampon A (Tris-HCl 50mM pH 7,2 ; NaCl 0,5M ; imidazole 20mM ; PMSF 1mM ; Benzamidine hydrochloride 1mM ; glycérol 5%) et cassées à la presse d'Eaton. L'ADN a été dégradé par incubation en présence de 0,1 U/ml de benzonase et 2 μ M de MgCl₂ pendant 1h à 4°C sous agitation douce. Le surnageant de cassage a alors été récupéré par centrifugation. La purification du *SvCYP121* a été réalisée par chromatographie d'affinité sur colonne Ni Sepharose HP 5 ml (GE Healthcare) à l'aide de l'appareil Äkta Purifier (GE Healthcare). L'échantillon a été chargé et lavé dans le tampon A à 1 ml/min jusqu'à ce que la DO à 280 nm devienne stable et basse. L'élution a été faite avec un gradient partant de 0 % de tampon B (Tris-HCl 50mM pH 7,2 ; NaCl 0,5M ; imidazole 500 mM, PMSF 1mM ; benzamidine hydrochloride 1mM ; glycérol 5%) et allant jusqu'à 100% de tampon B après 20 volumes de colonne. Les fractions contenant le cytochrome *SvCYP121* ont été groupées et dialysées dans un boudin de *cut-off* de 10 kDa incubé dans du tampon Tris-HCl 50 mM pH 7,2 / EDTA 1mM. Elles ont ensuite été purifiées par filtration en gel sur la colonne Hiprep 16/60 Sephacryl 120 ml (GE Healthcare) dans 100 ml de tampon C (Tris HCl 50mM pH 7,2 ; EDTA 1 mM ; NaCl 0,15 M) avec un flux de 0,8 ml/min, puis concentrées par ultrafiltration à l'aide d'une unité d'ultrafiltration par centrifugation Vivaspin 6 10 000 MWCO (Sartorius).

Les propriétés spectrales du cytochrome ont été étudiées en solution aqueuse à la concentration 5 μ M de protéine seule, avec 200 mM d'imidazole et avec 200 mM d'imidazole et 475 μ M de cYY. Son activité a été étudiée en réalisant les réactions dans le milieu réactionnel suivant : KH₂PO₄/K₂HPO₄

100 mM ; pH 7,5 ; cYY 1 mM ; *SvCYP121* 5 μ M ; ferrédoxine d'épinard 5 μ M ; ferrédoxine réductase d'épinard 0,2 U/ml ; NADPH 1 mM à 30°C pendant 2h. Le volume réactionnel était de 1 ml. Les réactions ont alors été arrêtées par acidification avec de l'acide formique à 2% final. Les mélanges réactionnels ont été analysés par LC-MS.

Analyse des surnageants de culture de la souche sauvage de *S. venezuelae*

Des cultures de 70 ml de milieu R2-S ont étéensemencées avec $1,4 \times 10^8$ spores de *S. venezuelae*. Les cultures ont été réalisées à 28°C dans des erlenmeyers à baffles de 500 ml sous agitation (180 rpm). Des prélèvements de 5 ml ont été réalisés à 24h, 48h, 72h, 96h et 168h de culture. Le surnageant a été séparé du culot bactérien par centrifugation, additionné d'acide trifluoroacétique à 2% final et analysé par LC-MS.

Surexpression du *cluster Sven_1864-5* chez *S. venezuelae*

Le vecteur pJWe07 a été introduit dans la souche sauvage de *S. venezuelae* par conjugaison avec *E. coli*. Les exconjugants ont été sélectionnés pour la résistance à l'apramycine ; leur ADN génomique extrait et l'intégration du plasmide confirmée par PCR. Des cultures de 50 ml de milieu R2-S ont étéensemencées avec $1,4 \times 10^8$ spores de *S. venezuelae* WT/pJWe07. Les cultures ont été réalisées à 28°C dans des erlenmeyers à baffles de 500 ml sous agitation (180 rpm). Au bout de 96h d'incubation, les cultures ont été centrifugées. Le surnageant a alors été acidifié avec l'acide trifluoroacétique à 2% final puis analysé par LC-MS ; le culot quant à lui a été repris dans 3 ml d'eau et cassé à la presse d'Eaton. Après centrifugation, le surnageant de cassage a été récupéré, additionné d'acide trifluoroacétique à 2% final puis analysé par LC-MS.

Expression hétérologue des *clusters Sven_1864-5* et *Rv2275-6* chez *S. coelicolor* M1154

Les vecteurs pJWe07 et pJWe19 ont été introduits dans la souche *S. coelicolor* M1154 par conjugaison *E. coli*/*Streptomyces*. Les exconjugants ont

été sélectionnés pour la résistance à l'apramycine ; leur ADN génomique extrait et l'intégration du plasmide confirmée par PCR.

Des cultures de 25 ml de milieu MP5 ont été ensemencées avec une suspension de spores et incubées 7 jours à 28°C sous agitation à 180 rpm dans des erlenmeyers à baffles de 500 ml. Au bout de 7 jours d'incubation, les cultures ont été centrifugées. Le surnageant a alors été acidifié avec l'acide trifluoroacétique à 2% final puis analysé par LC-MS.

Complémentation des mutants

Le vecteur pJWe05 a été utilisé pour compléter les mutants de délétion du gène *Sven_1864* ainsi que le mutant de délétion du *cluster* entier. Le vecteur pJWe06 a été utilisé pour compléter le mutant du gène *Sven_1865*. Le vecteur pJWe07 a été utilisé pour compléter le mutant de délétion du *cluster* entier. Tous ces vecteurs ont été insérés dans les souches respectives par conjugaison entre *E. coli* et *Streptomyces*. Les exconjugants ont été sélectionnés pour la résistance à l'apramycine ; leur ADN génomique extrait et l'intégration du plasmide confirmée par PCR.

Analyses LC-MS/MS

La chromatographie en phase liquide couplée à la spectrométrie de masse a été réalisée sur l'appareil Agilent 1100 HPLC avec une colonne Atlantis dC18 3µm, 4.6x150 mm (Waters) couplée au spectromètre de masse à piège à ions Esquire HTC (Bruker). L'ionisation a été effectuée par électro-ébulisaison en mode d'ionisation positive. Le flux était de 600 µl/min. Deux méthodes ont été utilisées : la méthode 1 où l'élution a été effectuée pendant 10 min avec 100% de la solution A (0,1% d'acide formique), puis un gradient croissant de la solution B (0,1% d'acide formique, 90% d'acétonitrile) a été appliqué pour atteindre 95% de la solution B au bout de 43 min ; la méthode 2 où l'élution a été effectuée pendant 5 min avec 100% de la solution A, puis un gradient croissant de la solution B a été appliqué pour atteindre 50% de B au bout de 55 minutes.

Analyse phénotypique des mutants

Le phénotype des souches de *S. venezuelae* WT, WT/pJWe07, Δ Sven_1864, Δ Sven_1864/pJWe05, Δ Sven_1865, Δ Sven_1865/pJWe06, Δ Sven_1864-5, Δ Sven_1864-5/pJWe05 et Δ Sven_1864-5/pJWe07 a été analysé sur quinze milieux solides : R2YE, HT, LB, SFM, BHI, 5006, R2M, TSB, ISP5, GPYE, GYM, Gruau, DAM5, SSG et MP5. Dans ce but un disque de cellophane (Focus Packaging & Design Ltd) a été déposé à la surface de boîtes de Petri contenant un milieu de culture solide. 5 μ l d'une suspension de spores à 10^6 spores/ μ l de la souche à caractériser ont été déposés au centre de la boîte. Les boîtes ont été incubées à 28°C pendant 72 h. Après ce temps la morphologie de la colonie a été évaluée. Le disque de cellophane a ensuite été enlevé, et 3 ml de milieu SNA additionné de la culture liquide en LB de *Micrococcus luteus* (à raison de 0,04 unités de DO₆₀₀ par ml) a été distribué de façon homogène à la surface des boîtes de Petri. Les boîtes ont alors été incubées à 37°C pendant 24 h et le diamètre du halo d'inhibition de croissance de la souche indicatrice a été mesuré. L'expérience a été réalisée en triplicat.

Résultats

Organisation génétique du *cluster* codant la SvCDPS et le SvCYP121 chez *S. venezuelae*

Comme nous pouvons l'observer sur figure III.2, l'environnement génétique du *cluster* CDPS-CYP121 de *M. tuberculosis* et *S. venezuelae* n'est pas conservé, ce qui suggère son acquisition par transmission horizontale chez l'une ou les deux espèces.

Notamment, la Gln385 et l'Arg386, deux des quatre acides aminés impliqués dans la formation du réseau de liaisons hydrogène autour du substrat, sont respectivement remplacées par une Met et une Lys. Parmi les sept résidus créant des liaisons van der Waals avec le substrat, trois sont modifiés (Met62 en Leu, Val78 en Ile et Phe280 en Ile).

Par conséquent, nous avons testé si SvCYP121 était capable de transformer le cYY en mycocyclosine malgré ces différences mises en évidence au niveau du site catalytique. Le cytochrome SvCYP121 a été exprimé de façon hétérologue chez *E. coli* et purifié.

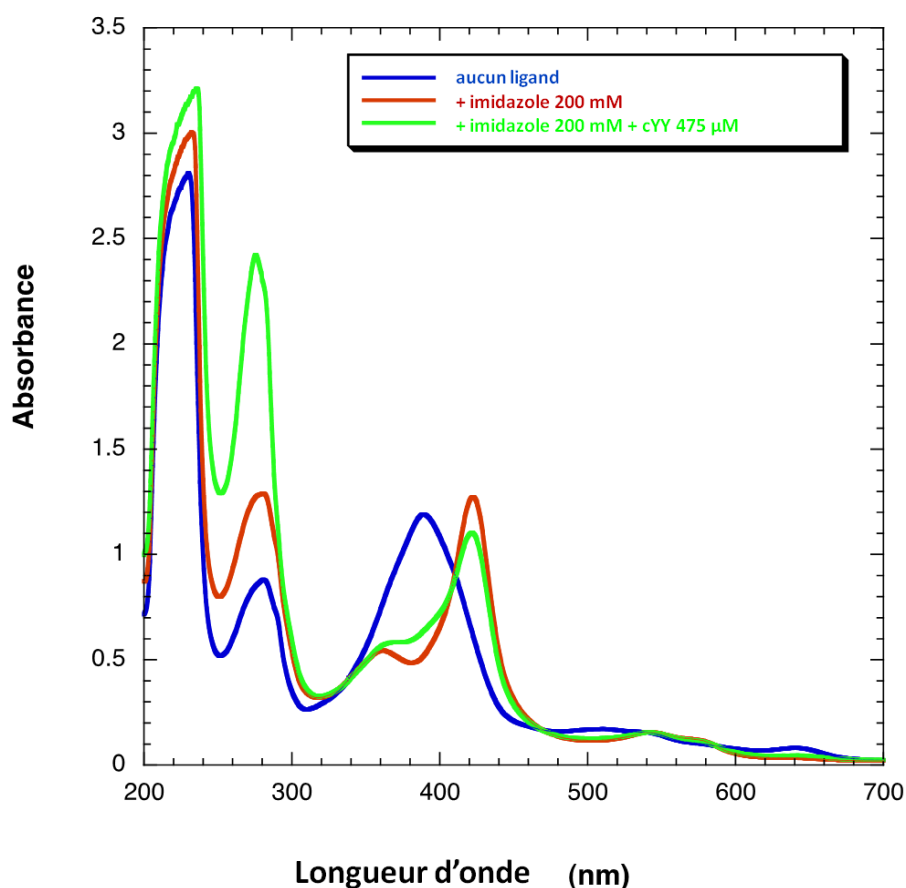


Figure III.4. Propriétés spectrales du cytochrome SvCYP121 seul, en présence d'imidazole et en présence d'imidazole et de cYY.

De manière surprenante, il se purifie sous forme high spin (bande de Soret à 389 nm). Le « spin » d'un cytochrome correspond à son état d'activation et se traduit par un pic d'absorbance à une longueur d'onde précise, appelé la bande de Soret. Les cytochromes P450 en général se

purifient sous forme low spin inactive et la transition vers la forme haut spin active se fait après la fixation du substrat. Cela permet d'éviter l'activité du cytochrome « à vide » qui pourrait engendrer la formation d'espèces réactives de l'oxygène. Dans le cas de SvCYP121, l'ajout d'imidazole permet de le déplacer vers la forme low spin (bande de Soret à 422 nm). L'ajout du substrat cYY permet d'observer un retour partiel vers la forme high spin ce qui indique que le cYY est un substrat naturel de l'enzyme (figure III.4). Un test d'activité enzymatique a ensuite été réalisé *in vitro*. L'analyse LC-MS du milieu réactionnel en utilisant la méthode 2 a permis de détecter dans les réactions incubées pendant 2h l'apparition d'un nouveau pic dont la valeur m/z (325) ainsi que la fragmentation sont cohérentes avec la fragmentation de la mycocyclusine, décrites précédemment par Belin et ses collaborateurs (2009) (figure III.5).

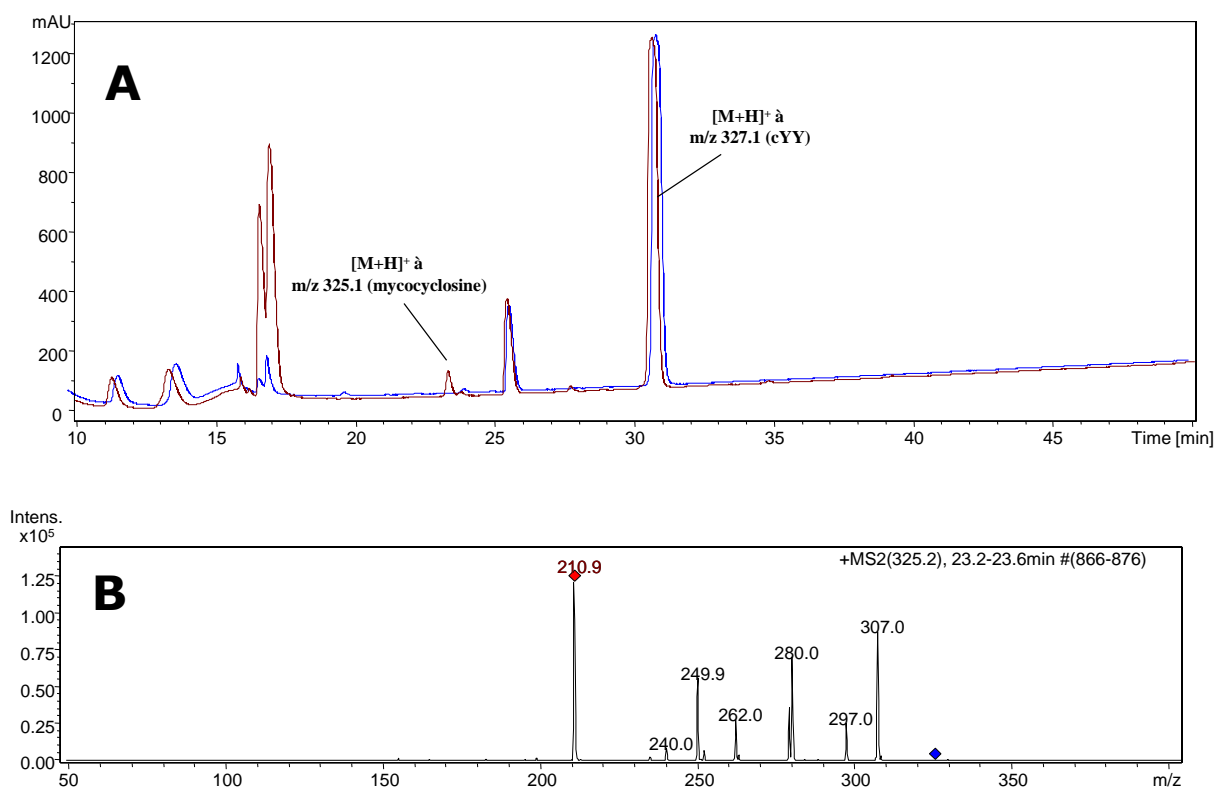


Figure III.5. Chromatogramme UV (214 nm) de la réaction *in vitro* du SvCYP121 sur cYY (A) et spectre de masse secondaire du pic correspondant à la mycocyclusine (B). Pour la figure A: en bleu – t=0, en rouge – t=120 min.

Ainsi, le cytochrome de *S. venezuelae* est capable de transformer *in vitro* le cYY en mycocyclosine comme cela avait été observé dans le cas de *M. tuberculosis* malgré la modification de certains résidus catalytiques.

Le cYY est détecté dans les surnageants de culture de *S. venezuelae* mais pas la mycocyclosine

Comme la *Su*CDPS catalyse la formation de cYY en expression hétérologue chez *E. coli* et comme le cytochrome *Su*CYP121 est capable de transformer le cYY en mycocyclosine *in vitro*, nous avons voulu tester la production de ce métabolite chez *S. venezuelae*, sachant qu'à notre connaissance la production *in vivo* de la mycocyclosine n'a jamais été mise en évidence.

Pour cela, le surnageant de culture de la souche sauvage de *S. venezuelae* en milieu R2-S a été analysé par LC-MS (méthode 1).

Un pic très faible au temps de rétention de 41 min (temps de rétention correspondant à celui du cYY dans ces conditions) a été observé sur le chromatogramme, suggérant la formation de cYY (figure III.6A). L'analyse LC-MS de ce pic a ensuite permis de confirmer qu'il s'agit bien de cYY : rapport m/z de 327,3 correspondant à celui du cYY dans ces conditions ; spectre de fragmentation secondaire de ce pic caractéristique de celui du cYY (figure III.6B), confirmant la production de cYY. L'analyse des courants ioniques extraits pour $m/z=327,3$ montre que l'on peut observer la formation de cYY à partir de 48h de culture et qu'à partir de 96h sa concentration n'évolue plus (figure III.6C). En revanche, la recherche spécifique d'un courant ionique extrait dont le rapport $m/z=325$ correspondrait à celui de la mycocyclosine n'a pas abouti à des résultats positifs.

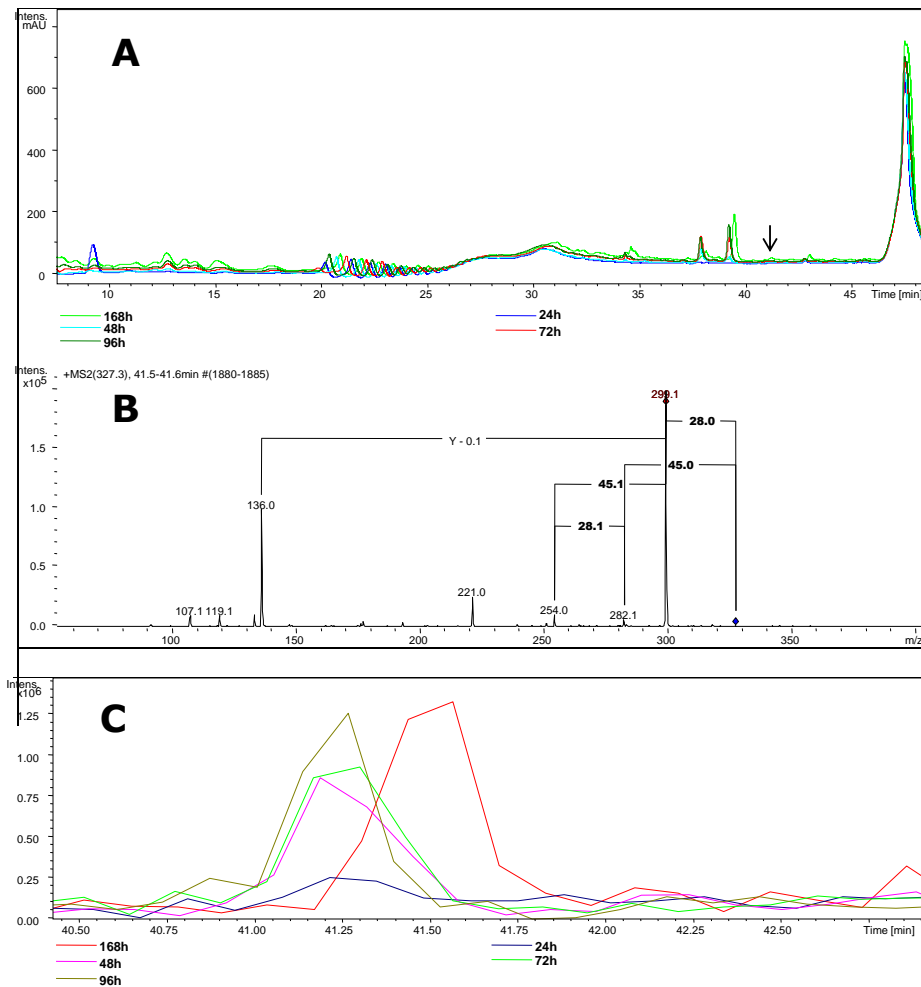


Figure III.6. Résultats de l'analyse par LC-MS du surnageant de culture de *S. venezuelae* WT. **A** – Chromatogramme UV (214 nm) du surnageant de culture en R2-S. Le pic correspondant au cYY est marqué par une flèche. **B** – Spectre de masse secondaire du pic à 41 min. La fragmentation correspond à celle du cYY. **C** – courants ioniques extraits pour la $m/z=327,3$ correspondant à cYY.

Dans les conditions de production du cYY, les gènes du *cluster* sont exprimés et forment un opéron

Comme nous avons détecté du cYY, mais pas de la mycocyclosine, nous nous sommes demandé si le gène codant le cytochrome *Su*CYP121 était exprimé dans nos conditions de culture. Ainsi, une analyse de l'expression des gènes *Sven_1864* et *Sven_1865* a été réalisée par RT-PCR sur des ARN extraits de la souche sauvage de *S. venezuelae* après 72h de culture en milieu R2-S (figure III.7).

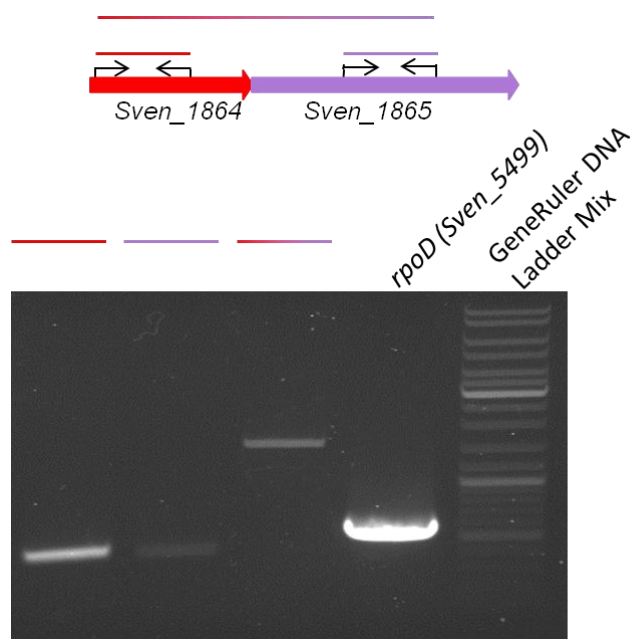


Figure III.7. RT-PCR sur les extraits d'ARN de la souche sauvage de *S. venezuelae* au bout de 72h de culture en milieu R2-S. La position des amorces est indiquée schématiquement en haut. Le transcrit du gène *rpoD* a été utilisé en tant que contrôle positif.

On observe l'expression du gène *Sven_1864*, un signal plus faible, mais visible pour le gène *Sven_1865* et on peut également observer la présence d'un amplicon correspondant aux deux gènes, ce qui démontre l'existence d'un ARN messager dicistronique. Ainsi, les gènes *Sven_1864-5* forment un opéron et sont co-transcrits.

La surexpression du *cluster Sven_1864-5* ne permet pas d'observer la mycocyclosine

Nous avons observé que le gène *Sven_1864* était exprimé. Le signal du gène *Sven_1865* est beaucoup plus faible, et ce n'est pas un artefact technique puisque le changement de couple d'amorces donne exactement le même résultat (données non montrées). Ainsi, nous pouvons émettre l'hypothèse que la protéine *Su*CYP121 est présente en trop faible quantité pour que l'on puisse observer son activité (surtout que son substrat, le cYY, est déjà présent en très faible quantité). Ainsi, nous avons décidé de surexprimer le *cluster*. Le *cluster Sven_1864-5* a été cloné sous contrôle du promoteur fort et constitutif *ermE**_p dans le plasmide intégratif pOSV666 puis intégré dans le génome de la souche sauvage de *S. venezuelae*. Au bout de 96h, le surnageant de culture ainsi que le surnageant de cassage des cellules de la souche résultante *S. venezuelae* WT/pJWe07 ont été analysés en LC-MS. La présence de cYY est observée dans le surnageant de culture. Cependant on ne retrouve pas parmi les courants ioniques extraits, de m/z correspondant à la mycocyclosine (données non montrées). Le cYY et la mycocyclosine n'ont pu être trouvés dans le surnageant de cassage des cellules.

Expression hétérologue du *cluster* CDPS-CYP121 de *S. venezuelae* et *M. tuberculosis*

Le *cluster Sven_1864-5* de *S. venezuelae* a été exprimé chez *S. coelicolor* M1154 sous contrôle du promoteur *ermE**_p en utilisant le plasmide intégratif pJWe07. Au bout de 7 jours de culture, l'analyse LC-MS du surnageant de culture en utilisant la méthode 2 permet de détecter des quantités abondantes de cYY, alors que la souche *S. coelicolor* M1154 contenant le vecteur vide pOSV666 n'en produit pas. La recherche spécifique de la m/z=325 en courants ioniques extraits en mode positif a permis de détecter une molécule qui pourrait être la mycocyclosine, le temps de rétention étant cohérent avec les données précédentes. Cependant le signal sortant à peine du bruit de fond, aucune fragmentation n'a pu être obtenue (figure III.8).

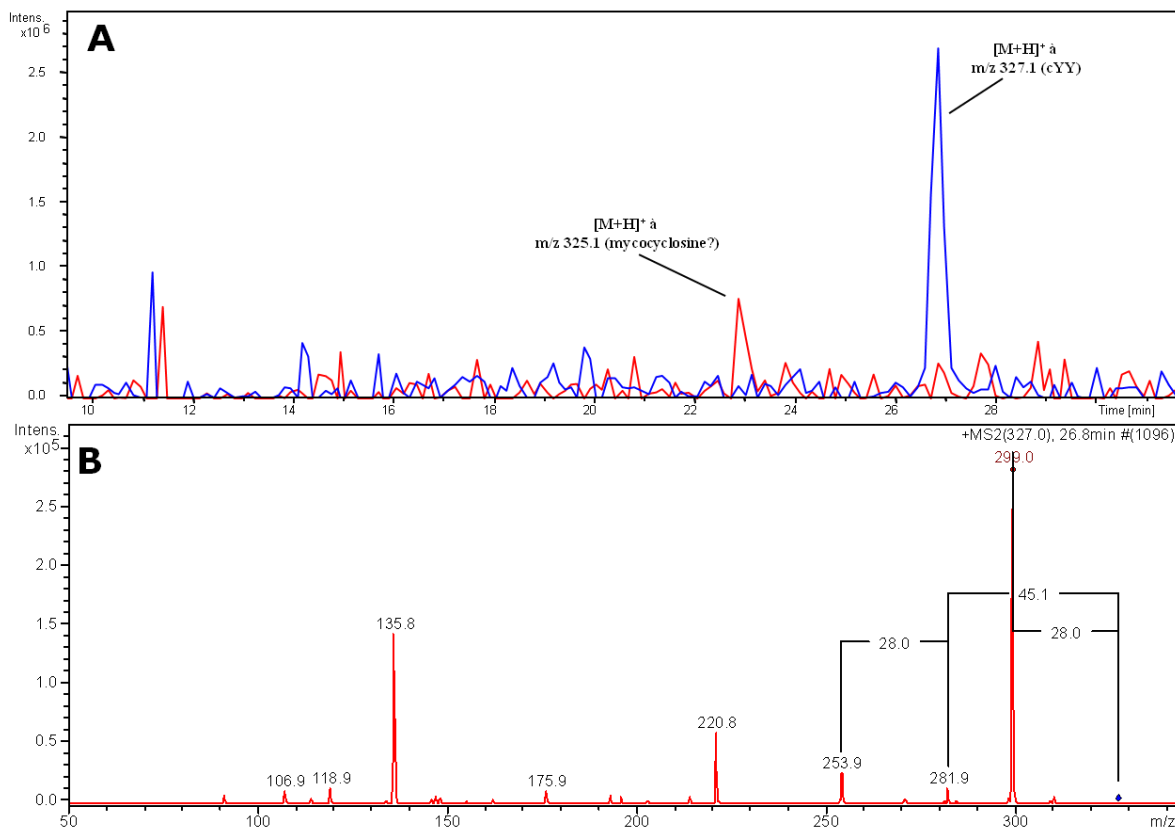


Figure III.8. Courants ioniques extraits de l'analyse LC-MS du surnageant de culture de *S. coelicolor* M1154/pJWe07 (exprimant le *cluster* *Sven_1864-5* sous contrôle de *ermE_p) (A) et le spectre de masse secondaire du pic correspondant au cYY (B). Pour la figure A: en bleu – courants ioniques extraits pour m/z=327 en mode positif (recherche spécifique du cYY), en rouge – courants ioniques extraits pour m/z=325 en mode positif (recherche spécifique de la mycrocyclisine).**

L'utilisation de la même technique d'expression pour le *cluster* *Rv2275-6* de *M. tuberculosis* a permis de confirmer la biosynthèse de cYY. Cependant aucun signal pouvant correspondre à de la mycrocyclisine n'a pu être détecté (données non montrées).

Aucun phénotype n'a pu être associé à la délétion ou surexpression des gènes *Sven_1864-5*

Afin d'étudier le rôle biologique du cYY ou des métabolites éventuellement produits à partir du cYY, la délétion en phase de la majeure partie de la séquence codante de chacun des gènes du *cluster* ainsi que celle du *cluster* complet a été réalisée par la technique de recombinaison

homologue utilisant une cassette de résistance à l'apramycine excisable. L'obtention des mutants a été facile, ce qui montre que le gène *Sven_1865* codant *SvCYP121* n'est pas essentiel, contrairement à son homologue chez *M. tuberculosis*. Les mutants de délétion ont aussi été complétés avec des vecteurs intégratifs exprimant les gènes appropriés sous contrôle du promoteur constitutif *ermE**_p. Une souche délétée pour l'opéron entier surexprimant uniquement la CDPS a également été construite. Cette étude a également inclus la souche sauvage surexprimant l'opéron. Pour toutes ces souches, la morphologie des colonies et la production de composés antibactériens actifs sur *M. luteus* ont été analysées sur quinze milieux solides de culture. Aucune différence notable entre toutes ces souches n'a pu être observée tant pour la morphologie des colonies que pour l'activité antibiotique (données non montrées).

Discussion

Le présent travail a démontré que le mutant délété pour le gène homologue au gène codant le cytochrome P450 CYP121 de *M. tuberculosis* est viable. De plus, aucun phénotype n'a pu être associé à cette délétion, alors que chez *M. tuberculosis* le gène codant CYP121 est essentiel (McLean *et al.* 2008). Bien que *SvCYP121* présente *in vitro* une activité identique à celle de *MtCYP121*, nous n'avons pas été non plus en mesure de démontrer formellement la production de la mycocyclosine (ni aucune autre forme de cYY transformée, comme cYY mono- ou dihydroxylés) *in vivo*, que ce soit dans la souche native ou en expression hétérologue. La conclusion primaire de ce travail est donc que *S. venezuelae* n'est pas un bon modèle pour l'étude du *cluster* mycocyclosine.

En effet, malgré les efforts, le présent travail n'a pas permis d'assigner un rôle biologique aux produits du *cluster* cyclodityrosine synthase – mycocyclosine synthase. La production de la mycocyclosine *in vivo* reste toujours une hypothèse non confirmée. Jamais trouvée dans les surnageants de culture des organismes producteurs, la mycocyclosine pourrait éventuellement être présente en état de traces lorsque le *cluster* de *S.*

venezuelae est exprimé de façon hétérologue. Afin d'avoir la confirmation que la molécule présentant la valeur $m/z=325$ et le temps de rétention de 23 min est effectivement de la mycocyclosine, il faudrait soit optimiser les conditions de culture afin que la souche produise plus de la molécule d'intérêt, soit développer un procédé de purification et concentration de ladite molécule à partir des surnageants de culture.

Les traces éventuelles de la mycocyclosine n'ont pas été détectées dans le cas d'expression hétérologue du *cluster* de *M. tuberculosis*. L'absence de mycocyclosine (due à l'absence de l'activité du cytochrome *MtCYP121*) pourrait éventuellement s'expliquer par la présence d'un site de fixation des ribosomes (RBS) de *M. tuberculosis* en amont du gène codant le *MtCYP121*, non reconnu par l'espèce de *Streptomyces*. La construction d'un vecteur exprimant les gènes à partir de deux promoteurs indépendants avec un RBS connu pour être efficace chez l'hôte hétérologue devrait permettre de trancher. Ceci étant dit, une expérience de ce type a été tentée pour le *cluster* de *S. venezuelae* sans pour autant donner des résultats positifs (Witwinowski *et al.* données non publiées). Une autre possibilité existe : le *MtCYP121* pourrait ne pas reconnaître les partenaires redox présents chez le hôte hétérologue.

Le cytochrome *SvCYP121* montre une activité *in vitro*. Elle est moins importante que pour *MtCYP121* (Belin *et al.* 2009), mais cela ne doit pas forcément indiquer une moindre activité intrinsèque, mais peut suggérer une interaction moins bonne avec les partenaires redox de l'épinard utilisés dans l'étude. Même si le fait que le cytochrome *SvCYP121* soit purifié sous la forme active (high spin), alors que typiquement les cytochromes sont purifiés sous la forme inactive (low spin) et sont activés par la liaison du substrat, est étonnant, ce n'est pas une situation sans précédent. C'est notamment le cas du CYP1A2 humain (Sandhu *et al.* 1994). Le fait qu'en présence de l'imidazole (basculant le *SvCYP121* en forme à low spin) le rajout de cYY ne permette pas de rebasculer complètement dans la forme high spin s'explique par la faible concentration du substrat, due à sa faible solubilité dans l'eau.

Finalement, on peut aussi émettre l'hypothèse que la mycocyclosine n'ait pas été détectée dans la majorité des expériences présentées dans ce

travail parce qu'elle était présente en quantité bien inférieure au seuil de détection. En effet, dans la seule expérience où un signal pouvant potentiellement correspondre à la mycocyclosine a été détecté, ledit signal sortait à peine du bruit du fond.

Concernant le fait que le signal correspondant à l'expression du gène *Sven_1865* en RT-PCR soit plus faible que le signal correspondant à l'expression du gène *Sven_1864*, alors qu'il semble que les gènes forment un opéron, on peut supposer une éventuelle dégradation de l'ARN messenger du côté 3'. Il serait intéressant de s'assurer par Western Blot que la protéine *SvCYP121* est effectivement présente dans les cellules.

Pour rendre les choses plus complexes et encore diminuer notre compréhension du fonctionnement de ce groupe de gènes, le *cluster* homologue de *Streptomyces sp.* TN58 en expression hétérologue mène à l'apparition de cyclo(L-tyrosyl-L-DOPA), alors que le CYP121 présente exactement les mêmes changements au niveau du site catalytique par rapport à celui de *M. tuberculosis* que celui de *S. venezuelae*. En étude *in vitro* il n'a pas été possible de détecter une activité, et l'analyse des surnageants de culture de l'organisme producteur n'a permis de trouver ni mycocyclosine ni cyclo(L-tyrosyl-L-DOPA) (Najah 2017). Afin de résoudre ce mystère, il faudrait réaliser des études structurales des deux cytochromes, afin de vérifier si les résidus catalytiques, inférés à partir de l'alignement des séquences *in silico*, sont effectivement positionnés de la même manière par rapport au substrat.

Au moment du début de l'étude, la présence du *cluster* composé des gènes codant pour la cyclodityrosyl synthase et le cytochrome P450 CYP121 était connue uniquement chez les espèces du complexe *M. tuberculosis* (*Mycobacterium bovis*, *Mycobacterium cannetti*, *M. tuberculosis*) et chez *S. venezuelae* (Belin *et al.* 2012). Cependant au moment où elle a touché à sa fin, les analyses bioinformatiques permettaient de trouver plus de cent quarante occurrences non redondantes de ce *cluster* parmi les actinobactéries appartenant à au moins six genres différents (*Actinomadura*, *Kitasatospora*, *Mycobacterium*, *Nocardiopsis*, *Streptacidiphilus* et *Streptomyces*). Le fait que l'environnement génétique autour du *cluster* ne

soit pas conservé, suggère que le *cluster* a été l'objet de nombreux transferts horizontaux. Vu la fréquence avec laquelle le *cluster* cyclodityrosine synthase – mycocyclosine synthase est retrouvé chez les actinobactéries, notre incapacité de lui assigner un rôle biologique *in vivo* est d'autant plus décevante.

Chapitre IV

Etude de la voie de biosynthèse de la bicyclomycine

Introduction

Deux raisons majeures justifient que nous ayons entrepris l'étude de la voie de biosynthèse de la bicyclomycine, une dicétopipérazine dérivée de cyclo-(L-isoleucyl-L-leucyl). La première raison est la relative complexité chimique de cette molécule (fig. I.10A) où le noyau DKP est fortement modifié avec une hydroxylation sur le carbone alpha de l'isoleucine et un pont éther formé entre le carbone distal de l'isoleucine et le carbone alpha de la leucine. De plus la chaîne latérale de l'isoleucine subit une déshydrogénation et celle de la leucine trois hydroxylation. Il était donc intéressant d'identifier les enzymes catalysant ces réactions et de comprendre comment des modifications aussi étendues peuvent avoir lieu.

La deuxième raison tient au fait que la bicyclomycine est une molécule utilisée en médecine comme antibiotique. Elle est la seule molécule connue pour inhiber le facteur de terminaison de la transcription Rho et des travaux récents suggèrent qu'elle peut avoir une activité bactéricide très prononcée quand elle est utilisée en synergie avec d'autres antibiotiques (Malik *et al.* 2014). Il s'agit d'une molécule qui connaît un regain d'intérêt pour son utilisation comme agent antibactérien. Connaître les gènes et la voie de biosynthèse de cette molécule peut permettre d'identifier des micro-organismes produisant des molécules proches et d'envisager la production de dérivés par voie biologique.

Dans l'équipe Microbiologie Moléculaire des Actinomycètes, Corinne POISIER, Laëtitia PHILIPPE, Sylvie LAUTRU, Antonio C. RUZZINI, Matthieu COUPET (un étudiant de M2 que j'ai encadré), ma co-encadrante et mon directeur de thèse ont été impliqués dans cette étude. Elle a été réalisée en collaboration avec l'équipe de Muriel GONDRY, en impliquant, outre Muriel, Mireille MOUTIEZ, Pascal BELIN, Emmanuel FAVRY et Jérôme SEGUIN et avec celle d'Olivier LEQUIN, en impliquant aussi Isabelle CORREIA et Lucrèce MATHERON.

Les résultats obtenus sont présentés ci-dessous sous forme de manuscrit en anglais, dans l'état actuel de l'avancement de la rédaction. Pour faciliter la lecture, les figures et les tableaux ont été insérés dans le

texte. Les tableaux trop volumineux pour être incorporés dans ce manuscrit ont conservé leur appellation de « Supplementary Table » et peuvent être téléchargés à l'adresse suivante : https://drive.google.com/open?id=0B49pyNXrf_9UVkdsakpSWU82NXM. L'archive à télécharger contient aussi le fichier Genbank du cluster de biosynthèse de bicyclomycine.

Bicyclomycin biosynthesis: multiple modifications of a cyclodipeptide synthesized in a tRNA-dependent CDPS manner.

Introduction

2,5-Diketopiperazines (DKP) are a class of molecules with various interesting biological properties, like antibiotic, antifungal, antiviral, antimetabolic, and enzyme inhibitors. Therefore DKPs, either of natural origin or obtained by chemical synthesis, are widely used in medicine, to treat various ailments as bacterial infections, erectile dysfunction or depression (Borthwick, 2012).

In this study we focus on the biosynthesis of one medically relevant diketopiperazine that is used in medicine as antibiotics: bicyclomycin (also known as aizumycin or under the trade name bicozamycin, Figure IV.1C). This molecule is particularly active against Gram-negative bacteria, and presents an original mechanism of action, as it is the only known specific inhibitor of the Rho termination factor (Zwiefka *et al.* 1993). Bicyclomycin does not impair the ATP or mRNA fixation by the Rho protein, but, putting itself near to the ATP fixation site, it blocks ATP hydrolysis and subsequent transcription termination (Ma *et al.*, 2016).

Bicyclomycin was discovered as a product of *Streptomyces cinnamoneus* (= *Streptomyces sapporonensis*) (Miyoshi *et al.*, 1972) or *Streptomyces aizunensis* (Miyamura *et al.* 1973). Recently this antibiotic was the object of a renewed interest as it was demonstrated that its concomitant use with other bacteriostatic molecules like tetracycline, chloramphenicol or rifampicin dramatically increases its bactericidal properties (Malik *et al.* 2014). At the beginning of this study it was known that the precursor of bicyclomycin is cyclo-(L-isoleucyl-L-leucyl) (cIL, Figure IV.1A) (Bradley *et al.* 1996). cIL is then heavily modified to generate bicyclomycin. It undergoes three hydroxylations of the leucine side chain, a hydroxylation of the diketopiperazine ring, an ether bridge formation between the diketopiperazine ring and the isoleucine side chain and one dehydrogenation

in the isoleucine moiety. Bradley and co-workers (1996) also remarked that the last precursor of the pathway was dihydrobicyclomycin (Figure IV.1B).

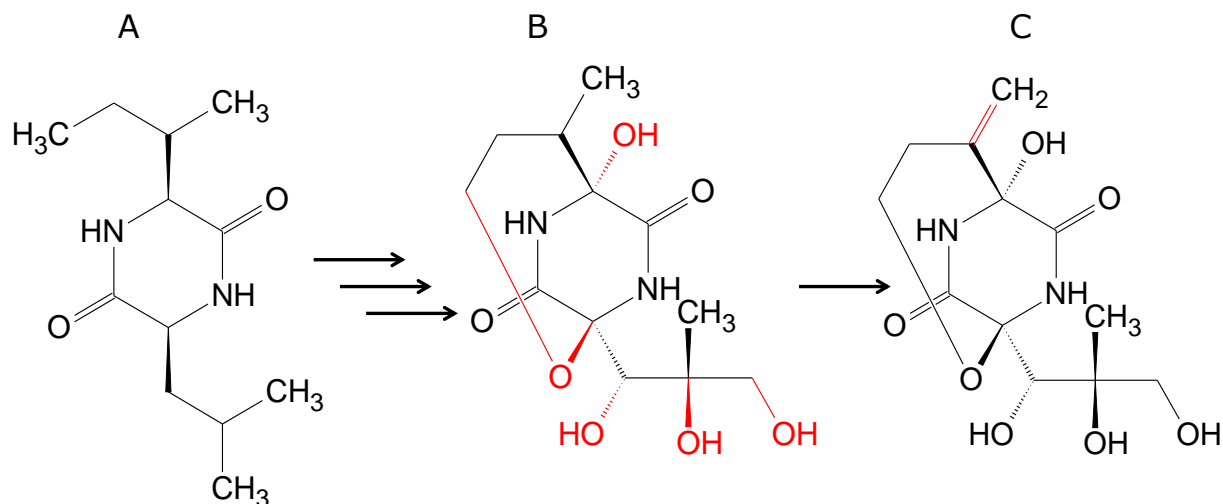


Figure IV.1. Bicyclomycin biosynthetic pathway according to Bradley *et al.* (1996). **A** – cyclo-(L-isoleucyl-L-leucyl), **B** – dihydrobicyclomycin, **C** – bicyclomycin. The modifications are highlighted in red.

The diketopiperazine core may be synthesized either by a Non-Ribosomal Peptide Synthetase (NRPS), directly incorporating proteinogenic or non-proteinogenic amino acids, or by a Cyclodipeptide Synthase (CDPS), a class of relatively small enzymes presenting structural homologies to class I aminoacyl-tRNA synthetases, particularly the tyrosyl-tRNA_{Tyr}-synthetase, using aminoacylated tRNA as substrate (Gondry *et al.* 2009; Gu *et al.*, 2013; Belin *et al.*, 2012). CDPS use a ping-pong catalytic mechanism, with the formation of aminoacyl-enzyme and dipeptidyl-enzyme covalent intermediates. Two pockets are responsible for substrate recognition, P1 and P2 (for a recent review, see Moutiez *et al.* 2017).

Hereby we demonstrate that bicyclomycin biosynthesis is directed by a gene cluster encoding a CDPS, six oxidoreductases (of which five are oxoglutarate/iron-dependent dioxygenases and one is a cytochrome P450 monooxygenase) and a protein belonging to Major Facilitator Superfamily 1, potentially acting as a transporter. We propose a complete biosynthetic pathway for bicyclomycin, the first known to date involving a CDPS and multiple oxidoreductases. Finally, we would like to emphasize that this is

also the first known case of a medically relevant molecule to be synthesized by a CDPS-dependent pathway.

Material and methods

Bacterial strains, culture conditions, strain manipulation and plasmids

Strain	Genotype	Plasmid	Source
<i>Escherichia coli</i> DH5α	<i>F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 ϕ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17($r_K^- m_K^+$), λ^-</i>	-	Promega
<i>Escherichia coli</i> BW25113pIJ790	<i>F- lacI⁺rrnB_{T14} ΔlacZ_{WJ16} hsdR514 ΔaraBAD_{AH33} ΔrhaBAD_{LD78} rph-1 Δ(araB-D)567 Δ(rhaD-B)568 ΔlacZ4787(::rrnB-3) hsdR514 rph-1</i>	pIJ790	Gust <i>et al.</i> 2003
<i>Escherichia coli</i> GM119	<i>F- dam-3 dcm-6 metBI galK2 galT22 lacY1 tsx-78 supE44 (thi-1 tonA31 mtl-I)?</i>	-	Arraj and Marinus 1983
<i>Escherichia coli</i> ET12567pUZ8002	<i>F- dam13::Tn9 dcm-6 hsdM hsdR zjj-202::Tn10 recF143 galK2 galT22 ara-14 lacY1 xyl-5 leuB6 thi-1 tonA31 rpsL136 hisG4 tsx-78 mtl-1 glnV44</i>	pUZ8002	Flett <i>et al.</i> 1997
<i>Escherichia coli</i> BL21-AI	<i>F^{ompT} hsdS_B ($r_B^- m_B^-$) gal dcm araB::T7RNAP-tetA</i>	-	Thermo Fisher Scientific
<i>Streptomyces aizunensis</i> ATCC 2196	WT	-	ATTC
<i>Streptomyces aizunensis</i> ΔbcmA	<i>ΔbcmA::Ωaac4</i>	-	This study
<i>Streptomyces aizunensis</i> Δbcm	<i>ΔbcmH-bcmG::aph2</i>	-	This study

Table IV.1 Bacterial strains used in this study

Bacterial strains are listed in Table IV.1. *Escherichia coli* strains were manipulated using standard laboratory procedures (Green and Sambrook 2012). Electrocompetent cells were prepared using the protocol established by Robey *et al.* (1996) unless otherwise stated. When needed, antibiotics were added to *E. coli* cultures at concentrations as follow: ampicillin: 100 mg/l, apramycin: 50 (liquid media) or 75 (solid media) mg/l, chloramphenicol: 30 mg/l, kanamycin: 25 mg/l, hygromycin: 50 (liquid media) or 100 (solid media) mg/l (cultures containing hygromycin were supplemented with 1 mM of IPTG, as in *E. coli* constructions the hygromycin resistance gene is expressed from the *lac* promoter), spectinomycin: 100 mg/l. *S. aizunensis* was manipulated using standard laboratory procedures (Kieser *et al.* 2000). Spores were prepared on Gruau medium (oatmeal flour: 30 g/l, glucose: 5 g/l, yeast extract: 5 g/l, pH adjusted to 7.2 with NaOH, agar: 20 g/l). Conjugations between *E. coli* ET12567pUZ8002

and *S. aizunensis* were carried by the modified Kieser *et al.* (2000) protocol, using the SFM medium complemented with 10 mM MgCl₂ and an antibiotic 3 ml overlay of SNA medium containing fosfomycin instead of nalidixic acid. Exconjugants were patched on SFM containing the selection antibiotic(s), fosfomycin and ampicillin. Antibiotic concentrations used with *Streptomyces* were as follow: ampicillin: 30 mg/l, apramycin 30 mg/l, fosfomycin 50 mg/l, kanamycin: 25 mg/l, hygromycin: 50 mg/l. All plasmids used in this study are listed in Table IV.2

DNA manipulations

PCR reactions for cloning applications were carried out using Phusion (Thermo Fischer Scientific) or Q5 (New England Biolabs) high fidelity polymerases according to manufacturers' protocol for GC-rich DNA, reaction mixtures being supplemented with 1 M of betaine. PCR reactions for verification of constructs were carried out using DreamTaq DNA Polymerase (Thermo Fischer Scientific) supplemented with 1x Q Solution (Qiagen). All primers are listed in Table IV.3. Primers were provided by either Integrated DNA Technologies or Eurofins. PCR products were purified using NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). Plasmid isolation from *E. coli* was performed using NucleoSpin Plasmid kit (Macherey-Nagel). Restriction enzymes and DNA-modifying enzymes (Klenow fragment, Shrimp Alkaline Phosphatase, Polynucleotide Kinase) were supplied by Thermo Fischer Scientific. Vector DNA was systematically dephosphorylated (with the exception of pGEM-T Easy). Digestion products were isolated on agarose gels and purified with NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel). Ligations were performed using T4 DNA Ligase (either Thermo Fischer Scientific or Promega). Chemocompetent *E. coli* DH5α cells were used for transformation by ligation products or plasmids unless stated otherwise. *Streptomyces* genomic DNA was extracted as described previously (Jacques *et al.* 2015), except for the DNA used for genome sequencing. DNA isolation for sequencing and the sequencing of *S. aizunensis* strain were carried out as described previously (Vingadassalon *et al.* 2015). Plasmid DNA was

Plasmid	Description	Source
pET15b	<i>E. coli</i> expression vector with N-terminal His-tag	Novagen / Merck-Millipore
pGEM-T Easy	<i>E. coli</i> cloning vector	Promega
pHP45Ω	Source of <i>ΩaadA</i> cassette	Prentki and Kirsch 1984
pIJ4070	Source of <i>ermE*</i> promoter	Bibb and Jansen 1986
pJDCyp	pET15b expressing <i>bcmD</i>	this study
pJDOG1	pET15b expressing <i>bcmB</i>	this study
pJDOG2	pET15b expressing <i>bcmC</i>	this study
pJDOG3	pET15b expressing <i>bcmE</i>	this study
pJDOG4	pET15b expressing <i>bcmF</i>	this study
pJDOG5	pET15b expressing <i>bcmG</i>	this study
pJWc9	pGEM-T Easy with JWm9/JWm10 PCR product cloned	this study
pJWc10	pGEM-T Easy with JWm11/JWm12 PCR product cloned	this study
pJWc13	pUC18 with a fragment of bicyclomycin cluster	this study
pJWc14	pUC18 with a fragment of bicyclomycin cluster	this study
pJWe10	pOSV666 with <i>bcmA</i> gene	this study
pJWe14.1	Bicyclomycin gene cluster cloned <i>HindIII</i> in pOSV668	this study
pJWe14.2	Bicyclomycin gene cluster cloned <i>HindIII</i> in pOSV668, differs from pJWe14.1.1 by the orientation of the cluster	this study
pJWe18	pJWe10 with a supplementary hygromycin resistance	this study
pJWe20	pOSV666 with the <i>bcmA</i> gene with its native promoter cloned <i>HindIII</i>	this study
pJWe21	pOSV806 with the <i>bcmH</i> gene with its native promoter cloned <i>Mph1103I/BcuI</i>	this study
pJWe22	pOSV806 with the <i>bcmA</i> gene with its native promoter cloned <i>Mph1103I/BcuI</i>	this study
pJWm04	Suicide vector for deletion of the <i>bcmA</i> gene in <i>Streptomyces aizunensis</i>	this study
pJWm07	Suicide vector for deletion of the whole bicyclomycin cluster in <i>Streptomyces aizunensis</i>	this study
pMC1	pJWe14.1 $\Delta bcmB::\Omega aadA$	this study
pMC2	pJWe14.1 $\Delta bcmB$ (in phase deletion)	this study
pMC3	pJWe14.1 $\Delta bcmC::\Omega aadA$	this study
pMC4	pJWe14.1 $\Delta bcmC$ (in phase deletion)	this study
pMC5	pJWe14.1 $\Delta bcmD::\Omega aadA$	this study
pMC6	pJWe14.1 $\Delta bcmD$ (in phase deletion)	this study
pMC7	pJWe14.1 $\Delta bcmE::\Omega aadA$	this study
pMC8	pJWe14.1 $\Delta bcmE$ (in phase deletion)	this study
pMC9	pJWe14.1 $\Delta bcmF::\Omega aadA$	this study
pMC10	pJWe14.1 $\Delta bcmF$ (in phase deletion)	this study
pMC11	pJWe14.1 $\Delta bcmG::\Omega aadA$	this study
pMC12	pJWe14.1 $\Delta bcmG$ (in phase deletion)	this study
pMC13	pJWe14.1 $\Delta bcmA::\Omega aadA$	this study
pMC14	pJWe14.1 $\Delta bcmA$ (in phase deletion)	this study
pMC15	pJWe14.1 $\Delta bcmH::\Omega aadA$	this study
pMC16	pJWe14.1 $\Delta bcmH$ (in phase deletion)	this study
pOSV215	Source of <i>Ωaac4</i> excisable cassette	Raynal <i>et al.</i> 2006
pOSV400	Suicide vector for gene deletion in <i>Streptomyces</i> by homologous recombination.	Vingadassalon <i>et al.</i> , unpublished data
pOSV408	pRT801 with $\Delta aac4::aph3$	Vingadassalon <i>et al.</i> 2015
pOSV666	pRT801 with <i>ermE*</i> _p , TipA RBS, <i>NdeI</i> , <i>PacI</i> and <i>HindIII</i> sites	this study
pOSV668	pRT801 with <i>ΩaadA</i> cassette from pHP45Ω cloned <i>BamHI</i>	this study
pOSV806	A ΦC31 integrative shuttle vector for gene expression in <i>Streptomyces</i> with hygromycin resistance	Aubry <i>et al.</i> , unpublished data
pRT801	<i>E. coli</i> - <i>Streptomyces</i> shuttle vector, replicative in <i>E. coli</i> , integrative (ΦBT1) in <i>Streptomyces</i>	Gregory <i>et al.</i> 2003
pUC18	<i>E. coli</i> cloning vector	Thermo Fischer Scientific

Table IV.2 Plasmids used in this study

Primer	Séquence	Description
JD1	GGGAATTCATATGACCATGTCCCGTGCACC	<i>bcmB</i> cloning into pET15b, adding <i>NdeI</i> site
JD2	CCGCTCGAGTCAGTGTCTCACGTCTGTTGT	<i>bcmB</i> cloning into pET15b, adding <i>XhoI</i> site
JD3	GGGAATTCATATGAGCACTGAGACGCTGCCG	<i>bcmC</i> cloning into pET15b, adding <i>NdeI</i> site
JD4	CCGCTCGAGTCATGCCCCTCACCCCGTCTT	<i>bcmC</i> cloning into pET15b, adding <i>XhoI</i> site
JD5	GGGAATTCATATGACCGCGCCCGCCACCC	<i>bcmD</i> cloning into pET15b, adding <i>NdeI</i> site
JD6	CCGCTCGAGTCAGTTCTGTCTCTTCGTGC	<i>bcmD</i> cloning into pET15b, adding <i>XhoI</i> site
JD7	GGGAATTCATATGGCGTCACCCGATCCGC	<i>bcmE</i> cloning into pET15b, adding <i>NdeI</i> site
JD8	CCGCTCGAGTCAGAGCGCACCGCTGTCTGT	<i>bcmE</i> cloning into pET15b, adding <i>XhoI</i> site
JD9	GGGAATTCATATGACGACGGTCGTCGACAA	<i>bcmF</i> cloning into pET15b, adding <i>NdeI</i> site
JD10	CCGCTCGAGTCAAATGCGGGAGTGGAACTCGTA	<i>bcmF</i> cloning into pET15b, adding <i>XhoI</i> site
JD11	GGGAATTCATATGAGCACGGCACAGGGATA	<i>bcmG</i> cloning into pET15b, adding <i>NdeI</i> site
JD12	CCGCTCGAGTCAGTACAGGCCGGCGGTGT	<i>bcmG</i> cloning into pET15b, adding <i>XhoI</i> site
JWc1	ATCTGCGAGTGTCCGTTTCGAGTGGCGGCTTGCGCCGATG	Cloning <i>ermE*</i> _p with adding the TipA RBS and <i>NdeI</i> , <i>PacI</i> and <i>HindIII</i> sites
JWc2	ATCAAGCTTTTAATTAACATATGTCCGCTCCCTTCTCTA AACTAGCATGCCGGTTCGACTCTAG	Cloning <i>ermE*</i> _p with adding the TipA RBS and <i>NdeI</i> , <i>PacI</i> and <i>HindIII</i> sites
JWc7	GGGAATTCATATGTCGCTAGAAGCGCAGCT	cloning the <i>bcmA</i> gene (adding the <i>NdeI</i> site)
JWc8	CCCAAGCTTTCATCGCCCTTTCAGGTGGA	cloning the <i>bcmA</i> gene (adding the <i>HindIII</i> site)
JWc9	CCCAAGCTTAGCAGCAGGAGACACGATGG	cloning the <i>bcm</i> cluster in fragments (adding <i>HindIII</i> site)
JWc10	AAGAACC CGCCGGAATTCAC	cloning the <i>bcm</i> cluster in fragments
JWc11	ACACCACTCCGCCTTCTTAC	cloning the <i>bcm</i> cluster in fragments
JWc12	TGAACGATCCGGGATACACC	cloning the <i>bcm</i> cluster in fragments
JWc13	CCGCGACTACCACGGCAAAC	cloning the <i>bcm</i> cluster in fragments
JWc14	GGGCGTCTCTCGATGTAGG	cloning the <i>bcm</i> cluster in fragments
JWc15	ACGTGATCCGGCGCTACTTC	cloning the <i>bcm</i> cluster in fragments
JWc16	CCTGTGCCGTGCTCATATGC	cloning the <i>bcm</i> cluster in fragments
JWc17	CGGCACCTTGCACGAGTACC	cloning the <i>bcm</i> cluster in fragments
JWc18	CCCAAGCTTCTTCGCCAAGCCACCTCC	cloning the <i>bcm</i> cluster in fragments (adding <i>HindIII</i> site)
JWc29	CCCAAGCTTAGGCGTCCAGTTGGATGCAG	cloning the <i>bcmA</i> gene with the native promoter (adding the <i>HindIII</i> site)
JWc30	CCAATGCATCACCCTCCGCTTCTTACC	<i>bcmH</i> cloning with its native promoter into pOSV806 adding the <i>Mph1103I</i> site
JWc31	GGACTAGTGGCTCACTCCGAGGAAACAG	<i>bcmH</i> cloning with its native promoter into pOSV806 adding the <i>BcuI</i> site
JWc32	CCAATGCATAGGCGTCCAGTTGGATGCAG	<i>bcmA</i> cloning with its native promoter into pOSV806 adding the <i>Mph1103I</i> site
JWc33	GGACTAGTTCATCGCCCTTTCAGGTGGA	<i>bcmA</i> cloning with its native promoter into pOSV806 adding the <i>BcuI</i> site
JWm9	AAGCTTGGCAACGGTGATGACCGTCT	upstream fragment for the <i>bcmA</i> deletion by homologous recombination adding <i>HindIII</i> site
JWm10	GATATCAGGCTCCATCAGCTGCGCTT	upstream fragment for the <i>bcmA</i> deletion by homologous recombination adding <i>Eco32I</i> site
JWm11	GATATCGTGGTGTCCCTCCACCTGAA	downstream fragment for the <i>bcmA</i> deletion by homologous recombination adding <i>Eco32I</i> site
JWm12	ACTAGTTGATGCCCGAGGCGCGGAAG	downstream fragment for the <i>bcmA</i> deletion by homologous recombination adding <i>BcuI</i> site
JWm19	GGATCTTACCTAGATCCTT	<i>aph2</i> amplification
JWm20	CCGGAATTCGGCAAGACCGATCCCCGGGG	<i>aph2</i> amplification adding <i>EcoRI</i> site
JWm23	CCCAAGCTTACCGGGACCCGTAAGTGAG	upstream fragment for the whole <i>bcm</i> cluster deletion by homologous recombination adding <i>HindIII</i> site
JWm24	CCGGAATTCAGACGCGGATCGGCCGCTCG	downstream fragment for the whole <i>bcm</i> cluster deletion by homologous recombination adding <i>EcoRI</i> site
JWm25	GACTAGTGGCGGGAGGTGCGGGTATG	downstream fragment for the whole <i>bcm</i> cluster deletion by homologous recombination

		adding <i>BcuI</i> site
JW-RT9	CGCTACAAGGCGGAGATAGG	Primers used for construction verification and sequencing
JW-RT10	TGAACGATCCGGGATACACC	Primers used for construction verification and sequencing
JW-RT12	TCGACAGACCGAAGTTGATG	Primers used for construction verification and sequencing
JWseq6	CCCTTTTTTGGCCTTGAAAT	Primers used for construction verification and sequencing
JWseq7	GAATTGTGAGCGGATAACAA	Primers used for construction verification and sequencing
MC1	ATCGATGCGAGGCAAGCTTATCGATG	<i>ΩaadA</i> cassette amplification adding <i>Bsu15I</i> restriction site
MC2	ATCGATTATCACGAGGCCCTTTCGTC	<i>ΩaadA</i> cassette amplification adding <i>Bsu15I</i> restriction site
MC3	TGAAAGGGCGATGACCATGTCCCCTGCACCCGGCAACA CGATCGATGCGAGGCAAGCTTA	<i>bcmB</i> deletion by PCR targeting
MC4	CGTGGTCGTCTCGTCGATGAGTTCCCGGGAGCCGCAG ACATCGATTATCACGAGGCCCT	<i>bcmB</i> deletion by PCR targeting
MC5	GACAGCGAGCTGAGCACTGAGACGCTGCGCCTCCAGA AGATCGATGCGAGGCAAGCTTA	<i>bcmC</i> deletion by PCR targeting
MC6	TCCCGAAGTCGGCGTAGCCGTCGCCACACGTCGTCG TTATCGATTATCACGAGGCCCT	<i>bcmC</i> deletion by PCR targeting
MC7	GGTGAGGGCATGACCGCGCCGCCACCCGCGGCT GTATCGATGCGAGGCAAGCTTA	<i>bcmD</i> deletion by PCR targeting
MC8	CGGGCCGGCGGGCCGGGGCATGGCGCGTGCCTCTTC AGATCGATTATCACGAGGCCCT	<i>bcmD</i> deletion by PCR targeting
MC9	AGCAGAAGTGTGGCGTACCCGATTCGCCACCCCTCC GGATCGATGCGAGGCAAGCTTA	<i>bcmE</i> deletion by PCR targeting
MC10	ACTCGCCACCGACTGGAGGGCCGGGCGTGCCTGTC TCATCGATTATCACGAGGCCCT	<i>bcmE</i> deletion by PCR targeting
MC11	AAGGACTGCGATGACGACGCTCGTCGACAACGAAGGAC ACATCGATGCGAGGCAAGCTTA	<i>bcmF</i> deletion by PCR targeting
MC12	AGAAGTCGCGGAAGCCCTGGTACTGTCGCAAGGTGCGG TCATCGATTATCACGAGGCCCT	<i>bcmF</i> deletion by PCR targeting
MC13	CCCGCATATGAGCACGGCACAGGGATACGGCTGGCAGA CGATCGATGCGAGGCAAGCTTA	<i>bcmG</i> deletion by PCR targeting
MC14	CGAGGAACGTGCCGAAGTTCGTCTCCAGGGGAGACCG GTATCGATTATCACGAGGCCCT	<i>bcmG</i> deletion by PCR targeting
MCc1	GGTTGGCTTTATGTCTGCTAGAAGCGCAGCTGATGGAGC CTATCGATGCGAGGCAAGCTTA	<i>bcmA</i> deletion by PCR targeting
MCc2	CGGGGCTCCGGGTGTTTTCCCTGAGCTATTTCCGAG AGATCGATTATCACGAGGCCCT	<i>bcmA</i> deletion by PCR targeting
MC-CI1	GAGATCACC AAGGTAGTCGG	Primers used for construction verification and sequencing
MCm1	GGCCCGGGGCTACCTCCGGCCCGTCCGGGCCATCGCC CTATCGATGCGAGGCAAGCTTA	<i>bcmH</i> deletion by PCR targeting
MCm2	GGCCCTGCCGAGCATCGGTACGACCTCGGCGCCTCGA CCATCGATTATCACGAGGCCCT	<i>bcmH</i> deletion by PCR targeting
MCpseq1	TGGACCCAGCCTGCGCGAG	Primers used for construction verification and sequencing
MCpseq2	ATAAGCCCTACACAAATTGG	Primers used for construction verification and sequencing
MCpseq3	ACCGGGCGGAAGAACTCGCG	Primers used for construction verification and sequencing
MCpseq4	TGCACCGCGTCGCCCGGCGAG	Primers used for construction verification and sequencing
MCpseq5	CCCGCCCGGACACGGGACAG	Primers used for construction verification and sequencing
MCpseq6	CCGGAGTCCGCCGACCGCTT	Primers used for construction verification and sequencing
MCpseq7	CCGGCCGGCTGTTCTCCGTG	Primers used for construction verification and sequencing
MCpseq8	CGTTGTGACGACCGTCGTC	Primers used for construction verification and sequencing
MCpseq9	GACCGAACGCTCGACCGAC	Primers used for construction verification and sequencing
MCpseq10	AACCTGGGTGACGCGTCCG	Primers used for construction verification and sequencing

Table IV.3. Primers used in this study

sequenced by Genewiz (formerly Beckman Coulter Genomics). RNA was isolated using the NucleoSpin RNA and NucleoSpin RNA/DNA Buffer Set kits (Macherey-Nagel).

Bioinformatic analyses

In silico manipulation of sequences and primer design were conducted using Clone Manager 9 Professional Edition (Scientific & Educational Software). Other analyses were performed using BLAST run locally or on the NCBI web server (Zhang *et al.* 2000a) and AntiSMASH 2 (Blin *et al.* 2013). Multiple sequence alignments were done with ClustalW (Combet *et al.* 2000).

bcmA mutant construction

Most of the *bcmA* coding sequence (except the first 33 bp and the last 33 bp) was replaced by the cassette $\Omega att1aac$, consisting of an apramycin resistance marker flanked by T4 terminators, by a double homologous recombination event. Regions upstream and downstream of the *bcmA* gene were amplified from the *S. aizunensis* genomic DNA using, respectively, the JWm9/JWm10 and JWm11/JWm12 primer couples. 3'-A overhangs were added by an incubation with Taq DNA polymerase and dNTP (Qiagen) at 72 °C during 10 min with no primers. The two amplicons were cloned into pGEM-T Easy, yielding respectively the two plasmids pJWc9 and pJWc10. The integrity of the insert was verified by sequencing. The upstream fragment was excised from pJWc9 by digestion with *HindIII/Eco32I*, and the downstream fragment was excised from pJWc10 by digestion with *Eco32I/BcuI*. The pOSV400 suicide vector was digested by *HindIII/BcuI* and the $\Omega att1aac$ cassette conferring apramycin resistance was obtained as a *Eco32I* fragment from pOSV215. All fragments were ligated together, yielding the plasmid pJWm04. pJWm04 was then transferred by conjugation to *S. aizunensis*. Exconjugants were selected for apramycin resistance. Hygromycin-sensitive and apramycin-resistant clones were then screened, their genomic DNA extracted and the replacement of most of the *bcmA* coding region by the $\Omega att1aac$ cassette was verified by PCR (primers: JW-RT9/JW-RT10).

Construction of the plasmid expressing *bcmA*

The *ermE** promoter followed by the *tipA* RBS site was amplified by PCR using the plasmid pIJ7090 as template and the primers JWc1 and JWc2 that added the *tipA* RBS and the appropriate restriction sites (*NdeI*, *PacI* and *HindIII*). The PCR product was then phosphorylated and cloned into pRT801 linearized by *Eco32I*, yielding the vector pOSV666. The *bcmA* coding sequence was amplified using primers JWc7 and JWc8 and *S. aizunensis* genomic DNA as template. The PCR product was then digested by *NdeI/HindIII* and cloned into pOSV666 digested by the same enzymes, yielding the plasmid pJWe10. The hygromycin resistance gene *hph* was then excised from pOSV400 by *VspI* digestion, blunt-ended with treatment by the Klenow enzyme, and cloned into the unique *ApaI* site of the pJWe10 plasmid, also blunt-ended with treatment by the Klenow enzyme, yielding the plasmid pJWe18. The integrity of all constructs' inserts was verified by sequencing.

Deletion of the whole *bcm* cluster.

The entire bicyclomycin biosynthetic gene cluster was replaced by a kanamycin resistance cassette. Regions upstream and downstream of the *bcm* cluster were amplified from the *S. aizunensis* genomic DNA using, respectively the primer couples JWm23/JWc10 and JWm24/JWm25. The *aphII* kanamycin resistance cassette was amplified using the primer couple JWm19/JWm20 and pOSV408 as template. The upstream fragment was digested by *HindIII/EcoRI*, the downstream fragment by *EcoRI/BcuI* and the cassette was digested by *EcoRI*. The suicide vector pOSV400 was digested by *HindIII/BcuI*. All fragments were ligated together, yielding the plasmid pJWm07. The integrity of the insert was verified by sequencing. pJWm07 was transferred by conjugation to *S. aizunensis*. Exconjugants were selected for kanamycin resistance. Hygromycin-sensitive and kanamycin-resistant clones were then screened, their genomic DNA extracted and the replacement of the cluster by the kanamycin resistance cassette was verified by PCR (primers: JW-RT9/JW-RT12).

Construction of a plasmid containing the *bcm* cluster

The plasmid pOSV668 was first constructed by cloning the spectinomycin resistance cassette $\Omega aadA$ from pHP45 Ω into pRT801 at the *Bam*HI site. The whole *bcm* cluster was amplified in five fragments by PCR from *S. aizunensis* genomic DNA, using the primer couples JWc9/JWc10 (fragment I), JWc11/JWc12 (fragment II), JWc13/JWc14 (fragment III), JWc15/JWc16 (fragment IV) and JWc17/JWc18 (fragment V). Fragment I was then digested by *Hind*III/*Eco*RI, fragment II by *Eco*RI/*Xba*I, fragment III by *Xba*I/*Mls*I, fragment IV by *Mls*I/*Nde*I and fragment V by *Nde*I/*Hind*III. Fragments I and II were then cloned into pUC18 digested by *Hind*III/*Xba*I yielding the plasmid pJWc13. Fragments III, IV and V were cloned into pUC18 digested by *Hind*III/*Xba*I yielding the plasmid pJWc14. The integrity of both inserts was verified by sequencing. Finally, the whole cluster was assembled by ligating fragments I-II excised by *Hind*III/*Xba*I from pJWc13 and III-V excised by *Hind*III/*Xba*I from pJWc14 into pOSV668 digested by *Hind*III, yielding the plasmids pJWe14.1 and pJWe14.2, which differ by the orientation of the insert.

Deletion of *bcm* genes on the pJWe14.1 plasmid

All the genes of the *bcm* cluster were deleted individually by a PCR-targeting approach. For this purpose, the $\Omega aadA$ cassette was first amplified using pHP45 Ω as a template and MC1 and MC2 primers adding *Bsu*15I (*Cl*aI) restriction site at each end of the amplicon. The obtained amplicon was isolated on agarose gel and used as template for a set of new PCR reactions with the primer couples MC3/MC4, MC5/MC6, MC7/MC8, MC9/MC10, MC11/MC12, MC13/MC14, MCc1/MCc2 and MCm1/MCm2., These primers present 40bp-long identical regions with the genes *bcmB*, *bcmC*, *bcmD*, *bcmE*, *bcmF*, *bcmG*, *bcmA* and *bcmH*. They were designed in order to replace, after homologous recombination, most of the coding sequence of each gene with the spectinomycin resistance cassette $\Omega aadA$, leaving 30 base pairs at the beginning of the coding sequence and 90 at the end. The PCR products were used in a PCR-targeting procedure on

pJWe14.1, following the protocol previously described (Gust *et al.* 2003), yielding respectively the plasmids pMC1, pMC3, pMC5, pMC7, pMC9, pMC11, pMC13 and pMC15. Those constructs were introduced in a non-methylating GM119 *E. coli* strain. Their integrity was verified by sequencing, and the *QaadA* cassette was removed by *Bsu15I* digestion followed by plasmid self-circularization, yielding respectively the plasmids pMC2, pMC4, pMC6, pMC8, pMC10, pMC12, pMC14 and pMC16.

Construction of plasmids expressing *bcmH* or *bcmA*

In order to complement the strains bearing an in-frame deletion of *bcmA* or *bcmH*, complementation vectors were constructed as follows. PCR amplifications using the primer couples JWc30/JWc31 and JWc32/JWc33 and the genomic DNA of *S. aizunensis* as a template were carried out. Each PCR amplified the gene with its native promoter. The obtained PCR products were digested by *Mph1103I/BcuI* and cloned into pOSV806 vector digested with the same enzymes, yielding respectively pJWe21 and pJWe22 vectors. The integrity of the inserts was verified by sequencing.

Construction of expression vector of the *bcmA* gene under its native promoter

In order to study the function of the BcmA protein, a vector expressing only this gene under the control of its native promoter was constructed as follows. A PCR amplification using primers JWc29 and JWc8 and the genomic DNA of *S. aizunensis* as a template was carried out. The obtained PCR product was digested by *HindIII* and cloned into pOSV668 digested with the same enzyme, yielding the plasmid pJWe20. The integrity of the insert was verified by sequencing.

Cloning of *bcm* genes into an *E. coli* expression vector

The *bcmB*, *bcmC*, *bcmD*, *bcmE*, *bcmF* and *bcmG* genes were amplified by PCR from *S. aizunensis* genomic DNA with, respectively the primer couples JD1/JD2, JD3/JD4, JD5/JD6, JD7/JD8, JD9/JD10 and

JD11/JD12. Each PCR product was digested by *NdeI/XhoI* and ligated into pET15b digested by the same enzymes, yielding respectively the plasmids pJDOG1, pJDOG2, pJDCyp, pJDOG3, pJDOG4 and pJDOG5. The integrity of the inserts was verified by sequencing.

Protein production and purification

The constructs pJDOG1, pJDOG2, pJDCyp, pJDOG3, pJDOG4 and pJDOG5 were introduced in *E. coli* BL21-AI by transformation. An overnight culture in 100 ml in LB medium with ampicillin grown at 37 °C was used to inoculate 1 liter cultures (in LB medium with ampicillin) at the initial optical density (OD) of 0,05 in Fernbach flasks. For the strain harboring pJDCyp, aminolevulinic acid was added to the medium at 100 mg/l. The cultures were grown at 37°C under orbital agitation (200 rpm) up to an OD between 0,3 and 0,4. They were cooled down to 20 °C during 30 min, the target gene expression was then induced with IPTG (1 mM) and L-arabinose (0.2%) and the cultures were grown overnight at 20 °C under orbital agitation (200 rpm). Cells were pelleted by centrifugation, resuspended in 30 ml of buffer A (0.5 M NaCl, 20 mM imidazole, 5% glycerol, 1 mM DTT, 1 mM PMSF, 50 mM HEPES, pH adjusted to 7.5 with NaOH) and broken with an Eaton press. The DNA was broken down by adding 2 µM of MgCl₂ and 0,1 U/ml of benzonase (Sigma Alrdich) and incubating for 45 min at 4 °C under gentle agitation. The lysate was centrifuged for 30 min at 20 000 g. The supernatant was used for purification on a 5 ml HisTrap HP (GE Healthcare) column using an Äkta Purifier (GE Healthcare) FPLC instrument. The injection was operated with buffer A at 1 ml/min until the OD at 280 nm was low and stable (c.a. 60 ml). The elution was made by a gradient from 100% buffer A up to 70% of buffer A and 30% of buffer B (0.5 M NaCl, 500 mM imidazole, 5% glycerol, 1 mM DTT, 1 mM PMSF, 50 mM HEPES, pH adjusted to 7.5 with HCl) over 10 column volumes followed by the isocratic elution at 30% buffer B over 10 column volumes. The fractions containing the protein were pooled and desalted by size-exclusion chromatography on an Äkta Purifier instrument (GE Healthcare), using five 5 ml HiTrap Desalting (GE Healthcare) columns assembled together. The elution was made with desalting buffer (0.1 M NaCl,

1 mM DTT, 1 mM PMSF, 20 mM HEPES, pH adjusted to 7.5 with NaOH) over 50 ml. The fractions containing protein were pooled and concentrated on Amicon Ultra-4 or Ultra-15 10000 NMWL filters (EMD Millipore). Glycerol was added to a final concentration 10% and protein extracts were stored at -80 °C prior to their use.

Chemical synthesis of cIL

All chemical products were purchased from Sigma Aldrich unless otherwise stated. Chemically synthesized cIL was obtained by adapting a procedure described by Jeedingunta *et al.* (2000) Briefly, 1 g (5.5 mmole) of H-L-Ile-OMeHCl (Novabiochem) was dissolved in 15 ml of dimethylformamide at room temperature and 935 μ l (5.5 mmole) of diisopropylamine was added. After freezing at 4 °C, 1.37 g (5.5 mmole) of N-Boc-L-Leu-OH was added, followed by 1.13 g (5.5 mmole) of dicyclohexylcarbodiimide. After 1h of incubation at 4 °C, the reaction mix was allowed to warm at room temperature and incubated overnight at room temperature. The rest of the procedure was performed according to Jeedingunta *et al.* (2000). The white precipitate obtained was washed several times and dried to give 528 mg of a white powder (2.32 mmole, yield 42%) that was found to be homogeneous by HPLC analysis (Atlantis, dC18, 150 x 4.6 mm, 3 μ m).

Enzymatic assays with purified proteins

All chemical products were purchased from Sigma Aldrich unless otherwise stated. The enzymatic assays with 2-oxoglutarate/iron dependent dioxygenases were performed at 30 °C in a reaction volume of 1 ml, and the reaction mix composition was as follows: 50 mM HEPES (pH 7.5), 100 μ M DTT, 50 μ M FeSO₄, 2 mM sodium L-ascorbate, 2 mM disodium 2-oxoglutarate, 100 mg/l catalase from bovine liver, 200 μ M of diketopiperazine substrate (cIL, unless stated otherwise), 2% DMSO (for cIL; DMSO was absent from reactions with other diketopiperazine substrates) and the tested enzyme at 1 μ M concentration. Aliquots of 100 μ l were collected at t=0, 1 min, 3 min, 10 min, 30 min and 60 min. For some samples a supplementary point was taken at 180 min or after overnight

incubation. The aliquots were inactivated by acidifying the sample with HCOOH to a final concentration of 2% (v/v) and incubating on ice.

Enzymatic assay with the putative cytochrome P450 monooxygenase BcmD was performed at 30 °C in a reaction volume of 1 ml, and the reaction mix composition was as follows: 51 mM potassium phosphate buffer (pH 7.5), 15 mM Tris-HCl buffer (pH 7.5), 10 μ M ferredoxin from *Spinacia oleracea*, 10^{-3} U/ml ferredoxin reductase from *S. oleracea*, 1 mM NADPH, 100 μ M purified compound **12** and 1 μ M BcmD. Reaction was incubated for 24h, then inactivated by acidifying the sample with HCOOH to a final concentration of 2% (v/v) and incubating on ice.

To scale-up the production of pathway intermediates for downstream NMR characterization, optimized approaches have been used. To produce a larger amount of compound **5** the concentration of enzyme BcmE was increased to 5 μ M, the concentration of cIL to 1 mM (and, as cIL was dissolved in DMSO, the DMSO concentration was raised to 10%). The reaction volume was increased to 20 ml and the reaction time was 5h. To produce a larger amount of compound **20**, the concentration of enzymes BcmE and BcmC was increased to 5 μ M, the concentration of cIL to 0.5 mM (and, as a consequence the DMSO concentration was increased to 5%). The reaction volume was increased to 40 ml and the reaction time was 5h. To produce a larger amount of compound **2**, the purified product **20** was used as a substrate at a concentration of approximately 2 mM (which eliminated the DMSO from the reaction mix), and the reaction volume was increased to 2.25 ml and the reaction time was 5h. The enzyme (BcmE) concentration was not changed (1 μ M). To produce a larger amount of compound **12**, the purified product **2** was used as a substrate at a concentration of approximately 2 mM (which eliminated the DMSO from the reaction mix) and the reaction volume was increased to 3 ml. The enzyme (BcmB) concentration was increased to 5 μ M and the 2-oxoglutarate concentration was increased to 6 mM. The reaction time was 2h. To produce larger amount of product **19**, the purified compound **12** was used as a substrate at a concentration of approximately 0.5 mM, and the reaction volume was increased to 4 ml. The enzyme (BcmD) concentration was not changed

(1 μ M). The reaction time was 24 h. The reaction mixes were always incubated in 50 ml Falcon tubes, at a maximum volume of 5 ml per tube at 30 °C under gentle agitation (60 rpm). Reactions were stopped by acidifying with HCOOH to a final concentration of 2% (v/v).

Production assays in *S. aizunensis* strains

All production experiments were carried out as follows. A 10 ml preculture in MP5 medium (Pernodet *et al.* 1993) was launched in 50 ml Falcon tube on an orbital shaker (180 rpm) at 28 °C. The preculture was inoculated with either spores or mycelium fragments. After 48h, 1 ml of preculture was used to inoculate 25 ml of MP5 medium in a 500 ml baffled Erlenmeyer flask, and the culture was incubated 7 days on an orbital shaker (180rpm) at 28 °C.

HPLC and LC-MS analysis of the *S. aizunensis* culture supernatants

Culture supernatants obtained after centrifugation of the culture were filtered on 0,2 μ m MiniUniPrep filter (Whatman) and analyzed on an Atlantis T3 column (250 mm x 4.6 mm, 5 μ m, temperature 30 °C) using an Agilent 1200 HPLC instrument equipped with a quaternary pump. Samples were eluted with isocratic 0.1% HCOOH in H₂O (solvent A) / 0.1% HCOOH in CH₃CN (solvent B) (95:5) at 1 ml/min for 5 min followed by a gradient to 100% solvent B over 45 min. The molecules present in the supernatant are detected by an Alltech 3300 evaporative light scattering detector (ELSD) (Grace).

If an analysis of the cell content was necessary, the mycelium pellet was kept after the centrifugation of 2 ml of culture, washed twice with 2 ml of water, resuspended in 400 μ l of water and broken down with glass beads using the FastPrep-24 apparatus (MP Biomedicals). A cell-free extract was obtained after centrifugation to remove cell debris.

LC-MS analyses were performed using an Agilent 1100 HPLC coupled via split system to an Esquire HCT ion trap mass spectrometer (Bruker) set in positive and negative modes. Multiple analyses were made, some of them were made using the same column and elution conditions as described

above (marked as condition “A” in Supplementary Table 5), other using the conditions described in Jacques *et al.* (2015) for CDPS 1-47 (Atlantis dC18 column; marked as condition “B” in Supplementary Table 5).

HPLC and LC-MS analysis of the enzymatic assays

Enzymatic assays were analyzed as described in Jacques *et al.* (2015) for CDPS 1-47 (Atlantis dC18 column) using VWR/Hitachi Elite LaChrom instrument. Interesting peaks were recuperated from the flow-through and injected to an Esquire HCT ion trap mass spectrometer (Bruker) set in positive and negative modes.

Preparative HPLC of *in vitro* enzymatic reaction products

Some products were purified on LiChroCART 250 x 10 Purospher STAR RP-18e (5 μ m) column (Merck) using VWR/Hitachi Elite LaPrep Σ instrument. Samples containing DMSO were loaded (5 ml maximum at once) and washed in 100% solvent A (0,1% HCOOH in water) over 30 minutes, then eluted in a gradient starting at 100% solvent A and 0% solvent B (0,1% HCOOH in 10 % water, 90 % acetonitrile). The flow was at 4,75 ml/min. The exact gradients were as follows: for the purification of product **5** from the reaction with BcmE on cIL alone: linear gradient up to 30% B over 30 minutes; for purification of product **20** from the reaction with BcmE and BcmC on cIL: linear gradient up to 10 % B over 5 minutes, then linear gradient up to 20% B over 20 minutes. Concerning the purification of the product **2** from the reaction with BcmG on compound **20**, as the reaction mix did not contain any DMSO, it was loaded and washed in 100% solvent A over 5 minutes, then eluted in a linear gradient up to 20% B over 20 minutes. As for the purification of the product **12** from the BcmB reaction on compound **2**, it was purified on Hypercarb 150 x 10 (5 μ M) column (Thermo Fisher Scientific). The reaction mixture was loaded and washed in 100 % solvent A over 5 minutes, then eluted in a linear gradient up to 25% B over 25 minutes; the collected peak contained a mixture of compounds **2** and **12**; it was lyophilized, dissolved in water and purified on LiChroCART 250 x 10 Purospher STAR RP-18e (5 μ m) column (Merck) using the same load and

elution conditions. Concerning the purification of the product **19** from the reaction with BcmD on compound **12**, it was purified on ACE Excel 3 C18-PFP (150 x 4.6 mm) column (Advanced Chromatography Technologies). The flow was at 0.6 ml/min and the gradient the same as for the previous purification. All the products obtained were lyophilized and their identity was confirmed by mass spectrometry analysis.

NMR Analysis

Samples of compounds cIL (10.6 mM), **5** (5 mM), **20** (20.9 mM), **2** (4.1 mM) and **12** (< 1 mM) were prepared in DMSO-*d*₆ (Eurisotop, Saint-Aubin, France) in 3 or 5 mm NMR tubes. NMR experiments were recorded on a 500 MHz Bruker Avance III spectrometer equipped with a 5-mm inverse TCI cryoprobe incorporating a Z-gradient coil. Spectra were recorded at 298.6 K. All data were processed and analyzed with Bruker TopSpin 3.2 program. ¹H and ¹³C resonances were assigned via the analysis of one-dimensional ¹H, one-dimensional ¹³C DEPTQ (Distortionless Enhancement by Polarization Transfer), two-dimensional ¹H-¹H COSY, two-dimensional ¹H-¹H TOCSY (Total Correlation Spectroscopy, mixing time of 66 ms), two-dimensional ¹H-¹H NOESY (Nuclear Overhauser Effect Spectroscopy, mixing time of 1.5 s), two-dimensional ¹H-¹³C HSQC (Heteronuclear Single Quantum Correlation), two-dimensional ¹H-¹³C HMBC (Heteronuclear Multiple-Bond Correlation). ¹H and ¹³C chemical shifts were referenced to the residual protiated DMSO solvent signal (δ ¹H 2.50 ppm) or deuterated DMSO solvent (δ ¹³C 39.5 ppm), respectively. Homonuclear $J_{H,H}$ couplings were measured on 1D ¹H spectra processed with Lorentz-Gauss apodization. 1D selective irradiations were applied to extract $J_{H,H}$ couplings from the most complex multiplets. Heteronuclear $J_{H,C}$ couplings were measured using 2D IPAP-HSQMBC (In-phase Anti-Phase Heteronuclear Single Quantum Multiple Bond Correlation) experiments, with an evolution delay $\Delta = 1/(2 \text{ } ^nJ_{CH})$ optimized for $J_{H,C}$ couplings of 8 Hz or 6 Hz (Gil *et al.* 2010).

HRMS analysis

For HRMS analysis lyophilized compounds were dissolved in methanol 50% and injected on LTQ-Orbitrap XL ETD (Thermo Fisher Scientific) high resolution mass and high mass accuracy spectrometer programmed in negative mode.

Results

Mining the *S. aizunensis* genome for a candidate bicyclomycin biosynthetic gene cluster

We sequenced the genome of the bicyclomycin producing strain *S. aizunensis* ATCC 2196. We obtained an assembly with 57 contigs and a total size of 6.7 Mbp with a GC content of 72.3%. To study the biosynthesis of bicyclomycin we first looked for the candidate cluster, thinking that the cluster should contain a gene coding for a protein synthesizing the cIL precursor. The Antismash search did not allow finding any NRPS susceptible of synthesizing this precursor. On the other hand, the tBLASTn search of the *S. aizunensis* genome using as queries the biochemically characterized CDPSs (Jacques *et al.* 2015) allowed finding a single putative CDPS gene. A BLAST search in the NCBI database permitted to find out that the most similar biochemically characterized CDPS was the one of *Pseudomonas aeruginosa* (e-value: 10^{-78} , query cover: 92%, identity: 52%). It was previously demonstrated that this CDPS synthesized mostly cIL, the precursor we looked for (Jacques *et al.* 2015). Moreover, not only the identity between these two proteins is rather high, but in addition the two substrate recognition pockets are particularly well conserved among the two CDPS. For the pocket P1 of the *S. aizunensis* CDPS we find the signature FGLGCIFM, which differs only by one amino acid (I instead of V) from the signature of the *P. aeruginosa* pocket. The pocket P2 presents the signature motif FDYKGFS, with two differences (D instead of N and K instead of H) when compared to the characterized *P. aeruginosa* CDPS. Those data indicate that this CDPS

was likely to synthesize the cIL precursor, previously demonstrated to be the starting point of bicyclomycin biosynthesis (Bradley *et al.* 1996).

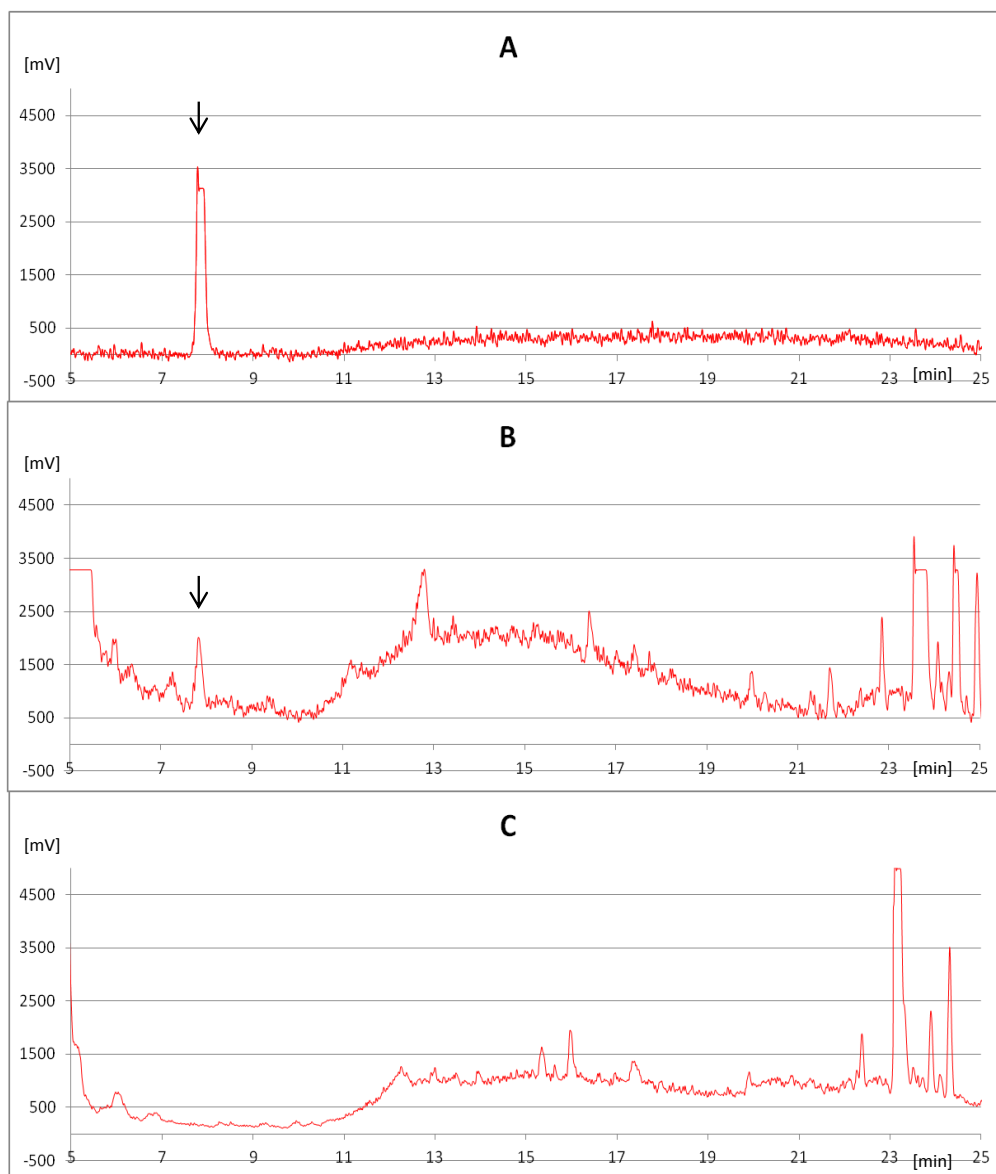


Figure IV.2. HPLC chromatograms of *S. aizunensis* culture supernatants. (A) – bicyclomycin standard, (B) – wild type strain, (C) – $\Delta bcmA::\Omega att1aac4$ (CDPS mutant). Cultures were realized in 25 ml of MP5 medium in a baffled 500 ml erlenmeyer flask on an orbital shaker at 28°C during 7 days. The peak corresponding to bicyclomycin is marked with an arrow.

The putative CDPS gene is involved in bicyclomycin biosynthesis.

Having found a candidate gene for the biosynthesis of cIL, we applied a double homologous recombination technique using a suicide vector to replace the putative CDPS-encoding gene (that we named *bcmA*) by an excisable cassette containing an apramycin resistance marker. In the mutant $\Delta bcma::\Omega att1aac$ that we obtained, bicyclomycin production was completely abolished (Figure IV.2).

The introduction of the *bcmA* gene cloned on the integrative plasmid pJWe18 in this mutant strain partially restored the bicyclomycin production (data not shown). The level of bicyclomycin production is low, however the compound produced has the same retention time, the same m/z and the same fragmentation pattern (m/z with peak intensities given in parentheses: 99 (4.6); 154.9 (13.1); 178.8 (5); 183.9 (100); 184.9 (17); 194.8 (18.1); 202.8 (5.9); 208.8 (15.1); 209.8 (12.2); 226.8 (65.7); 257 (7.4); 257.9 (4.6); 282.9 (44.8)) as the bicyclomycin standard. This proves that the putative CDPS BcmA is indispensable for bicyclomycin production

Bioinformatic analysis of the bicyclomycin biosynthetic gene cluster.

Having proved the involvement of *bcmA* in bicyclomycin biosynthesis, we analyzed the immediate vicinity of this gene to search for other genes possibly involved in bicyclomycin biosynthesis. We found that *bcmA* was part of a putative biosynthetic gene cluster (BGC) that also contains a gene (that we named *bcmH*) coding for a putative MFS1 class transporter (a potential resistance and/or export protein) as well as five genes (that we named *bcmBCEFG*) encoding putative 2-oxoglutarate/iron dependent dioxygenases and one gene (that we named *bcmD*) encoding a putative cytochrome P450 monooxygenase (Figure IV.3). The putative functions of the proteins encoded fit well with our expectations, as we predicted that there are six oxidoreduction reactions needed to transform cIL into bicyclomycin.

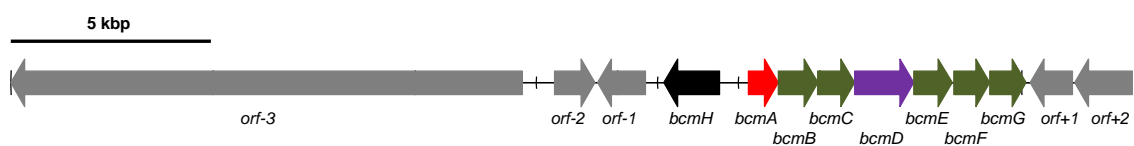


Figure IV.3. The bicyclomycin biosynthetic gene cluster of *S. aizunensis* and its genetic environment. Predicted functions of the proteins encoded by the genes: *orf-3* – polyketide synthase, *orf-2* – chorismatase, *orf-1* – phospho-2-dehydro-3-deoxyheptonate aldolase, *orf+1* – chitin binding protein, *orf+2* – chitinase; these five genes are not part of the cluster. *bcmH* – Major Facilitator Superfamily 1 transporter, *bcmA* – cyclodipeptide synthase, *bcmB*, *bcmC*, *bcmE*, *bcmF* and *bcmG* – 2-oxoglutarate-iron-dependent dioxygenases, *bcmD* – cytochrome P450.

A BLAST search (July 2017) allowed finding multiple homologous clusters presented on Figure IV.4. Nine actinobacterial homologues have the same overall organization as the *bcm* cluster and the deduced proteins are highly similar to those of the *bcm* cluster. In two of these clusters, there are differences for the *bcmH* homologous gene: in *Streptomyces ossamyceticus* it is missing and in *Actinokineospora spheciospongiae* it is found after *bcmG* in the same orientation as the other genes of the cluster. In *Williamsia herbipolensis* there is a stop codon inside the sequence of *bcmB* homolog. The overall organization of the homologous clusters in Proteobacteria is also similar. *P. aeruginosa*, *Tistrella mobilis* and *Burkholderia plantarii* have similar organization as *A. spheciospongiae*, however *bcmE* and *bcmF* have their position switched, and the predicted protein sequence identity is low for *bcmF* and *bcmH* homologues. *Photorhabdus temperata ssp. khanii* has the cluster split in two parts (one part containing *bcmAB* preceded by an additional *bcmE* homologue, the other one containing *bcmCDFEGH* genes) and its *bcmA* gene encodes a truncated protein. None of the cluster flanking genes is conserved between any pair of those species, suggesting that these genes are not involved in the biosynthetic pathways encoded by these clusters. The detailed percent identity matrix is given in Supplementary Table 4.

Concerning the limits of the cluster in *S. aizunensis*: first, the lack of conservation of the flanking genes between species that possess the cluster is a clear indication that this BGC is composed of the genes *bcmA-H*.

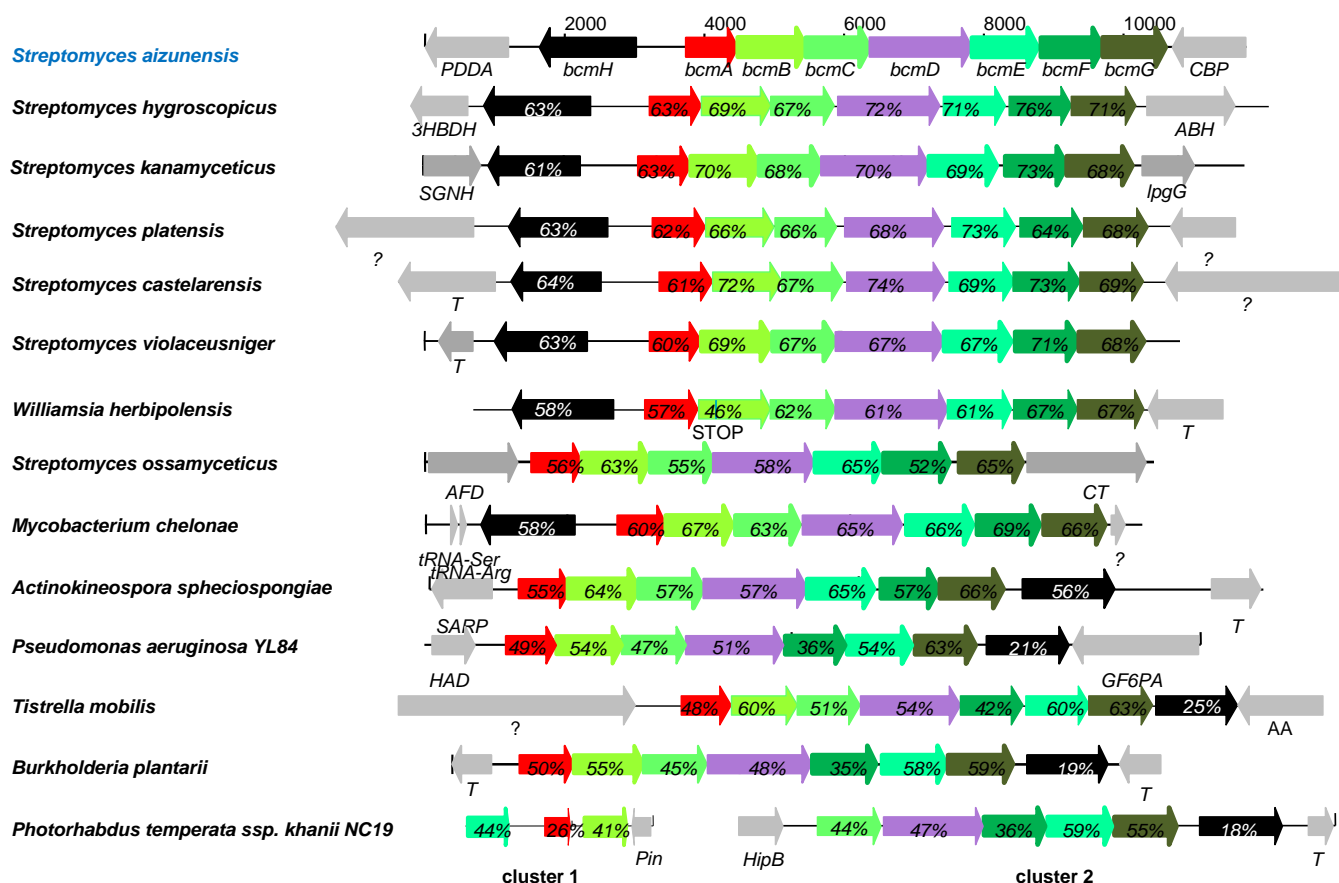


Figure IV.4. Homologs of the bicyclomycin BCG. Predicted functions of the proteins encoded by the genes adjacent to the clusters: PDDA: phospho-2-dehydro-3-deoxyheptonate aldolase, CBP – chitin binding protein, 3HBDH – 3-hydroxybutyrate dehydrogenase ABH – alpha-beta-hydrolase, SGNH – SGNH-hydrolase, LpgG – conserved protein with SIMPL domain, T – transposase, AFD – AMP-dependant ligase / bile acid-CoA ligase, CT – carbamoyl transferase, ? – hypothetical protein, SARP – SARP-family transcriptional regulator, HAD – HAD-family hydrolase, GF6PA – Glucose fructose-6-phosphate aminotransferase, AA – aminoacyl aminotransferase, Pin – DNA recombinase, HipB – XRE-family transcriptional regulator. The protein identity percentage with *S. aizunensis* enzymes is represented inside the arrows.

Moreover, we can hypothesize that genes upstream of the cluster, coding for an aldolase and a chorismatase are probably a part of an adjacent PKS cluster (the *orf-3* measuring over 12 kbp encodes a type I PKS). Genes downstream of the putative bicyclomycin BCG form an operon composed of genes coding for chitinase and chitin binding protein, well conserved among *Streptomyces*, for which role in bicyclomycin biosynthesis seems unlikely.

All biosynthetic enzymes within the *bcm* cluster are required for bicyclomycin biosynthesis

We wanted to study the role of all the genes of the cluster by creating mutant strains deleted for each of the genes. Our initial idea was to replace each gene by a excisable cassette, which would be excised in a second step, in order to create in-frame deletion, as we had done previously for the functional analysis of other BGCs (Nguyen *et al.* 2010). However, the experiments performed to inactivate *bcmA* showed that it was not easy to obtain gene replacement by a double homologous recombination event, even with regions of 3 kb upstream and downstream of the target gene. Moreover, we could not excise the cassette as previously described (Raynal *et al.* 2006), probably because of replication or expression problems encountered with the plasmids expressing the *xis* and *int* genes (data not shown). As the presence of the cassette could have a polar effect, it was necessary to find another strategy for the functional analysis of the *bcm* BGC and we decided to study this BGC in conditions of heterologous expression. For this purpose, we cloned the entire *bcm* cluster into an integrative vector, leading to pJWe14.1 and pJWe14.2 plasmids. These plasmids were introduced in standard heterologous hosts (*Streptomyces lividans* TK24, *Streptomyces coelicolor* M1154) but we have never been able to observe easily detectable levels of bicyclomycin in these conditions (data not shown). We hypothesized that this lack of production might be due to the absence, in the heterologous hosts, of a regulation system present in *S. aizunensis* but encoded by genes which were not clustered with the biosynthetic genes. To test this hypothesis we constructed a *S. aizunensis* mutant strain in which the whole *bcm* cluster was replaced by a kanamycin resistance marker in which all bicyclomycin production is abolished. When the vectors pJWe14.1 or pJWe14.2, carrying

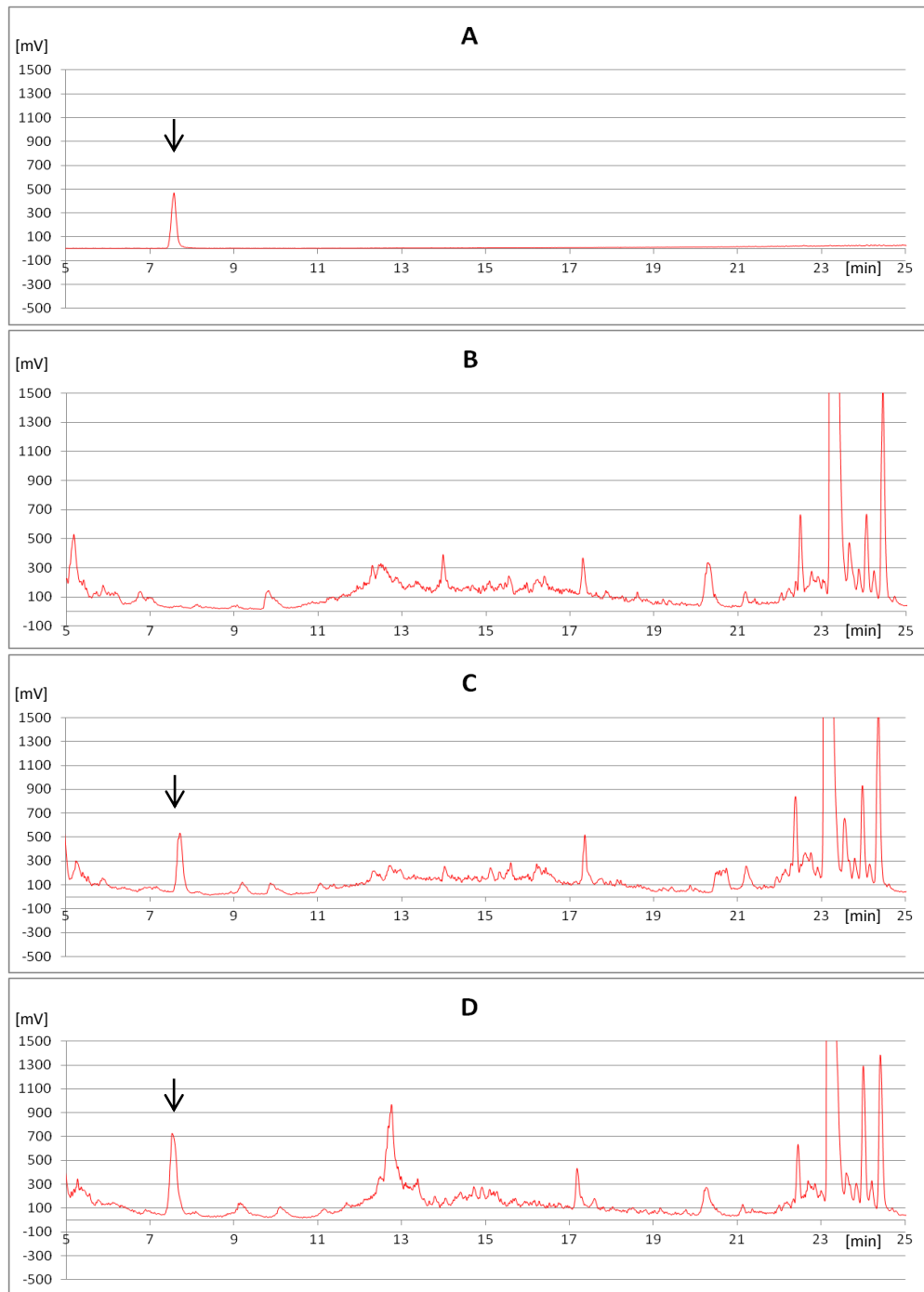


Figure IV.5. HPLC chromatograms of *S. aizunensis* culture supernatants. (A) – bicyclomycin (BCM) standard, **(B)** – $\Delta bcm::aph2$ (BCM BGC mutant strain), **(C)** – $\Delta bcm::aph2/pJWe14.1$ (BCM BGC mutant strain complemented with BCM cluster on an integrative plasmid), **(D)** – $\Delta bcm::aph2/pJWe14.2$ (the same as C, but the cluster orientation in the plasmid is inverted). Cultures were realized in 25 ml of MP5 medium in a baffled 500 ml erlenmeyer flask on an orbital shaker at 28°C during 7 days. The peak corresponding to BCM is pointed by an arrow.

the *bcm* cluster, were introduced into this mutant strain, they restored bicyclomycin production, independently of the direction in which the cluster was cloned (Figure IV.5). This result validated our approach. It was therefore possible to easily inactivate genes in the cloned cluster by PCR-targeting in *E. coli* and to study the effect of these deletion in *S. aizunensis*. In-frame deletion mutants were constructed for all the genes coding for biosynthetic enzymes in the cluster cloned in pJWe14.1, and the constructs obtained were introduced into the Δbcm mutant. The culture supernatant analysis showed that each time a gene coding a biosynthetic enzyme in the cluster was deleted, the bicyclomycin production was abolished (Figure IV.6), demonstrating that all the genes were implicated in the bicyclomycin production in *S. aizunensis*.

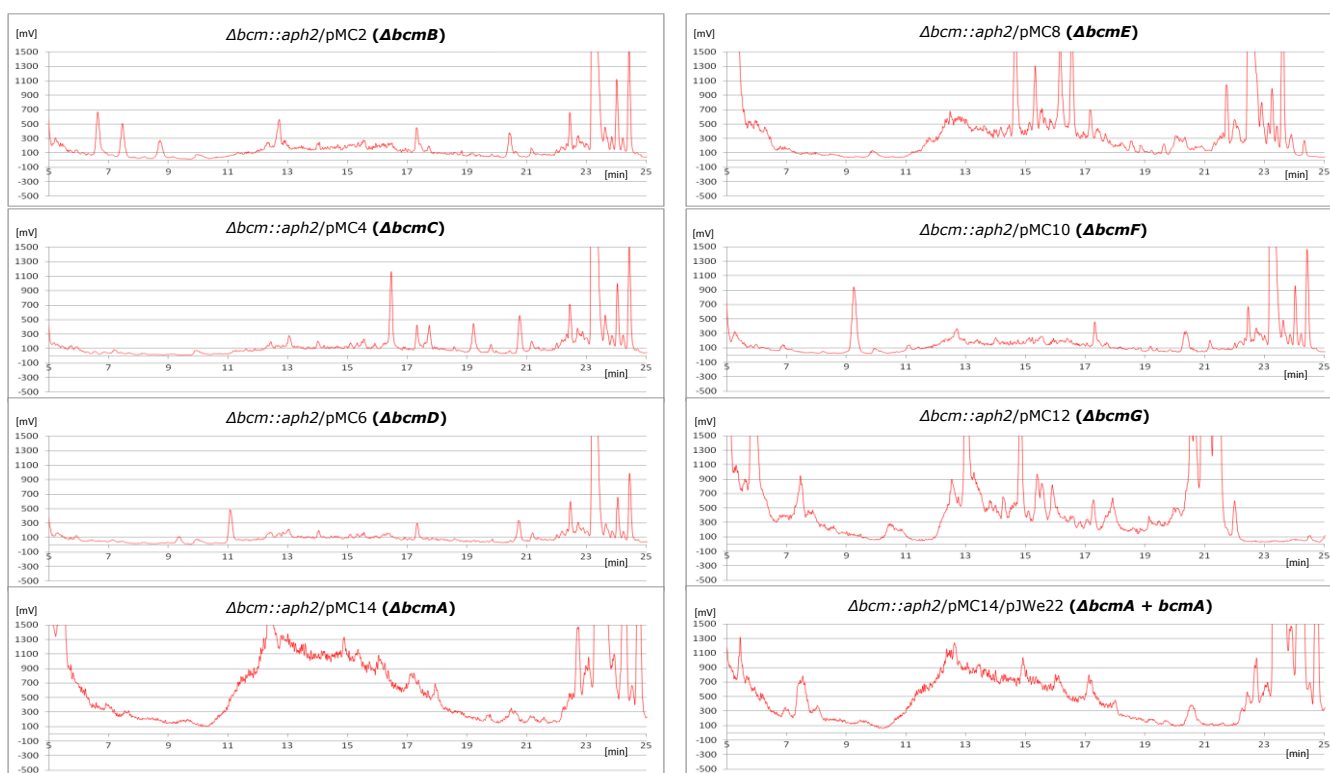


Figure IV.6. HPLC chromatograms of *S. aizunensis* culture supernatants. $\Delta bcm::aph2$ is the mutant with BCM BGC deleted. It was complemented by a series of integrative plasmids containing BCM cluster bearing an in-frame deletion of only one of each gene. Strain and plasmid are annotated over each chromatogram in normal script. In bold is precised the deleted gene. pJWe22 is an integrative vector with *bcmA* cloned and expressed under its native promoter. Cultures were realized in 25 ml of MP5 medium in a baffled 500 ml erlenmeyer flask on an orbital shaker at 28°C during 7 days.

BcmA is a functional CDPS that synthesizes cIL

We can already observe on the Figure IV.6 that when *bcmA* is deleted, not only the production of bicyclomycin is abolished, but we do not observe the apparition of any significant supplementary peaks, corresponding to potential biosynthesis intermediates, which is consistent with Bradley *et al.* (1996) findings that the first step of the bicyclomycin biosynthesis pathway is the synthesis of cIL, and with our hypothesis that BcmA is a CDPS synthesizing this precursor. Moreover, as now we disposed of an in-frame deletion mutant, we could attempt a proper complementation. The integration of the pJWe22 plasmid containing the *bcmA* gene restored bicyclomycin production in *bcm::aph2/pMC14* strain at a level similar to that of the wild type strain. (Figure IV.6).

In order to confirm the function of the putative CDPS BcmA, we constructed an integrative plasmid, named pJWe20, containing the *bcmA* gene under the control of its native promoter. We introduced it into the Δbcm mutant of *S. aizunensis*. Analysis of the culture supernatant by LC-MS allowed the detection of a single new peak associated with the presence of pJWe20 (chromatogram not shown). This peak had a retention time of 40 min (method B) or 23,5 min (method A) and a *m/z* of 227 in positive mode. It was identified as cIL by its fragmentation pattern. The fragments' *m/z* ratio and their relative intensity (in parentheses) were as follows: 69.3 (2.5); 86.1 (78.1); 98 (8.1); 114 (3.5); 153.9 (24.4); 170.8 (2.9); 181.9 (100); 198.9 (75). It was confirmed that this compound has the same retention time, *m/z* ratio and fragmentation pattern as a cIL authentic standard. We therefore know that the diketopiperazine scaffold is synthesized by the CDPS BcmA.

Functional characterization of the bicyclomycin BGC

Having characterized BcmA, we wanted to study the role of the tailoring enzymes present in the cluster, possibly responsible of the successive transformations leading from cIL to the final product, bicyclomycin. As visible on the Figure IV.6, analysis by HPLC of culture supernatants of *S. aizunensis* $\Delta bcm::aph2$ harboring *bcm* clusters with deletion of *bcmB*, *C*, *D*,

E, *F*, or *G* shows that not only the peak corresponding to bicyclomycin disappeared, but also novel peaks appeared which might correspond to bicyclomycin pathway intermediates or shunt products.

The putative function of the tailoring enzymes as well as the reaction order were at first deduced from the LC-MS data obtained with the supernatants of the Δbcm mutant complemented with the even-numbered plasmids pMC2 to pMC12 bearing in frame deletion of only one of each gene coding for tailoring enzymes (Supplementary Table 5). We hypothesized that the products observed in trace amounts in the condition where whole cluster is present ($\Delta bcm::aph2/pJWe14.1$) were real intermediates. Thus we selected these products to base upon them our reasoning (Table IV.4). The other molecules could possibly be unnatural by-products obtained because the tailoring enzymes present in a pathway are able to convert some substrate analogs even if they do not constitute their natural substrates (however the conversion is never complete, in spite of long reaction time).

Compound number	z0	2	19	z1	12	20	22	z2	5	x	y4	6
Retention time (condition B)	13,5 min	16,5 min	16,7 min	16,7 min	18,2 min	21,6 min	25,4 min	22,4 min	27,4 min	31,8 min	34,8 min	40,2 min
Molecular mass g/mol	288 g/mol	274 g/mol	304 g/mol	288 g/mol	288 g/mol	258 g/mol	256 g/mol	242 g/mol	242 g/mol	240 g/mol	256 g/mol	226 g/mol
pMC2 ($\Delta bcmB$)	-	3116	-	-	-	-	-	-	>100	EIC	-	56
pMC4 ($\Delta bcmC$)	-	-	-	-	-	-	-	-	3375	EIC	EIC	256
pMC6 ($\Delta bcmD$)	EIC	600	-	429	3154	-	-	EIC	1747	EIC	-	EIC
pMC8 ($\Delta bcmE$)	-	-	-	-	-	-	-	EIC	-	-	-	1359
pMC10 ($\Delta bcmF$)	-	EIC	7188	EIC	EIC	-	-	-	-	-	-	EIC
pMC12 ($\Delta bcmG$)	-	-	-	-	-	6519	10617	-	EIC	EIC	-	EIC

Table IV.4. Bicyclomycin pathway intermediates found in the culture supernatants of different mutants. The values in the table correspond to peak area in UV chromatogram, 214 nm. EIC means that the molecule could be observed only in extracted ionic currents and was not visible in UV chromatogram. The molecular mass is inferred from the m/z value. For the complete list of bicyclomycin-related molecules found in the culture supernatants of each mutant, please consult the Supplementary Table 6.

BcmF catalyzes the transformation of dihydrobicyclomycin into bicyclomycin, the last step of the biosynthetic pathway

When *bcmF* is deleted (complementation with pMC10), a product at 304 g/mol (compound **19**, Supplementary Table 5), corresponding to dihydrobicyclomycin, accumulates. This product has already been characterized by Bradley *et al.* (1996), so we can assume that BcmF is responsible for the last biosynthetic step, the dehydrogenation of the isoleucyl lateral chain.

All other tailoring enzymes act in a specific order

When *bcmD* is deleted (complementation with pMC6), we observe the accumulation of a product at 288 g/mol (compound **12**, Supplementary Table 5). We conclude therefore that BcmD adds a hydroxyl group and acts just before the action of BcmF. When *bcmB* is deleted (complementation with pMC2), we observe the accumulation of a product at 274 g/mol (compound **2**, Supplementary Table 5). The difference in molecular weight between this product and the next one is of 14 g/mol, which could correspond to a double reaction: the closing of the ether bridge and one hydroxylation. When *bcmG* is deleted (complementation with pMC12), we observe the accumulation of a product at 258 g/mol (compound **20**, Supplementary Table 5), which suggests that this enzyme adds a hydroxyl group. The same conclusion can be made for BcmC – accumulation of a product at 242 g/mol – when the gene coding for the enzyme is deleted which corresponds to the complementation with pMC4; compound **5**, Supplementary Table 5 – and BcmE – accumulation of unmodified **cIL**, when the gene coding for the enzyme is deleted which corresponds to complementation with pMC8; compound **6**, Supplementary Table 5.

Previous work (Bradley *et al.* 1996) showed that during the fragmentation of bicyclomycin, the first fragment to depart (m/z difference: 74) corresponds to the terminal diol of the leucyl moiety. The sole mutants in which we do not observe any product with this departure are those bearing

the deletions of *bcmC* and *G*. We can therefore propose that BcmC and BcmG add the two distal hydroxyl groups at the leucine lateral chain.

Concerning the cytochrome BcmD, we propose that it could add the hydroxyl group on the DKP ring (alpha carbon of isoleucyl moiety), according to Bradley's and her co-workers' (1996) preliminary results. Effectively, they reported that cytochrome P450 inhibitors lead to production of bicyclomycin analog lacking the hydroxylation of the DKP ring at the alpha carbon atom of isoleucine.

When the gene encoding BcmE was deleted, we never observed the closing of the ether bridge as no products with double dehydrogenation was detected. It also may suggest that BcmE adds on the side chain extremity of the isoleucine moiety, the hydroxyl group necessary for the closing of the secondary cycle. The last unattributed hydroxylation, on the beta atom of leucine moiety, could be attributed to BcmB. Moreover, it is logic that the enzyme, having correctly positioned the substrate in its catalytic pocket, would perform a supplementary reaction on an adjacent carbon, instead of completely repositioning the substrate to perform a reaction elsewhere. All these hypotheses led us to the hypothetical pathway presented on Figure IV.7.

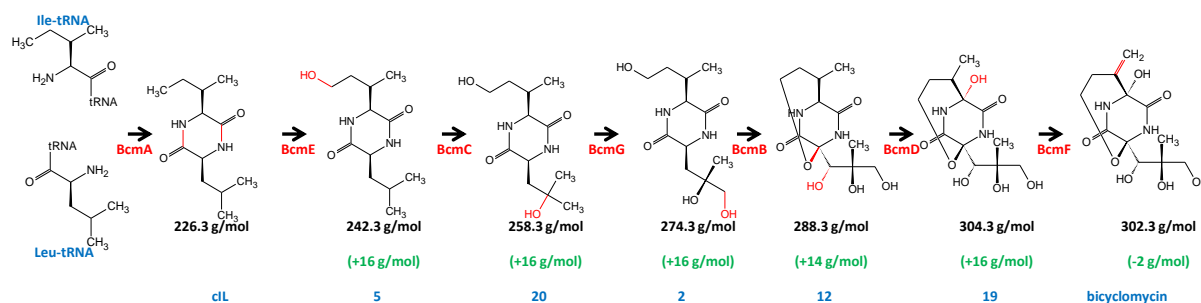


Figure IV.7. Bicyclomycin biosynthetic pathway. The modifications done by each enzyme are given in red. The enzyme names are annotated in red under the arrows. The molecular masses are annotated in black below molecular structures. The difference between the molecular mass of the compound and its immediate predecessor in the pathway, corresponding to the modification done by the upstream enzyme, are annotated in green. The blue numbers below correspond to compound numbers used throughout the paper.

In vitro characterization of the decoration enzymes

In order to test this hypothetical pathway, all the genes encoding tailoring enzymes were cloned into the pET15B vector. The recombinant proteins were produced using *E. coli* BL21-AI strain, purified, and enzymatic assays were undertaken. First, the enzymes were incubated in an isolated manner with the cIL precursor. The only enzyme that efficiently transformed cIL into a novel molecule in one hour was BcmE, which corroborates our hypothesis that it is the first tailoring enzyme acting on cIL (Figure IV.8B). Moreover, mass spectrometry data confirmed that the compound obtained is identical to the compound **5** from Supplementary Table 5 (observed when the gene encoding the next enzyme in our putative pathway, *bcmC*, is deleted; see Supplementary Table 6). The enzymes BcmB and F showed no activity on cIL (data not shown) and the enzymes BcmC and BcmG showed a very weak activity, with supplementary peak apparition, but no observable substrate depletion over one-hour reaction (Figure IV.8A and IV.8C).

Next, enzyme combinations were tested. When BcmE is combined with BcmC (the next enzyme in the tentative pathway), cIL is transformed into product **5**, and then efficiently converted into a novel, more polar product (Figure IV.9A), which is identified by mass spectrometry as compound **20** from Supplementary Table 5 (obtained when the gene encoding the next enzyme in our putative pathway, *bcmG*, is deleted, see Supplementary Table 6).

When BcmE is combined with BcmG, a similar conversion is observed. However the BcmE product is converted less efficiently (Figure IV.9B), which corroborates our assumption that BcmC catalyzes the next step in the pathway (novel products' mass spectra do not correspond to products observed in deletion mutants of the producing strain, data not shown). No transformation of the BcmE product is observed when this enzyme is combined with BcmB or BcmF (data not shown).

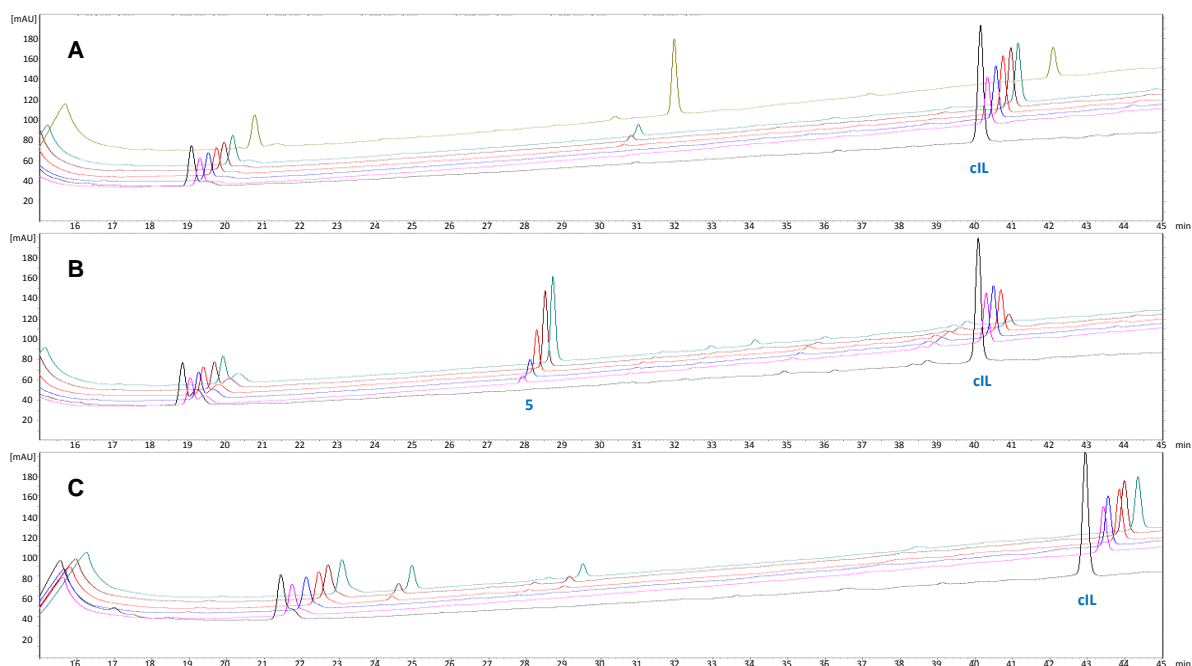


Figure IV.8. HPLC chromatograms of *in vitro* enzymatic assays on cIL. (A) – BcmC, (B) – BcmE, (C) – BcmG. Chromatograms are stacked from bottom to top, the lowest corresponding to the initial state ($t=0$), then to reaction times of 1, 3, 10, 30 and 60 minutes. For BcmC a supplementary chromatogram corresponding to an overnight incubation is shown. The position of the peaks corresponding to cIL and product 5 is marked on the chromatogram.

The combination of the first three enzymes in our tentative pathway (BcmE, C and G) leads to production of multiple peaks, some of which have already been observed when BcmE and BcmG were combined, and are possibly degradation products due to the hyperactivity of one of the enzymes (Figure IV.9C). One novel peak is observed at 17 min, and its mass spectrum is identical to that of the product **2** from Supplementary Table 5 (obtained when the gene encoding the next enzyme in our putative pathway, *bcmB*, is deleted; see Supplementary Table 6). The combination of the first four enzymes in our tentative pathway (BcmE, C, G and B) leads to a similar profile with the notable apparition of a single important novel peak with a RT=18,5 min (Figure IV.9D). The mass spectrum of this peak is identical with that of product **12** from Supplementary Table 5, accumulated when the gene encoding the next enzyme in our putative pathway, *bcmD*, is deleted (Supplementary Table 6).

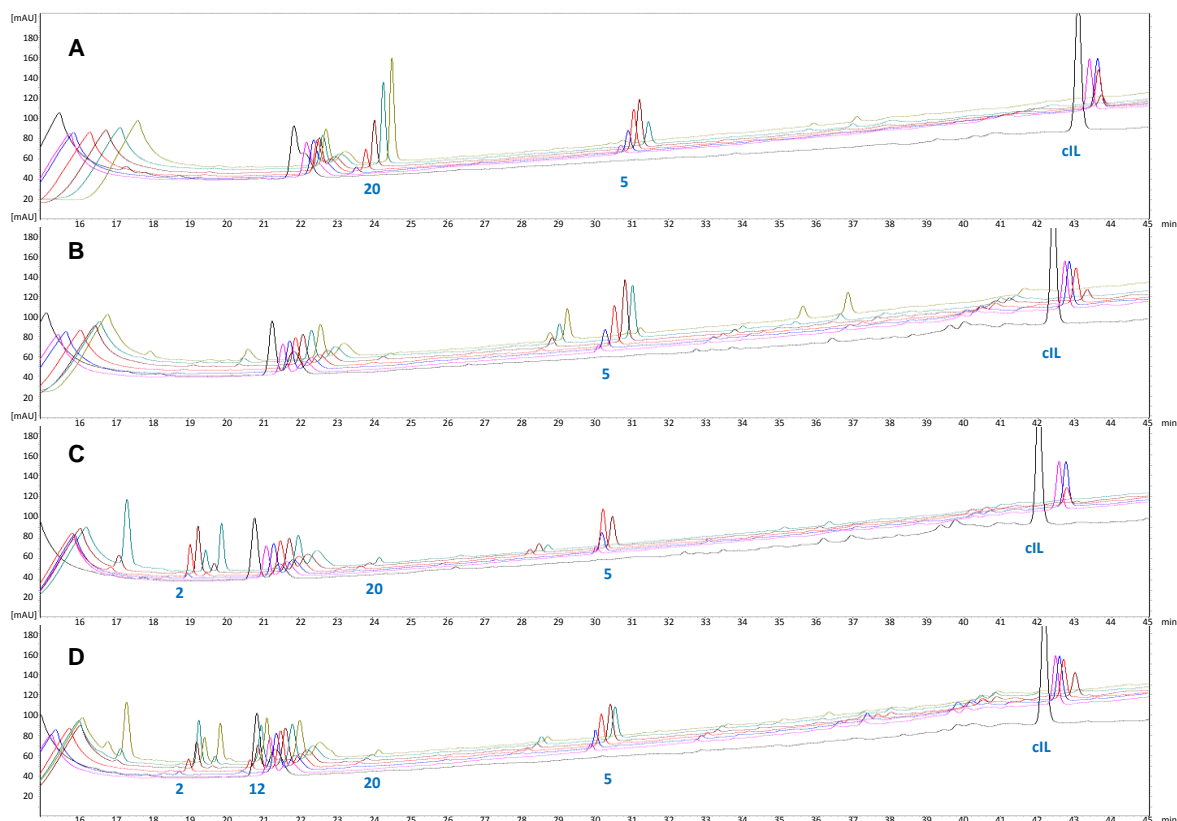


Figure IV.9. HPLC chromatograms of *in vitro* enzymatic assays on cIL. (A) – with BcmC+BcmE, **(B)** – with BcmE+BcmG, **(C)** – with BcmC+BcmE+BcmG, **(D)** – with BcmB+BcmC+BcmE+BcmG. Chromatograms are stacked from bottom to top, the lowest corresponding to the initial state ($t=0$), then to reaction times of 1, 3, 10, 30 and 60 minutes. The last chromatogram corresponds either to an overnight incubation (A) or to a 3h incubation (B, C and D). The position of the peaks corresponding to cIL and products 2, 5, 12 and 20 is marked on the chromatogram, their identity was verified by ESI-MS.

The reaction of the cytochrome BcmD on compound **12** leads to the apparition of a single new peak (Figure IV.10), with m/z value and fragmentation pattern consistent with the product **19** from Supplementary Table 5 (dihydrobicyclomycin, Supplementary Table 6).

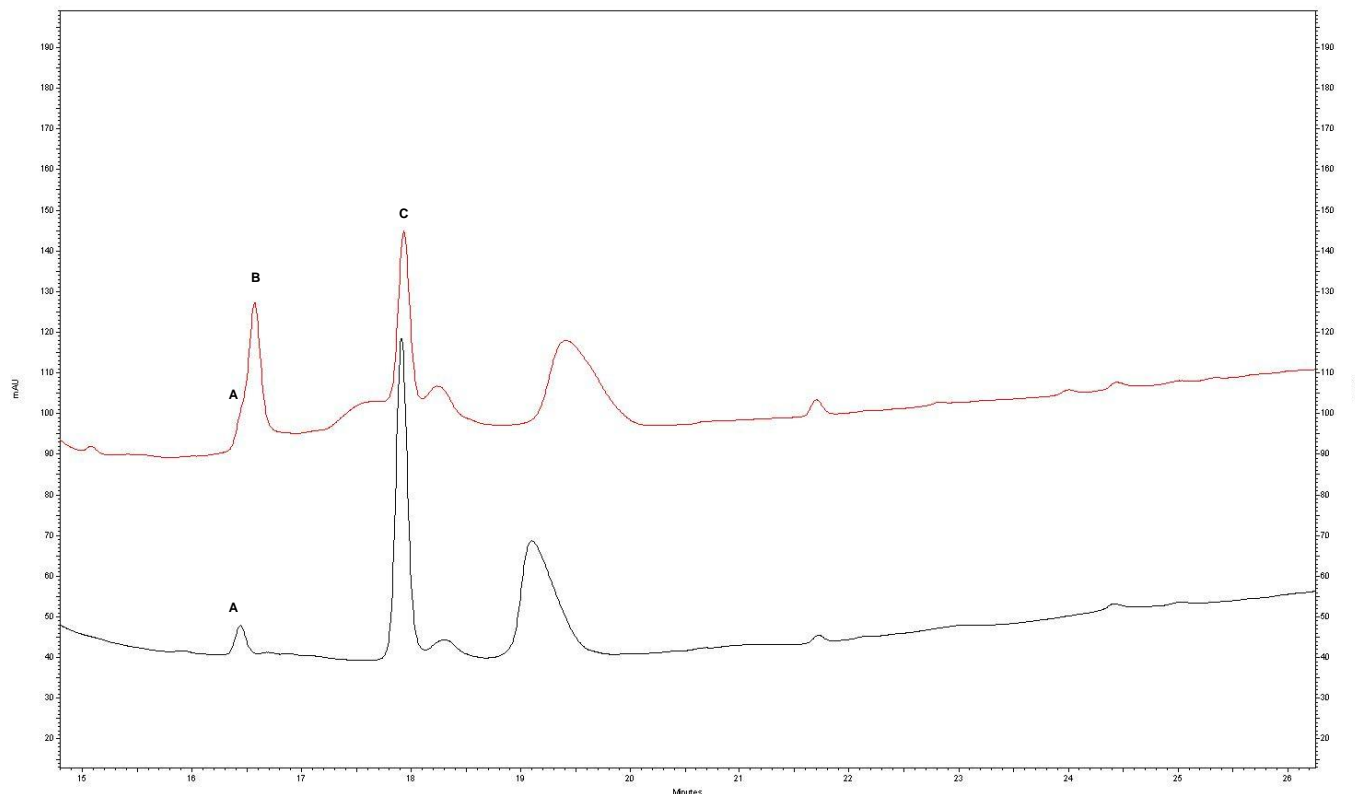


Figure IV.10. HPLC chromatograms of *in vitro* enzymatic assay of BcmD on product 12. In black – t=0. In red – 24h of incubation. **A** – trace amounts of product 2 and Z1, **B** – product 19 (dihydrobicyclomycin), **C** - product 12 (substrate). Identity of molecules was determined by ESI-MS.

The reaction of the enzyme BcmF on purified compound **19** leads to the apparition of a single new peak (Figure IV.11) with m/z value and fragmentation pattern consistent with bicyclomycin (Supplementary Table 6).

The *in vitro* reactions confirm the hypothetical pathway deduced from the product accumulation in deletion mutants up to the bicyclomycin itself. We therefore conclude that the enzymes BcmA-F are necessary and sufficient for bicyclomycin biosynthesis and that their preferential order of action is the one presented on Figure IV.7.

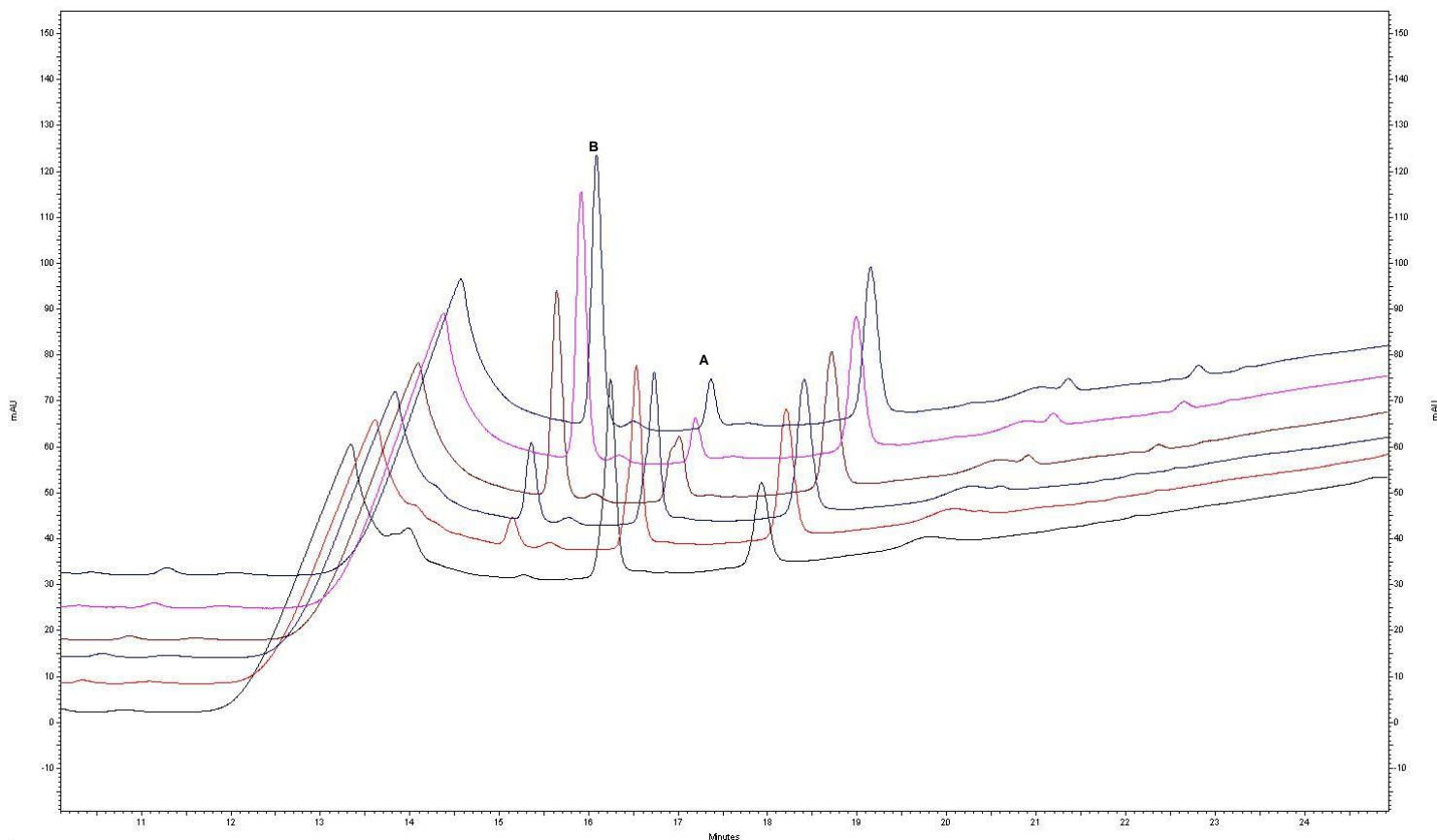


Figure IV.11. HPLC chromatograms of *in vitro* enzymatic assay of BcmF on product 19 (dihydrobicyclomycin). Chromatograms are stacked from bottom to top, the lowest corresponding to the initial state ($t=0$), then to reaction times of 1, 3, 10, 30 and 60 minutes. A – the mix of substrate (19, dihydrobicyclomycin) and the contaminants (products 2 and Z1). B – bicyclomycin. Identity of molecules was determined by ESI-MS.

BcmH is dispensible for bicyclomycin export in *S. aizunensis*

In order to study the role of the *bcmH* gene encoding a putative Major Facilitator Superfamily 1 transporter, we constructed a mutant strain in which this gene is deleted (*S. aizunensis* $\Delta bcm::aph2/pMC16$). We also complemented this mutant strain by the introduction of a copy of *bcmH* under the control of its native promoter (*S. aizunensis* $\Delta bcm::aph2/pMC16/pJWe21$). The presence of bicyclomycin in culture supernatants of these two strains was studied, as the presence of intracellular bicyclomycin in the mycelia. The results are presented in Table IV.5. In both cases LC-MS analysis of culture supernatant with method B allowed us to observe a peak with an area of about 3000 mAU which retention time (15,3 min), m/z value in negative mode (301,3) and

fragmentation pattern (m/z values with peak relative intensities inside parentheses: MS2 (301,4) : 99 (7.5); 154.9 (12.6); 183.9 (100); 184.8 (29.3); 194.8 (15.1); 208.9 (13.1); 209.9 (14.5); 226.9 (75); 227.8 (9.1); 258 (8.4); 282.9 (43.3); 283.9 (7.5)) are consistent with bicyclomycin. The same analysis was performed for acellular extracts of the mycelium from both strains. No differences in UV chromatogram could be observed and if bicyclomycin was present it was at concentrations below the detection level. However the search for extracted ionic currents for $m/z=301$ in negative mode allowed to find a clear signal in extracts from the strain $\Delta bcm::aph2/pMC16$ at the correct retention time and with correct fragmentation and no signal at all in extracts from the strain $\Delta bcm::aph2/pMC16/pJWe21$ (data not shown). This brings us to the conclusion that the BcmH is a bicyclomycin transporter. However other transporters are able to perform its function when it is inactive, with comparable efficiency (the only difference being the presence / absence of trace amounts of bicyclomycin inside the cells).

Strain	Bicyclomycin presence in culture supernatant	Bicyclomycin presence inside the cells
<i>S. aizunensis</i> $\Delta bcm::aph2/pMC16$	+++	+
<i>S. aizunensis</i> $\Delta bcm::aph2/pMC16/pJWe21$	+++	-

Table IV.5. Effects of the absence of the BcmH transporter.

The proposed structures of the bicyclomycin biosynthetic pathway intermediates are confirmed by NMR

The products **5**, **20**, **2** and **12** were obtained in large quantity by the *in vitro* enzymatic reaction scale-up, purified on semi-preparative HPLC and analyzed by the NMR. The ^1H , ^{13}C NMR assignments of the characterized compounds are given below:

cIL:

^1H NMR (500 MHz, DMSO): δ 8.14 (d, $J = 2.7$ Hz, 1H, H^{N} Leu), 8.01 (d, $J = 2.5$ Hz, 1H, H^{N} Ile), 3.76 (dddd, $J = 8.6$, $J = 4.7$, $J = 2.7$, $J = 1.3$ Hz, 1H, $\text{H}\alpha$ Leu), 3.69 (ddd, $J = 4$, $J = 2.5$, $J = 1.2$ Hz, 1H, $\text{H}\alpha$ Ile), 1.84 (m, 1H, $\text{H}\gamma$ Leu),

1.81 (m, 1H, H β Ile), 1.62 (ddd, $J = 13.5$, $J = 8.7$, $J = 4.7$ Hz, 1H, H β_3 Leu), 1.44 (m, 1H, H β_2 Leu), 1.42 (m, 1H, H γ_{12} Ile), 1.17 (ddq, $J = 13.4$, $J = 8.4$, $J = 7.5$ Hz, 1H, H γ_{13} Ile), 0.91 (d, $J = 7.2$ Hz, 3H, H γ_2 Ile), 0.88 (d, $J = 6.6$ Hz, 3H, H δ_2 Leu), 0.85 (d, $J = 6.6$ Hz, 3H, H δ_1 Leu), 0.849 (t, $J = 7.4$ Hz, 3H, H δ_1 Ile).

^{13}C NMR (125 MHz, DMSO): δ 168.3 (C' Leu), 166.8 (C' Ile), 58.9 (C α Ile), 52.3 (C α Leu), 43.6 (C β Leu), 38.2 (C β Ile), 24.3 (C γ_1 Ile), 23.5 (C γ Leu), 23.1 (C δ_2 Leu), 21.7 (C δ_1 Leu), δ 15.1 (C γ_2 Ile), 11.8 (C γ_1 Ile).

Compound 5:

^1H NMR (500 MHz, DMSO): δ 8.15 (d, $J = 2.6$ Hz, 1H, H N Leu), 8.01 (d, $J = 2.6$ Hz, 1H, H N Ile), 4.38 (t, $J = 5.2$ Hz, 1H, HO-C δ_1 Ile), 3.76 (dddd, $J = 8.6$, $J = 4.7$, $J = 2.7$, $J = 1.2$ Hz, 1H, H α Leu), 3.68 (ddd, $J = 4$, $J = 2.7$, $J = 1.2$ Hz, 1H, H α Ile), 3.46 (ddt, $J = 10.5$, $J = 7.4$, $J = 5.2$ Hz, 1H, H δ_{12} Ile), 3.37 (dtd, $J = 10.6$, $J = 7.2$, $J = 5.2$ Hz, 1H, H δ_{13} Ile), 2.04 (m, 1H, H β Ile), 1.85 (m, 1H, H γ Leu), 1.63 (ddd, $J = 13.5$, $J = 8.7$, $J = 4.7$ Hz, 1H, H β_3 Leu), 1.56 (dtd, $J = 13.4$, $J = 7.5$, $J = 4.4$ Hz, 1H, H γ_{13} Ile), 1.45 (ddd, $J = 13.5$, $J = 8.6$, $J = 5.4$ Hz, 1H, H β_2 Leu), 1.30 (dddd, $J = 13.4$, $J = 9.6$, $J = 6.9$, $J = 5.2$ Hz, 1H, H γ_{12} Ile), 0.91 (d, $J = 7.0$ Hz, 3H, H γ_2 Ile), 0.88 (d, $J = 6.6$ Hz, 3H, H δ_2 Leu), 0.85 (d, $J = 6.6$ Hz, 3H, H δ_1 Leu).

^{13}C NMR (125 MHz, DMSO): δ 168.3 (C' Leu), 166.8 (C' Ile), 59.1 (C α Ile), 58.7 (C δ Ile), 52.3 (C α Leu), 43.7 (C β Leu), 34.8 (C γ_1 Ile), 33.1 (C β Ile), 23.5 (C γ Leu), 23.1 (C δ_2 Leu), 21.7 (C δ_1 Leu), 15.5 (C γ_2 Ile).

Compound 20:

^1H NMR (500 MHz, DMSO): δ 8.07 (d, $J = 1.8$ Hz, 1H, H N Ile), 7.50 (d, $J = 1.5$ Hz, 1H, H N Leu), 4.95 (s, 1H, HO-C γ Leu), 4.36 (t, $J = 5.2$ Hz, 1H, HO-C δ_1 Ile), 4.10 (ddt, $J = 9.5$, $J = 3.1$, $J = 1.5$ Hz, 1H, H α Leu), 3.79 (dt, $J = 3.2$, $J = 1.8$ Hz, 1H, H α Ile), 3.44 (ddt, $J = 10.6$, $J = 7.6$, $J = 5.0$ Hz, 1H, H δ_{12} Ile), 3.35 (dtd, $J = 10.6$, $J = 7.2$, $J = 4.8$ Hz, 1H, H δ_{13} Ile), 2.10 (m, 1H, H β Ile), 2.03 (dd, $J = 14.2$, $J = 3.1$ Hz, 1H, H β_3 Leu), 1.59 (dd, $J = 14.3$, $J = 9.5$ Hz, 1H, H β_2 Leu), 1.56 (m, 1H, H γ_{12} Ile), 1.36 (m, 1H, H γ_{13} Ile), 1.19 (s, 3H, H δ_2 Leu), 1.15 (s, 3H, H δ_1 Leu), 0.93 (d, $J = 7.1$ Hz, 3H, H γ_2 Ile).

^{13}C NMR (125 MHz, DMSO): δ 168.9 (C' Leu), 166.7 (C' Ile), 69.4 (C γ Leu), 58.9 (C δ_1 Ile), 58.5 (C α Ile), 51.6 (C α Leu), 45.0 (C β Leu), 34.6 (C γ_1 Ile), 32.4 (C β Ile), 31.1 (C δ_2 Leu), 27.9 (C δ_1 Leu), 15.3 (C γ_2 Ile).

Compound 2:

^1H NMR (500 MHz, DMSO): δ 8.06 (d, $J = 1.6$ Hz, 1H, H $^{\text{N}}$ Ile), 7.48 (d, $J = 1.3$ Hz, 1H, H $^{\text{N}}$ Leu), 4.87 (t, $J = 5.7$ Hz, 1H, HO-C δ_1 Leu), 4.82 (s, 1H, HO-C γ Leu), 4.36 (t, $J = 5.2$ Hz, 1H, HO-C δ_1 Ile), 4.12 (ddt, $J = 9.6$, $J = 2.8$, $J = 1.4$ Hz, 1H, H α Leu), 3.80 (dt, $J = 2.8$, $J = 1.6$ Hz, 1H, H α Ile), 3.44 (ddt, $J = 10.4$, $J = 7.6$, $J = 5.2$ Hz, 1H, H δ_{12} Ile), 3.35 (m, 1H, H δ_{13} Ile), 3.25–3.17 (ABX, $J = 10.7$, $J_{\text{app}} = 5.5$, $J_{\text{app}} = 5.8$ Hz, 2H, H δ_{12} , H δ_{13} Leu), 2.11 (m, 1H, H β Ile), 2.06 (dd, $J = 14.4$, $J = 2.7$ Hz, 1H, H β_3 *pro-R* Leu), 1.58 (dd, $J = 14.4$, $J = 9.6$ Hz, 1H, H β_2 *pro-S* Leu), 1.56 (m, 1H, H γ_{12} Ile), 1.36 (dddd, $J = 13.4$, $J = 9.5$, $J = 7.1$, $J = 5.4$ Hz, 1H, H γ_{13} Ile), 1.08 (s, 3H, H δ_2 Leu), 0.93 (d, $J = 7.1$ Hz, 3H, H γ_2 Ile).

^{13}C NMR (125 MHz, DMSO): δ 169.0 (C' Leu), 166.6 (C' Ile), 71.9 (C γ Leu), 69.7 (C δ_1 Leu), 58.9 (C δ_1 Ile), 58.5 (C α Ile), 51.1 (C α Leu), 40.8 (C β Leu), 34.6 (C γ_1 Ile), 32.4 (C β Ile), 23.6 (C δ_2 Leu), 15.3 (C γ_2 Ile).

Compound 12:

^1H NMR (500 MHz, DMSO): δ 8.73 (d, $J = 1.5$ Hz, 1H, H $^{\text{N}}$ Leu), 8.40 (d, $J = 4.2$ Hz, 1H, H $^{\text{N}}$ Ile), 5.11 (s, 1H, HO-C γ Leu), 5.09 (d, $J = 7.7$ Hz, 1H, HO-C β Leu), 4.41 (t, $J = 5.7$ Hz, 1H, HO-C δ_1 Leu), 3.85 (d, $J = 7.7$ Hz, 1H, H β Leu), 3.73 (dd, $J = 13.5$, $J = 7.2$ Hz, 1H, H δ_{12} Ile), 3.57 (dd, $J = 13.7$, $J = 9.4$ Hz, 1H, H δ_{13} Ile), 3.52 (dt, $J = 4.2$, $J = 1.4$ Hz, 1H, H α Ile), 3.42 (dd, $J = 10.9$, $J = 5.6$ Hz, 1H, H δ_{12} Leu), 3.29 (dd, $J = 10.9$, $J = 5.6$ Hz, 1H, H δ_{13} Leu), 2.13 (m, 1H, H β Ile), 1.71 (ddd, 1H, $J = 16.0$, $J = 7.2$, $J = 4.6$ Hz, H γ_{12} Ile), 1.34 (dt, $J = 16.0$, $J = 10.0$ Hz, 1H, H γ_{13} Ile), 1.15 (s, 3H, H δ_2 Leu), 0.95 (d, $J = 7.0$ Hz, 3H, H γ_2 Ile).

^{13}C NMR (125 MHz, DMSO): δ 171.3 (C' Ile), 168.1 (C' Leu), 86.4 (C α Leu), 77.0 (C γ Leu), 70.2 (C β Leu), 66.7 (C δ_1 Leu), 61.3 (C δ_1 Ile), 58.9 (C α Ile), 33.5 (C γ_1 Ile), 40.0 (C β Ile), 23.9 (C δ_2 Leu), 19.5 (C γ_2 Ile).

The chemical structures of the different intermediates of the biosynthesis pathway of bicyclomycin were determined by NMR spectroscopy

and high-resolution MS analysis. ^1H and ^{13}C resonances of compounds **cIL**, **20**, **2**, and **12** were assigned using 1D ^1H , 1D ^{13}C DEPTQ, 2D ^1H - ^1H COSY and TOCSY, 2D ^1H - ^{13}C HSQC and HMBC experiments. The chemical structures were unambiguously identified through the analysis of scalar ^1H - ^1H correlations (2D COSY and TOCSY) and ^1H - ^{13}C correlations (2D HSQC and HMBC).

The configuration of the new stereogenic centers in compounds **20**, **2** and **12** relied on a NOE-based and J -based NMR strategy (Schmidt 2007, Bifulco *et al.* 2007, Matsumori *et al.* 1999). Homonuclear vicinal coupling constants ($^3J_{\text{H,H}}$) and heteronuclear coupling constants ($^3J_{\text{C,H}}$) follow a Karplus-type dihedral angle dependence which can be readily utilized for stereochemical analysis. In addition, two-bond carbon-proton coupling constants ($^2J_{\text{C,H}}$) also provide conformational and configurational information in oxygen-substituted two-carbon fragments. Indeed, when an oxygen substituent on a carbon atom is *gauche* to the nearby proton, $^2J_{\text{C,H}}$ is large (typically 5–6 Hz), and when it is *anti*, the $^2J_{\text{C,H}}$ value decreases (typically 0–3 Hz). These $^2J_{\text{C,H}}$ coupling constants were particularly useful for this stereochemical analysis as the new stereogenic centers in the compounds involved hydroxylation reactions. The heteronuclear $^{2/3}J_{\text{C,H}}$ coupling constants were extracted from two-dimensional ^1H - ^{13}C IPAP-HSQMBC (In-Phase Anti-Phase Heteronuclear Single quantum Multiple Bond Correlation) experiments.

Stereochemical analysis of compounds **2** and **12**

For the compound **2**, the first step in determining the relative configuration of the gamma carbon of the leucyl chain consisted in the conformational analysis of $\text{C}\alpha$ - $\text{C}\beta$ rotamer (Table IV.6). The large differences in $^3J_{\text{H}\alpha\text{-H}\beta}$ values (9.6 Hz and 2.8 Hz for upfield and downfield $\text{H}\beta$ protons, respectively) indicated the predominance of a major $\text{C}\alpha$ - $\text{C}\beta$ rotamer. The measurement of $^3J_{\text{H}\beta\text{-CO}}$ and $^3J_{\text{H}\alpha\text{-C}\gamma}$ provided unambiguous stereospecific assignment of $\text{H}\beta$ methylenic protons and determination of the χ_1 angle around -60° for $\text{C}\alpha$ - $\text{C}\beta$ rotamer. This analysis was further confirmed by

observed NOEs involving HN, H α and H β protons. Next, the conformational and configurational analysis in the C β -C γ fragment relied on the measurement of $^{2,3}J_{CH}$ coupling constants involving H β protons and C γ , C δ 1 and C δ 2 carbons, together with NOE analysis (Supplementary Table 7). The NMR data were consistent with a predominant rotamer around C β -C γ bond, with a χ^2 angle (C α -C β -C γ -O) around +60°, and the configuration of C γ atom was shown to be *R*.

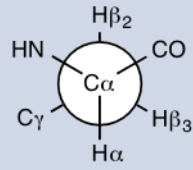
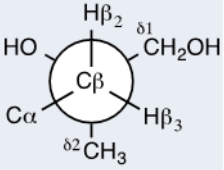
Fragment	Rotamer	$^3J_{H-H}$ (Hz)	Selected $^{2,3}J_{CH}$ (Hz)	Key NOEs
C α -C β		$^3J_{H\beta_2H\alpha} = 9.6$ $^3J_{H\beta_3H\alpha} = 2.8$	$^3J_{H\beta_2CO} = 1.9$ $^3J_{H\beta_3CO} = 2.5$ $^3J_{H\alpha C\gamma} = 1.5$	HN-H β_2 (m) HN-H β_3 (abs) HN-HO $_{C\gamma}$ (m) H α -H β_2 (abs) H α -H β_3 (s)
C β -C γ			$^2J_{H\beta_2C\gamma} = 4.6$ $^2J_{H\beta_3C\gamma} = 3.0$ $^3J_{H\beta_2C\delta_1} = 2.2$ $^3J_{H\beta_3C\delta_1} = 1.9$ $^3J_{H\beta_2C\delta_2} = 5.0$ $^3J_{H\beta_3C\delta_2} = 2.0$	H α -Me $_{\delta_2}$ (s) H β_3 -Me $_{\delta_2}$ (m) H α -OH (m) H β_2 -OH (m)

Table IV.6. Dominant rotamers for the leucyl chain of compound **2 along with their relative configurations.** NOE correlations are classified into strong (s), medium (m) and weak (w) intensities (abs = absent).

For the compound **12**, the null value of $^2J_{H\beta C\gamma}$ in the C β -C γ fragment indicated an *anti* arrangement of H β proton and the hydroxyl group on C γ carbon, while the small values of $^3J_{H\beta C\delta_1}$ and $^3J_{H\beta C\delta_2}$ allowed to establish the *gauche* arrangement of H β with respect to the methyl and CH₂OH substituents on C γ . The NOE correlation observed between the methyl group on C γ and the hydroxyl group on C β supported the rotamer shown in Table S2 and thus the *S* configuration for C β . The values of heteronuclear coupling constants involving H β together with the NOE correlations observed between the amide proton and the protons of the different groups on the carbon C γ were in agreement with a major rotamer around C α -C β with the *S* configuration for the C α atom (Table IV.7).

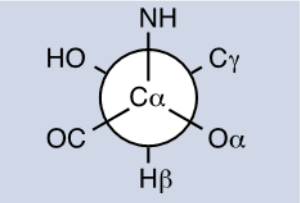
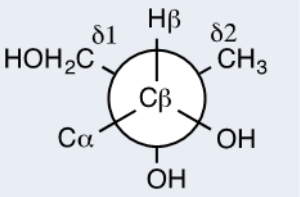
Fragment	Rotamer	Selected $^{2,3}J_{CH}$ (Hz)	Key NOEs
$C_{\alpha}-C_{\beta}$		$^2J_{H\beta C\alpha} = 4.2$ $^3J_{H\beta CO} = 2.1$	HN- H_{β} (m) HN- $H_{\delta 12}$ (m) HN- $H_{\delta 13}$ (m) HN-Me $_{\delta 2}$ (w) HN-HO $_{C\gamma}$ (w)
$C_{\beta}-C_{\gamma}$		$^2J_{H\beta C\gamma} \approx 0$ $^3J_{H\beta C\delta 1} = 1.2$ $^3J_{H\beta C\delta 2} = 1.5$	H_{β} -Me $_{\delta 2}$ (s) H_{β} - $H_{\delta 12}$ (s) H_{β} - $H_{\delta 13}$ (s) HO $_{C\beta}$ -Me $_{\delta 2}$ (m)

Table IV.7. Dominant rotamers for the leucyl chain of compound **12 along with their relative configurations.** NOE correlations are classified into strong (s), medium (m) and weak (w) intensities.

HMRS analysis of compound **19** and bicyclomycin

As bicyclomycin and dihydrobicyclomycin (compound **19**) are molecules that have been previously described, we did not perform a NMR analysis on them, but in complement to mass spectra already described above, we also performed a HRMS analysis on purified compounds obtained by *in vitro* reactions. Compound **19** gives an exact m/z ratio in negative mode of 303.1180, with a delta value of -2.1027 ppm as compared to the proposed crude formula of $C_{12}H_{19}O_7N_2^-$. Bicyclomycin gives an exact m/z ratio in negative mode of 301.1023, with a delta value of -2.3823 as compared to the proposed crude formula of $C_{12}H_{17}O_7N_2^-$.

All this data confirm that the structures of the pathway intermediates presented on Figure IV.7 (and hence the whole pathway) are correct.

Discussion

In this paper we demonstrated that bicyclomycin is synthesized by a CDPS-dependent pathway, which is the first known case of a medically significant molecule to be synthesized in this manner. This is also the first characterized CDPS-containing cluster of such complexity: eight genes

coding for eight proteins of which seven are directly involved in biosynthesis, as all previously characterized clusters encoded no more than three tailoring enzymes in addition to the CDPS (Gu *et al.* 2013). The largest CDPS cluster characterized to this day is the nocazine cluster studied in *Nocardiopsis dassonvillei* by Giessen and Marahiel (2013). Besides the CDPS encoding gene, it also contains two genes coding for methyltransferases, one methylating the DKP ring, the other the tyrosine lateral chain, and two genes coding for the two subunits of cyclodipeptide oxidase, an enzyme dehydrogenating the lateral chains. In the case of bicyclomycin BCG we have a CDPS gene followed by six genes coding for tailoring enzymes, acting in a specific, well established order.

Concerning the reactions that we propose to be catalyzed by the tailoring enzymes, hydroxylations are one of the canonical reactions in the repertoire of both oxoglutarate/iron-dependent dioxygenases and cytochrome p450 monooxygenases; desaturations are also well documented (Martinez and Hausinger 2015, Sono *et al.* 1996). The reaction that we propose to be catalyzed by BcmB (hydroxylation and the ether bridge formation) is more unusual, but not unheard of. For example, the 2-oxoglutarate-iron-dependent dioxygenase CAS, pivotal in the clavulanic acid biosynthesis in *Streptomyces clavuligerus*, is known to catalyze three distinct reactions: a hydroxylation, an ether bridge formation and a dehydrogenation (Zhang *et al.* 2000b). Loline biosynthesis by *Epichloë uncinata* involves the action of the enzyme LolO, belonging to this family of dioxygenases, which forms an ether bridge too (Pan *et al.* 2014). Moreover, the suggested reaction mechanism for ether bridge formation by oxoglutarate/iron-dependent dioxygenases involves an abstraction of a hydrogen atom, with a formation of carbon radical and a subsequent carbocation formation (Martinez and Hausinger 2015, Pan *et al.* 2015), the carbocation leading an attack on the hydroxyl group. This could explain the switch in stereochemistry on the alpha carbon of leucine moiety observed during the bicyclomycin biosynthesis (Bradley *et al.* 1996).

The identification of the genes involved in the pathway could lead to the optimization of the production of bicyclomycin by cluster refactoring and promoter replacement, for example.

The very low levels of production after complementation of the $\Delta bcm::\Omega att1aac$ mutant with the pJWe18 plasmid may be explained either by the polar effect of the two terminators present in the disruption cassette on the downstream genes, or by the fact that the *ermE** promoter's activity is low in *S. aizunensis*. We privilege the latter option, as no untransformed cIL could be detected upon LC-MS analysis.

Our inability to observe any heterologous expression in standard *Streptomyces* hosts when whole cluster is cloned on an integrative vector is probably due to the absence of appropriate transcriptional activator in the host strain.

Knowing that the bicyclomycin is efficiently exported from the bacterial cells even when the BcmH transporter is absent allows to explain why the cluster homologue in *S. ossamyceticus* possesses no homolog of the *bcmH* gene and why no homolog can be found anywhere in *S. ossamyceticus* genome. The active export in absence of the specific transporter is a phenomenon already documented amongst *Streptomyces*. For example, *Streptomyces venezuelae* is able to export jadomycin efficiently even when the gene coding for the specific MFS1 transporter within the cluster is disrupted (Forget *et al.* 2017).

A very interesting perspective is offered by the fact that homologous clusters could be found in many other bacteria. First, as the deduced protein sequence identity is pretty high for clusters found in Actinobacteria, we may potentially find organisms capable of producing bicyclomycin other than *S. aizunensis* and *S. cinnamoneus*. Moreover, the homologues found in *Proteobacteria* are very interesting, as the low sequence identity of the BcmF dioxygenase homologue and the even lower sequence identity of the BcmH transporter encoded on the cluster combined with alternative cluster organization suggest that the molecule produced may be different. Further study of these clusters could potentially allow to discover new natural analogs of bicyclomycin which could have pharmacological properties more

favorable than the original molecule. Such an opportunity cannot be easily dismissed in the present state of shortage in new antibiotic discovery combined with the apparition of multiresistant strains in the hospitals. This hypothesis is supported by the fact that BcmF is catalyzing the dehydrogenation at the isoleucine lateral chain and previous studies showed that this dehydrogenation was not essential for bicyclomycin activity (Park *et al.* 1995). Thus we can imagine that a different kind of modification is introduced on isoleucyl side chain by proteobacterial BcmF homologues. Moreover, we can imagine the use of all characterized enzymes to generate novel unnatural natural products by the combinatorial biosynthesis approach. Besides purely biological or chemical approaches, there exists synthetic biology with its methods like semisynthesis, mutasynthesis or combinatorial biology (Goss *et al.* 2012, Giessen and Marahiel, 2015). This kind of approach was already successful in developing plinabulin, a promising diketopiperazine anticancer agent currently under clinical trials, provoking a massive selective vascular collapse in the cancerous tissue. Plinabulin was obtained by combining enzymes from phenylalanine biosynthetic pathway from the fungus *Aspergillus ustus*, albonoursin biosynthetic pathway from the bacterium *Streptomyces noursei* and chemical modification of the final product (Bertelsen *et al.* 2011). One can easily imagine a similar approach using the genes from the bicyclomycin BGC.

As the cluster contains no transcription factor and knowing that heterologous expression attempts were insuccesfull, identifying the regulators influencing the cluster expression would be an interesting challenge. It could subsequently allow imagining the cluster refactoring.

To summarize this work, we present the first CDPS containing cluster of such complexity, and the first case of a medically relevant molecule to be synthesized by a CDPS-dependant pathway. Combining the approaches of deletion mutants in the producer strain and the *in vitro* characterization of biosynthetic enzymes, we were able to decipher entirely the bicyclomycin biosynthetic pathway.

Acknowledgements

We would like to thank Jordan Duarte for help with construction of *E. coli* expression vectors, Alain Lecoq for cIL chemical synthesis and Robert Thai, Nicolas Canu and Steven Dubois for help with LC-MS analysis. We would like thank DKP-COMBIBIO, UPSud JW doctoral fellowship and JPIAMR Evobiotics for funding this work.

Conclusion générale

Comme indiqué dans l'introduction, l'objectif de cette thèse était triple : explorer la diversité des CDPS et des produits qu'elles sont capables de synthétiser, étudier le rôle biologique de certaines dicétopipérazines chez les organismes producteurs et caractériser la voie de biosynthèse d'une dicétopipérazine fortement modifiée. Chacun de ces points fait l'objet d'un chapitre dans ce manuscrit et les résultats obtenus ont été discutés dans chaque chapitre. Cette conclusion résume brièvement les résultats obtenus, les perspectives qu'ils ouvrent et les questions qu'ils soulèvent.

Concernant le premier volet de la thèse, l'exploration de la diversité des CDPS, j'ai pu participer dans un travail collectif qui a abouti à la caractérisation d'une cinquantaine de nouvelles CDPS. Ce travail a permis de démontrer l'utilisation de dix-sept des acides aminés protéinogènes pour la biosynthèse de cyclodipeptides, notamment des acides aminés polaires et chargés (alors que les CDPS caractérisées auparavant incorporaient principalement des acides aminés hydrophobes). De plus, cette étude a mis en évidence deux familles de CDPS, tout en énonçant l'existence probable d'une troisième famille. Finalement, les résultats obtenus ont permis de jeter les bases de la prédiction des substrats utilisés par les CDPS pour la biosynthèse des dicétopipérazines à partir de la séquence protéique, grâce à l'identification de motifs de résidus responsables de la sélection du substrat. Un développement de ce travail pourrait conduire à un outil bioinformatique générant de telles prédictions sur la base de la séquence, comme cela se pratique couramment pour les NRPS et pour les PKS (Blin *et al.* 2017). Ce travail est aussi à la base de la génération d'analogues de dicétopipérazines naturelles par mutasynthèse, actuellement en cours (Canu *et al.* données non publiées) ainsi que de travaux de biologie structurale sur la famille XYP des CDPS (Seguin *et al.* données non publiées).

Concernant le deuxième objectif, à savoir l'étude du rôle biologique d'une dicétopipérazine chez l'organisme producteur, le modèle choisi était *Streptomyces venezuelae* et son *cluster* homologue du *cluster* de biosynthèse de la mycocyclosine de *Mycobacterium tuberculosis*. Malheureusement, même s'il les résultats obtenus indiquent qu'il est possible que ce *cluster* dirige la biosynthèse de la mycocyclosine, il n'a pas été possible de trouver dans quelles conditions ce composé est produit ni quel pourrait être son rôle biologique chez *S. venezuelae*. Ces résultats suscitent de nombreuses questions ; ils illustrent la difficulté qu'il y a à trouver les conditions d'expression d'un *cluster* cryptique et celle encore plus importante à établir son rôle biologique chez le producteur.

Quant au troisième objectif, l'élucidation de la voie de biosynthèse de la bicyclomycine, il a été atteint. Mon travail a permis d'élucider complètement cette voie, en identifiant le *cluster* de gènes dirigeant cette biosynthèse et en combinant ensuite deux approches complémentaires, d'une part l'analyse fonctionnelle du *cluster* et l'identification d'intermédiaires de biosynthèse accumulés par des mutants de délétion de chacun des gènes de biosynthèse, d'autre part l'étude *in vitro* des réactions catalysées par les enzymes purifiées et la caractérisation des différents produits de réaction. La voie de biosynthèse de la bicyclomycine est la voie dépendante d'une CDPS la plus complexe à ce jour, la seule faisant intervenir plusieurs enzymes d'oxydoréduction. Elle constitue le premier exemple de voie de biosynthèse d'une molécule d'utilité médicale dépendante d'une CDPS. Ceci permet de démontrer que les CDPS sont plus qu'une curiosité biochimique. J'ai également pu trouver chez différents micro-organismes des *clusters* associant un gène codant une CDPS prédite pour catalyser la formation de cIL avec plusieurs gènes codant des dioxygénases ou des cytochromes P450. Ceci ouvre la voie vers la caractérisation d'analogues naturels de la bicyclomycine. En effet, la faible similitude entre certaines enzymes, par exemple entre les homologues de BcmF, laisse supposer que ces enzymes pourraient catalyser des réactions conduisant à des modifications différentes de celles qui sont observées chez la bicyclomycine. A moyen terme, dans une approche de biologie combinatoire, des gènes caractérisés lors de ce travail

ou présents dans des clusters similaires pourraient être combinés de façon nouvelle et coexprimés pour générer de nouvelles molécules de la famille des dicétopipérazines potentiellement bioactives.

Bibliographie

- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:3389–3402.
- Alqahtani, N., Porwal, S.K., James, E.D., Bis, D.M., Karty, J.A., Lane, A.L. & Viswanathan, R. (2015) Synergism between genome sequencing, tandem mass spectrometry and bio-inspired synthesis reveals insights into nocardioazine B biogenesis. *Organic & Biomolecular Chemistry* 13:7177-7192.
- Aravind, L., de Souza, R.F. & Iyer, L.M. (2010) Predicted class-I aminoacyl tRNA synthetase-like proteins in non-ribosomal peptide synthesis. *Biology Direct* 5:48.
- Armirotti, A., Millo, E. & Damonte, G. (2007) How to discriminate between leucine and isoleucine by low energy ESI-TRAP MSn. *Journal of the American Society for Mass Spectrometry* 18:57–63.
- Arraj, J.A. & Marinus, M.G. (1983) Phenotypic reversal in dam mutants of *Escherichia coli* K-12 by a recombinant plasmid containing the *dam+* gene. *Journal of Bacteriology* 153:562-5.
- Bajad, S.U., Lu, W., Kimball, E.H., Peterson, C. & Rabinowitz, J.D. (2006) Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *Journal of Chromatography A* 1125:76–88.
- Barry, S.M., Kers, J.A., Johnson, E.G., Song, L., Aston, P.R., Patel, B., Krasnoff, S.B., Crane, B.R., Gibson, D.M., Loria, R. & Challis, G. (2012) Cytochrome P450-catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. *Nature Chemical Biology* 8:814-816.
- Belin, P., Moutiez, M., Lautru, S., Seguin, J., Pernodet, J.L. & Gondry, M. (2012) The nonribosomal synthesis of diketopiperazines in tRNA-dependent cyclodipeptide synthase pathways. *Natural Product Reports* 29:961-979.
- Belin, P., Le Du, M.H., Fielding, A., Lequin, O., Jacquet, M., Charbonnier, J.B., Lecoq, A., Thai, R., Courçon, M., Masson, C., Dugave, C., Genet, R., Pernodet, J.L. & Gondry, M. (2009) Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in *Mycobacterium tuberculosis*. *PNAS* 106(18):7426-7431.
- Bentley, J., Hyatt, L.S., Ainley, K., Parish, J.H., Herbert, R.B. & White, G.R. (1993) Cloning and sequence analysis of an *Escherichia coli* gene conferring bicyclomycin resistance. *Gene* 127:117-120.
- Bertelsen, L.B., Shen, Y.Y., Nielsen, T., Stødkilde-Jørgensen, H., Lloyd, G.K., Siemann, D.W. & Horsman, M.R. (2011) Vascular effects of plinabulin (NPI-2358) and the influence on tumour response when given alone or combined with radiation. *International Journal of Radiation Biology* 87:1126-1134.
- Bibb, M.J. & Jansen, G.R. (1986) *Fifth International Symposium on the Genetics of Industrial Microorganisms, Zagreb*, pp. 309-318.
- Bifulco, G., Dambrusco, P., Gomez-Paloma, L. & Riccio, R. (2007) Determination of relative configuration in organic compounds by NMR spectroscopy and computational methods. *Chemical Reviews* 107:3744-3779.
- Bissett, J. (1983) Notes on *Tolypocladium* and related genera. *Revue canadienne de botanique* 61:1311-1329.
- Blin, K., Wolf, T., Chevrette, M.G., Lu, X., Schwalen, C.J., Kautsar, S.A., Suarez-Duran, H.G., de Los Santos, E.L.C., Kim, H.U., Nave, M., Dickschat, J.S., Mitchell, D.A., Shelest, E., Breitling, R., Takano, E., Lee, S.Y., Weber, T. & Medema, M.H. (2017) antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acid Research* gkx319.

- Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E. & Weber, T. (2013) AntiSMASH 2.0 — a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acid Research* 41:W204-W212.
- Bonnefond, L., Arai, T., Sakaguchi, Y., Suzuki, T., Ishitani, R. & Nureki, O. (2011). Structural basis for nonribosomal peptide synthesis by an aminoacyl-tRNA synthetase paralog. *PNAS* 108:3912-3917.
- Borthwick, A.D. (2012) 2,5-diketopiperazines: synthesis, reactions, medicinal chemistry and bioactive natural products. *Chemical Reviews* 112:3641-3716.
- Bradley, E., Herbert, R., Lawrie, K., Khan, J., Moody, C. & Young, D. (1996) The biosynthesis of the *Streptomyces* Antibiotic Bicyclomycin. *Tetrahedron Letters* 37:6935-6938.
- Braud, S., Moutiez, M., Belin, P., Abello, N., Drevet, P., Zinn-Justin, S., Courçon, M., Masson, C., Dassa, J., Charbonnier, J.B., Boulain, J.C., Ménez, A., Genet, R. & Gondry, M. (2005) Dual expression system suitable for high-throughput fluorescence-based screening and production of soluble proteins. *Journal of Proteome Research* 4:2137–2147.
- Brock, G.B., McMahon, C.G., Chen, K.K., Costigan, T., Shen, W., Watkins, V., Anglin, G. & Whitaker, S. (2002) Efficacy and safety of tadalafil for the treatment of erectile dysfunction: results of integrated analyses. *The Journal of Urology* 168:1332-1336.
- Challis, G.L. (2008) Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology* 154:1555-1569.
- Challis, G.L., Ravel, J. & Townsend, C.A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemical Biology* 7: 211–224.
- Chater, K.F., Biró, S., Lee, K.J., Palmer, T. & Schrempf, H. (2010). The complex extracellular biology of *Streptomyces*. *FEMS Microbiology Reviews* 34:171-198.
- Chen, Y.H., Liou, S.E. & Chen, C.C. (2004) Two-step mass spectrometric approach for the identification of diketopiperazines in chicken essence. *European Food Research Technology* 218:589–597.
- Choulet, F., Aigle, B., Gallois, A., Mangeno, S., Gerbaud, C., Truong, C., Francou, F.X., Fourier, C., Guérineau, M., Decaris, B., Barbe, V., Pernodet, J.L. & Leblond, P. (2006) Evolution of the terminal regions of the *Streptomyces* linear chromosome. *Molecular Biology and Evolution* 23:2361-9
- Combet, C., Blanchet, C., Geourjon, C. & Deléage, G. (2000) NPS@: Network protein sequence analysis. *Trends in biochemical sciences* 25:147-150.
- Courtwright, D.T. (2009) Forces of habit drugs and the making of the modern world. *Harvard University Press*.
- Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Research* 14:1188–1190.
- Cryle, M.J., Bell, S.G. & Schlichting, I. (2010). Structural and biochemical characterization of the cytochrome P450 CypX (CYP134A1) from *Bacillus subtilis*. A cyclo-L-leucyl-L-leucyl dipeptide oxidase. *Biochemistry* 49: 7282-7296.
- Davies, J. & Ryan, K.S. (2011) Introducing the parvome : bioactive compounds in the microbial world. *ACS Chemical Biology* 7:252-259.
- Davies, J. (2013) Specialized microbial metabolites: functions and origins. *The Journal of Antibiotics* 66:361-364.
- Donovick, R., Pagano, J.F., Stout, H.A. & Weinstein, M.J. (1955) Thiostrepton, a new antibiotic. I. *In vitro* studies. *Antibiotics annual* 3:554-559.
- Dumas, V.G., Defelipe, L.A., Petruk, A.A., Turianski, A.G. & Marti, M.A. (2014) QM/MM study of the CAC coupling reaction mechanism of CYP121, an essential Cytochrome p450 of *Mycobacterium tuberculosis*. *Proteins Structure Function and Bioinformatics* 82:1004-1021.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, 1792–1797.
- Erickson, J.M. & Feeney, P. (1974) Sinigrin: a chemical barrier to the black swallowtail butterfly, *Papilio polyxenes*. *Ecology* 55:103-111.
- Falick, A.M., Hines, W.M., Medzihradszly, K.F., Baldwin, M.A. & Gibson, B.W. (1993) 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.
- Finking, R. & Marahiel, M. (2004) Biosynthesis of nonribosomal peptides. *Annual Review of Microbiology* 58:453–488.
- Fischer, G., Wenner, T., Decaris, B. & Leblond, P. (1998) Chromosomal arm replacement generates a high level of intraspecific polymorphism in the terminal inverted repeats of the linear chromosomal DNA of *Streptomyces ambofaciens*. *PNAS* 95:14296–301.
- Flårdh, K. & Buttner, M.J. (2009) *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nature Reviews Microbiology* 7:36–49.
- Flett, F., Mersinias, V. & Smith, C.P. (1997). High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting *streptomyces*. *FEMS Microbiology Letters* 155 : 223–229.
- Fonvielle, M., Le Du, M.H., Lequin, O., Lecoq, A., Jacquet, M., Thai, R., Dubois, S., Grach, G., Gondry, M. & Belin, P. (2013) Substrate and Reaction Specificity of *Mycobacterium tuberculosis* Cytochrome P450 CYP121: insights from biochemical studies and crystal structures. *The Journal of Biological Chemistry* 288:17347–17359.
- Forget, S.M., McVey, J., Vining, J.C. & Jakeman, D.L. (2017) *Streptomyces venezuelae* ISP5230 Maintains Excretion of Jadomycin upon Disruption of the MFS Transporter JadL Located within the Natural Product Biosynthetic Gene Cluster. *Frontiers in Microbiology* 8:432.
- Fukushima, K., Yazawa, K. & Tadashi, A. (1972) Biological activities of albonoursin. *The Journal of Antibiotics* 26:175–176.
- Giessen, T.W. & Marahiel, M.A. (2014) The tRNA-dependent biosynthesis of modified cyclic dipeptides. *International Journal of Molecular Sciences* 15:14610–14631.
- Giessen, T.W., von Tesmar, A.M. & Marahiel, M.A. (2013a) Insights into the generation of structural diversity in a tRNA-dependent pathway for highly modified bioactive cyclic dipeptides. *Chemistry & Biology* 20:828–838.
- Giessen, T. W., von Tesmar, A. M. & Marahiel, M. A. (2013b) A tRNA-dependent two-enzyme pathway for the generation of singly and doubly methylated ditryptophan 2,5-diketopiperazines. *Biochemistry* 52:4274–4283.
- Gil, S., Espinosa, J.F. & Parella, T. (2010) IPAP-HSQMBC: measurement of long-range heteronuclear coupling constants from spin-state selective multiplets. *Journal of Magnetic Resonance* 207:312–321.
- Gomez-Escribano, J.P. & Bibb, M.J. (2010) Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microbial Biotechnology* 4:207–215.
- Gondry, M., Lautru, S., Fusai, G., Meunier, G., Ménez, A. & Genet, R. (2001). Cyclic dipeptide oxidase from *Streptomyces noursei*: isolation, purification and partial characterization of a novel amino acyl α,β -dehydrogenase. *European Journal of Biochemistry* 268:1712–1721.
- Gondry, M., Sauguet, L., Belin, P., Thai, R., Amouroux, R., Tellier, C., Tuphile, K., Jacquet, M., Braud, S., Courçon, M., Masson, C., Dubois, S., Lautru, S., Lecoq, A., Hashimoto, S., Genet, R. & Pernodet, J.L. (2009) Cyclodipeptide synthases are a family of tRNA-dependent peptide bond-forming enzymes. *Nature Chemical Biology* 5:414–420.
- Goss, J.M., Shankar, S. & Abou Fayad, A. (2012) The generation of “unnatural” products: Synthetic biology meets synthetic chemistry. *Natural Product Reports* 29:870–889.
- Gottardi, E.M., Krawczyk, J.M., von Suchodoletz, H., Schadt, S., Mühlenweg, A., Uguru, G.C., Pelzer, S., Fiedler, H.P., Bibb, M.J., Stach, J.E.M. & Süssmuth, R.D. (2011) Abyssomicin Biosynthesis: Formation of an Unusual Polyketide, Antibiotic-Feeding Studies and Genetic Analysis. *ChemBiochem* 12:1401–1410.
- Gouy, M., Guindon, S. & Gascuel, O. (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* 27:221–224.
- Green, M.R. & Sambrook, J. (2012). Molecular Cloning : A Laboratory Manual. *Cold Spring Harbor Laboratory Press*.
- Gregory M.A., Till, R., & Smith, M.C. (2003) Integration site for *Streptomyces* phage phiBT1 and development of site-specific integrating vectors. *Journal of Bacteriology*: 185:5320–5323.

- Grove, D.C. & Randall, W.A. (1955) Assay methods of antibiotics. *Medical Encyclopedia Inc., New York*.
- Gu, B., He, S., Yan, X. & Zhang, L. (2013). Tentative biosynthetic pathways of some microbial diketopiperazines. *Applied Microbiology and Biotechnology* 97:8439-8453.
- Gulick, A.M. (2016) Structural insight into the necessary conformational changes of modular nonribosomal peptide synthetases. *Current Opinion in Chemical Biology* 35:89-96.
- Gust, B., Challis, G.L., Fowler, K., Kieser, T. & Chater, K.F. (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *PNAS* 100:1541-1546.
- Guzmán, G. (2005) Species diversity of the genus *Psilocybe* (Basidiomycotina, Agaricales, Strophariaceae) in the world mycobiota, with special attention to hallucinogenic properties. *International Journal of Medicinal Mushrooms* 7:305-331.
- Hamano, Y., Kuzuyama, T., Itoh, N., Furihata, K., Seto, H. & Dairi, T. (2002) Functional Analysis of Eubacterial Diterpene Cyclases Responsible for Biosynthesis of a Diterpene Antibiotic, Terpentecin. *The Journal of Biological Chemistry* 277:37098-37104.
- Haeder, S., Wirth, R., Herz, H. & Spiteller, D. (2009) Candicidin-producing *Streptomyces* support leaf-cutting ants to protect their fungus garden against the pathogenic fungus *Escovopsis*. *PNAS* 12:4742-4746.
- Healy, F.G., Wach, M., Krasnoff, S.B., Gibson, D.M. & Loria, R. (2000) The *txtAB* genes of the plant pathogen *Streptomyces acidiscabies* encode a peptide synthetase required for phytotoxin thaxtomin A production and pathogenicity. *Molecular Microbiology* 38:794-804.
- Jacques, I., Moutiez, M., Witwinowski, J., Darbon, E., Martel, C., Seguin, J., Favry, E., Thai, R., Lecoq, A., Dubois, S., Pernodet, J.L., Gondry, M., & Belin P. (2015). Analysis of 51 cyclodipeptide synthases reveals the basis for substrate specificity. *Nature chemical biology* 11:721-727.
- Jeedigunta, S., Krenisky, J.M. & Kerr, R.G. (2000) Diketopiperazines as advanced intermediates in the biosynthesis of ecteinascidins. *Tetrahedron* 56:3303-3307.
- Johnson, R.S., Martin, S.A., Biemann, K., Stults, J.T. & Watson, J.T. (1987) Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Analytical Chemistry* 59:2621-2625.
- Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. & Hopwood, D.A. (2000). Practical *Streptomyces* Genetics. *The John Innes Foundation Norwich*.
- Kim, K., Kim, N.J., Kim, S.Y., Kim, I.H. & Lee, G.R. (2015) Cyclo(Phe-Pro) produced by the human pathogen *Vibrio vulnificus* inhibits host innate immune responses through the NF- κ B pathway. *Infection and Immunity* 83:1150-1161.
- Kirby, R. (2011). Chromosome diversity and similarity within the Actinomycetales. *FEMS Microbiology Reviews* 319:1-10.
- Lautru, S., Gondry, M., Genet, R. & Pernodet, J.L. (2002). The albonoursin gene cluster of *S. noursei*: biosynthesis of diketopiperazine metabolites independent of nonribosomal peptide synthetases. *Chemistry & Biology* 9: 1355-1364.
- Lazos, O., Tosin, M., Slusarczyk, A.L., Boakes, S., Cortés, J., Sidebottom, P.J. & Leadlay, P.F. (2010) Biosynthesis of the putative siderophore erythrochelin requires unprecedented crosstalk between separate nonribosomal peptide gene clusters *Chemistry & Biology* 17:160-173.
- Letunic, I. & Bork, P. (2011) Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research* 39:W475-478.
- Letunic, I. & Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23:127-128.
- Li, Y., Lai, Y.M., Lu, Y., Yang, Y.L. & Chen, S. (2014) Analysis of the biosynthesis of antibacterial cyclic dipeptides in *Nocardiopsis alba*. *Archives of Microbiology* 196:765-774.
- Ma, C., Yang, X. & Lewis, P.J. (2016). Bacterial transcription as a Target for Antibacterial Drug Development. *Microbiology and Molecular Biology Reviews* 80:139-160.
- Malik, M., Li, L., Zhao, X., Kerns, R.J., Berger, J.M. & Drlica, K. (2014) Lethal synergy involving bicyclomycin: an approach for reviving old antibiotics. *Journal of Antimicrobial Chemotherapy* 69:3227-3235.

- Manteca, A., Claessen, D., Lopez-Iglesias, C. & Sanchez, J. (2007) Aerial hyphae in surface cultures of *Streptomyces lividans* and *Streptomyces coelicolor* originate from viable segments surviving an early programmed cell death event. *FEMS Microbiology Letters* 274:118-125.
- Martinez, S. & Hausinger, R.P. (2015) Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases. *The Journal of Biological Chemistry* 290:20702-20711.
- Matsuda, S. & Koyatu, S. (2000). Mechanism of action of cyclosporine. *Immunopharmacology* 47:119-125.
- Matsumori, N., Kaneno, D., Murata, M., Nakamura, H. & Tachibana, K. (1999) Stereochemical Determination of Acyclic Structures Based on Carbon-Proton Spin-Coupling Constants. A Method of Configuration Analysis for Natural Products. *Journal of Organic Chemistry* 64:866-876.
- McCormick, M.H., McGuire, J.M., Pittenger, R.C. & Stark, W.M. (1955) Vancomycin, a new antibiotic. I. Chemical and biologic properties. *Antibiotics Annual* 3:606-611.
- McLean, K.J., Carroll, P., Lewis, D.G., Dunford, A.J., Seward, H.E., Neeli, R., Cheesman, M.R., Marsollier, L., Douglas, P., Smith, W.E., Rosenkrands, I., Cole, S.T., Leys, D., Parish, T. & Munro, A.W. (2008). Characterization of active site structure in CYP121 a cytochrome P450 essential for viability of *Mycobacterium tuberculosis* H37Rv. *Journal of Biological Chemistry* 283:33406-33416.
- Miyamura, S., Ogasawara, N., Otsuka, H., Niwayama, S., Tanaka, H., Take, T., Uchiyama, T. & Ochiai, H. (1973) Antibiotic 5879 produced by *Streptomyces aizunensis* identical with bicyclomycin. *The Journal of Antibiotics* 26:479-484.
- Miyoshi, T., Miyairi, N., Aoki, H., Koshaka, M., Sakai, H.I. & Imanaka, H. (1972) Bicyclomycin, a new antibiotic. I. Taxonomy, isolation and characterization. *The Journal of Antibiotics* 10:569-575.
- Moutiez, M., Belin, P. & Gondry, M. (2017) Aminoacyl-tRNA-utilizing enzymes in natural product biosynthesis. *Chemical Reviews* 117:5578-5618.
- Moutiez, M., Seguin, J., Fonvielle, M., Belin, P., Jacques, I.B., Favry, E., Arthur, M. & Gondry, M. (2014a) Specificity determinants for the two tRNA substrates of the cyclodipeptide synthase AlbC from *Streptomyces noursei*. *Nucleic Acids Research* 42:7247-7258.
- Moutiez, M., Schmitt, E., Seguin, J., Thai, R., Favry, E., Belin, P., Mechulam, Y. & Gondry, M. (2014b) Unravelling the mechanism of non-ribosomal peptide synthesis by cyclodipeptide synthases. *Nature Communications* 5:5141.
- Najah, S. (2017) Etude de la souche *Streptomyces* sp. TN58 : séquençage et analyse du génome, étude de la biosynthèse de métabolites secondaires de la famille des acyl-rhamnopyranosides et des dicétopipérazines. *Thèse de doctorat en préparation, Université Paris-Saclay*.
- Narahashi, T., Moore, J.W. & Scott, W.R. (1964) Tetrodotoxin blockage of sodium conductance increase in lobster giant axons. *The Journal of General Physiology* 47:965-974.
- Nichols, W.G., Steel, H.M., Bonny, T., Adkison, K., Curtis, L., Millard, J., Kabeya, K. & Clumeck, N. (2008) Hepatotoxicity observed in clinical trials of aplaviroc (GW873140). *Antimicrobial Agents and Chemotherapy* 52:858-865.
- Nguyen, H.C., Karray, F., Lautru, S., Gagnat, J., Lebrihi, A., Huynh, T.D. & Pernodet, J.L. (2010) Glycosylation steps during spiramycin biosynthesis in *Streptomyces ambofaciens*: involvement of three glycosyltransferases and their interplay with two auxiliary proteins. *Antimicrobial Agents and Chemotherapy* 57:2830-2839.
- Onodera, H., Hasegawa, A., Tsumagari, N., Nakai, R., Ogawa, T. & Kanda, Y. (2004) MPC1001 and Its Analogues: New Antitumor Agents from the Fungus *Cladorrhinum* Species. *Organic Letters* 6:4101-4104.
- Pan, J., Bhardwaj, M., Faulkner, J.R., Nagabhyru, P., Charlton, N.D., Higashi, R.M., Miller, A.F., Young, C.A., Grossman, R.B. & Schardl, C.L. (2014) Ether bridge formation in loline alkaloid biosynthesis. *Phytochemistry* 98:60-68.
- Papayannopoulos, I.A. (1995) The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrometry Reviews* 14, 49-73.
- Park, D.K., Lee, K.E., Baek, C.H., Kim, I.H., Kwon, J.H., Lee, W.K., Lee, K.H., Kim, B.S., Choi, S.H. & Kim, K.S. (2006). Cyclo(Phe-Pro) Modulates the Expression of *ompU* in *Vibrio* spp. *Journal of Bacteriology* 188(6):2214-2221.

- Park, H.G., Zhang, X., Moon, H.S., Zwiefka, A., Cox, K., Gaskell, S.J., Widger, W.R. & Kohn, H. (1995) Bicyclomycin and Dihydrobicyclomycin Inhibition Kinetics of *Escherichia coli* rho-Dependent Transcription Termination Factor ATPase Activity. *Archives of Biochemistry and Biophysics* 323:447-454.
- Pernodet, J.-L., Alègre, M.-T., Blondelet-Rouault, M.-H. & Guérineau, M. (1993). Resistance to spiramycin in *Streptomyces ambofaciens*, the producer organism, involves at least two different mechanisms. *J Gen Microbiol* 139:1003–101.
- Pollack, J.R. & Neillands, J.B. (1970) Enterobactin, an iron transport compound form from *Salmonella typhimurium*. *Biochemical and Biophysical Research Communications* 38:989-992.
- Prasad, C. (1995). Bioactive cyclic dipeptides. *Peptides* 16(1): 151-164.
- Prentki, P. & Kirsch H.M. (1984) *In vitro* insertional mutagenesis with a selectable DNA fragment. *Gene* 29:303-313.
- Randazzo, P., Aubert-Frambourg, A., Guillot, A. & Auger, S. (2016) The MarR-like protein PchR (YvmB) regulates expression of genes involved in pulcherrimic acid biosynthesis and in the initiation of sporulation in *Bacillus subtilis*. *BMC Microbiology* 16:190.
- Raynal, A., Karray, F., Tuphile, K., Darbon-Rongère, E. & Pernodet, J.L. (2006) Excisable Cassettes: New Tools for Functional Analysis of *Streptomyces* Genomes. *Applied and Environmental Microbiology* 72:4839-4844.
- Rebets, Y., Tokovenko, B., Lushchik, I., Rückert, C., Zaburannyi, N., Bechthold, A., Kalinowski, J. & Luzhetskyy, A. (2014) Complete genome sequence of producer of the glycopeptide antibiotic Aculeximycin *Kutzneria albida* DSM 43870T, a representative of minor genus of Pseudonocardiaceae. *BMC Genomics* 15:885.
- Remmert, M., Biegert, A., Hauser, A. & Soding, J. (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature methods* 9:173–175.
- Robiquet, P.J. & Boutron-Charland, A. (1880) Nouvelles expériences sur les amandes amères et sur l'huile volatile qu'elles fournissent. *Annales de chimie et de physique* 44:352-382
- Robey, R.B., Osawa, H., Printz, R.L. & Granner, D.K. (1993) Preparation of Electro-Competent *E. coli* Using Salt-Free Growth Medium *BioTechniques* 20:42-44.
- Roepstorff, P. & Fohlman, J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 11, 601.
- Sánchez, L. & Braña, A.F. (1996). Cell density influences antibiotic biosynthesis in *Streptomyces clavuligerus*. *Microbiology* 142:1209-1220.
- Sandhu, P., Guo, Z.Y., Baba, T., Martin, M.V., Turkey, R.H. & Guengerich, F.P. (1994) Expression of Modified Human Cytochrome P450 1A2 in *Escherichia coli*: Stabilization, Purification, Spectral Characterization, and Catalytic Activities of the Enzyme. *Archives of Biochemistry and Biophysics* 309:168-177.
- Sauguet, L., Moutiez, M., Li, Y., Belin, P., Seguin, J., Le Du, M.H., Thai, R., Masson, C., Fonville, M., Pernodet, J.L., Charbonnier, J.B. & Gondry, M. (2011). Cyclodipeptide synthases, a family of class-I aminoacyl-tRNA synthetase-like enzymes involved in non-ribosomal peptide synthesis. *Nucleic Acid Research* 39:4475-4489.
- Schäffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V. & Altschul, S.F. (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Research* 29:2994–3005.
- Schäffer, A., Konrad, R., Kuhnigk, T., Kämpfer, P., Hertel, H. & König, H. (1996) Hemicellulose-degrading bacteria and yeasts from the termite gut. *Journal of Applied Bacteriology* 80:471-478.
- Schmidt, J.M. (2007) Asymmetric Karplus curves for the protein side-chain ³J couplings. *Journal of Biomolecular NMR* 37:287-301.
- Schneider, T.D. & Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research* 18:6097–6100.
- Schulz, A.W., Oh, D.C., Carney, J.R., Williamson, R.T., Udvary, D.W., Jensen, P.R., Gould, S.J., Fenical, W. & Moore, B.S. (2008) Biosynthesis and structures of cyclomarins and cyclomarazines, prenylated cyclic peptides of marine actinobacterial origin. *Journal of the American Chemical Society* 130:4507-4516.

- Shinghal, R., Barnes, A., Mahar, K.M., Stier, B., Giancaterino, L., Condreay, L.D., Black, L. & McCallum, S.W. (2013) Safety and efficacy of epelsiban in the treatment of men with premature ejaculation: a randomized double-blind, placebo-controlled, fixed-dose study. *The Journal of Sexual Medicine* 10:2506-2517.
- Shirling, E.B. & Gottlieb, D. (1966). Methods for characterization of *Streptomyces* species. *International Journal of Systematic Bacteriology* 16 :313-340.
- Sing, B. & Mitchison, D.A. (1954) Bactericidal Activity of Streptomycin and Isoniazid Against Tubercle Bacilli. *British Medical Journal* 1:130-132.
- Sono, M., Roach, M.P., Coulter, E.D. & Dawson, J.H. (1996) Heme-containing oxygenases. *Chemical Reviews* 96:2841-2888.
- Stachelhaus, T., Mootz, H.D. & Marahiel, M A. (1999) The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemical Biology* 6:493-505.
- Stark, T. & Hofmann, T. (2005) Structures, sensory activity, and dose/response functions of 2,5-diketopiperazines in roasted cocoa nibs (*Theobroma cacao*). *Journal of Agricultural and Food Chemistry* 53:7222-7231.
- Staunton, J. & Weissman, K.J. (2001) Polyketide biosynthesis: a millennium review. *Natural Product Reports* 18:380-416.
- Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* 41:207-234.
- Svarstad, H., Bugge, H.C. & Dhillon, S.S. (2000) From Norway to Novartis: Cyclosporin from *Tolypocladium inflatum* in an open access bioprospecting regime. *Biodiversity and Conservation* 9:1521-1541.
- Tang, M.R., Sternberg, D., Behr, R.K., Sloma, A. & Berka, R.M. (2006). Use of transcriptional profiling & bioinformatics to solve production problems. *Industrial Biotechnology* 2:66-74.
- Tholl, D. (2015) Biosynthesis and biological functions of terpenoids in plants. *Advances in Biochemical Engineering / Biotechnology* 148:63-106.
- Toyomasu, T., Nakaminami, K., Toshima, H., Mie, T., Watanabe, K., Ito, H., Matsui, H., Mitsuhashi, W., Sassa, T. & Oikawa, H. (2014) Cloning of a Gene Cluster Responsible for the Biosynthesis of Diterpene Aphidicolin, a Specific Inhibitor of DNA Polymerase α . *Bioscience, Biotechnology and Biochemistry* 68:146-152.
- Umezawa, H., Aoyagi, T., Ogawa, K., Obata, T., Iinuma, H., Naganawa, H., Hamada, M. & Takeuchi, T. (1985) Foroxymithine, a new inhibitor of angiotensin-converting enzyme, produced by Actinomycetes. *The Journal of Antibiotics* 12:1813-1815.
- Vetting, M.W., Hedge, S.S. & Blanchard, J.S. (2011). The structure and mechanism of the *Mycobacterium tuberculosis* cyclodityrosine synthetase. *Nature Chemical Biology* 6:797-799.
- Vingadassalon, A., Lorieux, F., Juguet, M., Le Goff, G., Gerbaud, C., Pernodet, J.L. & Lautru, S. (2015) Natural Combinatorial Biosynthesis Involving Two Clusters for the Synthesis of Three Pyrrolamides in *Streptomyces netropsis*. *ACS Chemical Biology* 10:601-610.
- Waller, G.R. & Nowacki, E.K. (1978) Alkaloid biology and metabolism in plants. *Springer*.
- Weber, T. (2014) In silico tools for the analysis of antibiotic biosynthetic pathways. *International Journal of Medical Microbiology* 304:230-235.
- Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Brucoleri, R., Lee, S.Y., Fischbach, M.A., Müller, R., Wohlleben, W., Breitling, R., Takano, E. & Medema, M.H. (2015) AntiSMASH 3.0. – A comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research* 43:W237-W243.
- Withering, W. (1785) An account on the foxglove and some of its medicinal uses. *M. Swinney*
- World Health Organisation (2015) Global Tuberculosis report 2016. ISBN 978 92 4 156539 4.
- Yanai, K. & Murakami, T. (2004) The kanamycin biosynthetic gene cluster from *Streptomyces kanamyceticus*. *The Journal of Antibiotics* 57:351-354.
- Yang, C.C., Tseng, S.M. & Chen, C.W. (2015). Telomere-associated proteins add deoxynucleotides to terminal proteins during replication of the telomeres of linear chromosomes and plasmids in *Streptomyces*. *Nucleic Acid Research* 43:6373-6383.
- Yao, Y., Tian, L., Li, J., Cao, J. & Pei, Y. (2009) Cytotoxic piperazine-2,5-dione derivatives from marine fungus *Gliocladium* sp. *Pharmazie* 64:616-618.

- Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. (2000a) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* 7:203-214.
- Zhang, Z., Ren, J., Stammers, D.K., Baldwin, J.E., Harlos, K. & Schofield, C.J. (2000b) Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nature Structural Biology* 7:127-133.
- Zwiefka, A., Kohn, H. & Widger, W.R. (1993) Transcription termination factor rho: The site of bicyclomycin inhibition in *Escherichia coli*. *Biochemistry* 32:3564-3570.

Titre : Caractérisation de voies de biosynthèse de dicétopipérazines chez les actinobactéries

Mots clés : dicétopipérazines, cyclodipeptides, actinobactéries, *Streptomyces*

Résumé : Les dicétopipérazines (DKP : diketopiperazines) sont une classe de produits naturels largement utilisée en médecine. Chez les organismes producteurs, elles peuvent servir de facteurs de virulence, de molécules de signalisation ou encore être impliquées dans la concurrence entre (micro-)organismes ou la défense contre les prédateurs, mais souvent leur rôle est inconnu. Elles peuvent être synthétisées par deux voies de biosynthèse principales : soit par des voies dépendant de synthétases de peptides non ribosomiques (NRPS: Non Ribosomal Peptide Synthetase), des complexes enzymatiques de grande taille utilisant comme substrats les acides aminés protéinogènes ou non, soit par des voies dépendant des cyclodipeptide synthases (CDPS), de petites enzymes utilisant comme substrats des ARNt chargés. Je me suis intéressé à ce deuxième type de voie de biosynthèse. J'ai d'abord participé à une étude visant à caractériser la diversité des CDPS et des DKP synthétisées par ces enzymes. Ce travail a permis d'élargir le champ des CDPS caractérisées expérimentalement, de démontrer l'incorporation par ces enzymes de

17 acides aminés dans des DKP et de poser les bases de la prédiction des DKP synthétisées grâce à l'identification dans la séquence protéique des CDPS de motifs de résidus responsables de la sélection du substrat. J'ai ensuite étudié chez *Streptomyces venezuelae* un cluster de gènes homologues à des gènes essentiels impliqués chez *Mycobacterium tuberculosis* dans la biosynthèse d'une dicétopipérazine. J'ai enfin identifié chez *Streptomyces aizunensis* les gènes dirigeant la biosynthèse de la bicyclomycine, un antibiotique de la famille des DKP utilisé en médecine. Je présente la caractérisation complète de la voie de biosynthèse de la bicyclomycine. Cette biosynthèse fait intervenir une CDPS, cinq dioxygénases dépendantes du fer et du 2-oxoglutarate et une monooxygénase de la famille des cytochromes P450. C'est la voie la plus complexe dépendante de CDPS jamais décrite, la seule contenant plusieurs enzymes d'oxydoréduction et la première voie de biosynthèse d'une molécule utilisée en médecine dépendante d'une CDPS.

Title : Characterization of diketopiperazine biosynthetic pathways in Actinobacteria

Keywords : diketopiperazines, cyclodipeptides, Actinobacteria, *Streptomyces*

Abstract : Diketopiperazines (DKP) are a class of natural products widely used in medicine. In producer organisms, they may act as virulence factors, signaling molecules or may be involved in competition between (micro-) organisms or defense against predators, but often their role is unknown. They can be synthesized by two main biosynthetic pathways: either by non-ribosomal peptide synthetase (NRPS) dependent pathways, NRPS being large enzyme complexes using proteinogenic or non-proteinogenic amino acids as substrates, or by Cyclodipeptide synthases (CDPS), small enzymes using loaded tRNAs as substrates. I was interested in this second type of biosynthetic pathway. I first participated in a study having as a goal to characterize the diversity of CDPSs and DKPs synthesized by these enzymes. This work enabled to widen the field of experimentally characterized CDPSs, to demonstrate the incorporation by these enzymes of 17 amino

acids in DKPs and to lay the bases for the prediction of the DKPs synthesized thanks to the identification in the protein sequence of CDPSs of residue patterns responsible for substrate selection. I then studied in *Streptomyces venezuelae* a cluster of genes homologous to essential genes involved in *Mycobacterium tuberculosis* in the biosynthesis of a diketopiperazine. I finally identified in *Streptomyces aizunensis* the genes directing the biosynthesis of bicyclomycin, an antibiotic of the DKP family used in medicine. I present the complete characterization of the bicyclomycin biosynthetic pathway. This biosynthesis involves a CDPS, five iron-2-oxoglutarate-dependent dioxygenases and a cytochrome P450 monooxygenase. It is the most complex CDPS-dependent pathway ever described, the only one containing several redox enzymes and the first CDPS-dependent biosynthetic pathway for a molecule used in medicine.

