



HAL
open science

Caractérisation structurale et fonctionnelle d'enzymes
impliquées dans la maturation d'alcaloïdes
pyrrolizidiniques issus de voies NRPS chez des
 γ -protéobactéries pathogènes

Andrea Calderari

► To cite this version:

Andrea Calderari. Caractérisation structurale et fonctionnelle d'enzymes impliquées dans la maturation d'alcaloïdes pyrrolizidiniques issus de voies NRPS chez des γ -protéobactéries pathogènes. Biochimie, Biologie Moléculaire. Université de Lorraine, 2024. Français. NNT : 2024LORR0086 . tel-04820731

HAL Id: tel-04820731

<https://hal.univ-lorraine.fr/tel-04820731v1>

Submitted on 5 Dec 2024

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

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



**UNIVERSITÉ
DE LORRAINE**

**BIBLIOTHÈQUES
UNIVERSITAIRES**

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



**UNIVERSITÉ
DE LORRAINE**

BioSE



IMoPA
Ingénierie moléculaire, cellulaire & physiopathologie

Thèse

Présentée et soutenue publiquement pour l'obtention du titre de

DOCTEUR DE L'UNIVERSITÉ DE LORRAINE

Mention : « Sciences de la Vie et de la Santé »

par **Andrea CALDERARI**

sous la direction de la **Pr. Kira J. WEISSMAN** et du **Dr. Arnaud GRUEZ**

**Caractérisation structurale et fonctionnelle d'enzymes
impliquées dans la maturation d'alcaloïdes pyrrolizidiniques
issus de voies NRPS chez des γ -protéobactéries pathogènes**

le 8 juillet 2024

Membres du jury :

Directeurs de thèse :	M^{me} Kira J. WEISSMAN	Professeure, Université de Lorraine, Nancy
	M. Arnaud GRUEZ	Maître de conférences, Université de Lorraine, Nancy
Président du jury :	M. Thierry OSTER	Professeur, Université de Lorraine, Nancy
Rapporteurs :	M^{me} Muriel GONDRY	Directrice de recherche, CEA, Saclay
	M. Pascal ARNOUX	Directeur de recherche, CEA, Cadarache
Examineur :	M^{me} Claudine MAYER	Professeure, Université Paris-Cité, Paris

Au terme de ce travail, je tiens à exprimer ma sincère gratitude envers la Pr. Kira Weissman et le Dr. Arnaud Gruez, mes directeurs de thèse, pour m'avoir témoigné leur confiance indéfectible et pleinement formé à la recherche scientifique. Je les remercie de m'avoir transmis les connaissances fondamentales et les compétences techniques essentielles pour accomplir avec succès cet ambitieux projet multi-disciplinaire. Je leur suis reconnaissant pour m'avoir orienté dans l'analyse des résultats expérimentaux afin de les interpréter avec clairvoyance et les discuter en détail.

J'adresse mes profonds remerciements à la Dr. Muriel Gondry et au Dr. Pascal Arnoux pour m'avoir fait l'honneur d'être rapporteurs de cette thèse. Je désire remercier les membres du comité de suivi, la Pr. Claudine Mayer et le Pr. Thierry Oster, pour m'avoir prodigué des conseils réfléchis et des encouragements salutaires afin de m'aider à atteindre les objectifs de recherche, ainsi que pour avoir volontiers accepté d'être examinateurs de ce travail final. À cet égard, je remercie grandement les jurés de leur intérêt pour ma thèse.

Je souhaite aussi remercier les membres de l'équipe S2BEAM (*Structural and Synthetic Biology, Enzymology, and Applications of Megaenzymes*), en particulier le Dr. Christophe Jacob, le Dr. Jean-Malo Massicard, l'Ing. Sabrina Collin et le Dr. Benjamin Chagot, pour avoir échangé leurs points de vue constructifs autour de mon sujet et m'apporté leur appui technique précieux à certaines reprises. Je remercie tout autant mes proches collègues de travail, Romain, Serge, Arnaud et Magali, pour les nombreux moments appréciables en leur compagnie, que ce soit au sein du laboratoire ou en dehors, notamment à la 10^{ème} école RéNaFoBiS de biologie structurale sur l'île d'Oléron et au 47^{ème} congrès international de la FEBS à Tours.

Je remercie l'Agence Nationale de la Recherche d'avoir financé ces recherches et le laboratoire IMoPA de m'avoir permis de les réaliser dans des conditions enviables.

TABLE DES MATIÈRES

LISTE DES FIGURES ET TABLEAUX

LISTE DES ABRÉVIATIONS

INTRODUCTION BIBLIOGRAPHIQUE.....	1
SECTION I. Classification, production et application des métabolites spécialisés.....	1
I.1. Définition, structures chimiques et activités biologiques.....	1
I.2. Production de métabolites spécialisés par les micro-organismes et les plantes.....	2
I.2.A. Groupe des alcaloïdes.....	3
I.2.B. Groupe des phénylpropanoïdes.....	3
I.2.C. Groupe des terpénoïdes.....	4
I.2.D. Groupe des carbohydrates.....	4
I.2.E. Groupe des polycétides et peptides non-ribosomiques.....	5
I.3. Applications biomédicales et agro-industrielles.....	8
I.3.A. Administration de médicaments et de nanoparticules.....	8
I.3.B. Utilisation de pesticides et de régulateurs de croissance.....	8
SECTION II. Identification, caractérisation et production de composés bioactifs.....	9
II.1. Exploration computationnelle des génomes bactériens.....	9
II.1.A. Identification des BGC par des programmes de criblage à haut-débit.....	9
II.1.B. Prédiction des structures chimiques par des outils spécifiques.....	11
II.1.C. Corrélation entre les données génomiques et métabolomiques.....	12
II.1.D. Exploration des métabolomes bactériens.....	13
II.2. Optimisation des conditions de culture bactérienne et activation des BGC cryptiques.....	14
II.2.A. Modulation des conditions de croissance par des facteurs biotiques et abiotiques.....	15
II.2.B. Expression hétérologue de BGC cryptiques au sein de souches châssis.....	17
II.2.B.α. Refactorisation des BGC par des outils <i>in silico</i> et des méthodes d'assemblage... 	17
II.2.B.β. Construction et adaptation de souches châssis par ingénierie génomique.....	21
II.3. Production d'analogues structuraux bioactifs.....	22
SECTION III. Compréhension des voies de biosynthèse NRPS canoniques ou atypiques.....	23
III.1. Éluclidation du mécanisme catalytique de synthèse non-ribosomique.....	23
III.1.A. Expérimentations préliminaires.....	23
III.1.B. Présentation des types de biosynthèse et du mécanisme catalytique canonique.....	25
III.2. Organisation et composition des NRPS de type I, II et III.....	27
III.2.A. NRPS de type I modulaires : exemple de la biosynthèse des β-lactamines	28
III.2.B. NRPS de type I itératives : exemple de la biosynthèse des quinoxalines	31

III.2.C. NRPS de type II autonomes : exemple de la biosynthèse de la congocidine	33
III.2.D. NRPS de type III non-linéaires.....	35
III.2.D.α. NRPS avec un domaine A absent : cas de la biosynthèse de la capréomycine.....	35
III.2.D.β. NRPS avec un domaine C atypique : exemple de la biosynthèse de l'andrimide.	37
III.2.E. NRPS-PKS hybride : exemple de la biosynthèse de la bléomycine.....	38
SECTION IV. Relations structure-fonction des domaines catalytiques.....	41
IV.1. Activation du domaine PCP par une phosphopantéthéinyltransférase.....	41
IV.2. Liaison des substrats et des intermédiaires peptidiques par le domaine PCP.....	43
IV.3. Sélection et activation d'un monomère spécifique par le domaine A.....	45
IV.3.A. Analyse structurale et fonctionnelle.....	45
IV.3.B. Utilisation de sondes chimiques dans les systèmes de didomains A-PCP.....	50
IV.4. Formation de liaisons peptidiques par le domaine C.....	52
IV.4.A. Analyse structurale et fonctionnelle	52
IV.4.B. Évolution structurale	57
IV.4.B.α. N-acylation par le domaine $C_{starter}$	58
IV.4.B.β. Formation de liaisons esters ou isopeptidiques par le domaine C.....	59
IV.4.B.γ. Formation d'un cycle β-lactame par le domaine C_{BL}	61
IV.4.B.δ. Cyclisation de Pictet-Spengler par le domaine PS	62
IV.4.B.ε. Recrutement d'enzymes auxiliaires par le domaine X	63
IV.4.B.ζ. Libération de chaîne par le domaine C_T terminal	64
IV.5. Communication et interaction entre les sous-unités par les domaines de <i>docking</i>	65
IV.5.A. Communication intermodulaire par les domaines <i>COM</i>	66
IV.5.B. Interactions intermoléculaires par les domaines <i>βhD</i> et <i>PAX</i>	67
IV.6. Fonctionnalisation terminale du NRP par le domaine TE	71
IV.6.A. Analyse fonctionnelle	71
IV.6.B. Analyse structurale	73
IV.7. Intervention de domaines catalytiques subsidiaires	75
IV.7.A. N-acylation par les acyl-CoA ligases.....	76
IV.7.B. N-formylation par le domaine F	78
IV.7.C. Méthylation par le domaine MT	80
IV.7.D. β-Hydroxylation par les cytochromes P450	83
IV.7.E. Oxydation par les enzymes NAD ⁺ /FAD-dépendantes.....	85
IV.7.F. Épimérisation par le domaine E.....	87
IV.7.G. Hétérocyclisation par le domaine Cy.....	89
IV.7.H. Réduction par le domaine R.....	91

IV.7.I. Modification par d'autres domaines rares.....	94
IV.8. Maturation post-NRPS des produits par des enzymes auxiliaires.....	96
MODELES D'ÉTUDE ET OBJECTIFS.....	100
I. Identification, caractérisation et dérivatisation des AP bactériens.....	100
I.1. Identification des BGC AP chez des γ -protéobactéries et actinobactéries.....	100
I.2. Caractérisation de la structure chimique et de l'activité biologique des AP bactériens.....	102
I.3. Dérivatisation chimique des AP bactériens.....	105
II. Caractérisation des systèmes biosynthétiques modèles chez des γ-protéobactéries	107
II.1. Organisation de la voie de biosynthèse commune aux AP bactériens.....	107
II.2. Maturation post-assemblage des produits et réactions non-enzymatiques.....	109
II.3. Fonction écophysiological de la pyrrolizwilline et des azétidomonamides.....	110
III. Élucidation des relations structure-fonction des enzymes biosynthétiques.....	110
MATÉRIELS ET MÉTHODES.....	111
SECTION I. Matériels biologiques.....	111
I.1. Souches bactériennes.....	111
I.2. Plasmides d'expression.....	111
SECTION II. Méthodes d'étude des enzymes biosynthétiques XhpG et AzeD.....	113
II.1. Clonage des séquences géniques dans les plasmides d'expression.....	113
II.1.A. Amplification des séquences par PCR et ligation des inserts dans les plasmides.....	113
II.1.B. Préparation et transformation de bactéries <i>E. coli</i> compétentes par choc thermique...	115
II.1.C. Préparation d'ADN plasmidique par lyse alcaline.....	115
II.1.D. Analyse des plasmides recombinants par restriction et séquençage.....	116
II.1.E. Modification de séquences géniques par PCR mutagène.....	116
II.2. Production et purification des protéines recombinantes.....	117
II.2.A. Expression des protéines chez la bactérie <i>E. coli</i> par un réactif inducteur.....	117
II.2.B. Lyse de la biomasse bactérienne par sonication	117
II.2.C. Purification des protéines par chromatographie en phase liquide.....	117
II.2.C.α. Purification par chromatographie d'affinité et protéolyse du <i>tag</i> protéique	117
II.2.C.β. Purification par chromatographie d'exclusion stérique.....	118
II.3. Analyse des échantillons par diffusion dynamique de la lumière.....	118
II.4. Résolution des structures tridimensionnelles par cristallographie aux rayons X.....	120
II.4.A. Cristallisation des protéines recombinantes par diffusion de vapeur.....	120
II.4.B. Diffraction des cristaux par un rayonnement synchrotron	123
II.4.C. Détermination des phases ϕ_{hkl} par remplacement moléculaire.....	126

II.4.D. Construction et amélioration des modèles par des cycles itératifs d'affinement.....	128
II.4.E. Validation des structures cristallographiques et soumission dans la PDB	131
II.4.F. Modélisation des complexes binaires par <i>docking</i> moléculaire.....	131
II.5 Analyse des échantillons par diffusion des rayons X aux petits angles.....	132
SECTION III. Méthodes d'étude du domaine C2 de la NRPS AzeB	137
III.1. Conception et construction des plasmides d'expression originaux.....	139
III.1.A. Conception <i>in silico</i> des plasmides pSZ avec le logiciel SnapGene.....	139
III.1.B. Construction des plasmides pSZ par un assemblage de type Golden Gate.....	140
III.2. Clonage des séquences géniques dans les plasmides d'expression.....	143
III.3. Production et purification des protéines recombinantes.....	144
III.4. Caractérisation des interactions protéine-protéine par titration calorimétrique isotherme.	145
III.5. Analyse des échantillons par diffusion d'ondes électromagnétiques.....	146
III.6. Cristallisation des complexes intermoléculaires par diffusion de vapeur.....	146
RÉSULTATS ET DISCUSSION	147
SECTION I. Élucidation des relations structure-fonction de l'hydrolase XhpG	147
I.1. Résolution de la structure cristallographique.....	149
I.1.A. Description du repliement α/β hydrolase.....	150
I.1.B. Organisation structurale du site actif et liaison du substrat.....	151
I.2. Comparaison avec la structure cristallographique d'une enzyme homologue.....	153
I.2.A. Organisation structurale du sous-domaine α -hélicoïdal et de la région charnière.....	153
I.2.B. Dimérisation de l'enzyme en solution et modélisation de conformères.....	155
I.2.C. Configuration spatiale et composition du site actif.....	157
I.2.D. Conservation de la triade catalytique et liaison des ligands au sein du site actif.....	159
SECTION II. Élucidation des relations structure-fonction de la déshydratase AzeD	161
II.1. Résolution de la structure cristallographique.....	164
II.1.A. Description du repliement global.....	165
II.1.B. Organisation structurale du site actif et liaison du substrat.....	167
II.2. Comparaison avec la structure cristallographique de domaines homologues.....	169
II.2.A. Organisation structurale des éléments interlobulaires et conformation en solution....	169
II.2.B. Investigation du mécanisme catalytique	172
SECTION III. Caractérisation biophysique des complexes PCP-C de la NRPS AzeB	173
CONCLUSION ET PERSPECTIVES	179
ANNEXES	
BIBLIOGRAPHIE	

LISTE DES FIGURES ET TABLEAUX

INTRODUCTION BIBLIOGRAPHIQUE

SECTION I. Classification, production et application des métabolites spécialisés

Figure 1 : Représentation de structures chimiques et utilisation de produits naturels

Figure 2 : Représentation de structures chimiques et utilisation de NRP

Tableau I : Production de NRP au sein de divisions et classes bactériennes

SECTION II. Identification, caractérisation et production de composés bioactifs

Figure 3 : Exploration bio-informatique des génomes bactériens

Figure 4 : Représentation de structures chimiques et utilisation potentielle de produits de NRPS/PKS cryptiques

Figure 5 : Optimisation des conditions de culture bactérienne et activation des BGC cryptiques

Figure 6 : Isolation de la teixobactine à partir de la technologie micro-puce iChip

Figure 7 : Conception *in silico* de parties génétiques et optimisation de l'expression génique d'un BGC

Tableau II : Utilisation de méthodes d'assemblage d'ADN

SECTION III. Compréhension des voies de biosynthèse NRPS canoniques ou atypiques

Figure 8 : Organisation de la biosynthèse des NRP

Figure 9 : Réactions canoniques de synthèse de peptides non-ribosomiques

Figure 10 : Constitution des NRPS de type I, II et III

Figure 11 : Représentation de la voie de biosynthèse des β -lactamines

Figure 12 : Représentation de la voie de biosynthèse de l'échinomycine

Figure 13 : Représentation de la voie de biosynthèse de la congocidine

Figure 14 : Représentation de la voie de biosynthèse de la capréomycine

Figure 15 : Représentation de la voie de biosynthèse de l'andrimide

Figure 16 : Représentation de structures chimiques et utilisation d'hybrides NRP-polycétides

Figure 17 : Représentation de la voie de biosynthèse de la bléomycine

SECTION IV. Relations structure-fonction des domaines catalytiques

Figure 18 : Représentation de la structure tridimensionnelle du complexe de la PPTase Sfp avec un domaine PCP en présence de coenzyme A

Figure 19 : Représentation de la structure cristallographique du domaine PCP PltL en présence du cofacteur 4'-Ppant

Figure 20 : Représentation de la structure cristallographique du domaine A autonome Thr1

Figure 21 : Caractérisation du code non-ribosomique de domaines A

Figure 22 : Rotation des domaines A1 et PCP1 de la NRPS LgrA lors de la réaction d'adénylation

Figure 23 : Représentation de la structure cristallographique du domaine SlgN1

Figure 24 : Représentation de l'interface au sein de domaines A autonomes

Figure 25 : Utilisation de sondes chimiques dans les complexes de didomains A-PCP

Figure 26 : Représentation de la structure cristallographique du complexe autonome PltF-PltL

Figure 27 : Représentation de la structure cristallographique du domaine C1 de la NRPS Txo2

Figure 28 : Réaction de condensation peptidique par le domaine C

Figure 29 : Comparaison des structures cristallographiques de domaines C homologues

Figure 30 : Représentation de la structure cristallographique du complexe PCP-C de la NRPS LgrA

Figure 31 : Lipoinitiation par le domaine $C_{starter}$ du module SrfAA

Figure 32 : Formation d'une liaison ester par le domaine C6 du module terminal de la NRPS CrpD

Figure 33 : Formation de liaisons isopeptidiques par les domaines C de modules NRPS

Figure 34 : Déshydratation et cycloaddition par le domaine C de la NRPS NocB

Figure 35 : Cyclisation intramoléculaire de Pictet-Spengler par le domaine PS de la NRPS SfmC

Figure 36 : Cyclisation intramoléculaire par des CYP450 au sein du système NRPS Tcp9–12

Figure 37 : Libération de NRP par le domaine C_T terminal de systèmes NRPS

Figure 38 : Organisation des domaines de *docking* au sein d'une NRPS hypothétique

Figure 39 : Représentation de la structure cristallographique du didomaine COM^A-C7 de la NRPS SrfAC

Figure 40 : Représentation de la structure RMN du complexe des domaines de *docking* Kj12B–C

Figure 41 : Caractérisation du code de *docking* des domaines *SLiM-βhD*

Figure 42 : Représentation de la structure RMN du complexe des domaines de *docking* PaxB–C

Figure 43 : Libération de NRP par le domaine TE de systèmes NRPS

Figure 44 : Représentation de la structure cristallographique du domaine TE de la NRPS SrfAC

Figure 45 : Liaison de la fengycine au sein du site actif du domaine TE de la NRPS FenB

Figure 46 : Modification de NRP par les domaines non-canoniques de systèmes NRPS

Figure 47 : *N*-acylation par les acyl-CoA ligases

Figure 48 : Représentation de l'interface de la NRPS LgrA

Figure 49 : Réaction de *N*-formylation par le domaine F1 de la NRPS LgrA

Figure 50 : Réactions de méthylation par les domaines MT de la NRPS TioS et de l'enzyme TioN

Figure 51 : Représentation de la structure cristallographique du didomaine A-MT de la NRPS TioS

Figure 52 : Représentation de la voie de biosynthèse de la skyllamycine A

Figure 53 : Représentation de la structure cristallographique du complexe CYP163B3-PCP7 au sein du système NRPS Sky29–31

Figure 54 : Représentation de la voie de biosynthèse de la pentabromopseudilline

Figure 55 : Représentation de la structure cristallographique du complexe autonome Bmp3-Bmp1

Figure 56 : Représentation de la structure cristallographique du domaine E1 de la NRPS TycA

Figure 57 : Représentation de la voie de biosynthèse de l'épothilone

Figure 58 : Représentation de la structure cristallographique du domaine Cy de la NRPS EpoB

Figure 59 : Représentation de la voie de biosynthèse de la myxalamide S

Figure 60 : Représentation de la structure cristallographique du domaine R de la NRPS MxaA

Figure 61 : Liaison du cofacteur NADPH au sein du site actif du domaine R de la NRPS MxaA

Figure 62 : Représentation de la voie de biosynthèse de la vioprolide B

Figure 63 : Représentation de la voie de biosynthèse de la paenilamicine B1

Figure 64 : Représentation de structures chimiques de (lipo)-glycopeptides

Figure 65 : Maturation de la chloroérémomycine et de la vancomycine par les glycosyltransférases

Figure 66 : Représentation de la structure cristallographique de l'*O*-glycosyltransférase GtfD

Tableau III : Classification évolutive des domaines C

MODÈLES D'ÉTUDE ET OBJECTIFS

Figure 67 : Organisation des BGC des AP chez des γ -protéobactéries et actinobactéries

Figure 68 : Représentation de structures chimiques et utilisation potentielle d'AP bactériens et analogues naturels

Figure 69 : Dimérisation de la bohémamine A en présence de méthanal endogène

Figure 70 : Création d'une banque de bohémamines en présence de radicaux linéaires ou cycliques

Figure 71 : Représentation de la voie de biosynthèse des AP bactériens

Figure 72 : Maturation post-assemblage de la pyrrolizwilline et du PA496

MATÉRIELS ET MÉTHODES

SECTION I. Matériels biologiques

Figure 73 : Représentation de la carte du plasmide d'expression pBG102

SECTION II. Méthodes d'étude des enzymes biosynthétiques XhpG et AzeD

Figure 74 : Conditions expérimentales d'amplification en chaîne par polymérase

Figure 75 : Mutagenèse site-spécifique d'un plasmide

Figure 76 : Analyse d'un échantillon par DLS

Figure 77 : Sursaturation, nucléation et croissance des cristaux protéiques

Figure 78 : Cristallisation d'une protéine par diffusion de vapeur en système hermétique

Figure 79 : Traitement d'un patron de diffraction avec les programmes XDS, pointless et aimless

Figure 80 : Fonction de densité électronique et équation complexe du facteur de structure F_{hkl}

Figure 81 : Équation complexe du facteur de structure F_{hkl}

Figure 82 : Observation de cartes de densité électronique de structures protéiques avec le programme WinCoot.

Figure 83 : Configuration de la ligne de lumière SWING au synchrotron SOLEIL

Figure 84 : Analyse d'un échantillon par SAXS

Figure 85 : Réduction des trames et moyennation en une courbe SAXS avec le programme FOXTROT

Figure 86 : Analyse d'une courbe de diffusion radiale avec le programme PRIMUS

Figure 87 : Utilisation des programmes de la suite ATSAS 3.0

Tableau IV : Composition du milieu réactionnel de PCR et séquences des amorces

Tableau V : Enregistrement des données cristallographiques des enzymes XhpG et AzeD

Tableau VI : Affinement des données cristallographiques des enzymes XhpG et AzeD

Tableau VII : Validation des structures des enzymes XhpG et AzeD avec le programme MolProbity

SECTION III. Méthodes d'étude du domaine C2 de la NRPS AzeB

Figure 88 : Configuration expérimentale des domaines chimériques et structuration heptamérique d'une paire de SYNZIP

Figure 89 : Représentation de la carte du plasmide d'expression pSZ17WLC

Figure 90 : Construction des plasmides pSZ par assemblage de type Golden Gate

Figure 91 : Analyse d'un échantillon par ITC

Tableau VIII : Composition du milieu réactionnel de PCR et séquences des amorces/oligonucléotides

Tableau IX : Conditions expérimentales d'une réaction de PCR des séquences d'assemblage

Tableau X : Composition du milieu réactionnel d'une réaction de Golden Gate

Tableau XI : Conditions expérimentales d'une réaction de Golden Gate

Tableau XII : Composition du milieu réactionnel de PCR et séquences des amorces

Tableau XIII : Conditions expérimentales d'une réaction de PCR des séquences PCP1-C2 et PCP2

RÉSULTATS ET DISCUSSION

SECTION I. Élucidation des relations structure-fonction de l'hydrolase XhpG

Figure 92 : Alignement séquentiel des homologues de l'hydrolase XhpG

Figure 93 : Purification de l'hydrolase XhpG S98A mutante par chromatographie liquide

Figure 94 : Analyse de l'hydrolase XhpG S98A mutante par DLS et SAXS

Figure 95 : Reproduction des cristaux de l'hydrolase XhpG S98A mutante par diffusion de vapeur

Figure 96 : Représentation de la structure cristallographique de l'hydrolase XhpG

Figure 97 : Liaison du substrat oxopyrrolizidinacétamide dans le site actif de l'hydrolase XhpG par modélisation moléculaire

Figure 98 : Superposition structurale du sous-domaine α -hélicoïdal des enzymes XhpG et PcaD

Figure 99 : Dimérisation de l'hydrolase XhpG en solution et modélisation d'un conformère

Figure 100 : Comparaison de la configuration spatiale du site actif entre les enzymes homologues

Figure 101 : Comparaison de la composition en résidus du site actif entre les enzymes homologues

Figure 102 : Conservation de la triade catalytique au sein des enzymes homologues

Figure 103 : Comparaison du mode de liaison des ligands au sein du site actif entre les enzymes homologues

SECTION II. Élucidation des relations structure-fonction de la déshydratase AzeD

Figure 104 : Classification phylogénétique de domaines homologues à la déshydratase AzeD

Figure 105 : Alignement séquentiel des homologues de la déshydratase AzeD

Figure 106 : Purification de la déshydratase AzeD par chromatographie liquide

Figure 107 : Analyse de la déshydratase AzeD par DLS

Figure 108 : Cristallisation de la déshydratase AzeD par diffusion de vapeur

Figure 109 : Représentation de la structure cristallographique de la déshydratase AzeD

Figure 110 : Configuration spatiale du site actif de la déshydratase AzeD

Figure 111 : Liaison du substrat L-azétidomonamide dans le site actif de la déshydratase AzeD par modélisation moléculaire

Figure 112 : Superposition structurale des éléments interlobulaires de la déshydratase AzeD et du domaine AmbE_C_{modAA}

Figure 113 : Conformation de la déshydratase AzeD en solution et modélisation d'un conformère

SECTION III. Caractérisation biophysique des complexes PCP-C de la NRPS AzeB

Figure 114 : Amplification des séquences d'assemblage des plasmides pSZ par PCR

Figure 115 : Détection du plasmide pSZ17WLC par PCR sur colonies de souches *E. coli* DH5 α recombinantes

Figure 116 : Digestion des plasmides pSZ recombinants par les enzymes *Bam*HI/*Eco*RI

Figure 117 : Purification des domaines chimériques par chromatographie liquide.

Figure 118 : Analyse des complexes PCP1-C2-PCP2 par ITC

Figure 119 : Interaction des multidomains PCP1-C2-SZ17C et PCP2-SZ18C et conformation d'un complexe par modélisation moléculaire

Tableau XIV : Enregistrement et traitement des données SAXS des complexes PCP1-C2-PCP2

CONCLUSION ET PERSPECTIVES

Figure 120 : Cristallisation de la BVMO AzeC par diffusion de vapeur

LISTE DES ABRÉVIATIONS

A : domaine d'adénylation	F : domaine formylase
A_{core} : sous-domaine N-terminal du domaine A	FAAL : domaine ligase d'acide gras
A_{sub} : sous-domaine C-terminal du domaine A	FAD : flavine adénine dinucléotide
Å : ångström	FMN : flavine mononucléotide
A_{600nm} : absorbance à 600 nm	GF : chromatographie d'exclusion stérique
ACP : domaine porteur d'acyle	L-Hpg : hydroxyphénylglycine.
ADN : acide désoxyribonucléique	H₄F : tetrahydrofolate
AG : acide gras	IMAC : chromatographie d'affinité sur ions nickel immobilisés
AMP : adénosine monophosphate	IPTG : isopropyl β-D-1-thiogalactopyranoside
AP : alcaloïdes pyrrolizidiniques	ITC : titration calorimétrique isotherme
AT : domaine acyltransférase	K_a/K_d : constantes d'association/dissociation
ATP : adénosine triphosphate	LB : bouillon lysogène
ARN : acide ribonucléique	LC : chromatographie en phase liquide
AVS : adénosine vinylsulfonamide	MgCl₂ : chlorure de magnésium
AZC : acide azétidine-2-carboxylique	MIR : remplacement isomorphe multiple
BGC : cluster de gènes biosynthétiques	MLP : protéine de type MbtH
BVMO : mono-oxygénase de Baeyer-Villiger	MS : spectrométrie de masse
bZIP : <i>basic-region leucine ZIPper</i>	MT : domaine méthyltransférase
C : domaine de condensation	NaCl : chlorure de calcium
CDS : séquence codante	NADP : nicotinamide adénine dinucléotide phosphate
CoA : coenzyme A	NaOH : hydroxyde de sodium
COM : domaine de communication	NRP : peptide non-ribosomique
Cy : domaine d'hétérocyclisation	NRPS : synthétase de NRP
CYP450 : mono-oxygénase à cytochrome P450	NSD : disparité spatiale normalisée
Da : dalton	OX : domaine oxydase
DD : domaine de <i>docking</i>	PEG : polyéthylène glycol
DH : domaine déshydratase	PCP : domaine porteur de peptidyle
DLS : diffusion dynamique de la lumière	PCR : réaction de polymérisation en chaîne
D_{max} : distance interatomique maximale	PDB : banque de données protéiques
dNTP : désoxyribonucléoside triphosphate	PKS : synthase de polycétides
L-dSer : déshydroalanine	PMSF : phénylméthylsulfonyl
DTT : dithiothréitol	PPant : bras 4'-phosphopantéthéine
E : domaine épimérase	
EDTA : acide éthylène-diamine-tétraacétique	

PPTase : phosphopantéthéinyltransférase

PPi : pyrophosphate

q : vecteur d'onde

R : domaine réductase

RBS : site de liaison des ribosomes

R_g : rayon de giration

R_h : rayon hydrodynamique

RMN : résonance magnétique nucléaire

RMSD : écart quadratique moyen

ROS : espèce réactives de l'oxygène

SAD : diffusion anormale à

une longueur d'onde

SAH : *S*-adénylhomocystéine

SAM : *S*-adénylméthionine

SAXS : diffusion des rayons X aux petits angles

SDR : déshydrogénases et réductases
à chaîne courte

SDS : dodécylsulfate de sodium

SLiM : motif linéaire court

SPR : résonance plasmonique de surface

SYNZIP : bZIP synthétiques

TE : domaine thioestérase

UTR : région transcrite non-traduite

INTRODUCTION BIBLIOGRAPHIQUE

Au cours de l'évolution naturelle, le fonctionnement des enzymes du métabolisme primaire a été progressivement optimisé par le biais de cycles itératifs de mutation, sélection et amplification, pour maximiser le rendement de l'unique produit de biosynthèse. Si celui-ci confère un avantage sélectif aux micro-organismes, alors un tel processus d'adaptation aurait été favorisé afin de répondre à des besoins spécifiques, liés à la prolifération ou à d'autres fonctions. Les voies secondaires, quant à elles activées en réponse à des signaux génétiques ou environnementaux, conduisent de manière fréquente à la biogenèse d'une seule molécule, mais parfois à de multiples métabolites à l'origine d'une grande diversité structurale et chimique.

Dans ce contexte, l'introduction bibliographique détaille d'abord la classification, la production et l'application des métabolites spécialisés, y compris des peptides de synthèse non-ribosomiques (NRP) (section I.), puis expose les nombreuses avancées technologiques mises en œuvre à la fois dans l'identification, la caractérisation et la modification des composés bioactifs synthétisés (section II.). Cette présentation met ensuite en évidence l'organisation et la composition de systèmes NRPS étudiés (section III.), ainsi que les relations structure-fonction élucidées des domaines catalytiques essentiels et additionnels (section IV.).

SECTION I. Classification, production et application des métabolites spécialisés

I.1. Définition, structures chimiques et activités biologiques

Les métabolites spécialisés constituent une source considérable de produits biosynthétisés par les organismes, tels que les bactéries, les champignons, les plantes et les animaux, qui ne sont pas directement impliqués dans leur développement essentiel (Oladipo *et al.*, 2022). Ces produits jouent un rôle crucial dans l'adaptation à l'environnement, la communication entre les micro-organismes ou la défense des ressources au sein de niches écologiques spécifiques (Hemmerling et Piel, 2022). Plus de 2 millions de produits naturels se classent dans divers groupes majeurs, à savoir les alcaloïdes, les phénylpropanoïdes (phénols), les terpénoïdes, les carbohydrates, les polycétides, les peptides non-ribosomiques (NRP) et les peptides modifiés à un niveau post-traductionnel (RiPP), selon leurs voies de biosynthèse, structures chimiques ou activités biologiques (Hemmerling et Piel, 2022).

I.2. Production de métabolites spécialisés par les micro-organismes et les plantes

Les produits naturels trouvent leur source au sein de milieux extrêmement diversifiés, incluant les environnements du sol et aquatiques (*e.g.* sédiments marins), les microbiotes symbiotiques (*e.g.* nez humain, insectes) et endophytaires (*i.e.* végétaux), ou encore des habitats extrêmes (*e.g.* déserts, zones abyssales) (Hammerling et Piel, 2022). En raison de leurs rôles écophysiologiques variés, les composés issus des micro-organismes possèdent des structures chimiques ayant été optimisées au fil de l'évolution afin de se lier sélectivement à des cibles cellulaires (Hammerling et Piel, 2022). Une telle diversité d'écosystèmes et de producteurs de métabolites bioactifs associés est à la base de la découverte d'une vaste gamme de médicaments (**Figure 1**). Depuis « l'âge d'or des antibiotiques » (1940–1960) jusqu'aux programmes modernes, la plupart des thérapies dérivées de bactéries se sont développées à partir d'un nombre restreint de taxons telluriques, facilement accessibles, tels que les actinomycètes filamenteux (*e.g.* *Streptomyces*) (Hammerling et Piel, 2022).

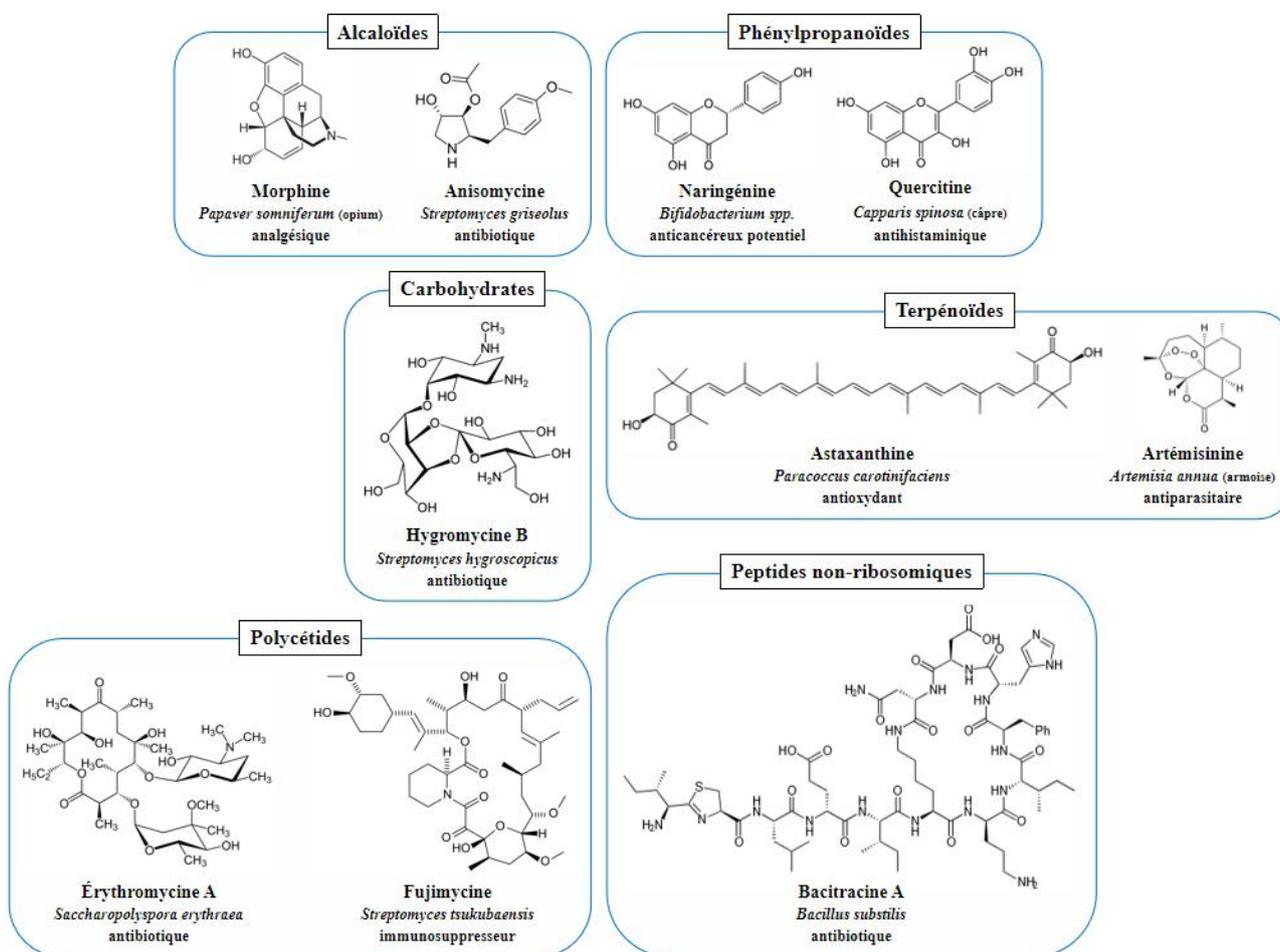


Figure 1. Représentation de structures chimiques et utilisation de produits naturels. Les métabolites spécialisés sont principalement isolés des bactéries, champignons et plantes. Ils regroupent des classes variées au niveau de la structure chimique (alcaloïdes, phénylpropanoïdes, terpénoïdes, carbohydrates, polycétides et NRP) et de la fonction biologique. Ils sont principalement exploités à des fins médicales, par exemple comme antibiotiques (*e.g.* anisomycine, hygromycine, érythromycine A, bacitracine), anticancéreux, immunosuppresseurs (*e.g.* fujimycine), antihistaminiques (*e.g.* quercitine), analgésiques (*e.g.* morphine), antioxydants (*e.g.* astaxanthine) et antiparasitaires (*e.g.* artémisinine).

I.2.A. Groupe des alcaloïdes

Les alcaloïdes sont produits chez les plantes et leurs microbiomes, et dérivent des acides aminés aromatiques, chargés positivement (*i.e.* L-Lys, L-Arg et L-His), non-protéino-gène (L-Orn) ou d'autres acides donnés (*e.g.* anthranilique, nicotinique, *etc.*) (Roberts et Wink, 1998 ; Brielmann *et al.*, 2006). Ces hétérocycles azotés présentent le noyau caractéristique correspondant au précurseur (pipéridine, β -carboline, imidazole, pyrrolizidine, *etc.*). Celui-ci est modifié par des enzymes spécifiques en une structure chimique variée dont dépend leur fonctionnalité (Roberts et Wink, 1998 ; Brielmann *et al.*, 2006). En effet, ces composés servent d'éléments de réserve pour la synthèse protéique et, du fait de leur caractère hautement toxique, contribuent aux mécanismes de défense de l'hôte contre l'action de prédateurs ou pathogènes (Brielmann *et al.*, 2006). Certains demeurent sujets à des études cliniques avancées (*e.g.* anisomycine de la bactérie *Streptomyces griseolus*), et d'autres sont même employés en tant qu'anticancéreux (*e.g.* vincristine), anti-inflammatoires (*e.g.* colchicine), analgésiques (*e.g.* morphine du pavot *Papaver somniferum*), anticholinergiques (*e.g.* scopolamine), ou antiparasitaires (*e.g.* quinine, chloroquine) (Devereaux *et al.*, 2018 ; Thirumurugan *et al.*, 2018 ; Mohammadi *et al.*, 2020 ; Shao *et al.*, 2023) (**Figure 1**).

I.2.B. Groupe des phénylpropanoïdes

Les phénylpropanoïdes sont biosynthétisés par la voie de l'acide shikimique et présentent une structure chimique composée d'unités phénols (C6), comme les dérivés de l'acide hydroxybenzoïque, ainsi que de polymères à longue chaîne, tels que les mélanines de catéchol, les lignines ou les tanins condensés (Agostini-Costa *et al.*, 2012). De même, les stilbénoloïdes et les flavonoïdes présentent une structure chimique formée de deux ou trois cycles aromatiques (C6), respectivement, reliés par une chaîne carbonée (Agostini-Costa *et al.*, 2012). Dans l'ensemble, les dérivés phénoliques protègent les végétaux face aux infections microbiennes et aux stress environnementaux, en particulier contre l'oxydation engendrée par les radicaux libres de l'oxygène (ROS) et les irradiations causées par les rayonnements UV (Agostini-Costa *et al.*, 2012 ; Del Valle *et al.*, 2020). La naringénine, quant à elle extraite des espèces *Bifidobacterium* spp. du microbiote intestinal, inhibe la prolifération cellulaire de lignées du cancer de l'ovaire, ce qui laisse présager un traitement potentiel (Lin *et al.*, 2022).

I.2.C. Groupe des terpénoïdes

Les terpénoïdes trouvent également leur origine chez les micro-organismes et végétaux associés (Mahmoud *et al.*, 2002). Ils dérivent d'unités isoprènes à cinq atomes de carbone (C5), assemblées et modifiées en des structures linéaires ou cycliques et de tailles variées, jusqu'au caoutchouc naturel composé d'environ 20 000 monomères (Mahmoud *et al.*, 2002). À l'instar des phénylpropanoïdes, les terpénoïdes régulent la croissance et le développement de certains tissus spécifiques, et agissent comme divers signaux défensifs ou attractifs (pollinisateurs) (Sharkey et Yeh, 2001). De nombreuses molécules aromatiques se révèlent ainsi être des produits bioactifs, tels que le menthol, le linalol, et le géraniol (monoterpènes à deux unités isoprènes (C10)), et le caryophyllène (C15), chacun utilisé en tant qu'odorant dans l'industrie agro-alimentaire ou en parfumerie (Agostini-Costa *et al.*, 2012). Les propriétés pharmaceutiques de ces composés restent en cours d'investigation, certains étant déjà exploités comme antioxydants (*e.g.* astaxanthine de la bactérie *Paracoccus carotinifaciens*) ou encore antiparasitaires (*e.g.* artémisinine de la plante *Artemisia annua*) (Mohammadi *et al.*, 2020 ; Mussagy et Dufossé, 2023) (**Figure 1**).

I.2.D. Groupe des carbohydrates

Les carbohydrates sont des composés organiques majoritairement extraits des micro-organismes, et présentent une structure chimique composée d'un oligosaccharide et d'un aglycone (McCranie et Bachmann, 2014). Des sous-familles de produits bioactifs, les orthosomycines, les moénomycines, les saccharomicines et les acarviostatines sont biosynthétisées par des voies métaboliques distinctes, et diffèrent en fonction de la nature et du nombre des unités constitutives du squelette (McCranie et Bachmann, 2014). Les carbohydrates ciblent un large éventail de processus biologiques, notamment la perturbation de la biogenèse de la paroi cellulaire et de la traduction bactérienne, ou l'inhibition de l' α -amylase humaine (McCranie et Bachmann, 2014). À ce titre, ils ouvrent la voie à différents traitements antibiotiques et hypolipémiants potentiels (McCranie et Bachmann, 2014). Malgré cela, ces molécules d'une haute complexité demeurent vraisemblablement difficiles à isoler et à identifier, et sont également moins stables, ce qui se traduit par une faible occurrence dans les bases de données (McCranie et Bachmann, 2014). Un des rares métabolites à effet antibactérien notoire utilisé comme tel est l'hygromycine B, une orthosomycine de type II (aminotrisaccharide) produite par la bactérie *S. hygrosopicus* (Borovinskaya *et al.*, 2008) (**Figure 1**).

I.2.E. Groupe des polycétides et peptides non-ribosomiques

Les polycétides, NRP et molécules hybrides sont produits à partir de clusters (opérons) de gènes biosynthétiques (BGC) par des méga-complexes modulaires (Yang *et al.*, 2018). Cette promiscuité moléculaire résulte du transfert horizontal de gènes entre espèces microbiennes, et de leur évolution constante par mutation et recombinaison pour générer une diversité chimique caractéristique (Shi *et al.*, 2019). Un tel processus d'adaptation génique contribue à modifier la spécificité de substrat dans les chaînes d'assemblage et/ou à conférer des propriétés particulières aux métabolites (Yoshikuni *et al.*, 2006). À cet égard, les voies des polycétides synthases (PKS) et des synthétases de peptides non-ribosomiques (NRPS), similaires au niveau de l'architecture linéaire et de la catalyse enzymatique, sont à l'origine de la variété exceptionnelle des produits naturels découverts (Weissman, 2015).

Les polycétides sont biosynthétisés chez les bactéries, les champignons et les plantes, et dérivent d'unités acétyl- ou malonyl-CoA condensées par des PKS de type I, II ou III, puis modifiées par des domaines accessoires et des enzymes auxiliaires post-assemblage (Weissman, 2015). Les structures chimiques diversifiées résultantes confèrent ainsi à ces métabolites une activité biologique dans des processus essentiels et de multiples propriétés médicamenteuses (Weissman, 2015). Par exemple, en termes de mode d'action, la fujimycine isolée de la bactérie *S. tsukubaensis*, inhibe les lymphocytes T effecteurs et, de ce fait, est mise à profit lors de transplantations d'organes (Demain, 1999).

Les NRP regroupent des métabolites spécialisés généralement produits par de nombreux micro-organismes, en particulier les bactéries et les champignons (Walsh *et al.*, 2013 ; Süssmuth et Mainz, 2017) (**Tableau I**). Parmi les bactéries, les producteurs prolifiques correspondent aux divisions des Actinobactéries (*e.g. Streptomyces, Mycobacterium*) et Firmicutes (*e.g. Bacillus*), ainsi qu'à la classe des γ -protéobactéries (*e.g. Pseudomonas, Photorhabdus*) (Süssmuth et Mainz, 2017). Les composés fongiques sont issus des Ascomycètes (*e.g. Typocladium*) ou des Basidiomycètes, mais ces *phylums* demeurent moins explorés à cause d'un génome de grande taille, de la présence d'introns dispersés dans les BGC et d'outils génétiques peu développés (Süssmuth et Mainz, 2017).

Au contraire des RiPP générés par les ribosomes et modifiés à un niveau post-traductionnel, les NRP proviennent de multiples monomères itérativement condensés et modifiés par des NRPS et/ou des enzymes auxiliaires (Süssmuth et Mainz, 2017). Les produits de biosynthèse présentent soit une structure chimique linéaire, de tripeptides (*e.g. pénicilline* ou *céphalosporine*) jusqu'à 25 monomères (*e.g. syringopeptine*), soit cyclique (*e.g. gramicidine S*) (Süssmuth et Mainz, 2017). Celles-ci peuvent aussi être ramifiées de groupements comme un disaccharide glucose-vancosamine (*e.g. vancomycine*) ou un décanol (*e.g. daptomycine*) (Süssmuth et Mainz, 2017). Ces structures chimiques extrêmement diversifiées leur confèrent des propriétés bioactives dans des processus écophysologiques, telles la communication inter-cellulaire et la signalisation moléculaire, ou d'autres fonctions de sidérophores, de facteurs de virulence ou de toxines (Walsh *et al.*, 2013 ; Süssmuth et Mainz, 2017).

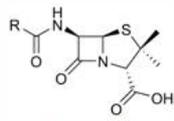
Tableau I. Production de NRP au sein de divisions et classes bactériennes (adapté de Süssmuth et Mainz, 2017).

Division/classe+	Genre	Produits naturels
Actinobactéries	<i>Streptomyces</i>	Multiples, e.g. antibiotiques glycopeptidiques
	<i>Mycobacterium</i> *	Myobactine (sidérophore)
Cyanobactéries	<i>Mycrocystis, Planktothrix, Anabaena</i>	Cyanotoxines, e.g. microcystines, cyanopeptolines
Firmicutes	<i>Bacillus</i>	Lipocyclopeptides, e.g. surfactine (tensioactif)
	<i>Staphylococcus</i> *	Aureusimine
	<i>Streptococcus</i> *	Mutanobactine (sidérophore)
β-Protéobactéries+	<i>Burkholderia</i> *	Malléobactine (sidérophore)
γ-Protéobactéries+	<i>Pseudomonas</i> *	Syringomycine, pyoverdine
	<i>Escherichia</i> *, <i>Salmonella</i> *, <i>Vibrio</i> *	Sidérophores, e.g. entérobactine, salmochéline
	<i>Photobacterium</i>	Multiples peptides linéaires et cycliques
δ-Protéobactéries+	Myxobactéries (ordre)	Anticancéreux e.g. Argyrine, NRP-polycétide

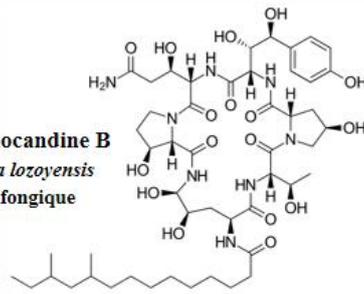
Le symbole + réfère à une classe bactérienne et * à un pathogène humain

Environ 30 structures chimiques de NRP natives, incluant des antibiotiques (e.g. pénicilline G), antibactériens (e.g. vancomycine), antifongiques (e.g. pneumocandine), antiviraux (e.g. daptomycine), anticancéreux (e.g. bléomycine) ou immunosuppresseurs (e.g. cyclosporine A) contribuent à des profits de milliards d'euros dans l'industrie pharmaceutique (Süssmuth et Mainz, 2017) (**Figure 2**). Les médicaments en question font l'objet d'études approfondies sur les relations structure-activité et gagnent le marché sous forme de produits semi-synthétiques. Les NRP microbiens sont également exploités en tant que tensioactifs (e.g. surfactine), pesticides (e.g. bassianolide) et insecticides (e.g. destruxine) par les industries agro-alimentaires (Martínez-Núñez et López, 2016).

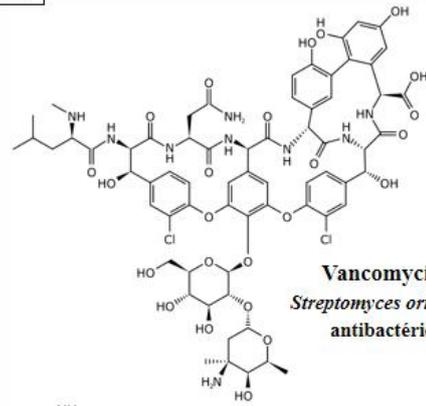
Peptides non-ribosomiques



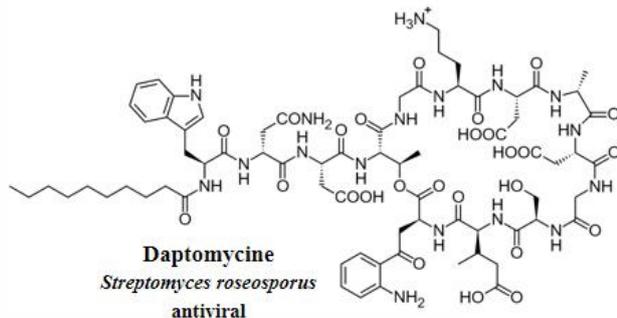
Pénicilline
Penicillium glaucum
antibiotique



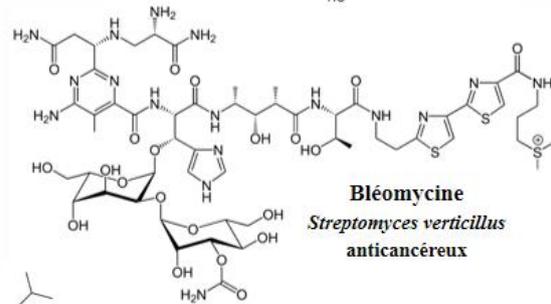
Pneumocandine B
Glarea lozoyensis
antifongique



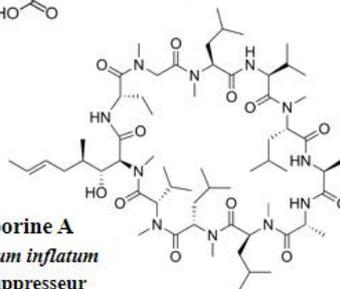
Vancomycine
Streptomyces orientalis
antibactérien



Daptomycine
Streptomyces roseosporus
antiviral



Bléomycine
Streptomyces verticillus
anticancéreux



Cyclosporine A
Tolypocladium inflatum
immunosuppresseur

Figure 2. Représentation de structures chimiques et utilisation de NRP. Les peptides non-ribosomiques incluent des classes diversifiées en termes de structures chimiques et d'activités biologiques. Ils sont essentiellement utilisés dans un but thérapeutique, comme antibiotiques (e.g. pénicilline, vancomycine, pneumocandine), antiviraux (e.g. daptomycine), anticancéreux (e.g. bléomycine) ou immunosuppresseurs (e.g. cyclosporine A), entre autres.

I.3. Applications biomédicales et agro-industrielles

I.3.A. Administration de médicaments et de nanoparticules

Avec l'avènement de nouvelles techniques d'isolation, d'identification et de caractérisation des composés, ayant fait suite à la découverte fortuite de la pénicilline chez le champignon *Penicillium notatum*, les scientifiques ont été incités à exploiter les micro-organismes en raison de leur capacité biosynthétique accrue à produire des métabolites spécialisés à haute valeur ajoutée (Fleming, 1929 ; Demain et Fang, 2000). En effet, comme cela a été décrit précédemment, nombre d'espèces génèrent une gamme diversifiée de produits naturels, ce qui ouvre la voie au développement des applications médicales et au traitement des maladies avec des substances bioactives ou des nanoparticules.

Dans le domaine de la nanomédecine, des métabolites spécialisés aptes à accumuler des ions ont été utilisés pour contrôler la topologie et la morphologie de vecteurs de libération ciblée (Govindappa *et al.*, 2018 ; Khandel et Sashi, 2018). Des nanoparticules d'or ont par exemple été préparées à partir d'extraits de champignons *Cladosporium* sp. et ont exercé un effet anticancéreux significatif contre la lignée cellulaire MCF-7 (Munawer *et al.*, 2020).

I.3.B. Utilisation de pesticides et de régulateurs de croissance

Les micro-organismes produisent de nombreux métabolites employés dans le secteur agricole en raison de leurs fonctions de pesticides, y compris des insecticides (*e.g.* nikkomycine), fongicides (*e.g.* kasugamycine), herbicides (*e.g.* phosphonitricine) et parasitocides (*e.g.* ivermectine), ou même des promoteurs de croissance, afin d'améliorer la qualité des cultures de plantes importantes sur le plan économique (Thirumurugan *et al.*, 2018 ; Ek-Ramos *et al.*, 2019). Parmi ces molécules exogènes, les actinomycètes synthétisent des phénylpropanoïdes et des lipopeptides naturellement favorables à la croissance des plantes et la gestion des maladies des cultures (Ek-Ramos *et al.*, 2019).

De plus, certains micro-organismes de la rhizosphère produisent des antibiotiques directement utilisés par les plantes associées pour lutter contre les stress biotiques (*i.e.* bactéries pathogènes) et environnementaux, dus à des variations de température ou à l'accumulation de métaux lourds (Watts *et al.*, 2023). En outre, les souches microbiennes trouvent une utilité en agriculture, sylviculture et horticulture (Paulitz et Bélanger, 2001). Par exemple, des genres bactériens (*e.g.* *Streptomyces*, *etc.*) et fongiques (*e.g.* *Trichoderma*) servent d'agents de biocontrôle pour améliorer la croissance végétale et ainsi maximiser le rendement agricole (Paulitz et Bélanger, 2001 ; Watts *et al.*, 2023).

La recherche réalisée sur les NRP met en avant des intérêts majeurs dans l'identification de ces produits de synthèse bioactifs, l'exploration de la diversité microbienne, la compréhension de voies métaboliques, la caractérisation de nouvelles molécules et dans les applications biotechnologiques.

SECTION II. Identification, caractérisation et production de composés bioactifs

Les travaux appliqués à la recherche scientifique ont pris une ampleur grandissante en raison des progrès techniques significatifs réalisés dans le séquençage des génomes bactériens, la synthèse de l'ADN et l'édition génomique, l'optimisation de conditions de culture et l'expression hétérologue, ainsi que dans la purification des métabolites en phase liquide et l'analyse spectroscopique (Li *et al.*, 2021). Les efforts entrepris ont considérablement amélioré les prédictions des structures chimiques de produits naturels et de leurs activités biologiques à partir de sources bactériennes inconnues par des approches intégratives de génomique fonctionnelle, criblage computationnel et métabolomique (Hammerling et Piel, 2022) (section II.1–2.). Ces études ont progressivement contribué à la fois à l'isolation, l'identification et la caractérisation de divisions bactériennes peu explorées auparavant et des molécules associées (section II.2.). Elles ont également fourni les connaissances nécessaires afin de produire des dérivés originaux par dérivatisation fonctionnelle, héli-synthèse chimique ou ingénierie biosynthétique (Atanasov *et al.*, 2021) (section II.3.).

II.1. Exploration computationnelle des génomes bactériens

L'approche d'exploration génomique, appelée *genome mining*, consiste d'abord à identifier des BGC non-caractérisés au sein de (méta-)génomés d'organismes séquencés, puis à analyser les voies enzymatiques encodées et à prédire les métabolites biosynthétisés (Singh *et al.*, 2017). Différentes approches bio-informatiques ont ainsi été développées dans cette intention (**Figure 3**).

II.1.A. Identification des BGC par des programmes de criblage à haut-débit

Les connaissances approfondies liées à la biosynthèse des produits naturels tels que les NRP ou les polycétides permettent d'identifier de nouveaux BGC et également de les annoter avec plus de précision (Hammerling et Piel, 2022). Compte tenu des intérêts de la recherche sur les métabolites spécialisés, les algorithmes d'identification des clusters dévoilent pléthore de nouveaux composés et contribuent *a posteriori* à la déréplication des mélanges complexes (Hammerling et Piel, 2022). En plus, ces analyses *in silico* mettent en lumière les producteurs prolifiques parmi l'immense diversité de bactéries au sein des microbiomes (Hammerling et Piel, 2022). Dans tous les cas, les algorithmes mettent à profit la règle de colinéarité entre l'architecture multimodulaire des chaînes d'assemblage NRPS/PKS, dont dépend la spécificité de substrat et la stéréochimie des monomères ajoutés, et la structure du produit naturel biosynthétisé (Staunton et Wilkinson, 1997 ; Singh *et al.*, 2017).

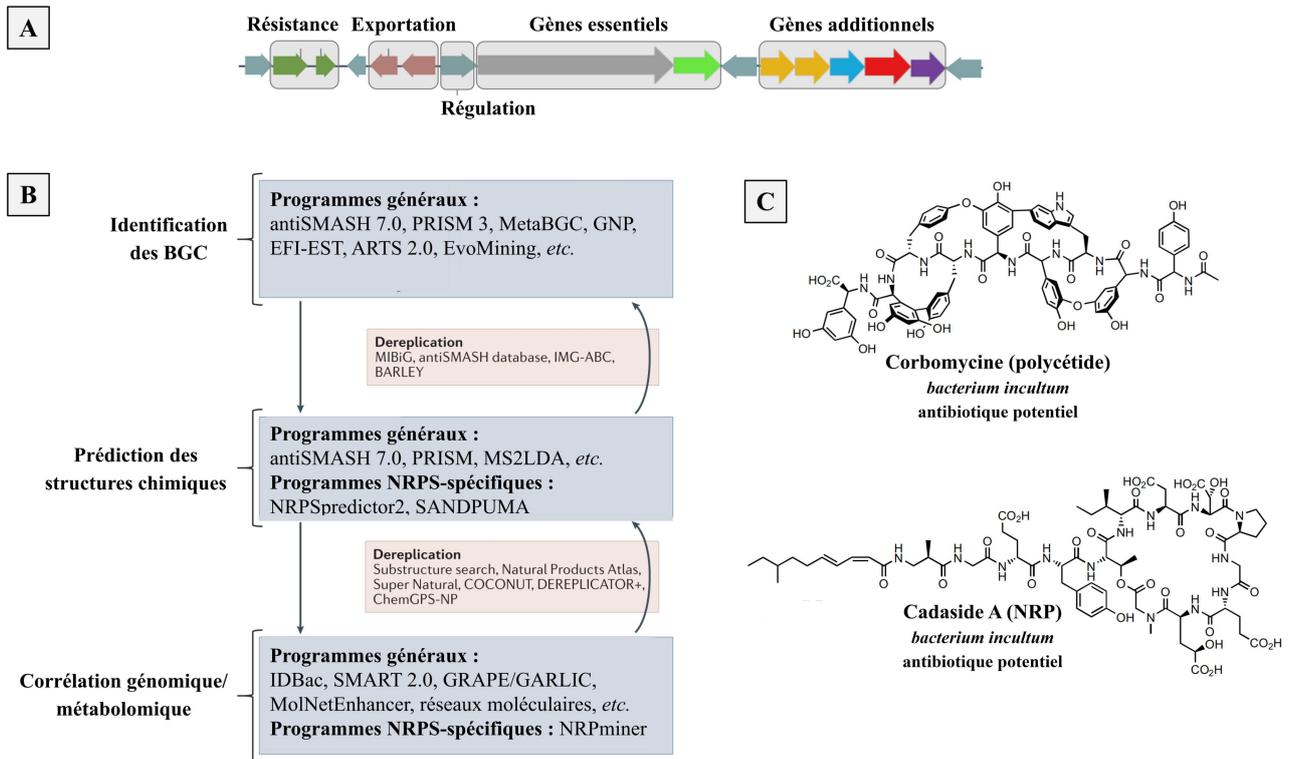


Figure 3. Exploration bio-informatique des génomes bactériens (adaptée de Hemmerling et Piel, 2022). (A) Les clusters des génomes bactériens comportent les gènes biosynthétiques essentiels à l'assemblage du squelette du produit naturel et divers gènes additionnels. Ceux-ci codent des enzymes responsables de la synthèse de certains précurseurs, de la modification du métabolite primaire, de l'auto-résistance et de l'export transmembranaire, par exemple. (B) Le flux de travail relatif à la découverte de produits naturels bioactifs comprend trois étapes successives : l'identification des BGC, la prédiction de la structure chimique du métabolite, et la corrélation avec les résultats analytiques. Il requiert différents programmes et outils bio-informatiques généraux ou spécialisés dans une des classes de métabolites. Entre chacune de ces étapes, le processus de déréplication consiste à ne conserver que les produits naturels nouveaux. (C) Ces approches de *genome mining* ont par exemple permis la découverte du polycétide corbomycine A (Culp *et al.*, 2020) et du NRP cadaside A (Wu *et al.*, 2019) au sein de bactéries non-cultivées, puis de les associer chacun à une activité antibiotique. BGC : cluster de gènes biosynthétiques, NRPS : synthétase de peptides non-ribosomiques.

Un programme bio-informatique couramment employé, antiSMASH 7, permet de cribler à haut-débit des génomes pour y identifier les BGC des principaux groupes de métabolites encodés (Blin *et al.*, 2023). Le module ClusterBlast intégré recherche des BGC phylogénétiquement apparentés chez d'autres organismes, ce qui facilite la déréplication des souches et produits ciblés. L'outil PRISM 3 (*Prediction Informatics for Secondary Metabolomes*) permet quant à lui de reconnaître des BGC liés à 22 classes de métabolites différentes, y compris ceux issus de systèmes non-linéaires (Skinnider *et al.*, 2017). Les NRPS de type non-linéaire seront décrites ultérieurement. L'outil MetaBGC permet de détecter les BGC des séquences métagénomiques, sans nécessiter au préalable l'assemblage des contigs correspondants (Sugimoto *et al.*, 2019). Initialement mises en place sur une base de génome unique, les stratégies d'exploration computationnelle sont désormais étendues à des genres entiers, à des collections de souches et à des microbiomes.

Dans l'ensemble, les programmes bio-informatiques présentent néanmoins certaines limites afin de préciser l'organisation des gènes constitutifs des BGC, et la position du cadre de lecture ouvert ou des séquences régulatrices, ainsi que pour délimiter les bornes des domaines catalytiques (Singh *et al.*, 2017). Des serveurs *in silico* tels que NCBI viewer, ORF finder, BPROM, ou CDD analysis, permettent d'en compléter l'analyse (Singh *et al.*, 2017).

II.1.B. Prédiction des structures chimiques par des outils spécifiques

Outre l'identification de BGC originaux, les programmes antiSMASH et PRISM sont en mesure de traduire les informations génétiques en structures chimiques *via* la règle de colinéarité applicable à la majorité des systèmes NRPS/PKS étudiés (Skinnider *et al.*, 2017 ; Blin *et al.*, 2023). Dans le cas des NRPS, les outils NRSPredictor2 et SANDPUMA intégrés permettent de prédire les spécificités des domaines d'adénylation, responsables de la sélection et de l'activation du substrat d'acide aminé au sein de chaque module, et également la structure peptidique finale (Röttig *et al.*, 2011 ; Chevette *et al.*, 2017). Ces outils offrent ainsi la possibilité de procéder directement à la synthèse chimique de produits computationnels pour outrepasser les difficultés relatives à l'isolation à partir d'une culture bactérienne et tester leur bioactivité respective (Hemmerling et Piel, 2022).

Par exemple, une recherche d'homologie basée sur la composition des résidus du site actif de domaines d'adénylation conservés au sein de la division des Actinomycètes a permis de prédire la biosynthèse du NRP coélichéline (sidérophore) de la souche *Streptomyces coelicolor* A3(2) (Challis, 2008). Cette première réussite de prospection génomique a ouvert la voie à des approches étendues, en particulier dans la découverte d'antibiotiques, comme les lipopeptides gacamide (NRSPredictor2) et cadaside (SANDPUMA) issus de voies non-ribosomiques chez *Pseudomonas* sp. et *Streptomyces* sp., respectivement, au moyen des programmes évoqués (Jahanshah *et al.*, 2019 ; Wu *et al.*, 2019).

Une stratégie alternative intéressante à la prédiction chimique *de novo* consiste à comparer des clusters de gènes putatifs identifiés et des bases de données de BGC expérimentaux (Hemmerling et Piel, 2022). Celles-ci incluent notamment MIBiG 2.0 (*Minimum Information about a BGC*), IMG-ABC v.5.0. (*Integrated Microbial Genomes Atlas of BGC*) et BiG-FAM (*BGC FAMILies database*) (Kautsar *et al.*, 2020 ; Palaniappan *et al.*, 2020 ; Kautsar *et al.*, 2021). Dans ce cadre, une recherche phylogénétique a conduit à l'identification du glycopeptide corbomycine, antibiotique potentiel avec un mode d'action original, à partir de nouvelles espèces de *Streptomyces* (Culp *et al.*, 2020).

II.1.C. Corrélation entre les données génomiques et métabolomiques

Compte tenu du fait que les BGC associés aux produits naturels caractérisés soient généralement inconnus, un cluster génique d'intérêt est susceptible d'être relié à une structure chimique déjà isolée auparavant (Hemmerling et Piel, 2022). Le processus de déréplication implique alors d'identifier les composés connus à l'aide de bases de données métaboliques, avant de se focaliser sur des produits potentiellement sujets à des études prioritaires. Parmi les référentiels comptent *The Natural Products Atlas* avec 30 000 molécules, Super Natural II en lien avec leur toxicité et dégradation, COCONUT (*COLlection of Open NatUral producTs*), ou NP-MRD relatif aux spectres RMN obtenus (Banerjee *et al.*, 2015 ; van Santen *et al.*, 2019 ; Sorokina et Steinbeck, 2020 ; Wishart *et al.*, 2022).

Suite à la déréplication, la méthode d'isolation de produits nouveaux nécessite de comparer les structures chimiques attendues avec les spectres de masse et/ou de résonance magnétique nucléaire (RMN) d'extraits de culture ou fractions brutes (Hemmerling et Piel, 2022). Les données spectrales *in silico* sont notamment obtenues grâce à la plateforme bio-informatique IDBac (MALDI-TOF) et à l'outil d'apprentissage automatique SMART (*Small Molecule Accurate Recognition Technology*) (Clark *et al.*, 2018 ; Reher *et al.*, 2020). Les analyses permettent ainsi de comparer les différences métaboliques et de mettre en évidence les signaux caractéristiques des produits naturels originaux, ainsi que d'établir des réseaux moléculaires (Atanasov *et al.*, 2021). Des méthodes plus complexes requièrent l'utilisation de précurseurs isotopiques, l'expression hétérologue de BGC, la comparaison avec des souches mutantes et/ou des analyses métabolomiques en relation avec les essais biologiques (Atanasov *et al.*, 2021 ; Hemmerling et Piel, 2022).

Des méthodes intégratives ont ainsi été développées afin de mettre directement en corrélation les données génomiques et métabolomiques. Le moteur de prédiction d'assemblage rétro-biosynthétique GRAPE et l'outil d'alignement GARLIC combinés permettent d'associer les clusters de gènes à des métabolites déjà connus ou à des nouvelles structures chimiques (Dejong *et al.*, 2016). De plus, les algorithmes de déréplication, tels que DEREPLICATOR+, font partie intégrante des pipelines bio-informatiques, ce qui rationalise le processus de découverte des produits et en améliore l'affectation dans les réseaux moléculaires (Mohimani *et al.*, 2018). Ceux-ci sont établis à travers les plateformes correspondantes GNP (*Genomes-to-Natural Products*) ou GNPS (*Global Natural Products Social Molecular Networking*) (Johnston *et al.*, 2015 ; Wang *et al.*, 2016 ; Hemmerling et Piel, 2022). Par ailleurs, le programme NRPminer permet également d'identifier des produits à partir d'ensembles de données (méta-)génomiques et analytiques, mais en plus d'en détecter diverses modifications post-assemblage natives et d'en cribler la bioactivité *in silico* (Behsaz *et al.*, 2021). Grâce à ces flux de travail optimisés et aux métadonnées disponibles, des études de microbiomes entiers sont possibles, avec par exemple l'identification de familles originales de NRP à fonction d'antiparasitaire chez les γ -protéobactéries *Xenorhabdus* et *Photorhabdus* (XPF) (Behsaz *et al.*, 2021). En outre, les réseaux

métabolomiques peuvent être associés à des données transcriptomiques, protéomiques, et à des cribles basés sur la cytométrie ou l'imagerie de cellules uniques (Atanasov *et al.*, 2021).

II.1.D. Exploration des métabolomes bactériens

Les approches d'exploration génomique exposées précédemment aboutissent à la découverte de produits naturels issus de classes déjà caractérisées au sein des réseaux moléculaires, mais s'avèrent moins favorables à la recherche de BGC isolés sur le plan évolutif (Hammerling et Piel, 2022). Dans ce contexte, l'algorithme ClusterFinder et l'outil d'apprentissage automatique DeepBGC permettent d'identifier de nouveaux types de BGC (divergents) au sein des génomes bactériens et également de les regrouper en fonction de la bioactivité potentielle des métabolites associés (Cimermanic *et al.*, 2014 ; Hannigan *et al.*, 2019). De plus, le moteur BiG-SCAPE sert à établir des réseaux interactifs de similarité de séquences des clusters, tandis que le programme CORASON en tandem avec celui-ci dans un pipeline bio-informatique, contribue à élucider les relations évolutives entre les familles de BGC au sein d'un microbiome donné (Navarro-Muñoz *et al.*, 2020). Par ailleurs, des variantes de cette stratégie créent aussi de tels réseaux, soit *via* des sous-classes d'enzymes biosynthétiques et de leurs fonctions propres (EFI-EST), soit à travers des ensembles de domaines catalytiques (modules) rares constitutifs des systèmes NRPS/PKS (CO-ED) (Gerlt *et al.*, 2015 ; de Rond *et al.*, 2021). Par exemple, l'outil EFI-EST a permis la découverte de la brévicidine, NRP actif contre des pathogènes à Gram négatif (Li *et al.*, 2018).

Les avancées accomplies dans les technologies computationnelles méta-omiques ont fourni les moyens adéquats afin d'accéder à la diversité cryptique des produits naturels et mettre en évidence leur potentiel majeur dans la santé humaine (Hammerling et Piel, 2022). Parallèlement, les progrès significatifs entrepris dans l'apprentissage automatique sont apparus primordiaux pour en prédire les cibles biologiques et les modes d'action, ainsi que pour concevoir des médicaments (Atanasov *et al.*, 2021 ; MULLOWNEY *et al.*, 2023).

Toutefois, ces approches d'exploration génomique doivent être suivies de stratégies d'extraction des métabolites. Ces expérimentations requièrent l'activation des BGC silencieux *via* l'optimisation empirique des conditions de culture bactérienne, l'expression hétérologue des clusters au sein d'un châssis idéal et/ou la régulation de leur expression génique, en vue de la caractérisation des produits par des techniques analytiques ou par profilage métabolique (Atanasov *et al.*, 2021).

II.2. Optimisation des conditions de culture bactérienne et activation des BGC cryptiques

Historiquement, les approches de cultures bactériennes se concentrent sur les bactéries simples à manipuler et à faire croître, ce qui a induit un biais dans la découverte des produits naturels extraits des souches étudiées au premier plan (*e.g. Streptomyces, Bacillus, etc.*) (Atanasov *et al.*, 2021). La diversité métabolique ayant été largement sous-évaluée, les approches récentes se portent davantage vers des genres moins explorés, mais aussi capables de biosynthétiser des produits d'une haute valeur ajoutée (Pidot *et al.*, 2014 ; Challinor et Bode, 2015) (**Figure 4**). Par exemple, les γ -protéobactéries pathogènes *Xenorhabdus*, *Photorhabdus* ou *Pseudomonas* sont chacune des organismes producteurs de composés bioactifs en quantités abondantes (Tobias *et al.*, 2017 ; Saati-Santamaría *et al.*, 2022). Néanmoins, plus de 99 % des espèces bactériennes ne peuvent théoriquement pas être cultivées avec réussite dans des conditions de laboratoire classiques, et environ 90 % de leurs BGC sont cryptiques et doivent être activés, ce qui limite fortement l'accès à leur capacité biosynthétique (Pham et Kim, 2012 ; Dejong *et al.*, 2016 ; Atanasov *et al.*, 2021). Des travaux intégratifs ont progressivement mis au jour celle-ci à la faveur de l'identification de centaines de BGC silencieux au sein des génomes, de leur activation *in vivo*, et de la caractérisation des produits en termes de structures chimiques et d'activités biologiques (Saati-Santamaría *et al.*, 2022 ; Hemmerling et Piel, 2022).

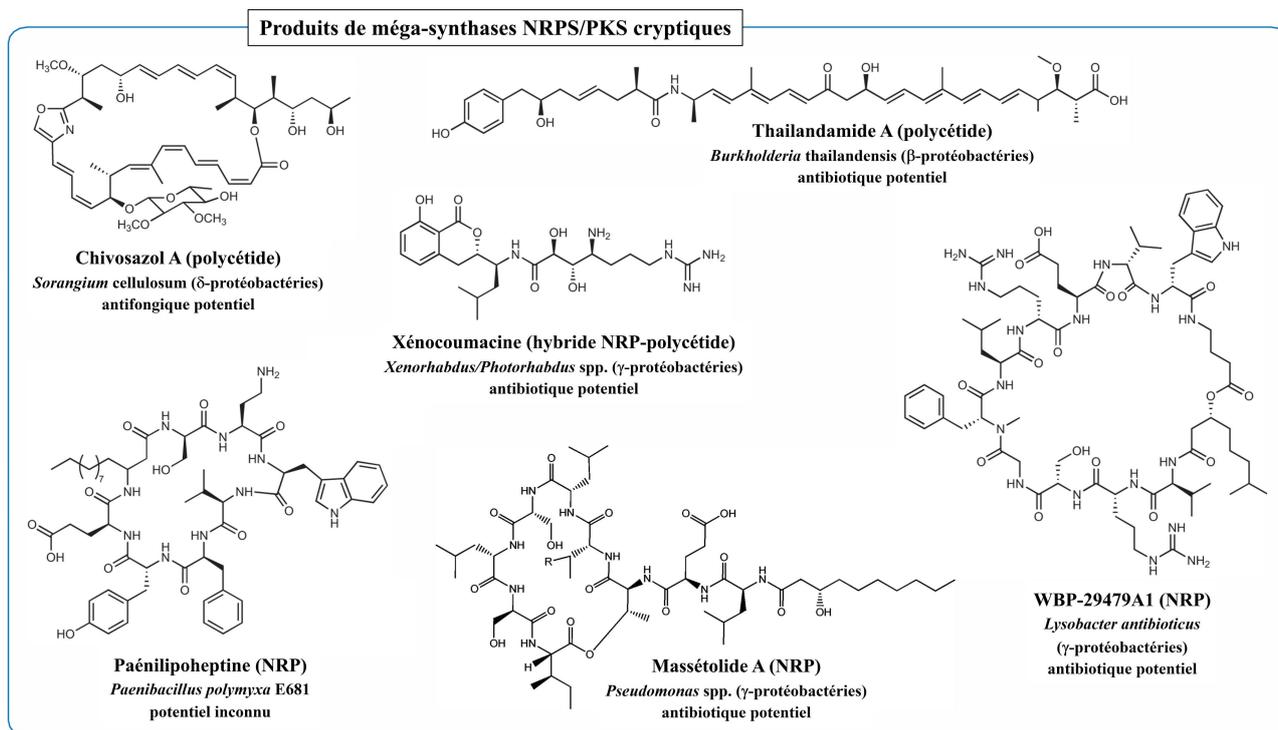


Figure 4. Représentation de structures chimiques et utilisation potentielle de produits de NRPS/PKS cryptiques (adaptée de Hemmerling et Piel, 2022). Les produits naturels tels que les peptides non-ribosomiques et les polycétides cryptiques sont découverts grâce à des approches combinées d'exploration génomique *in silico* et d'isolation *in vivo*. Ils trouvent leur source chez des ordres bactériens jusqu'alors peu étudiés, y compris les γ -protéobactéries, et couvrent un large spectre d'activités, en l'occurrence des antibiotiques potentiels.

En particulier, diverses stratégies de culture bactérienne et d'activation des BGC cryptiques ont été développées afin de promouvoir la prolifération des micro-organismes dans un environnement naturel simulé et d'améliorer au possible la détection des produits naturels biosynthétisés (Ren *et al.*, 2020 ; Atanasov *et al.*, 2021 ; Hemmerling *et al.*, 2022) (**Figure 5**).

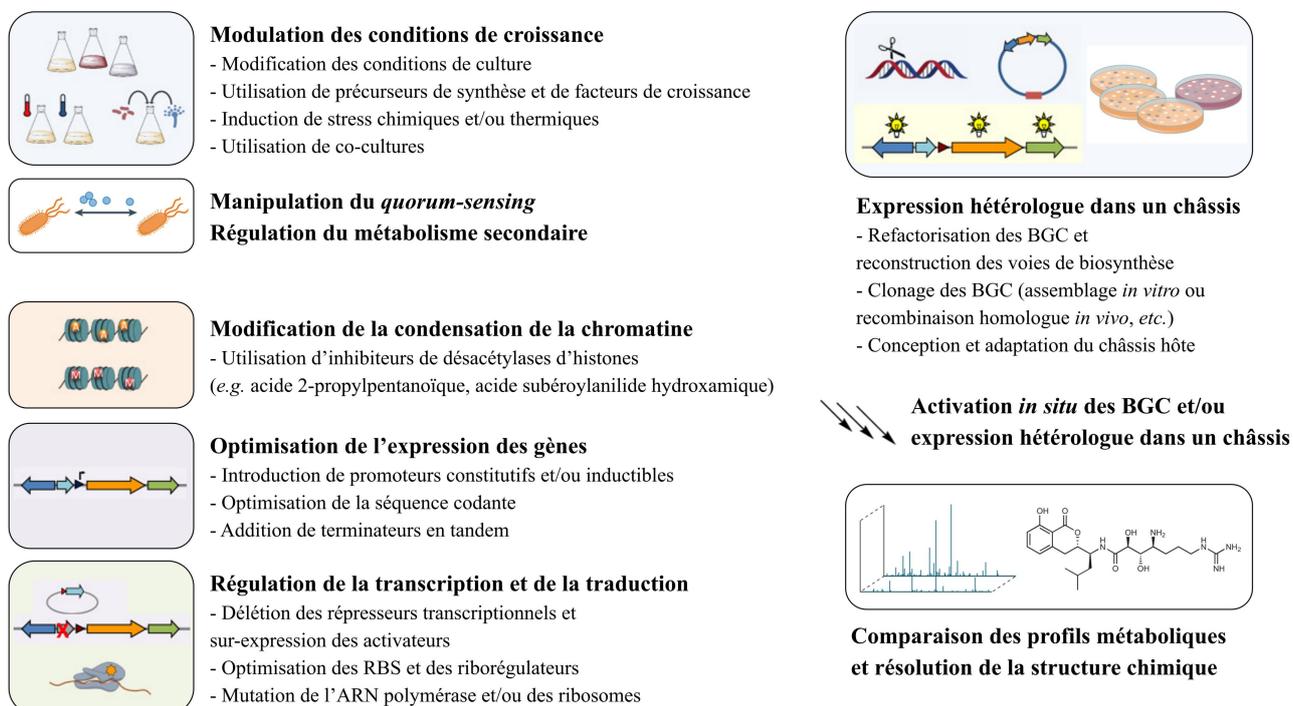


Figure 5. Optimisation des conditions de culture bactérienne et activation des BGC cryptiques. Les stratégies d'isolation des produits naturels nécessitent soit la modulation des conditions de prolifération de l'hôte natif suivie de l'activation *in situ* des BGC régulés, soit l'expression hétérologue de BGC refactorisés (réorganisés) au sein d'un châssis optimisé. Les stratégies de profilage métabolique assurent ensuite une corrélation entre l'expression d'un BGC donné et la production du métabolite prédit *in silico*, préalablement à l'élucidation de la structure chimique. BGC : clusters de gènes biosynthétiques, RBS : site de liaison des ribosomes.

II.2.A. Modulation des conditions de croissance par des facteurs biotiques et abiotiques

L'approche la plus simple à mettre en œuvre, désignée par le terme OSMAC (*One Strain Many Compounds*), consiste à moduler les conditions de culture bactérienne testées, telles que les sources de nutriments, le pH et la température, ou le degré d'hygrométrie (Challinor et Bode, 2015). Cette approche originelle est fréquemment combinée à des stratégies alternatives, y compris l'utilisation de précurseurs de synthèse, comme la L-proline présente dans l'hémolymphe des insectes et requise aux bactéries entomopathogènes (*Xenorhabdus* spp.), des facteurs de croissance (e.g. co-enzymes, sidérophores, etc.), des régulateurs globaux du métabolisme secondaire (e.g. agent bactériostatique triméthoprime), ou des micro-organismes symbiotiques (e.g. champignons ascomycètes) (Crawford *et al.*, 2010 ; D'Onofrio *et al.*, 2010 ; Seyedsayamdost *et al.*, 2014 ; Moussa *et al.*, 2019). Dans le cas de la culture anaérobie de la bactérie *Clostridium cellulolyticum*, des extraits aqueux du sol ont été directement ajoutés dans le milieu de croissance, ce qui a permis d'isoler le NRP closthioamide à potentiel antibiotique (Lincke *et al.*, 2010).

Des stratégies de miniaturisation ont également été développées dans l'intention de reproduire les conditions exactes de milieux naturels avec, entre autres, le dispositif micro-puce iChip (Nichols *et al.*, 2010). Cette plateforme de criblage nécessite d'abord de diluer les échantillons bactériens étudiés en des cellules uniques, avant d'incuber celles-ci dans des milliers de chambres à membrane semi-perméable (Nichols *et al.*, 2010). Suite à l'ensemencement, le dispositif analytique est enfoui dans l'environnement originel, ce qui permet de cultiver *in situ* les souches et de les isoler (Nichols *et al.*, 2010). Cette technologie novatrice se révèle parfaitement adaptée pour explorer la diversité microbienne et investiguer la phylogénie des espèces éloignées (Nichols *et al.*, 2010). Initialement appliquée aux communautés du sol, la micro-puce iChip a permis de découvrir le NRP teixobactine, un antibiotique produit chez la β -protéobactérie *Eleftheria terrae* (Ling *et al.*, 2015) (**Figure 6**).

- Utilisation du dispositif iChip avec des échantillons microbiens dilués et incubés dans leur milieu originel (sol)
- Extraction de 10 000 isolats de la micro-puce iChip et réalisation d'essais de criblage de l'activité contre des pathogènes courants
- Identification d'un extrait d'une nouvelle espèce bactérienne, *Eleftheria terrae*, avec une activité accrue contre le pathogène *Staphylococcus aureus*
- Éluclidation de la structure chimique de l'antibiotique teixobactine par RMN

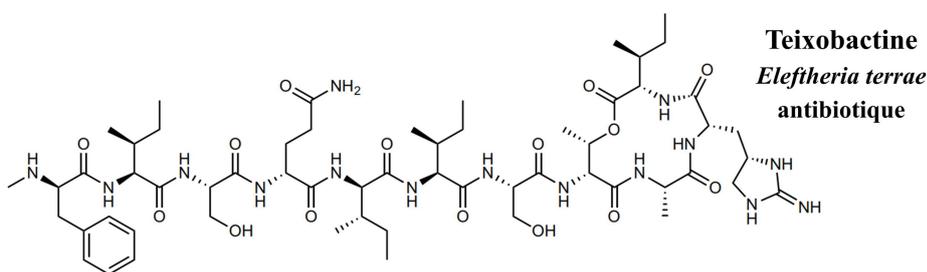


Figure 6. Isolation de la teixobactine à partir de la technologie micro-puce iChip. Le dispositif iChip se compose de milliers de chambres de diffusion, au sein desquelles prolifèrent les cellules uniques sous l'influence de conditions de leur milieu originel. Un isolat d'un nouveau genre, *Eleftheria*, produit la teixobactine aux propriétés antibiotiques contre les pathogènes multi-résistants *Staphylococcus aureus* et *Mycobacterium tuberculosis*. Ce métabolite spécialisé se lie spécifiquement à des motifs structuraux des précurseurs lipidiques du peptidoglycane, ce qui empêche la formation de la paroi bactérienne et conduit à l'élimination des espèces bactériennes à Gram positif ciblées.

En dépit des nombreuses avancées entreprises dans les stratégies de culture de micro-organismes, les conditions optimales de croissance ne permettent pas toujours de dévoiler le spectre des produits naturels biosynthétisés au sein des environnements extrêmement complexes (Atanasov *et al.*, 2021). Néanmoins, la miniaturisation des équipements analytiques, à l'instar du dispositif iChip, contribue à améliorer la détection des métabolites directement dans la biosphère (Atanasov *et al.*, 2021).

II.2.B. Expression hétérologue de BGC cryptiques au sein de souches châssis

De nombreuses souches nouvellement isolées restent difficiles à cultiver en milieux optimisés, ce qui se traduit par la faible disponibilité des médicaments issus des voies de biosynthèse étudiées (Ren *et al.*, 2020). Dans ce contexte, les BGC correspondants sont alors refactorisés (réorganisés) à façon au niveau des gènes, cistrons et éléments régulateurs, puis clonés dans des vecteurs réplicatifs ou intégratifs, à l'image des plasmides, cosmides, fosmidés, ou chromosomes artificiels (BAC) (Liu *et al.*, 2021). Ceux-ci sont ensuite transférés au sein de micro-organismes adaptés à la manipulation génétique, à l'expression hétérologue, ainsi qu'à la fermentation industrielle (Ren *et al.*, 2020). En particulier, *Streptomyces* spp. jouent un rôle crucial dans l'identification des produits bioactifs, avec plus de 20 % du génome associé au métabolisme spécialisé et environ 100 000 composés à fonction d'antibiotique potentiel (Watve *et al.*, 2001 ; Baltz, 2019). Divers outils et méthodes ont donc été développés afin de réorganiser et activer les BGC cryptiques, et ce, en vue d'augmenter à terme la production des métabolites (Ren *et al.*, 2020 ; Liu *et al.*, 2021).

II.2.B.α. Refactorisation des BGC par des outils *in silico* et des méthodes d'assemblage

Sur le plan expérimental, un BGC d'intérêt gagne à être modifié *via* l'introduction de promoteurs constitutifs et/ou inductibles, l'optimisation des séquences codantes et l'addition de terminateurs en tandem (Ren *et al.*, 2020) (**Figure 7**). En termes d'analyse des parties génétiques, des algorithmes d'apprentissage automatique, comme BacPP ou BPROM, permettent à la fois de déterminer la force d'un promoteur endogène à partir du facteur σ correspondant et de reconnaître les sites d'initiation de la transcription du BGC ciblé (Solovyev et Salamov, 2011 ; de Avila *et al.*, 2016). Des modèles en cours de développement visent désormais à concevoir rationnellement les motifs spécifiques des promoteurs σ^{70} -dépendants avec le taux de transcription souhaité (LaFleur *et al.*, 2022). Le serveur Présyncodon sert quant à lui à diminuer les biais de redondance des codons et le contenu en bases GC des CDS, mais aussi à éliminer les structures en tiges-boucles potentielles des ARNm transcrits (Tian *et al.*, 2018). Parmi les méthodes mentionnées, la substitution des promoteurs s'avère efficace avec, par exemple, 10 promoteurs forts de la souche *S. albus* modèle (J1074) ayant été utilisés dans une plateforme *plug-and-play* afin d'activer un BGC silencieux (altéramide) chez plusieurs hôtes apparentés (Luo *et al.*, 2015). Par ailleurs, la base de données WebGeSter relatives aux terminateurs ρ -indépendants et le programme FindTerm aident à concevoir les structures élémentaires idéales (*e.g.* faibles nombres de mésappariements, présence d'une queue poly-T, *etc.*), voire des tiges-boucles en tandem (Mitra *et al.*, 2011 ; Solovyev et Salamov, 2011).

En outre, le BGC ciblé peut être régulé au niveau transcriptionnel *via* la délétion des répresseurs et/ou la sur-expression des activateurs (Liu *et al.*, 2021) (**Figure 7**). Dans une étude d'inactivation génique d'un répresseur transcriptionnel (*mmyR*) à l'aide du système CRISPR-Cas9, un tel contrôle positif a permis d'identifier l'acide sclérique, NRP bioactif produit chez un proche hôte hétérologue

(*Streptomyces* sp.) (Alberti *et al.*, 2019). Parallèlement à cette régulation interne au cluster de gènes, le processus d'équilibration global des flux métaboliques consiste parfois en la décondensation de la chromatine en présence d'inhibiteurs des histones désacétylases (*e.g.* acide 2-propylpentanoïque), en des mutations favorables de la sous-unité β de l'ARN polymérase RpoB (*e.g.* H437Y) ou encore en l'induction de la réponse stringente (alarhone guanosine pentaphosphate) (Artsimovich *et al.*, 2004 ; Gomez-Escribano et Bibb, 2011 ; Zhou et Troyanskaïa, 2015). Enfin, le cluster étudié peut être ré-assemblé de sorte que les ARNm transcrits présentent des régions 5'-UTR pré-définies (Ren *et al.*, 2020) (**Figure 7**). Pour cela, les programmes RBS calculator et Toeholder sont capables de générer des séquences synthétiques optimales de liaison aux ribosomes (*i.e.* RBS) ou à divers effecteurs (*e.g.* *riboswitches*), ainsi que de prédire le taux d'initiation de la traduction correspondant (Salis, 2011 ; Cisneros *et al.*, 2023). De surcroît, des mutations ponctuelles de la protéine ribosomale S12 (RpsL) (*e.g.* K88E) de la sous-unité 30S peuvent s'avérer bénéfiques (Shima *et al.*, 1996 ; Ochi et Hosaka, 2013). Par exemple, une méthode TTO (*i.e.* *Transcription-Translation in One*) d'expression en *trans* des gènes *rpoB/rpsL* mutants, à l'aide d'un système de plasmides modulaires, a permis d'activer des BGC cryptiques chez *Streptomyces* sp. et ainsi d'en modifier le profil métabolique avec la détection de polycétides aromatiques à activité antibiotique (Zhang *et al.*, 2020).

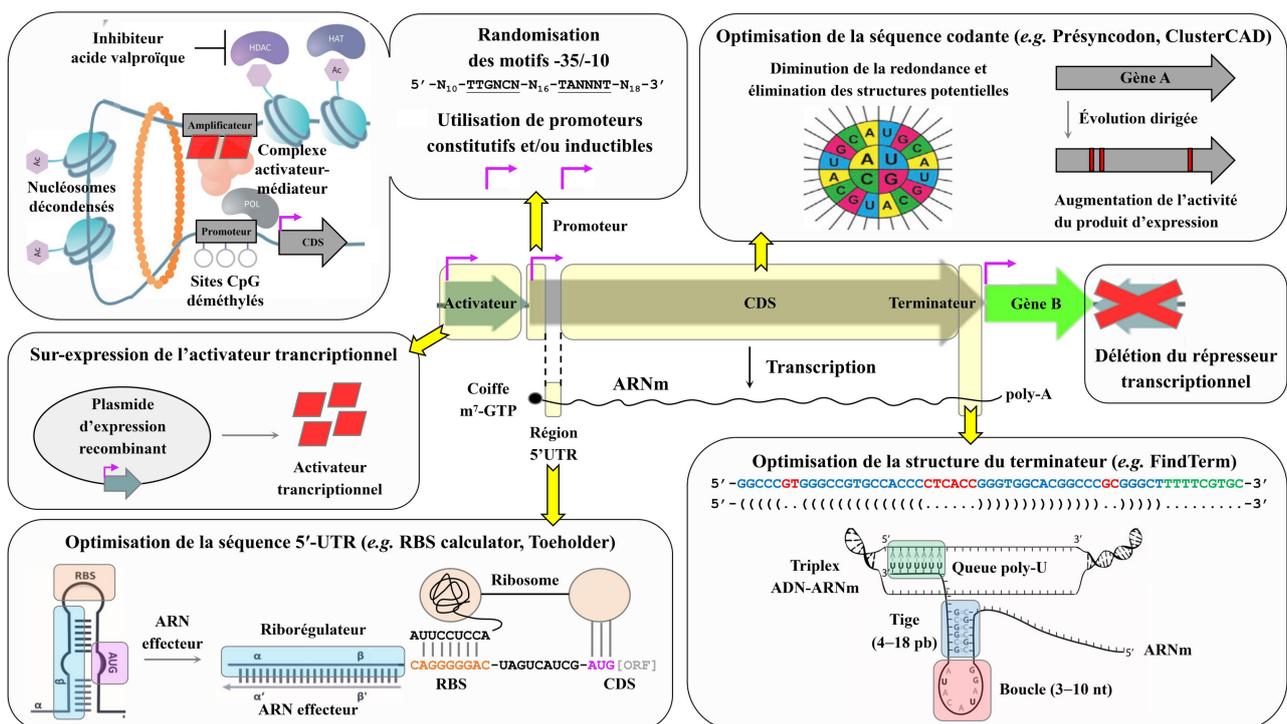


Figure 7. Conception *in silico* de parties génétiques et optimisation de l'expression génique d'un BGC. Des parties génétiques d'un BGC donné tels que les promoteurs et les sites de liaison aux ribosomes (RBS), les séquences codantes (CDS), ainsi que les terminateurs, sont susceptibles d'être optimisées en termes de séquences à l'aide d'outils d'analyse *in silico* afin d'activer l'expression génique. En plus de ces modifications, les gènes codants des régulateurs positifs ou négatifs de la transcription peuvent être sur-exprimés *via* des plasmides ou au contraire supprimés du BGC.

Suite au processus d'exploration génomique, les BGC identifiés ne contiennent généralement pas l'ensemble des gènes nécessaires à la biosynthèse, certains étant situés en dehors des clusters, alors les stratégies de rétrosynthèse apparaissent pertinentes afin de concevoir des voies artificielles avec des gènes orthologues (Hadadi et Hatzimanikatis, 2015). En effet, un produit naturel donné peut être théoriquement scindé en ses propres précurseurs et des enzymes biosynthétiques impliquées dans les réactions en cause peuvent être rationnellement exploitées pour reconstruire l'ensemble du système étudié (Hadadi et Hatzimanikatis, 2015). Actuellement, l'un des principaux programmes disponibles, BioCAT (*Biosynthesis Cluster Analysis Tool*), combine le logiciel antiSMASH 7 et l'algorithme de rétrosynthèse rBAN des NRP (Ricart *et al.*, 2019 ; Konanov *et al.*, 2022). Bien que cette méthode ait été appliquée avec succès sur plusieurs classes de produits naturels, elle ne reste que rarement testée sur le plan expérimental à cause de la difficulté à identifier des enzymes biosynthétiques avec une spécificité adéquate (Cravens *et al.*, 2019). En conséquence, des outils capables de les classer selon la similarité et/ou la promiscuité de substrat sembleraient mieux adaptés à cette fin (Ren *et al.*, 2020).

Les éléments génétiques du BGC conçu à façon sont ensuite assemblés entre eux et clonés dans des vecteurs répliatifs ou intégratifs (Chao *et al.*, 2015). Diverses méthodes ont été développées et peuvent être divisées en différents groupes, basés sur la restriction enzymatique, sur l'homologie de séquence et/ou sur l'hybridation d'oligonucléotides (Chao *et al.*, 2015) (**Tableau II**). Les techniques BioBrick, Golden Gate ou MASTER, entre autres, établies sur l'usage d'endonucléases de restriction de type IIP/IIS et de l'ADN ligase, permettent d'assembler des constructions d'une taille importante dans un ordre pré-défini, mais les parties génétiques doivent être exemptes des sites en question, ce qui nécessite souvent la mutagenèse préalable de leurs séquences (Chao *et al.*, 2015). Compte tenu du fait que ces méthodes dépendent de l'appariement complémentaire d'extrémités cohésives (4 nt), cela limite l'efficacité du clonage à 3–10 fragments d'ADN (Chao *et al.*, 2015). De par le principe inhérent aux techniques d'assemblage liées à l'homologie *in vitro* (e.g. Gibson, SLiCE, LIC, SIRA, CATCH) ou *in vivo* (e.g. TAR/miCRISTAR, LCHR/LLHR, ExoCET) et en contrepartie de schémas complexifiés, celles-ci permettent de reconstruire des voies de biosynthèse complètes en une unique étape (Chao *et al.*, 2015). À titre d'exemple, la méthode DiPaC (*Direct Pathway Cloning*) a permis d'activer le cluster silencieux responsable de la biosynthèse de nostamide A, composé bioactif issu d'une voie NRPS de la cyanobactérie *Noctos punctiforme* non-cultivée (Greunke *et al.*, 2018).

Tableau II. Utilisation de méthodes d'assemblage d'ADN.

Méthode	Utilisation
Restriction enzymatique	
BioBrick/BglBrick (Anderson <i>et al.</i> , 2010)	Assemblage séquentiel d'éléments génétiques standards.
Golden Gate (Engler <i>et al.</i> , 2008)	Clonage modulaire et directionnel de parties d'ADN avec des extrémités non-palindromiques en présence d'enzymes de type IIS spécifiques (<i>BbsI</i> , <i>BsaI</i> et <i>BsmBI</i>)
MASTER (<i>Methylation-Assited Tailorable Ends Rational</i>) (Chen <i>et al.</i> , 2013)	Utilisation d'amorces méthylées et de l'enzyme de type IIS MspJI. Reconnaissance spécifique des sites ^m CNNR
Homologie séquentielle <i>in vitro</i>	
OE-PCR (<i>Overlap Extension PCR</i>) (Horton <i>et al.</i> , 1989)	Assemblage de parties d'ADN avec des régions homologues par PCR. Absence de séquence résiduelle.
SLIC (<i>Sequence and Ligation-Independant Cloning</i>) (Li et Elledge, 2007)	Utilisation des activités 3'-exonucléase et polymérase de l'ADN polymérase T4.
Gibson (Gibson <i>et al.</i> , 2009)	Incubation des inserts et du vecteur en présence d'enzymes spécifiques (exonucléase T5, ADN polymérase Phusion et ligase T4) à 50 °C.
CPEC (<i>Circular Polymerase Extension Cloning</i>) (Quan et Tian, 2009)	Extension des régions d'homologie entre les inserts et le vecteur par l'ADN polymérase Phusion.
SLiCE (<i>Seamless Ligation Cloning Extract</i>) (Zhang <i>et al.</i> , 2012)	Utilisation d'extraits d' <i>E.coli</i> sous forme de BAC.
NE-LIC (<i>Nicking Endonucleases Ligation-Independent Cloning</i>) (Wang <i>et al.</i> , 2013).	Création des régions d'homologie par digestion des inserts avec des enzymes de restriction. Assemblage des parties d'ADN par hybridation à 37 °C.
SIRA (<i>Serine-Integrase Recombinational Assembly</i>) (Colloms <i>et al.</i> , 2014 ; Du <i>et al.</i> , 2015)	Recombinaison site-spécifique <i>attP/attB</i> par des sérine-intégrases (ϕ BT1, ϕ C31).
CATCH (<i>Cas9-Assited Targeting of CHromosome segments</i>) (Jiang <i>et al.</i> , 2015)	Excision du BGC du chromosome bactérien par la nucléase Cas9 et assemblage de Gibson avec le vecteur en une seule étape.
DiPaC (<i>Direct Pathway Cloning</i>) (Greunke <i>et al.</i> , 2018)	Amplification des BGC par <i>long-range</i> PCR et assemblage de Gibson avec le vecteur.
Homologie séquentielle <i>in vivo</i>	
TAR (<i>Transformation-Associated Recombination</i>) miCRISTAR (Orr-Weaver <i>et al.</i> , 1981 ; Kim <i>et al.</i> , 2019) <i>DNA assembler</i> (Shao <i>et al.</i> , 2009)	Incubation de l'ADN génomique avec des enzymes de restriction de type II (TAR) ou de type V (miCRISTAR). Co-transformation des fragments générés ou de parties d'ADN pré-amplifiées (<i>DNA assembler</i>) et du vecteur chez l'hôte intermédiaire <i>Saccharomyces cerevisiae</i> .
LLHR/LCHR (<i>Linear-Linear/Circular Homologous Recombination</i>) (Fu <i>et al.</i> , 2012)	Co-transformation du BGC linéaire ou circulaire et du vecteur chez l'hôte intermédiaire <i>E. coli</i> .
ExoCET (<i>Exonuclease Combined with RecET recombination</i>) (Wang <i>et al.</i> , 2018)	Utilisation de l'activité 3'-exonucléase de la polymérase T4 <i>in vitro</i> . Recombinaison entre l'ADN linéaire et le vecteur chez l'hôte intermédiaire <i>E.coli</i> .
Hybridation d'oligonucléotides	
LCR (<i>Ligase Cycling Reaction</i>) (Kok <i>et al.</i> , 2014)	Répétition de cycles dénaturation-hybridation-ligature avec des oligonucléotides et une ADN ligase thermostable.

Les outils *in silico* et méthodes expérimentales associés à la refactorisation des BGC fournissent une plateforme pratique afin de découvrir des produits naturels avec des bioactivités intéressantes à une échelle sans précédent (Li *et al.*, 2021). Cependant, des souches châssis appropriées et des outils d'ingénierie génomique élaborés se révèlent parfois indispensables à l'expression hétérologue.

II.2.B.β. Construction et adaptation de souches châssis par ingénierie génomique

Les châssis hétérologues optimaux doivent : (1) être faciles à cultiver en conditions de laboratoire et à l'échelle de bioréacteurs, (2) présenter un réseau de régulation génétique et métabolique connu, (3) avoir un faible bruit de fond en termes de métabolites secondaires, (4) disposer des précurseurs et cofacteurs nécessaires à la biosynthèse de produits complexes, (5) posséder un système d'export fonctionnel, et (6) être sujets à des éditions génomiques avec les outils disponibles (Liu *et al.*, 2021). Un panel d'espèces telles que *E. coli*, *Bacillus subtilis*, *Pseudomonas putida* ou *Streptomyces* spp., a ainsi été caractérisé et standardisé pour exprimer des BGC d'intérêt dans un contexte spécifique (Li *et al.*, 2021). Ces hôtes reconnus ont été modélisés à partir de leurs réseaux métaboliques à l'échelle du génome et optimisés de manière combinatoire à travers des modifications génétiques pertinentes afin d'améliorer les titres de production des métabolites biosynthétisés (Martin-Pascual *et al.*, 2021).

Les réseaux métaboliques computationnels (GEM) établis permettent notamment d'identifier les gènes essentiels, y compris les régulateurs pléiotropiques, le flux des ressources, ainsi que le taux de croissance ou le rendement de production du micro-organisme étudié (Martin-Pascual *et al.*, 2021). Les régions chromosomiques instables et/ou non-nécessaires (*e.g.* multiples BGC) sont sujettes à des réarrangements, alors leur suppression contribue à diminuer les interférences avec les produits ciblés en phase de détection (Liu *et al.*, 2021). À titre indicatif, la délétion de 11 BGC (229 kb) dans des souches de *S. lividans* a permis de générer un châssis à génome réduit et d'y produire des nouveaux composés de différentes classes (Ahmed *et al.*, 2020). De plus, les gènes pléiotropiques codant des protéines régulatrices du métabolisme spécialisé offrent un moyen additionnel afin de reprogrammer les réseaux internes à l'hôte hétérologue (Liu *et al.*, 2021). Dans ce cadre, la perturbation génique du régulateur global AdpA a permis d'activer la biosynthèse d'un polycétide cryptique (oviédomycine) et simultanément d'en réprimer celle du NRP nikkomycine chez la bactérie *S. ansochromogenes* (Xu *et al.*, 2017). Par ailleurs, les cofacteurs des enzymes biosynthétiques ou les précurseurs des produits du BGC peuvent faire défaut au sein de la souche châssis, alors l'inactivation de voies métaboliques clés (*e.g.* glycolyse) contribue à rediriger le flux de ressources, soutenir la production hétérologue et donc en accroître le rendement (*e.g.* polycétides actinorhodine, rapamycine) (Borodina *et al.*, 2008 ; Dang *et al.*, 2017). Alternativement, un hôte comme *Pseudomonas putida* KT2440 fournit une large gamme de substrats indispensables à la formation de peptides non-ribosomiques (*e.g.* serrawettine), polycétides (*e.g.* 2,4-diacétylphloroglucinol) ou métabolites hybrides (*e.g.* prodigiosine) cryptiques, tel que cela a été mis en évidence à maintes reprises (Loeschcke et Thies, 2015). En outre, les outils

d'édition génomique visent actuellement à insérer de multiples sites de transposition (*e.g.* Tn5, Tn7), de recombinaison (*e.g.* LoxP (Cre)) ou d'intégration (*e.g.* *attB* (λ , ϕ C31, Bxb1)) dans le chromosome du châssis afin d'amplifier le BGC ciblé (*i.e.* en augmenter le nombre de copies), et d'en assurer leur expression stable (Liu *et al.*, 2021 ; Martin-Pascual *et al.*, 2021). La technologie CRAGE (*Chassis-independent Recombinase-Assisted Genome Engineering*) a par exemple permis de modifier à haut-débit le génome de 25 espèces de γ -protéobactéries avec des cassettes de transposition (Himar-1) et de recombinaison (Cre-LoxP), et d'y activer des clusters donnés, avant l'analyse des produits NRPS précédemment non-caractérisés (mévalagmapeptide B et ririwpeptide A) (Wang *et al.*, 2019). Enfin, l'introduction de pompes à efflux, comme LmrA ou MdfA, est susceptible de réduire la toxicité des métabolites et d'en améliorer les rendements d'extraction (Li *et al.*, 2021).

II.3. Production d'analogues structuraux bioactifs

Les avancées scientifiques et techniques décrites ouvrent la voie à la découverte de médicaments basés sur des produits naturels pour lutter contre les maladies infectieuses ou cancéreuses. En effet, la recherche médicale travaille à optimiser les classes de métabolites connues au moyen de stratégies de dérivatisation fonctionnelle, héli-synthèse chimique ou ingénierie biosynthétique (Atanasov *et al.*, 2021). En dépit de l'usage dûment approuvé de nombreux médicaments issus de composés bioactifs, plusieurs problèmes récurrents ont amené les entreprises pharmaceutiques à en réduire tout ou partie du programme d'identification/caractérisation au profit de leur modification chimique (Atanasov *et al.*, 2021). En premier lieu, les difficultés relatives à la culture de nouvelles bactéries en condition de laboratoire et à l'expression hétérologue de BGC originaux imposent de développer des méthodes alternatives afin d'assurer la biosynthèse des produits ciblés. De plus, ces métabolites peuvent ne pas présenter une biodisponibilité suffisante lors de l'extraction, ne pas être compatibles avec les essais analytiques réalisés ou ne pas démontrer l'activité potentielle attendue (Atanasov *et al.*, 2021). Par ailleurs, les produits naturels bioactifs possèdent souvent des propriétés d'absorption, de distribution, de métabolisme, d'excrétion et de toxicité (*i.e.* ADMET) sous-optimales (Atanasov *et al.*, 2021). En outre, ces composés ne peuvent à ce jour pas être brevetés, contrairement à leurs dérivés en mesure d'être protégés à travers des droits de propriétés intellectuelles (Atanasov *et al.*, 2021).

Bien que la diversité structurelle et la complexité des structures chimiques des produits naturels puissent être avantageuses afin de générer des banques d'analogues structuraux avec des bioactivités intéressantes, le manque de données au sujet des voies de biosynthèse correspondantes peut s'avérer préjudiciable. Dans ce contexte, l'étude biochimique et structurale des enzymes biosynthétiques clés, en l'occurrence des systèmes NRPS, revêt une importance considérable.

SECTION III. Compréhension des voies de biosynthèse NRPS canoniques ou atypiques

Les NRPS sont des complexes enzymatiques multifonctionnels d'architecture variable, dont le mécanisme séquentiel est à l'origine de la biosynthèse de produits naturels d'une grande diversité moléculaire et d'un large spectre d'activités dans la santé humaine.

Dans ce contexte, cette section rapporte en premier lieu l'élucidation du mécanisme catalytique historique *thiotemplate* à travers les expérimentations préliminaires (section III.1.A.) et les différents modes de biosynthèse (section III.1.B.). Par la suite, elle expose l'organisation et la composition de NRPS de types I, II et III, y compris les méga-synthèses hybrides, responsables de la biogenèse de métabolites d'intérêt majeur, avec des exemples représentatifs dans chaque cas (section III.2.).

III.1. Éluclidation du mécanisme catalytique de synthèse non-ribosomique

III.1.A. Expérimentations préliminaires

La recherche scientifique avait à peine décrypté le mécanisme de biosynthèse protéique dans les années 1950–1960 que, parallèlement, des données ont indiqué que la production de peptides ayant une structure différente faisait intervenir un mécanisme enzymatique complètement distinct (Mach *et al.*, 1963). Dans ce cadre, les systèmes NRPS ont été initialement étudiés, avec l'exemple de la biogenèse de la gramicidine J, la tyrocidine ou la subtiline chez des espèces de *Bacillus* (Mach *et al.*, 1963). En effet, ces travaux ont démontré que la machinerie ribosomale était inhibée par certains antibiotiques, tels que le chloramphénicol, mais que cela n'affectait pas la production simultanée du décapeptide tyrocidine A chez la bactérie *B. brevis*, ce qui soutenait leurs observations préliminaires (Mach *et al.*, 1963). À cet égard, plusieurs hypothèses avaient été formulées : les polypeptides de ce groupe pourraient être biosynthétisés le long d'un modèle linéaire, comme le sont les protéines, sans intervention d'ARN et avec l'addition enzymatique séquentielle d'acides aminés précurseurs, voire même générés à partir de la dégradation de composants cellulaires (Mach *et al.*, 1963). De plus, des travaux réalisés sur un NRP putatif, la polymyxine B, ont prouvé que la production de ce composé n'était pas annihilée par l'utilisation d'inhibiteurs du ribosome ou de RNAses (Daniels, 1968). Les premières études ayant été publiées quant à ces oligopeptides d'origine inconnue ont par conséquent révélé l'existence d'un mécanisme de biosynthèse indépendant des ribosomes.

Les travaux réalisés par l'équipe de Lipmann ont en outre démontré que les enzymes impliquées dans la biosynthèse de gramicidine S chez la bactérie *B. brevis* assuraient la transcarboxylation des acides aminés, y compris la L-Orn non-protéino-gène, en une unique réaction dépendante de l'ATP (Gevers *et al.*, 1968). Cela conduisait à leur activation spécifique et suggérait alors l'existence d'un transporteur au sein de la NRPS (Gevers *et al.*, 1968). Des expériences complémentaires opérées en présence d'une forme radiomarquée à l'isotope ^{14}C du premier acide aminé utilisé au sein de NRPS modèles ont mis en évidence l'élongation directionnelle du peptide lié à l'*holo*-enzyme (Roskoski *et*

al., 1970). Cela attestait que les peptides étaient biosynthétisés par l'incorporation progressive des résidus constitutifs de la chaîne en cours d'extension et que celle-ci restait attachée à la matrice.

L'analyse de la stabilité chimique de la liaison covalente définie entre l'enzyme et l'acide aminé activé supposait que celui-ci était thioestérifié (Felnagle *et al.*, 2008). Les hypothèses développées sur la nature du site de liaison concernaient un résidu cystéine et/ou un cofacteur thiolé (Felnagle *et al.*, 2008). Elles ont été appuyées par les travaux réalisés sur la biosynthèse des acides gras qui ont établi une corrélation entre l'utilisation d'intermédiaires thioestérifiés et la présence du cofacteur 4'-phosphopantéthéinyle (4'-Ppant) (Felnagle *et al.*, 2008). Les équipes de Laland et Lipmann ont par la suite conjointement prouvé qu'un unique groupement 4'-Ppant était utilisé par chaque module de sous-unités catalytiques et qu'environ 70 kDa de protéines étaient associés par acide aminé activé (Gilhuus-Moe *et al.*, 1970 ; Kleinkauf *et al.*, 1970 ; Lee *et al.*, 1973). Cela indiquait qu'un unique monomère était incorporé dans le peptide pour chaque module de sous-unités catalytiques intégré dans la chaîne d'assemblage. De plus, ces expériences ont démontré que la fonction thioester liée au cofacteur 4'-Ppant réalisait une transpeptidation de l'oligopeptide en cours de synthèse, à partir du précurseur pré-activé, et que l'addition des autres acides aminés se poursuivait dans un ordre prescrit jusqu'à la terminaison du processus (Lipmann, 1973).

Cette modularité combinée aux réactions d'échanges d'intermédiaires thioesterifiés a donné lieu à la proposition du mécanisme catalytique historique, dit *thiotemplate*, ayant été supplanté par un mécanisme alternatif, dit *multiple carrier*, représentatif des voies de biosynthèse canoniques des NRPS (Stein *et al.*, 1996). En effet, les efforts réalisés dans le séquençage des BGC de NRPS chez le genre *Bacillus* a rendu compte que ces méga-complexes contenaient généralement des séquences protéiques répétées qui correspondaient exactement au nombre d'acides aminés activés au cours de la biosynthèse (Felnagle *et al.*, 2008). L'analyse des séquences associées a de plus fait remarquer que chaque module était responsable d'une multitude de modifications structurales des antibiotiques (Schnell *et al.*, 1988 ; Stein *et al.*, 1996).

Fait intéressant, les recherches ultérieures réalisées sur ces composés à haute valeur ajoutée se sont principalement concentrées sur les voies de biosynthèse des NRP qui exploitent la modularité des systèmes, où chaque module active et couple donc un acide aminé à une chaîne peptidique en cours d'extension. Ce principe de colinéarité régit les différents types de biosynthèse de peptides non-ribosomiques *via* un mécanisme commun.

III.1.B. Présentation des types de biosynthèse et du mécanisme catalytique canonique

Les chaînes d'assemblage adoptent majoritairement une architecture linéaire et multimodulaire constituée de groupes répétés de différents domaines canoniques, qui incluent les domaines porteurs de peptidyle (PCP/T), les domaines d'adénylation (A) et de condensation (C), ainsi que de domaines supplémentaires, qui modifient la stéréochimie des produits naturels (Payne *et al.*, 2017) (**Figure 8**).

Tout d'abord, le domaine *apo*-PCP inactif nécessite la fixation d'un cofacteur 4'-Ppant sur un résidu sérine conservé par une phosphopantéthéinyltransférase (PPTase), ce qui donne la forme *holo*-PCP active (Beld *et al.*, 2014) (**Figure 9A**). Ensuite, le domaine A lie le monomère activé sur le domaine *holo*-PCP par des réactions d'adénylation et de thiolation, ce qui forme l'aminoacyl-*S*-PCP (Corpuz *et al.*, 2022) (**Figure 9B**). De manière alternative, des NRPS monomodulaires autonomes, fréquemment exprimées sous forme de didomains A-PCP, se chargent d'un précurseur (Süssmuth et Mainz, 2017). Celui-ci est possiblement modifié par des enzymes accessoires à activités *trans*-catalytiques et subséquentment transféré vers la chaîne principale (*e.g.* échinomycine, syringomycine) ou libéré en tant que produit mature (*e.g.* nikkomycine) (Süssmuth et Mainz, 2017) (**Figure 8**).

Dans tous les cas, les complexes aminoacyl-/peptidyl-PCP chargés se lient aux sites donneurs ou accepteurs du domaine C, lequel catalyse alors la formation d'une liaison amide entre les substrats thioestérifiés et permet ainsi l'extension de la chaîne (Payne *et al.*, 2017) (**Figure 9C**). Par ailleurs, les domaines additionnels incorporés au sein des modules NRPS peuvent assurer diverses réactions de formylation (F), de méthylation (MT), de β -hydroxylation, d'oxydation (OX) ou d'épimérisation (E) (voir la section **IV.7.**) (Corpuz *et al.*, 2022). Le complexe peptidyl-PCP transfère finalement la chaîne peptidique ainsi élongée à un domaine thioestérase (TE) qui hydrolyse ou cyclise le produit, ce qui en assure la libération (**Figure 9D**). En outre, les NRP biosynthétisés sont aussi sujets à être modifiés par des enzymes discrètes qui agissent en *trans* au cours de leur formation et/ou suite à leur libération de la NRPS modulaire (Süssmuth et Mainz, 2017) (**Figure 8**).

L'organisation modulaire de certaines NRPS permet donc de déduire la structure chimique du produit de synthèse par l'analyse fonctionnelle de la linéarité des domaines et de la sélectivité des domaines A (Payne *et al.*, 2017). Les modifications fonctionnelles des NRP synthétisés démontrent la variété des enzymes disponibles pour assurer une interaction transitoire avec les domaines PCP et une transformation naturelle des produits au sein des voies métaboliques (Corpuz *et al.*, 2022).

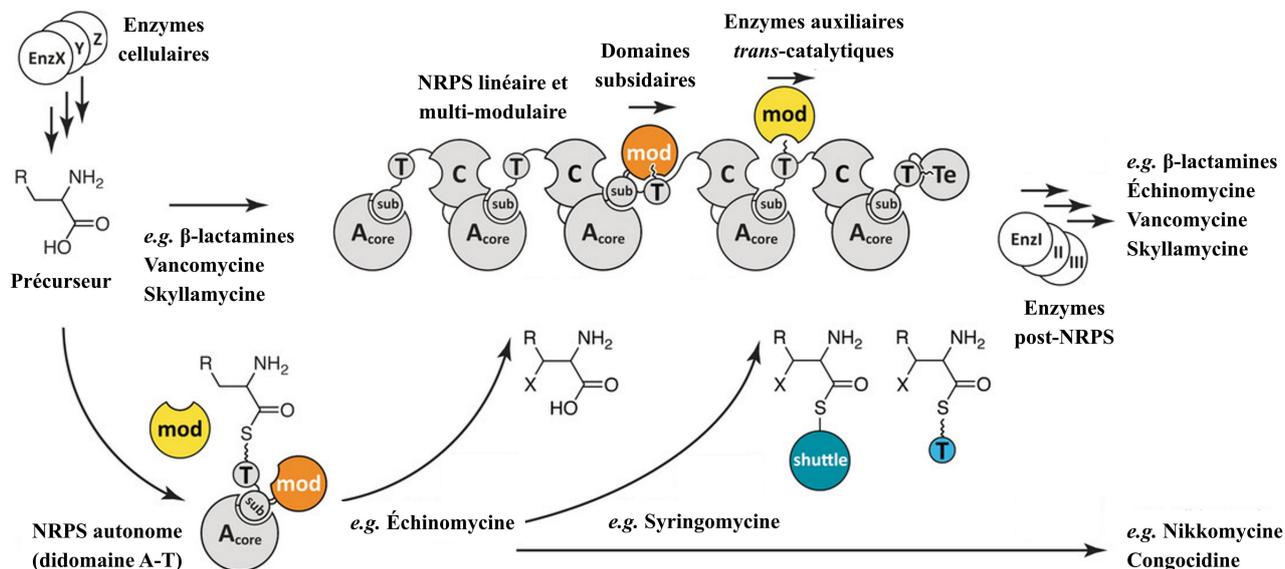


Figure 8. Organisation de la biosynthèse des NRP (adaptée de Süssmuth et Mainz, 2017). Les substrats sont pris en charge par des NRPS autonomes ou modulaires, et modifiés par divers domaines subsidiaires et/ou enzymes auxiliaires en des NRP. A_{core}/A_{sub} : sous-domaines du domaine d'adénylation, C : domaine de condensation, T : domaine porteur de peptidyle, TE : domaine thioestérase. mod : domaine ou enzyme de modification.

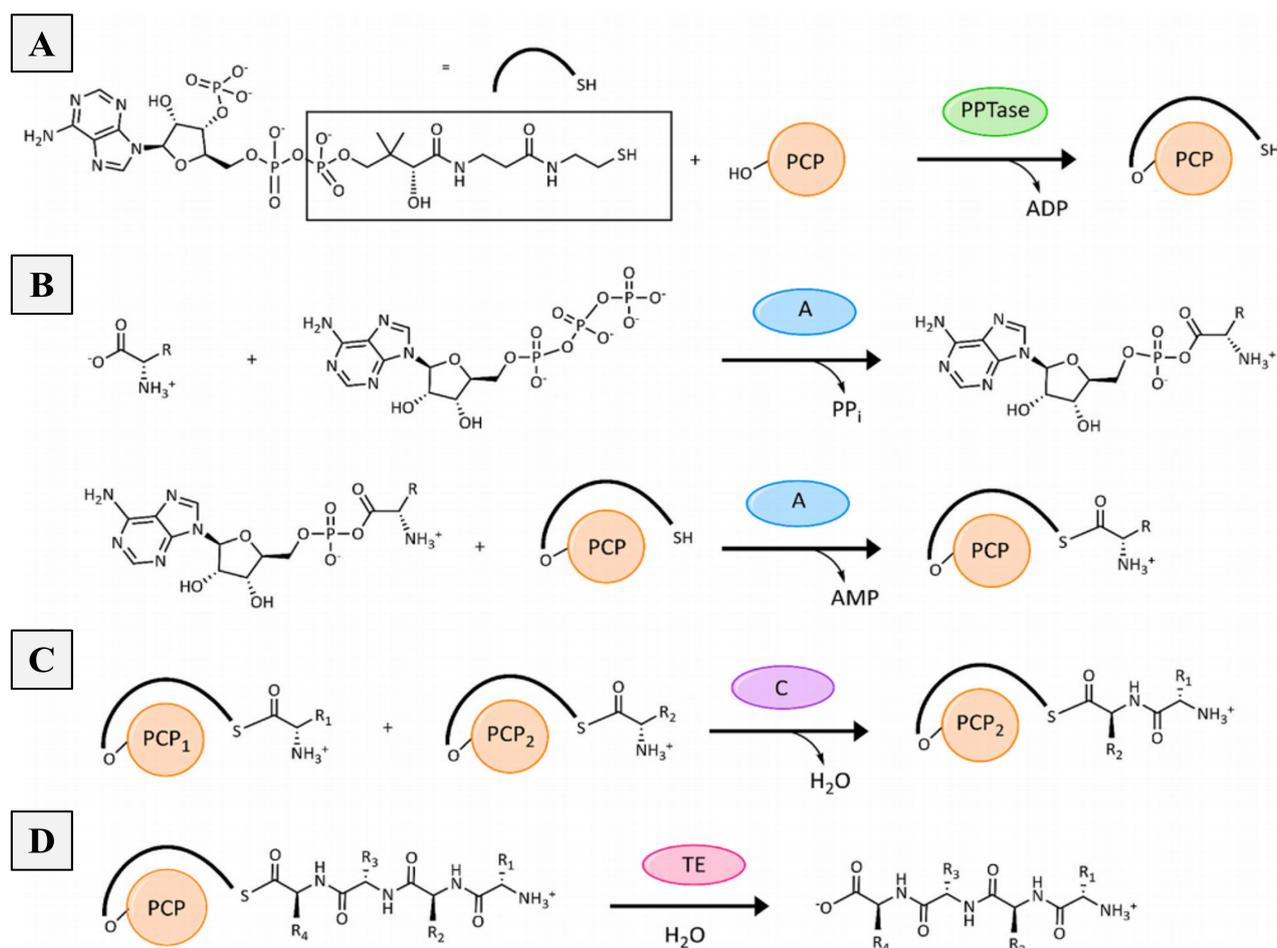


Figure 9. Réactions canoniques de synthèse de peptides non-ribosomiques (adaptée de Corpuz *et al.*, 2022). Les réactions de phosphopantéthénylation (A), d'adénylation et thiolation (B), de condensation (C) et de thioestérification (D) sont catalysées dans cet ordre par une PPTase, et par les domaines minimaux A, C et TE, respectivement. PPTase : phosphopantéthényltransférase, A : domaine d'adénylation, C : domaine de condensation, TE : domaine thioestérase.

III.2. Organisation et composition des NRPS de type I, II et III

Un module contient au minimum les domaines *cis*-canoniques A-PCP-C dans cet ordre et assure l'activation et la thioestérification d'un précurseur, ainsi que la formation de liaisons amides entre les monomères au cours de la biosynthèse des NRP fonctionnels (Felnagle *et al.*, 2008). Par ailleurs, comme évoqué préalablement, des domaines accessoires peuvent intégrer certains modules NRPS et alors modifier la chaîne peptidique en cours d'extension. Les NRPS de type I utilisent une unique fois (type I modulaire) ou plusieurs fois (type I itératif) chacun des modules, tandis que les systèmes NRPS de type II font de plus intervenir des unités catalytiques autonomes (modules et/ou domaines) (Felnagle *et al.*, 2008) (**Figure 10AB**). Les NRPS de type III non-linéaires exploitent des domaines à de multiples reprises lors de la biosynthèse (Felnagle *et al.*, 2008) (**Figure 10C**).

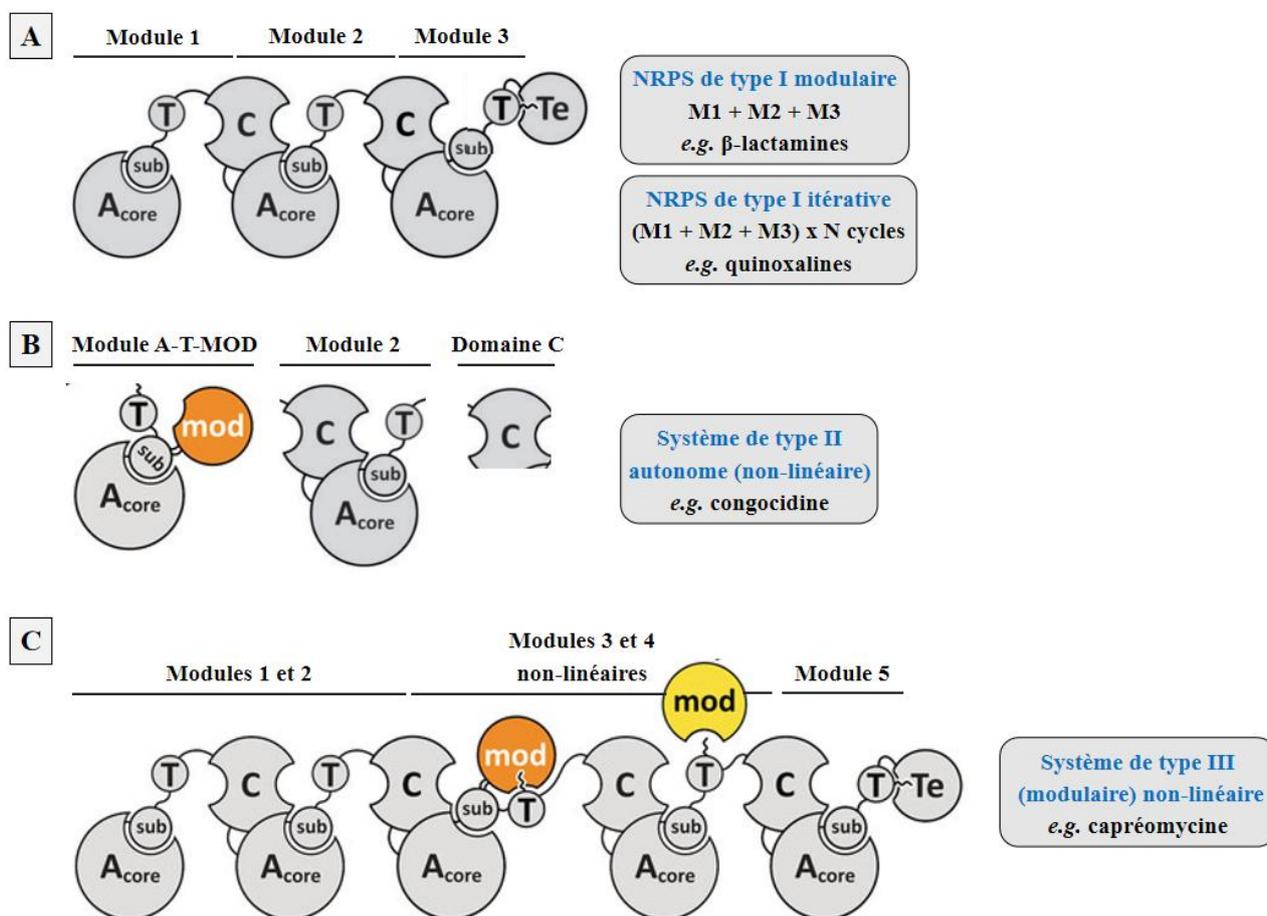


Figure 10. Constitution des NRPS de type I, II et III. (A) Les lignes d'assemblage NRPS présentent majoritairement une architecture linéaire et multimodulaire, et exploitent une ou plusieurs fois (type I modulaire ou itératif) chacun des modules. (B) Le systèmes de type II contiennent uniquement des sous-unités catalytiques autonomes, bien qu'une telle disposition soit assez peu commune. Les systèmes hybrides qui impliquent à la fois une NRPS multimodulaire et des didomains autonomes sont plus fréquents. (C) Les systèmes de type III non-linéaires se caractérisent également par une organisation modulaire, mais les sous-unités catalytiques diffèrent entre elles au niveau de la composition en domaines. A : domaine d'adénylation, C : domaine de condensation, T : domaine porteur de peptidyle, mod : domaine ou enzyme de modification, TE : domaine thioestérase.

L'architecture canonique organisée en modules constitués des domaines essentiels A-PCP-C est naturellement exploitée pour générer une diversité structurale considérable de NRP. Celle-ci repose donc sur le nombre de modules utilisés, les monomères activés par les domaines A et la présence de domaines catalytiques supplémentaires utilisés. De plus, les peptides biosynthétisés peuvent différer s'ils sont cyclisés et, dans ce cas, les résidus engagés dans cette réaction peuvent varier, comme cela sera expliqué à la section **IV.6**. Par ailleurs, ces produits sont susceptibles d'être à nouveau sujets à des modifications fonctionnelles réalisées par diverses enzymes auxiliaires, qualifiées de post-NRPS, telles que des hydrolases, des oxydoréductases ou des glycosyltransférases. Cela accroît la diversité des structures chimiques et donc la bioactivité des NRP matures (Süssmuth et Mainz, 2017).

III.2.A. NRPS de type I modulaires : exemple de la biosynthèse des β -lactamines

Les NRPS de type I correspondent aux synthèses linéaires et multimodulaires historiquement associées à la biosynthèse des NRP avec des modules distribués sur plusieurs chaînes protéiques et impliqués dans l'initiation, l'extension ou la terminaison du processus (Felnagle *et al.*, 2008). Par exemple, les systèmes SevAB de *Paenibacillus* ou PhsABC de *Streptomyces* sont répartis sur deux (A, B) ou trois sous-unités protéiques (A, B, et C) et impliqués dans la biosynthèse des tripeptides linéaires sevacidine ou bialaphos, respectivement (Süssmuth et Mainz, 2017). De même, les NRPS CepABC, TycABC ou FegQRS sont composées de 7, 10 ou 13 modules, eux-mêmes ordonnés sur trois chaînes protéiques, et impliqués dans la biosynthèse des peptides vancomycine, tyrocidine A ou feglymycine (Süssmuth et Mainz, 2017). Dans les systèmes les plus complexes caractérisés, une unique NRPS Kol géante d'une masse moléculaire de 1,8 MDa est impliquée dans la biogenèse du pentadécapeptide kolossine A (Bode *et al.*, 2015). Autre exemple exceptionnel, le système SypABC tripartite est associé au plus grand NRP connu à ce jour avec 25 résidus (syringopeptine) chez des γ -protéobactéries (Scholz-Schroeder *et al.*, 2003). D'un point de vue évolutif, les lignes d'assemblage extrêmement étendues apparaissent rares à cause d'une vraisemblable moindre efficacité catalytique au cours de l'élongation du peptide (Bode *et al.*, 2015 ; Süssmuth et Mainz, 2017).

L'exemple typique d'un système de type I correspond à la NRPS trimodulaire impliquée dans la biogenèse d'isopénicilline (Townsend, 2016). Les β -lactamines sont une classe d'antibiotiques très diversifiés, notamment avec les pénicillines et les céphalosporines, produits biosynthétisés par des genres de champignons (*Penicillium*, *Cephalosporium*, etc.) et bactéries (*Streptomyces*) (Demain et Elander, 1999). Indépendamment de l'organisme producteur, l'ACV synthétase catalyse d'abord la formation du tripeptide L- α -aminoadipate-L-Cys-D-Val (**Figure 11**). En effet, le module d'initiation et le second module, organisés en domaines A-PCP-C, activent et thioestérifient les acides aminés L- α -aminoadipate et L-Cys, puis le domaine C2 catalyse la formation de la liaison amide entre ces monomères greffés au domaine PCP2 (Townsend, 2016). Le module terminal active le résidu L-Val, catalyse la formation du produit au niveau du domaine PCP3, épimérise le résidu en la stéréochimie D-, et hydrolyse enfin le tripeptide linéaire de la NRPS (Felnagle *et al.*, 2008) (**Figure 11**).

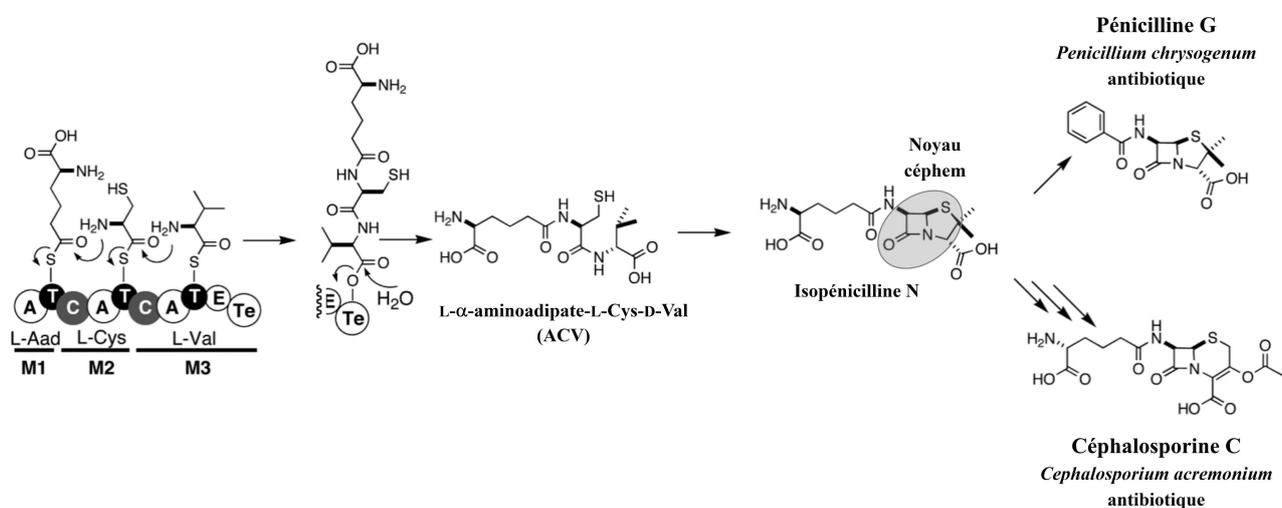


Figure 11. Représentation de la voie de biosynthèse des β -lactamines (adaptée de Felnagle *et al.*, 2008). La NRPS trimodulaire assemble le tripeptide ACV, puis l'isopénicilline synthase (IPNS) catalyse la formation du noyau β -lactame, avant que des enzymes indépendantes ne génèrent la pénicilline G ou la céphalosporine C. A : domaine d'adénylation, C : domaine de condensation, T : domaine porteur de peptidyle, TE : domaine thioestérase, L-Aad : L- α -aminoadipate.

Dès la libération de ce précurseur commun, l'oxygénase IPNS dépendante du fer non-héminique assure la conversion en isopénicilline N (Townsend, 2016). En effet, cette enzyme catalyse la cyclisation du noyau β -lactame et celui-ci fusionne *via* l'atome de soufre avec le cycle thiazolidine (Townsend, 2016). Bien que les oxygénases apparaissent dans un certain nombre de réactions lors de la biosynthèse des antibiotiques β -lactamines, cette voie spécifique est la seule connue où l'une assure la formation de l'anneau bicyclique correspondant (Townsend, 2016). De plus, la maturation du composé requiert diverses enzymes post-NRPS, en particulier, une acétyltransférase substitue le résidu L- α -aminoacétate par des fonctions phényl(oxy)acétate et génère ainsi les pénicillines G et V, respectivement (Demain et Elander, 1999). Dans la voie alternative, une épimérase convertit le résidu L- α -Aad en la configuration D- et le noyau thiazolidine en dihydrothiazine, puis une hydroxylase et une acétyltransférase produisent alors la céphalosporine C (Townsend, 2016) (**Figure 11**). Dans tous les cas, le mécanisme d'action de ce sous-groupe de composés bioactifs consiste en l'inactivation de la réaction de transpeptidation au cours de la biosynthèse du peptidoglycane de la paroi bactérienne (Demain et Elander, 1999). À ce sujet, la modification rationnelle de la structure chimique de NRP a permis d'en améliorer l'usage contre les micro-organismes pathogènes (Walsh, 2003).

III.2.B. NRPS de type I itératives : exemple de la biosynthèse des quinoxalines

Des NRPS de type I se différencient par un fonctionnement itératif de l'ensemble des modules au cours d'un cycle biosynthétique complet. Ces systèmes nécessitent de ce fait un positionnement d'attente des intermédiaires réactionnels sur les domaines PCP ou TE du module terminal (Süssmuth et Mainz, 2017). Le produit monomérique fixé à sa position de stockage au terme du premier cycle d'élongation subit de 2 à 5 autres cycles jusqu'à une longueur critique, ce qui provoque la libération ou la macrocyclisation de ce peptide avec une symétrie moléculaire (Süssmuth et Mainz, 2017). Des produits naturels de ce groupe incluent les chromophores quinoxalines, cyclodepsipeptides produits par l'ordre des Actinomycétales, qui présentent une activité anticancéreuse approuvée (échinomycine) et un supposé potentiel anti-VIH (luzopeptines) (Felnagle *et al.*, 2008).

Les recherches associées à l'échinomycine, par exemple, ont élucidé la spécificité de substrat et le mécanisme catalytique de la NRPS à l'origine de sa biosynthèse (Watanabe *et al.*, 2006) (**Figure 12**). En effet, le domaine A discret Ecm1 active tout d'abord un dérivé du monomère L-tryptophane, l'acide quinoxaline-2-carboxylique, sous forme d'un intermédiaire acyl-AMP, et le lie au domaine porteur autonome FabC (Watanabe *et al.*, 2006). Le module initial de la sous-unité Ecm6 active et thioestérifie l'acide aminé L-Ser, puis le condense avec le chromophore et épimérise le résidu sérine, tandis que le module suivant incorpore le résidu L-Ala (Watanabe *et al.*, 2006). Les modules de la NRPS Ecm7 ajoutent ensuite au tripeptide produit les acides aminés L-Cys et L-Val. Des domaines méthyltransférases catalysent la *N*-méthylation de ces résidus (Watanabe *et al.*, 2006) (**Figure 12**).

Le domaine TE maintient le peptide linéaire sur la matrice, le domaine porteur étant ainsi rendu disponible pour un autre cycle, et catalyse la cyclisation intramoléculaire entre la fonction hydroxyle du résidu D-Ser du premier peptide et le groupement carbonyle du résidu L-Val du second, avant la libération du dimère produit (Watanabe *et al.*, 2006) (**Figure 12**). Ce cyclopeptide devient enfin le substrat de l'oxydoréductase Ecm17, laquelle génère un pont disulfure entre des groupements thiols, et de la méthyltransférase Ecm18, responsable de la formation d'un pont thioacétal (Watanabe *et al.*, 2006) (**Figure 12**). Au niveau du mécanisme d'action, cette classe de métabolites s'insère dans des séquences d'ADN spécifiques par l'intermédiaire de leur structure plane, ce qui inhibe la progression des polymérases lors de la réplication (Felnagle *et al.*, 2008).

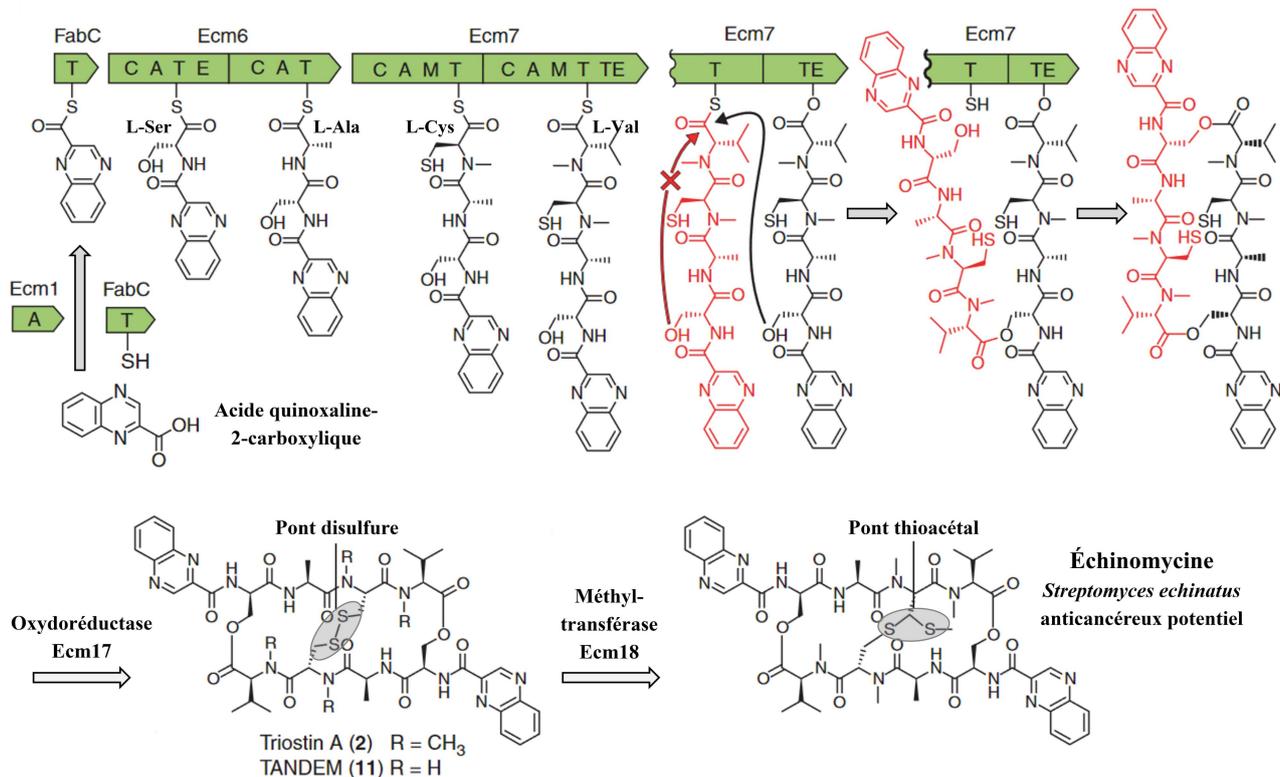


Figure 12. Représentation de la voie de biosynthèse de l'échinomycine (adaptée de Watanabe *et al.*, 2006). Le système des NRPS itératives Ecm6-Ecm7 synthétise deux copies d'un peptide linéaire, puis le domaine TE les assemble en cyclodepsipeptide. Le produit dimérique libéré est le substrat de l'oxydoréductase Ecm17 et de la méthyltransférase Ecm18 impliquées dans la formation du pont thioacétal de l'échinomycine. A : domaine d'adénylation, C : domaine de condensation, M : domaine de méthylation, T : domaine porteur de peptidyle, TE : domaine thioestérase.

Une des raisons plausibles à même d'expliquer l'utilisation de mécanismes itératifs réside dans la nécessité de diminuer le coût énergétique de la synthèse de systèmes NRPS de taille conséquente et d'éviter les problèmes de repliement résultants (Süssmuth et Mainz, 2017). Par exemple, la NRPS PF1022 (350 kDa) produit un octadepsipeptide, dont la biosynthèse linéaire devrait théoriquement exiger une unique chaîne protéique d'une masse moléculaire de 1 MDa ou des sous-unités distinctes avec des éléments de reconnaissance intermodulaires spécifiques (Süssmuth et Mainz, 2017). Bien que les NRPS étudiées suivent principalement la logique modulaire, environ 10–15 % des systèmes font néanmoins intervenir des modules ou domaines autonomes (Wang *et al.*, 2014).

III.2.C. NRPS de type II autonomes : exemple de la biosynthèse de la congocidine

Les systèmes NRPS de type II se caractérisent par l'absence d'une architecture multimodulaire spécifique et présentent des sous-unités catalytiques autonomes (Dell *et al.*, 2022). Cependant, dans ces voies biosynthétiques, les systèmes hybrides avec à la fois une NRPS multimodulaire (type I) et des enzymes autonomes (type II), tels que des didomains (A-PCP, C-PCP) ou domaines classiques (C, A, PCP), sont plus fréquemment observés (Fortin *et al.*, 2007). De nombreux exemples ont été rapportés, parmi lesquels la biosynthèse de la congocidine (ou nétropsine) à activité antivirale chez la bactérie *S. netropsis* (Al-Mestarihi *et al.*, 2015).

Au sein de cette voie métabolique, le domaine A indépendant Cgc3 active et lie le monomère acide acétamidopyrrole-carboxylique au domaine PCP discret Cgc19, puis l'amidohydrolase Cgc14 désacétyle ce précurseur en aminopyrrole (Al-Mestarihi *et al.*, 2015) (**Figure 13**). Parallèlement, le domaine A de la NRPS monomodulaire Cgc18 active et lie l'acide guanidino-acétique au domaine PCP adjacent. Ensuite, le domaine C de cette sous-unité catalyse la condensation de l'aminopyrrole et de l'acide guanidinoacétique liés à leurs domaines porteurs, puis un domaine C autonome (Cgc16) catalyse à son tour l'extension du NRP avec un nouveau monomère d'aminopyrrole (Al-Mestarihi *et al.*, 2015) (**Figure 13**). Un autre domaine autonome (Cgc2) catalyse la condensation d'une molécule exogène de 3-aminopropionamidine avec le groupement carbonyle de l'intermédiaire thioestérifié à la matrice (Cgc19). L'enzyme Cgc15 greffe enfin des groupements *N*-méthyles au niveau des cycles pyrroles du métabolite, ce qui forme la congocidine mature (Al-Mestarihi *et al.*, 2015) (**Figure 13**).

Dans tous les cas, les domaines et multidomains de ces systèmes hybrides émergent comme de précieux points de départ au sein des voies artificielles qui viseront à incorporer divers précurseurs inhabituels à des étapes définies de la biosynthèse. En outre, de nombreuses chaînes d'assemblage NRPS de type I/II s'écartent de la règle de colinéarité.

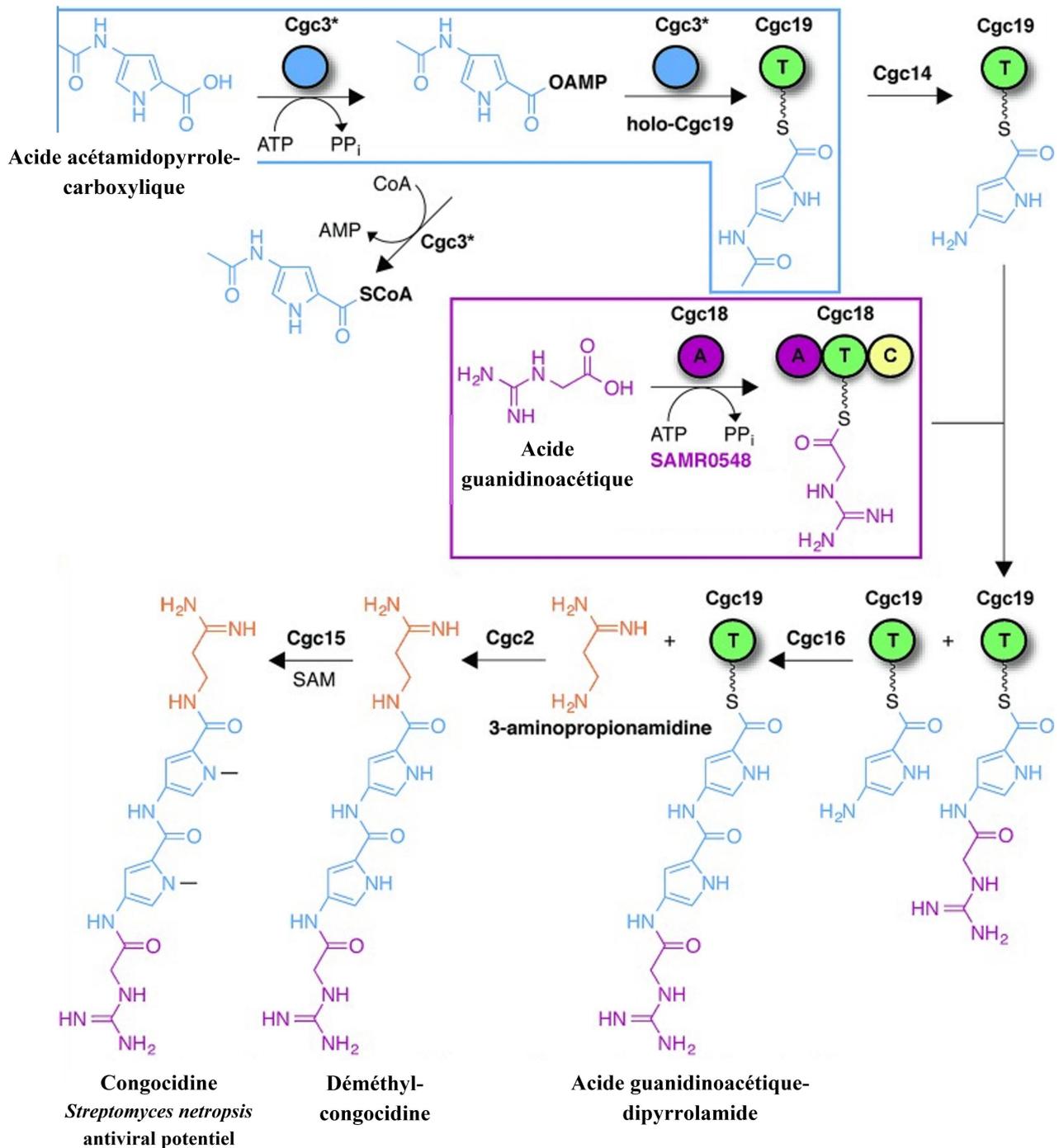


Figure 13. Représentation de la voie de biosynthèse de la congocidine (adaptée de Al-Mestarihi *et al.*, 2015). Le système NRPS de type II impliqué dans la biogenèse de congocidine requiert des domaines A (Cgc3) et PCP autonomes (Cgc19), une NRPS monomodulaire Cgc18 d'architecture A-PCP-C, une aminohydrolase (Cgc14), deux domaines C discrets (Cgc16 et Cgc2), ainsi qu'une *N*-méthyltransférase indépendante. A : domaine d'adénylation, C : domaine de condensation, T : domaine porteur de peptidyle, AMP/ATP : adénosine mono/triphosphate, CoA : coenzyme A, PPi : pyrophosphate, SAM : *S*-adénosylméthionine.

III.2.D. NRPS de type III non-linéaires

III.2.D.α. NRPS avec un domaine A absent : cas de la biosynthèse de la capréomycine

Le système NRPS de type III impliqué dans la biosynthèse de l'antibiotique capréomycine chez la bactérie *S. capreolus* fait office d'exemple concret de la diversité structurale des NRP issus d'une unique voie (Johansen *et al.*, 2006). Cette ligne d'assemblage d'enzymes indépendantes utilise trois acides aminés non-protéinogènes (L-diaminopropionate, β -uréidodéhydroalanine et capréomycidine) et exploite les propriétés particulières de certains domaines (Felnagle *et al.*, 2007) (**Figure 14**). En effet, le module 4 CmnI ne présente pas de domaine A, ce qui suppose alors qu'un autre domaine, probablement celui du module CmnF d'initiation, charge le domaine PCP de la sous-unité incomplète (Felnagle *et al.*, 2007). De plus, le module 2 CmnA comporte un domaine qui assure la conversion entre les résidus L-Ser et L-Ala à cette position du noyau pentapeptidique (Felnagle *et al.*, 2007). Par ailleurs, l'enzyme auxiliaire CmnJ catalyse la déshydratation de l'un des monomères de L-Dap (Felnagle *et al.*, 2007). En outre, la sous-unité CmnG comporte un domaine C modifié, à la place du domaine TE, responsable de la cyclisation et de la libération du NRP (Felnagle *et al.*, 2007). Enfin, l'enzyme post-NRPS CmnL catalyse la carbamylation du résidu L-Ala en β -uréidodéhydroalanine, avant que le didomaine A-PCP CmnO et le domaine C autonome CmnM n'incorporent un acide β -aminé (lysine) sur la fonction amine secondaire du troisième résidu (L-Dap) (Felnagle *et al.*, 2007). Les composés capréomycine IA/IB produits ou, en l'absence de cette modification, les analogues IIA/IIB correspondants, ciblent les ARNr 16S et 23S du pathogène *M. tuberculosis*, ce qui inhibe la progression des ribosomes lors de la traduction (Johansen *et al.*, 2006 ; Felnagle *et al.*, 2007). Ces composés bioactifs s'utilisent en association avec d'autres antibiotiques à effets synergiques afin de traiter la tuberculose induite par l'agent infectieux (Johansen *et al.*, 2006).

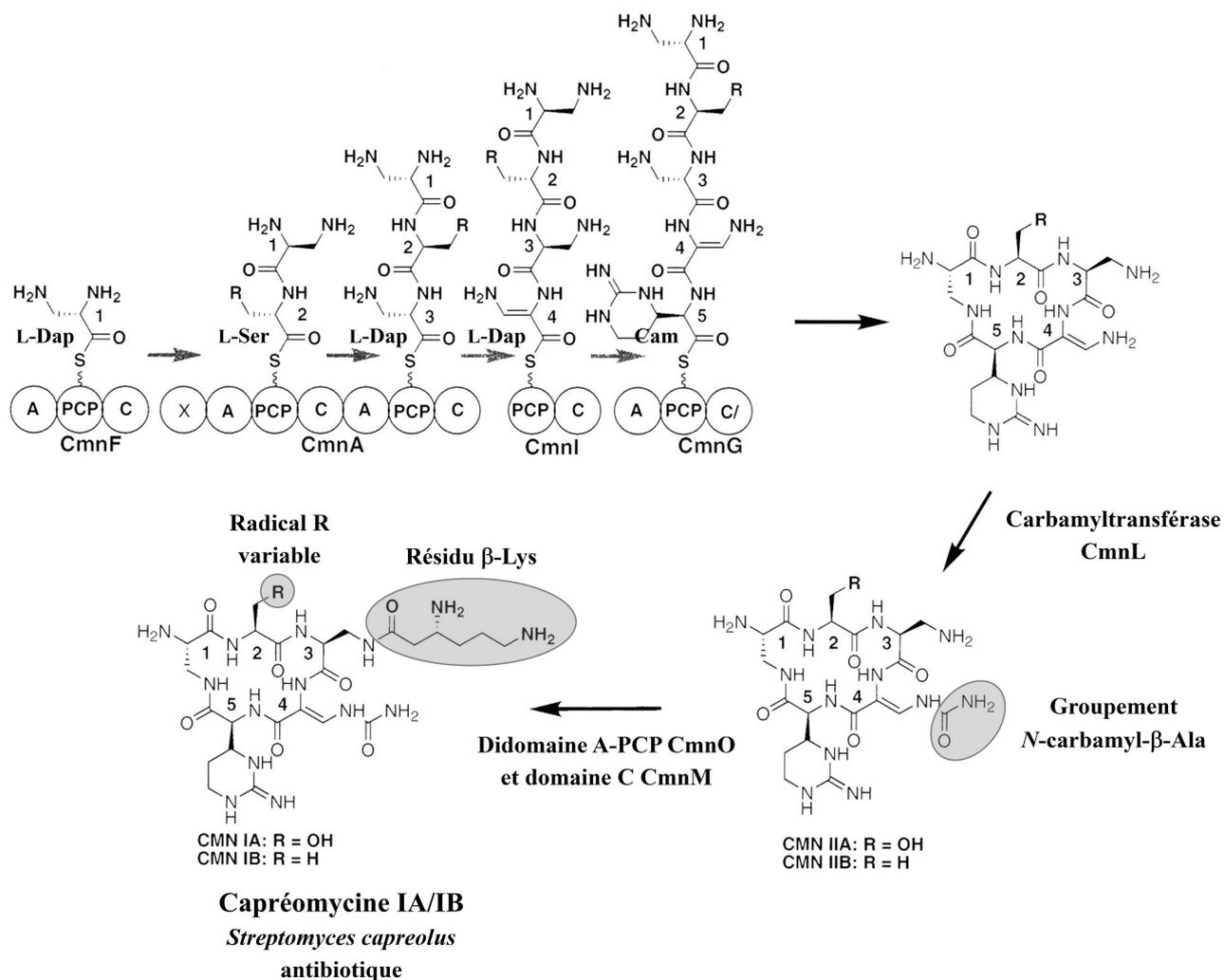


Figure 14. Représentation de la voie de biosynthèse de la capréomycine (adaptée de Felnagle *et al.*, 2007). Le système NRPS de type III impliqué dans la biogenèse de la capréomycine intègre notamment des domaines particuliers. En effet, un domaine déshydratase putatif assure la conversion entre les résidus L-Ser et L-Ala au sein de la sous-unité CmnA, et un domaine C terminal (noté C/) permet la cyclisation du pentapeptide primaire. De plus, le module 4 CmnI ne contenant pas de domaine A, le domaine A1 du module initial CmnF charge vraisemblablement le monomère L-Dap sur le domaine PCP4. Par ailleurs, l'ornithine carbamoyltransférase CmnL assure la conversion du résidu L-Ala en β -uréidodéhydroalanine. En outre, le didomaine A-PCP CmnO et le domaine C CmnM associés ajoutent l'acide β -Lys au niveau du groupement amine du troisième résidu (L-Dap) des capréomycines IA et IB. A : domaine d'adénylation, C : domaine de condensation, PCP : domaine porteur de peptidyle, Cam : capréomycidine, L-Dap : L-2,3-diaminopropionate.

III.2.D.β. NRPS avec un domaine C atypique : exemple de la biosynthèse de l'andrimide

Divers systèmes NRPS de type III ne font pas usage des domaines C classiques pour former des liaisons amides, mais utilisent des protéases ou d'autres enzymes dépendantes de l'ATP (Goswami et Van Lanen, 2015). En effet, la transglutaminase AdmF catalyse la formation des liaisons amides du composé andrimide à activité antibiotique, produit chez la bactérie *Vibrio coralliilyticus* (Fortin *et al.*, 2007) (**Figure 15**). De même, des enzymes de trois familles différentes, une transglutaminase, une carboxylate-amine ligase dépendante de l'ATP et une aminoglycoside acyltransférase assurent l'assemblage du squelette du NRP closthioamide aux propriétés antibiotiques, biosynthétisé chez la bactérie *Ruminiclostridium cellulolyticum* (Goswami et Van Lanen, 2015).

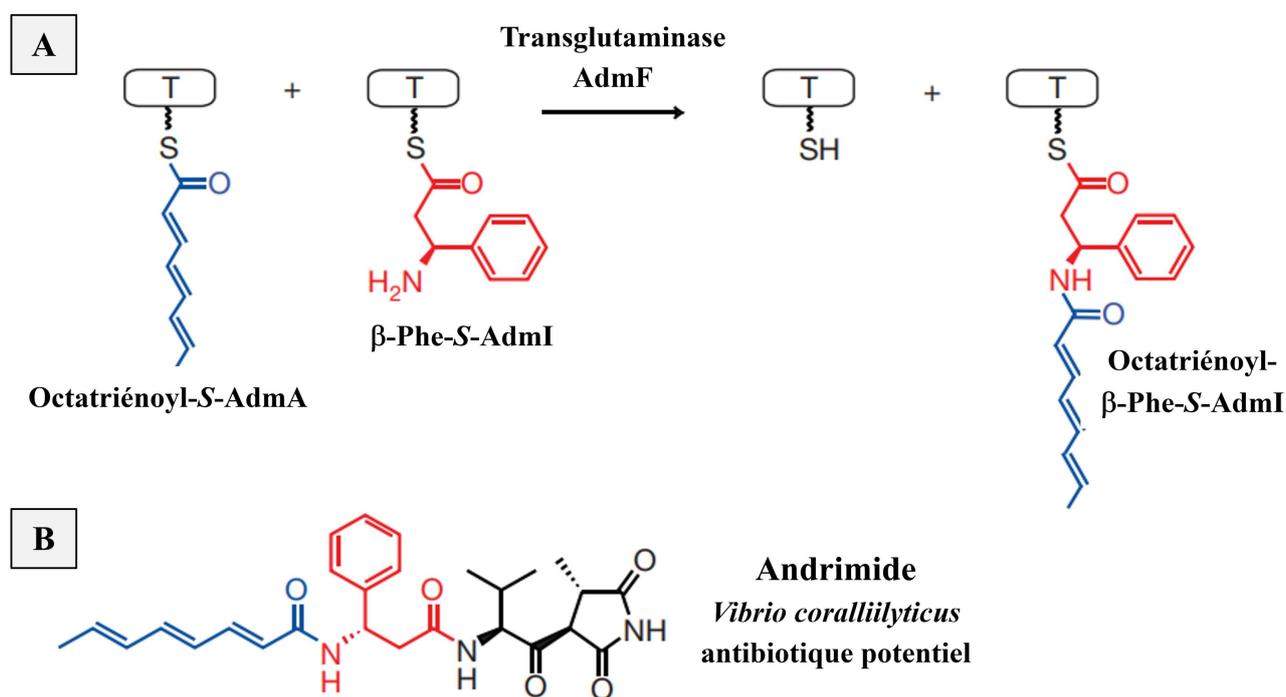


Figure 15. Représentation de la voie de biosynthèse de l'andrimide (adaptée de Fortin *et al.*, 2007). (A) La NRPS de type III impliquée dans la biosynthèse d'andrimide recrute la transglutaminase AdmF, laquelle catalyse la formation d'une liaison amide entre les monomères octatriénoyl et β -Phe thioestérifiés aux domaines porteurs (T) AdmA et AdmI, respectivement (B) La structure chimique de l'antibiotique potentiel andrimide est représentée.

III.2.E. NRPS-PKS hybride : exemple de la biosynthèse de la bléomycine

Des initiatives d'exploration génomique ont rapporté la présence de plus de 3300 BGC de mégasynthèses au sein d'un millier de micro-organismes étudiés, dont 34 % (1147) de clusters de gènes hybrides NRPS-PKS (Wang *et al.*, 2014). Les machineries correspondantes sont en mesure d'assurer en commun la biosynthèse de structures d'une importante complexité chimique et d'une bioactivité capitale, dont l'anticancéreux bléomycine A2 (Figure 16).

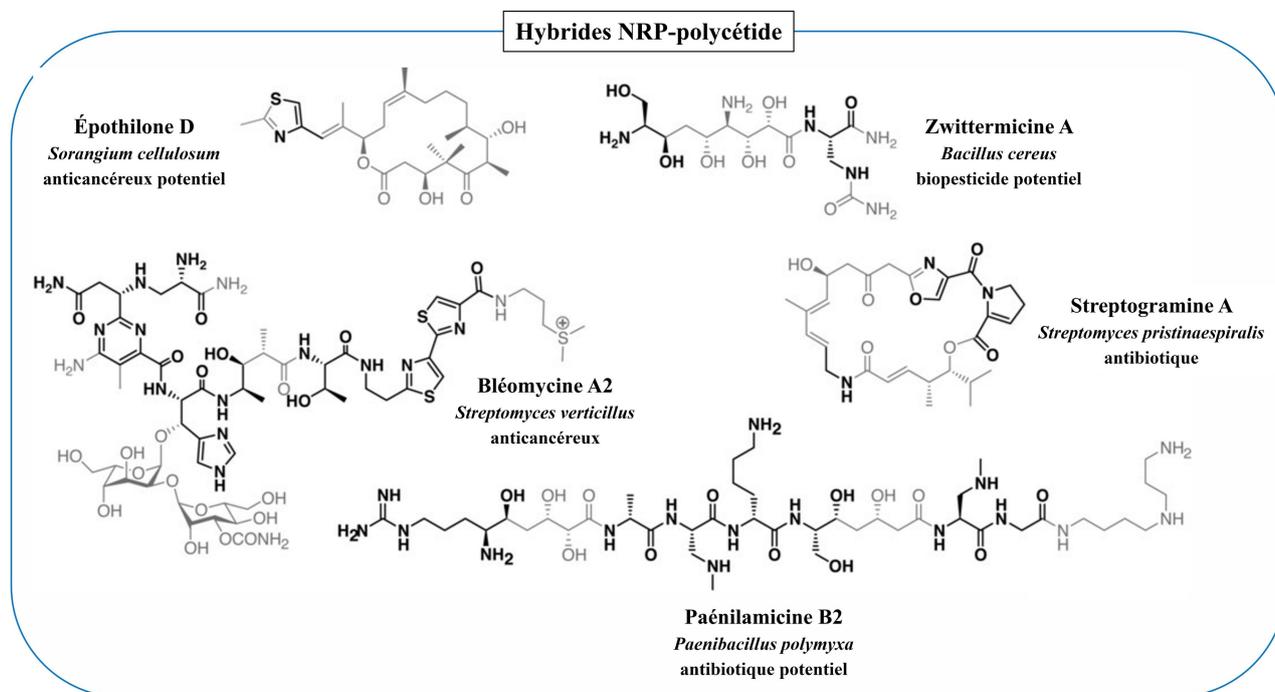


Figure 16. Représentation de structures chimiques et utilisation d'hybrides NRP-polycétides (adaptée de Süßmuth et Mainz, 2017). Les hybrides NRP-polycétides comprennent des classes diversifiées en termes de structures chimiques et de bioactivités. Ils sont utilisés comme antibiotiques (*e.g.* streptogramine A) ou anticancéreux (*e.g.* bléomycine A2) et également étudiés en vertu de leur potentiel thérapeutique (*e.g.* paénilamicine B2, épothilone D) ou agrochimique (*e.g.* zwittermicine A). Les structures chimiques partielles de type peptides ou polycétides sont représentées en noir et en gris, respectivement.

Le système de type III impliqué dans la biogenèse de la bléomycine A2/B2, synthétisé chez la bactérie *S. verticillus* et utilisé dans le traitement des lymphomes ou des carcinomes épidermoïdes, constitue un exemple singulier d'une telle méga-synthase hybride (Du *et al.*, 2000). Les recherches associées à cette molécule ont dévoilé un système avec 11 modules NRPS dispersés sur plusieurs chaînes protéiques (*i.e.* M1-M2-M3, M4-M5, M9-M10) ou isolés (didomaine PCP3-C'3, M6, M8 et M11), dont 4 présentent une activité non-canonique, et avec un module PKS (M7) retrouvé au sein de la voie (Du *et al.*, 2000) (**Figure 17**). Elles en ont également révélé la spécificité de substrat et le mécanisme catalytique à la base de leur biosynthèse.

Le premier module contient des domaines homologues à une acyl-CoA ligase responsables de la formation de la partie β -aminoalaninamide. En effet, les modules 2 et 3 activent et thioestérifient les acides aminés L-Ser et L-Asn, tandis que le domaine C2 catalyse la déshydratation du résidu sérine en L-Ala, ce qui rend l'atome de carbone situé en position C3 sujet à une attaque nucléophile par la fonction α -amino du résidu L-Asn adjacent (Felnagle *et al.*, 2008). L'activité d'aminolyse propre au module initial libère ensuite le dipeptide du domaine PCP3. Le module suivant incorpore à nouveau un acide aminé L-Asn, avant que le domaine C'3 du didomaine PCP-C' ne cyclise le térapeptide en un noyau pyrimidine caractéristique de la structure des bléomycines (Felnagle *et al.*, 2008) (**Figure 17**). Les modules 5 et 6 suivants ajoutent à leur tour les acides aminés L-His et L-Ala au peptide en cours d'extension, alors que le module PKS catalyse l'acétylation, à partir d'un précurseur malonyl-CoA, et la *N*-méthylation de l'atome de carbone situé en position β (**Figure 17**). Les modules 8 et 9 incorporent quant à eux les monomères L-Thr et β -Ala. Le module 10 lie deux acides aminés L-Cys au domaine PCP du module 11 *trans*-actif, dépourvu de domaine A fonctionnel, et assure également l'hétérocyclisation du noyau bithiazoline du produit synthétisé (Felnagle *et al.*, 2008). Par ailleurs, un domaine OX supplémentaire, inclus au sein de ce dernier module, catalyse l'oxydation des cycles thiazolines en thiazoles, et une enzyme accessoire catalyse la β -hydroxylation de l'atome de carbone du résidu L-His (Felnagle *et al.*, 2008) (**Figure 17**). Un domaine C indépendant permet la libération de la molécule cyclisée de la NRPS (Felnagle *et al.*, 2008). Enfin, des transférases assurent l'addition des groupements L-gulose et 3-*O*-carbamoyl-D-mannose au niveau de l'extrémité N-terminale des composés bléomycines A2/B2 (Felnagle *et al.*, 2008). En termes de mode d'action, ces métabolites se complexent avec une molécule de dioxygène, ce qui génère des ROS et provoque donc la rupture double-brins de l'hélice d'ADN au niveau des bases pyrimidines (Felnagle *et al.*, 2008).

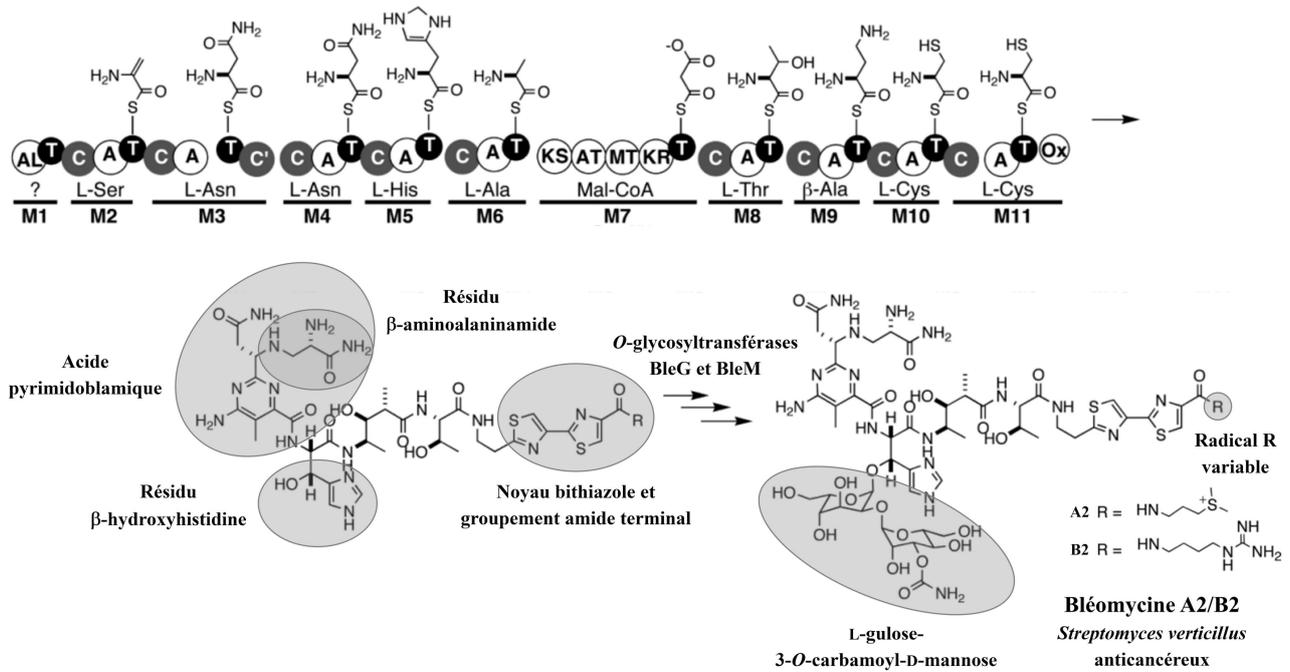


Figure 17. Représentation de la voie de biosynthèse de la bléomycine (adaptée de Felnagle *et al.*, 2008). La NRPS hybride assemble un produit primaire et des *O*-glycosyltransférases fonctionnalisent celui-ci en bléomycine A2/B2. Les groupements pyrimidoblamique et glucidique de la région N-terminale du métabolite constituent le site de complexation à l'ion Fe^{2+} . Les groupements bithiazole et amide de la région C-terminale permettent quant à eux la liaison à l'ADN. AL : acyl-CoA ligase, A : domaine d'adénylation, C : domaine de condensation, PCP : domaine porteur de peptidyle, AT : domaine acyltransférase, KS : domaine cétoynthase, KR : domaine cétoréductase, OX : domaine oxydase.

SECTION IV. Relations structure-fonction des domaines catalytiques

Au cours des dernières décennies, des travaux de biologie structurale intégrative opérés sur des modules intacts et sur les domaines constitutifs ont fourni un aperçu détaillé des éléments suivants : (1) l'organisation linéaire des domaines canoniques A-PCP-C au sein de modules NRPS donnés, (2) les interactions protéines-protéines entre les sous-unités, (3) la spécificité de substrat des domaines d'adénylation, (4) les conformations plausibles au cours d'un cycle catalytique et (5) l'intervention de domaines ou protéines agissant en *trans* (Tanovic *et al.*, 2008 ; Bloudoff *et al.*, 2013 ; Süssmuth et Mainz, 2017 ; McErlean *et al.*, 2019 ; Corpuz *et al.*, 2022).

Dans ce contexte, cette section détaille la sélection, l'activation, la liaison et la condensation des monomères par les domaines A, PCP et C (section **IV.1–4.**), puis expose la communication entre les modules par les domaines de *docking* et la fonctionnalisation terminale du NRP par les domaines TE (section **IV.5–6.**). Elle met ensuite en évidence les modifications potentielles du squelette peptidique par les domaines ou autres protéines accessoires (section **IV.7–8.**).

IV.1. Activation du domaine PCP par une phosphopantéthéinyltransférase

Les PPTases réalisent une modification post-traductionnelle des protéines porteuses d'acyle des synthèses d'acide gras (FAS), de polycétides et de peptides non-ribosomiques (Corpuz *et al.*, 2022). Elles convertissent le domaine porteur inactif (*apo*-protéine) en un domaine actif (*holo*-protéine) par la connexion de la partie 4'-Ppant (phosphopantéthéine), originaire de la coenzyme A, sous forme d'une liaison thioester sur un résidu Ser conservé situé au début de l'hélice $\alpha 2$ (Corpuz *et al.*, 2022). La structure cristallographique de la PPTase Sfp isolée de la bactérie *B. subtilis* a été déterminée en complexe avec la coenzyme A et le domaine PCP de la NRPS TycC3 (Tufar *et al.*, 2014) (**Figure 18**). L'analyse de l'interface protéique a révélé un mode de fixation conservé au sein de nombreuses voies métaboliques : la région C-terminale de la PPTase assure une liaison hydrogène avec le résidu Gln40, situé au niveau de la boucle $\alpha 2$ - $\alpha 3$ du domaine PCP, et des interactions hydrophobes avec les résidus Leu46 et Met49, localisés dans les premiers tours de l'hélice $\alpha 2$ (Tufar *et al.*, 2014). De plus, des études complémentaires de mutagenèse dirigée ont démontré l'importance de ces résidus dans l'interaction avec le domaine PCP, ce qui a appuyé l'activité promiscuitaire de la PPTase Sfp envers les protéines porteuses (Tufar *et al.*, 2014).

Grâce à leur capacité naturelle à fixer des substrats sur les domaines porteurs et à leur activité promiscuitaire, tant au niveau des protéines ciblées que des groupements activés, les PPTases ont été exploitées dans des applications biotechnologiques, notamment pour greffer des fluorophores ou des agents de réticulations chimiques (Beld *et al.*, 2014 ; Corpuz *et al.*, 2022). Les études réalisées dans ce cadre ont démontré que même les peptides synthétiques ayant une structure simple en hélice α et permettant des interactions hydrophobes, sont chargés avec un cofacteur 4'-Ppant marqué (Yin *et al.*, 2005 ; Zhou *et al.*, 2007). De plus, des approches computationnelles et biochimiques basées sur cette observation ont rendu possible la fonctionnalisation de protéines données en présence de PPTases de différentes spécificités (Tallorin *et al.*, 2018).

Phosphopantéthéinyltransférase Sfp

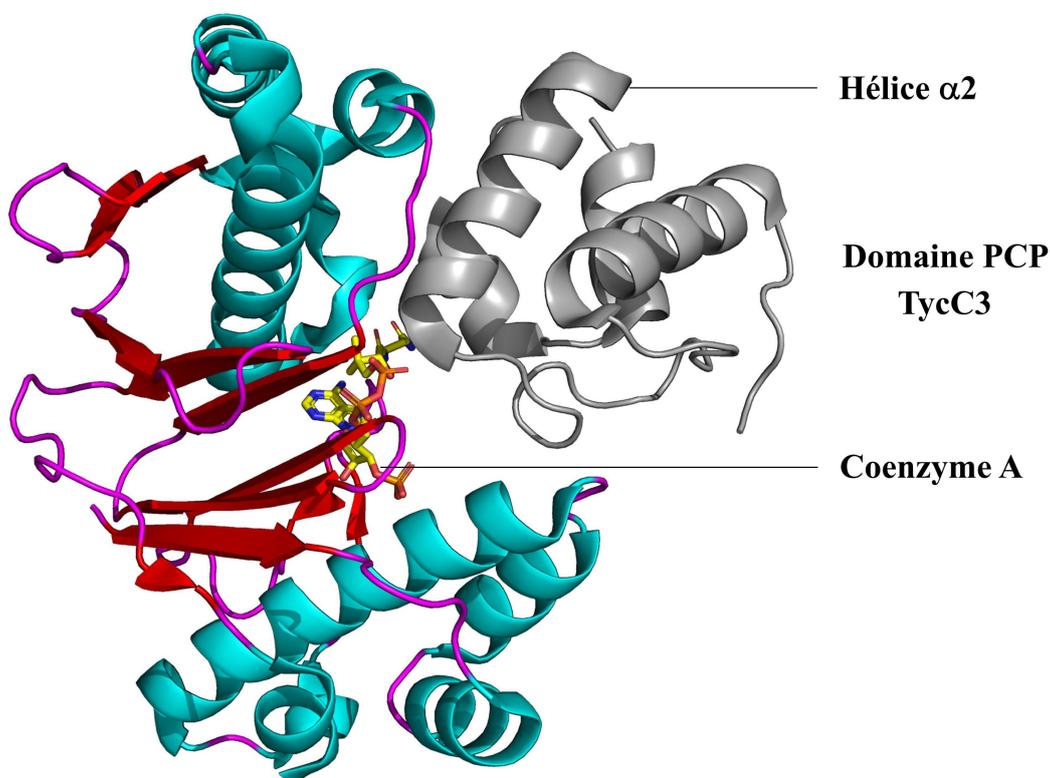


Figure 18. Représentation de la structure tridimensionnelle du complexe de la PPTase Sfp avec un domaine PCP en présence de coenzyme A. La phosphopantéthéinyltransférase Sfp (PDB : 4MRT) adopte une structure dite pseudo-dimérique, dont chaque monomère comprend un noyau en feuillet β à trois brins et trois hélices α (Tufar *et al.*, 2014). L'enzyme interagit avec l'hélice $\alpha 2$ du domaine PCP partenaire au moyen de contacts hydrophobes spécifiques. Pour obtenir cette structure tridimensionnelle, le résidu sérine d'ancrage du cofacteur 4'-Ppant a été muté en alanine, ce qui empêche la liaison et stabilise le complexe avec la coenzyme A au sein du site actif. Les hélices α , les brins β , et les boucles de la PPTase sont représentés en cyan, rouge, et magenta, respectivement, et le domaine PCP en gris.

IV.2. Liaison des substrats et des intermédiaires peptidiques par le domaine PCP

Le domaine *holo*-PCP transporte successivement les substrats et peptides en cours d'extension entre les domaines catalytiques responsables de la formation des liaisons amides, de la modification chimique et de la libération du produit primaire (Mercer et Burkart, 2007). En l'absence de réaction, le cofacteur 4'-Ppant présente une conformation repliée, proche du site d'ancrage ; dès l'interaction du domaine PCP avec le domaine A, ce groupement prosthétique s'étend jusqu'à 18 Å de longueur dans le tunnel du site actif où se produit la thiolation du substrat activé (Drake *et al.*, 2016).

Les domaines PCP adoptent chacun un faisceau de quatre hélices, dont les hélices $\alpha 1$, $\alpha 2$ et $\alpha 4$ sont assez longues et plutôt parallèles, tandis que l'hélice $\alpha 3$ est plus courte et s'étend aux axes des trois autres (Miller et Gulick, 2016) (**Figure 19**). Le résidu Ser du motif d'ancrage conservé GxxS correspond au site de modification et se trouve dans le premier tour de l'hélice $\alpha 2$, précédée d'une longue boucle dont la séquence diverge entre les homologues (Miller et Gulick, 2016). La structure RMN du domaine PCP de la sous-unité TycC3 a initialement démontré que ces domaines partagent le repliement canonique des protéines porteuses impliquées dans la biosynthèse des polycétides et acides gras (Weber *et al.*, 2000). La structure cristallographique de domaines PCP intégratifs (type I) ou autonomes (type II, *e.g.* BlmI) a révélé que le cœur α -hélicoïdal présente des caractéristiques plutôt rigides, avec des variations minimales selon les différents états catalytiques du domaine A de la machinerie NRPS correspondante (Lohman *et al.*, 2014 ; Izoré et Cryle, 2018).

Les structures cristallographiques de domaines PCP résolues en des formes complexées avec les domaines catalytiques exposent les régions de l'interface sous-jacentes à leur association spécifique, comme cela sera détaillé aux sections **IV.3–4**. (Miller et Gulick, 2016). Le cofacteur 4'-Ppant, les hélices $\alpha 2$ et $\alpha 3$, ainsi que la boucle $\alpha 1$ - $\alpha 2$, déterminent la nature des interactions établies avec les partenaires de liaison, notamment les domaines A et C (Miller et Gulick, 2016). Des expériences de mutagenèses aléatoires de ces régions essentielles, suivies de criblages fonctionnels *in vivo*, ont à maintes reprises corroboré ces observations (Corpuz *et al.*, 2022).

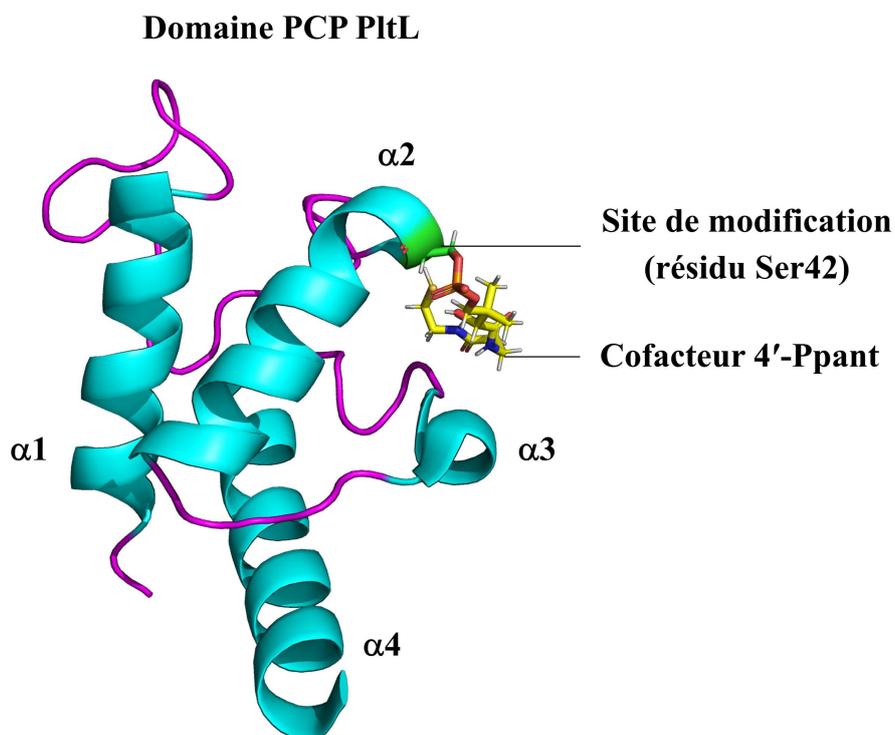


Figure 19. Représentation de la structure cristallographique du domaine PCP PtlL en présence du cofacteur 4'-Ppant. Le domaine *holo*-PCP autonome PtlL (PDB : 2N5H) est impliqué dans la biosynthèse de la pyolutéorine chez la bactérie *Pseudomonas protegens* (Jaremko *et al.*, 2015). La structure se compose de quatre hélices α , dont l'une ($\alpha 3$) adopte une conformation unique. Le résidu Ser42 fait office de site de modification à l'ajout du cofacteur 4'-Ppant. Les hélices α sont représentées en cyan et les boucles en magenta. Le cofacteur 4'-Ppant est illustré en jaune.

IV.3. Sélection et activation d'un monomère spécifique par le domaine A

IV.3.A. Analyse structurale et fonctionnelle

Le domaine d'adénylation active et fixe un monomère spécifique sur le domaine PCP adjacent, préalablement à son incorporation dans le produit final (Corpuz *et al.*, 2022). La capacité d'activer ces unités indépendamment de la machinerie ribosomale permet d'élargir la gamme de précurseurs disponibles, avec plus de 500 identifiés comme substrat pour le domaine A (Payne *et al.*, 2017). En effet, les groupements les plus souvent retrouvés dans les acides aminés non-protéinogènes incluent les fonctions β -hydroxyle et β -méthyle, des chaînes latérales aliphatiques ou aromatiques modifiées, voire des acides halogénés (Süssmuth et Mainz, 2017). De plus, des domaines additionnels peuvent naturellement être incorporés dans des domaines A, un processus de diversification supplémentaire des monomères sélectionnés, comme cela sera détaillé pour un domaine méthyltransférase (MT) à la section IV.7.C (Mori *et al.*, 2018).

Le domaine A se compose d'un cœur N-terminal, appelé A_{core} (résidus $\sim 1-400$), impliqué dans la liaison du substrat, de l'ATP et de l'ion Mg^{2+} , et d'un sous-domaine C-terminal (A_{sub}) (résidus 400–500), connectés entre eux par une séquence flexible composée d'environ cinq résidus (Gulick, 2009) (**Figure 20**). Les motifs distinctifs A1–A10 des domaines d'adénylation ont été caractérisés au sein de plusieurs structures cristallographiques et jouent un rôle structural ou fonctionnel, ce qui confère aux domaines A une spécificité propre et détermine le code non-ribosomique (Süssmuth et Mainz, 2017) (**Figure 21**). En effet, la poche de liaison des ligands varie selon le monomère fixé entre les différents homologues, mais présente dans tous les cas des résidus aspartate (motif A1) et lysine (motif A10), conservés dans les signatures caractéristiques, indispensables à l'interaction ou à la réaction d'adénylation, respectivement (Süssmuth et Mainz, 2017 ; Corpuz *et al.*, 2022). Des algorithmes bio-informatiques ont ainsi été développés afin de prédire les résidus ciblés et, de ce fait, les structures chimiques des NRP biosynthétisés, tel que cela a été vu à la section II.1.B.

Des recherches réalisées sur des multidomains excisés ont mis en évidence que le code non-ribosomique des domaines A est dégénéré et que les domaines C adjacents déstabilisent leurs états conformationnels, ce qui se traduit fréquemment par des profils d'une faible spécificité de substrat (Meyer *et al.*, 2016 ; Süssmuth et Mainz, 2017). Ceci contribue en conséquence à la diversité des produits biosynthétisés à partir d'un unique système NRPS, comme dans un cas extrême à la centaine d'hépatotoxines microcystines assemblées chez les cyanobactéries (Meyer *et al.*, 2016).

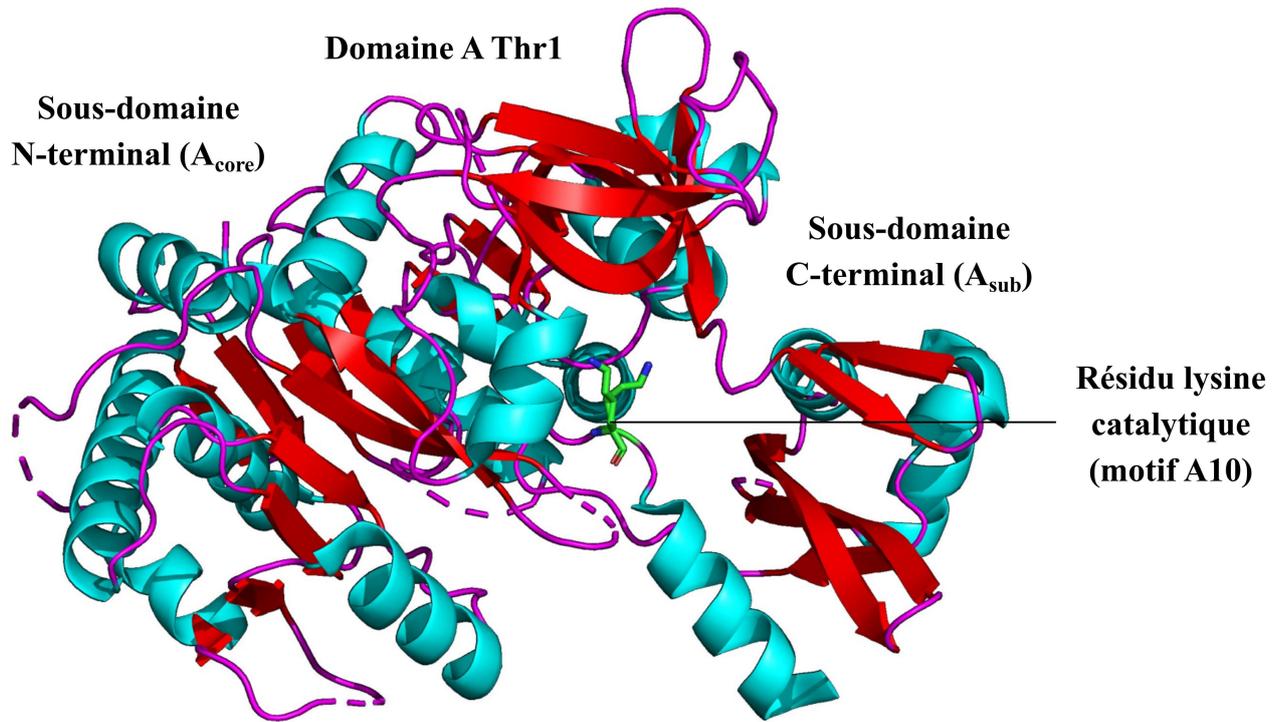


Figure 20. Représentation de la structure cristallographique du domaine A autonome Thr1. Le domaine A discret Thr1 (PDB : 5N9W) est impliqué dans la biosynthèse de 4-chlorothréonine chez la souche *Streptomyces* sp. OH-5093 (Scaglione *et al.*, 2017). La structure se présente dans un état d'adénylation avec les sous-domaines N- et C- terminaux distincts situés de part et d'autre du résidu lysine catalytique (motif A10). Celui-ci assure la première demi-réaction, préalablement au mouvement conformationnel caractéristique du nouvel état de thiolation. Les hélices α , les brins β , et les boucles sont représentés en cyan, rouge, et magenta, respectivement.

A

B

Domaine A	Signature spécifique	Substrat	PDB
GrsA	1 2 3 4 5 6 7 8 9 10 D A W T I A A I C K	Phe	1amu
ApnA-A1	D V E S I G V I A K	Arg/Tyr	4d57/56
CytC1	D F w n i G g V F K	D-Val	3vns
DltA	D L M t l C t v a K	D-Ala	3dhv
AB3403	D I l q l g v V W K	Gly	4zxi
SidNA3	D V L G G G V I G K *	cis-AMHO	3ite
VinN	D F k p t S m K R K	β -MeAsp	3wv5
DhbE	N Y S a q G V - V K	DHBA	1md9
BadA	A Y g y s A g H I K	BA	4eat
Alb1-A1	A V K Y V A N D A K	pABA	---
BlmIV-A1	V D W V I S L A D K	β -Ala	---
ESyn-A1	G A L M V V G S I K	α -HIV	---
Vlm1-A1	A A L W I A V S G K	α -KIV	---

Figure 21. Caractérisation du code non-ribosomique de domaines A (adaptée de Süssmuth et Mainz, 2017). (A) La poche de liaison du domaine A de la NRPS GrsA est représentée. Le substrat L-Phe accommodé est ainsi reconnu par l'ensemble des résidus de la séquence caractéristique DAWTIAAIC, dont le résidu aspartate (motif A1) situé à l'entrée du canal hydrophobe. (B) La signature spécifique de domaines A et leurs substrats respectifs sont indiqués. Les résidus impliqués dans la liaison du substrat sont représentés en lettres capitales. Les résidus Asp et Lys conservés (motifs A1 et A10, respectivement) sont mis en évidence en rouge.

Dans un processus en trois étapes, le domaine A sélectionne d'abord l'acide aminé à ajouter à la chaîne peptidique en cours de croissance, puis catalyse la réaction d'adénylation *via* le résidu lysine, ce qui forme un intermédiaire aminoacyl-AMP dans le site actif et libère un pyrophosphate (Izoré et Cryle, 2018). Un motif additionnel LPxP, localisé en aval du motif A10, interagit avec cette boucle du sous-domaine C-terminal, ce qui conditionne la position de celui-ci et du domaine PCP dans une conformation favorable à la réaction (Izoré et Cryle, 2018).

Le sous-domaine A_{sub} subit alors un changement conformationnel qui correspond à une rotation d'un angle de 140° le long d'une région charnière située au niveau du motif A8, comme cela a été mis en évidence dans plusieurs structures cristallographiques de domaines A activés (Reimer *et al.*, 2019). Cette rotation partielle de la région C-terminale modifie donc l'organisation du site actif, qui passe ainsi à l'état de thiolation, mais pas la position ni l'orientation du substrat, et laisse libre une surface d'interaction avec le domaine PCP correspondant (Corpuz *et al.*, 2022) (**Figure 22**). En fait, le sous-domaine A_{sub} et le domaine porteur interagissent par l'intermédiaire d'une courte séquence peptidique, ce qui associe leurs mouvements lors de cette étape d'adénylation (Gulick, 2009). Ceci explique la capacité des substrats acylés à accéder au site actif des domaines catalytiques au cours de la biosynthèse (Payne *et al.*, 2017).

Au cours de la liaison du domaine *holo*-PCP avec le domaine A dans un état de thiolation, le cofacteur 4'-Ppant s'insère au niveau de l'intermédiaire aminoacyl-AMP (Gulick, 2009). Le domaine A catalyse alors le transfert du monomère activé vers la fonction thiol du groupement prosthétique, ce qui forme alors une liaison covalente avec l'acide aminé (Gulick, 2009). L'AMP et le domaine PCP chargé avec le substrat se dissocient enfin, et le domaine A retourne à son état original pour catalyser le prochain cycle de réactions enzymatiques (Miller et Gulick, 2016).

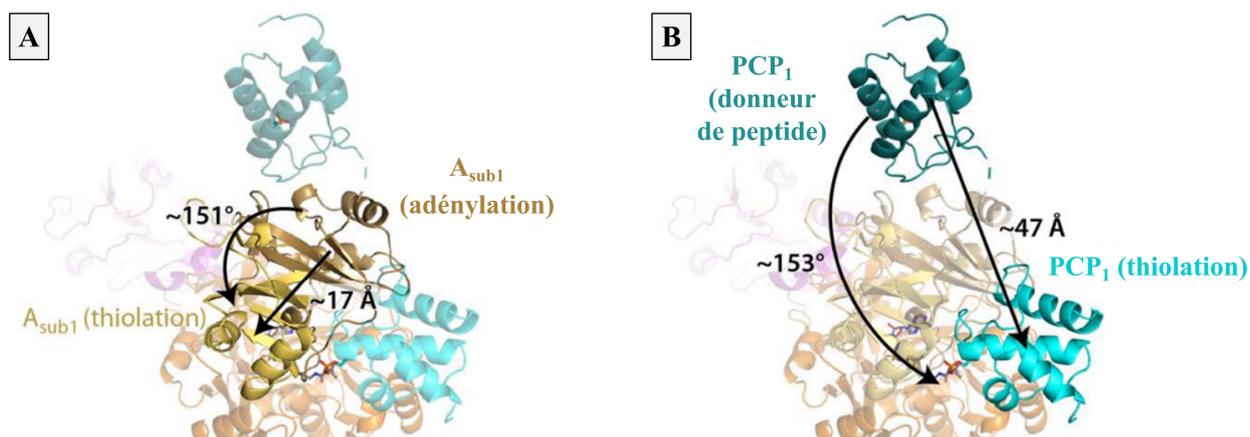


Figure 22. Rotation des domaines A1 et PCP1 de la NRPS LgrA lors de la réaction d'adénylation (adaptée de Reimer *et al.*, 2019). Le module initial de la NRPS LgrA (PDB : 6MFZ) est impliqué dans la biogenèse de gramicidine cyclique chez la bactérie *Brevibacillus brevis* (Reimer *et al.*, 2019). Lors de la liaison du substrat, le domaine PCP1 est dans l'état donneur de peptide et le domaine A1 dans l'état d'adénylation. Suite à la réaction enzymatique, le sous-domaine A_{sub} (A) et le domaine PCP1 (B) pivotent chacun d'un angle précis (151° , 153° , respectivement) et déplacent leur centre de masse d'une certaine distance (17 Å, 47 Å, respectivement) le long d'une région charnière au niveau du motif A8, ce qui forme l'état de thiolation.

En outre, de nombreux clusters NRPS expriment une protéine d'environ 70 résidus qui exerce un rôle dans l'activation du domaine A, et semble primordiale dans certains systèmes afin de former l'intermédiaire acyl-adénylate correspondant (Baltz, 2011). Ces protéines accessoires sont désignées comme protéines de type MbtH (MLP, *MbtH-Like Proteins*), telles que MbtH, PA2412 ou SlgN1 (Miller et Gulick, 2016). Les MLP adoptent une structure organisée en forme de flèche avec trois feuillets β antiparallèles centraux suivis d'une hélice α droite (Miller et Gulick, 2016) (**Figure 23**). Deux résidus tryptophane conservés apparaissent essentiels à l'activité de cette famille de protéines, mais les mécanismes moléculaires qui régissent leur fonctionnement catalytique et déterminent leur activité biologique demeurent toujours en cours d'élucidation (Corpuz *et al.*, 2022). Une possibilité pourrait être la capacité à réaliser une régulation allostérique de l'alternance des sous-domaines A_{core} et A_{sub} , et donc la régénération de la NRPS à chaque cycle catalytique (Payne *et al.*, 2017).

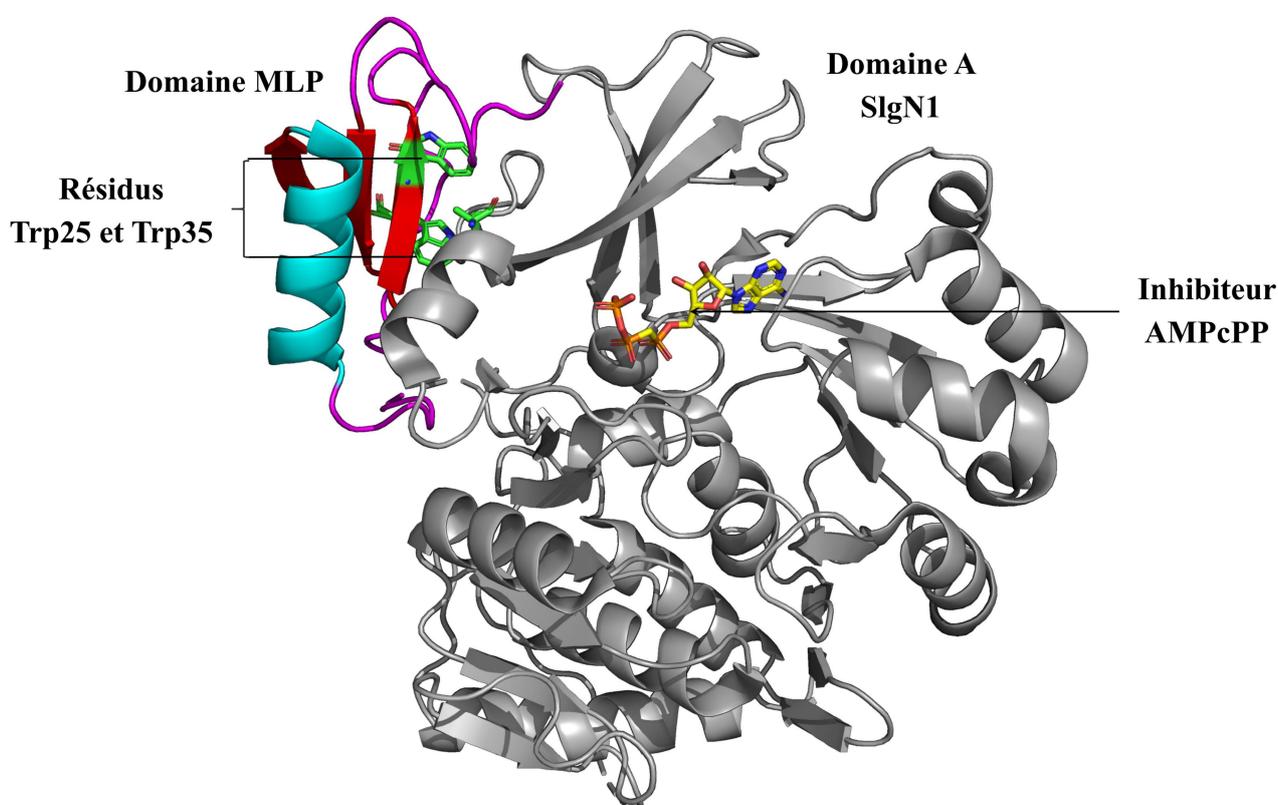


Figure 23. Représentation de la structure cristallographique du domaine SlgN1. Le domaine A autonome SlgN1 (PDB : 4GR5) est impliqué dans la biosynthèse de streptolydigue chez la bactérie *Streptomyces lydicus* (Herbst *et al.*, 2013). La structure est fusionnée à un domaine MLP à l'extrémité N-terminale, dont des résidus conservés contribuent de manière cruciale à cette interaction stable. En effet, le domaine MLP se lie à un petit motif α - β hydrophobe du sous-domaine A_{core} , entouré de régions en boucle et des motifs A6 et A7. Le brin β de trois résidus de ce motif d'ancrage comprend le résidu Ala433, dont la chaîne latérale s'insère dans une fente hydrophobe formée par les résidus conservés Trp25, Pro32 et Trp35. Le résidu aspartate du motif A7, localisé à proximité, interagit avec les groupements hydroxyles du ribose de l'ATP, en l'occurrence l'inhibiteur non hydrolysable qui piège le complexe natif dans une conformation d'adénylation. AMPcPP : méthylène-adenosine triphosphate, MLP : protéine de type MbtH.

Les premières structures cristallographiques de domaines A résolues correspondent au domaine PheA de la NRPS GrsA (gramicidine S), au domaine autonome DhbE (bacillibactine) et au domaine EntE de la NRPS EntF (entérobactine) (Conti *et al.*, 1997 ; May *et al.*, 2002 ; Sundlov *et al.*, 2012). Les structures se présentent dans un état d'adénylation avec l'acide dihydroxybenzoïque et l'AMP dans le site actif de PheA, par exemple, ou dans un état de thiolation avec un inhibiteur dans le site actif de EntE (**Figure 24**). Bien que la majeure partie du site actif soit située dans le sous-domaine A_{core}, le résidu lysine catalytique présent dans la boucle A10 du sous-domaine A_{sub} interagit avec les groupements fonctionnels de l'AMP pour assurer la formation de l'intermédiaire réactionnel (Miller et Gulick, 2016) (**Figure 24**). Certaines régions N-terminales s'avèrent également d'une importance considérable, en particulier la boucle A3 qui oriente les groupements phosphates β et γ de l'ATP, ainsi que le résidu aromatique du motif A4 ou le résidu aspartate du motif A7 qui interagissent avec les groupements phosphoriques et hydroxyles du nucléotide, respectivement (Miller et Gulick, 2016). Dès la formation de l'intermédiaire aminoacyl-AMP et la libération du pyrophosphate, le changement conformationnel prépare le site actif à la réaction de thiolation (Miller et Gulick, 2016). Dans toutes ces structures, les positions des interfaces sont cohérentes, mais les types d'interactions qui assurent la liaison au domaine *holo*-PCP sont variables (Corpuz *et al.*, 2022).

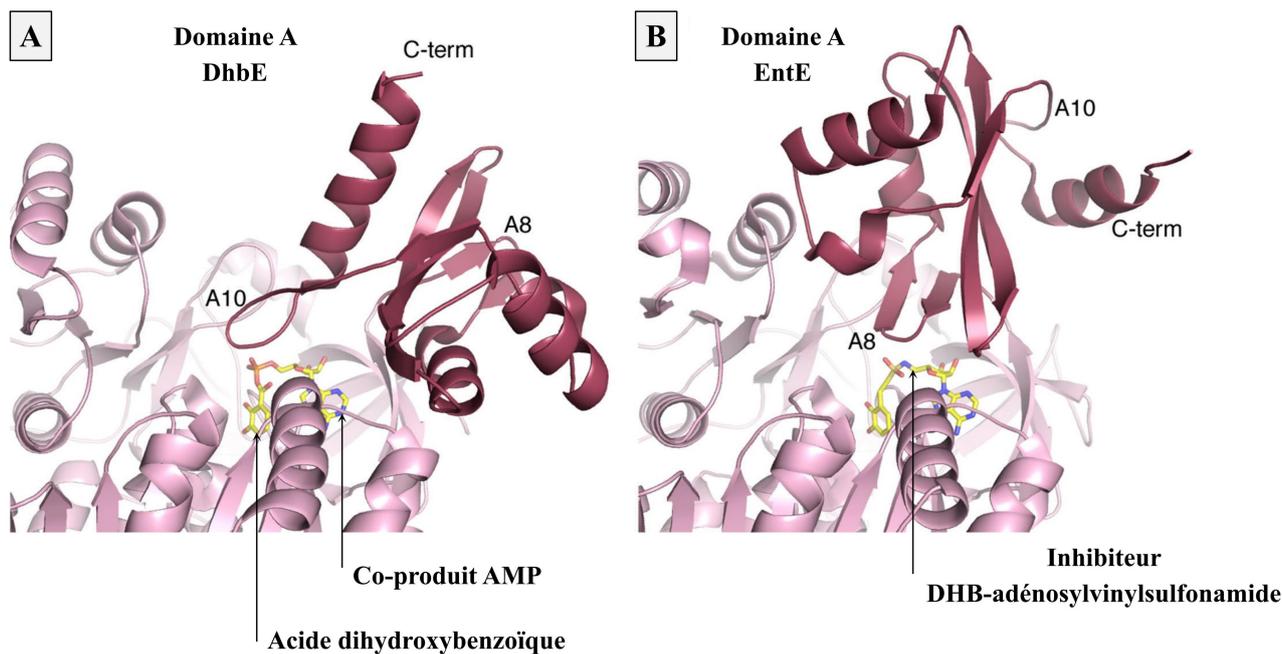


Figure 24. Représentation de l'interface au sein de domaines A autonomes (adaptée de Miller et Gulick, 2016). (A) Le domaine discret DhbE (PDB : 1MDB) adopte une conformation d'adénylation avec le motif A10 du sous-domaine C-terminal dirigé vers le site actif. (B) Le domaine indépendant EntE (PDB : 4IZ6) est par contre dans la conformation de thiolation avec le motif A8 présent à proximité du site actif. AMP : adénosine monophosphate.

IV.3.B. Utilisation de sondes chimiques dans les systèmes de didomains A-PCP

Les études menées sur les didomains A-PCP ont nécessité l'application de sondes chimiques pour déterminer la base moléculaire des interactions protéiques responsables de la liaison entre ces domaines et de la translocation du substrat (Corpuz *et al.*, 2022). Le principal agent chimique utilisé correspond à l'inhibiteur adénosine vinylsulfonamide (AVS), initialement élaboré comme analogue de l'intermédiaire aminoacyl-adénylate (Corpuz *et al.*, 2022) (**Figure 25**). Cet inhibiteur incorpore en plus un piège électrophile sous forme d'un alcène (éthylène) lié à un groupement attracteur (*e.g.* salicylate, valyl, séryl, prolyl, *etc.*), attaqué par la fonction thiol du cofacteur 4'-phosphopantéthéine (Corpuz *et al.*, 2022). Cela crée ainsi une liaison covalente avec le domaine PCP, alors que dans un contexte natif, celui-ci n'interagit que de manière transitoire avec l'état de thiolation du domaine A.

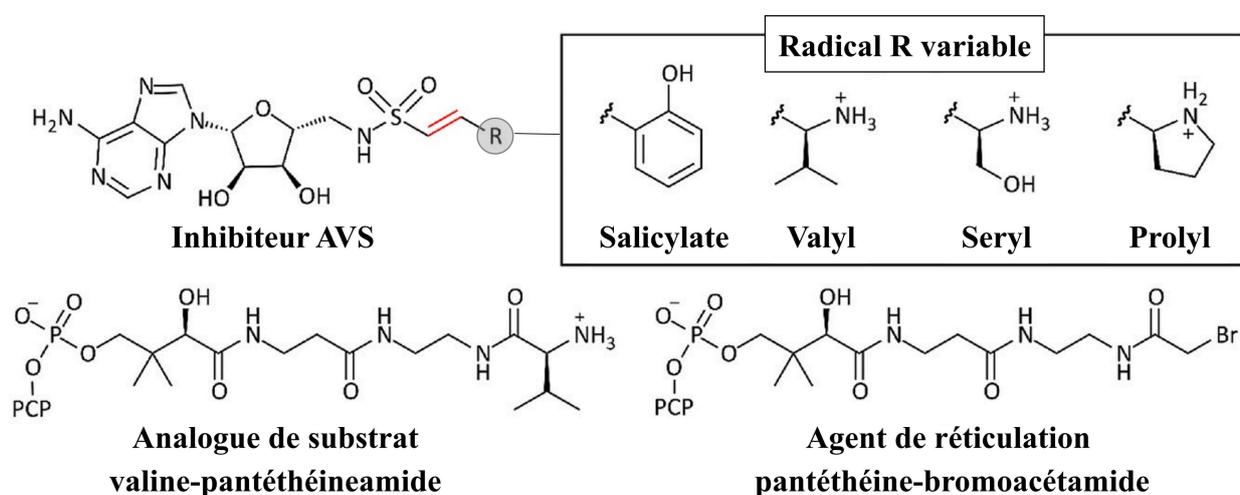


Figure 25. Utilisation de sondes chimiques dans les complexes de didomains A-PCP (adaptée de Corpuz *et al.*, 2022). Les inhibiteurs AVS, les analogues de substrat ou les agents de réticulation ont tous pour fonction de figer et/ou de stabiliser le complexe des domaines A et PCP dans des états donnés par l'intermédiaire de pièges électrophiles. Dans le cas du pantéthéinamide, un groupement amide remplace la fonction thioester labile, ce qui empêche le transfert de l'intermédiaire réactionnel. Dans le cas de l'agent de réticulation, le centre adjacent à l'atome de Br est électropositif, ce qui permet une attaque nucléophile et la formation d'une liaison croisée.

Par exemple, la structure cristallographique du complexe des domaines autonomes PltL et PltF (PCP-A), impliqués dans la biosynthèse de pyolutéorine chez la bactérie *Pseudomonas protegens*, a été résolue en présence d'un inhibiteur prolyl-AVS (Corpuz *et al.*, 2020) (**Figure 26A**). Cette étude a mis en évidence un mode de liaison comparable aux structures analogues déterminées, mais avec un rôle amoindri de l'hélice α_2 du domaine PCP PltL dans les interactions hydrophobes à l'interface (Corpuz *et al.*, 2020) (**Figure 26B**). En effet, seul le résidu Met43, situé juste en aval du résidu sérine conservé de cette structure secondaire, contacte le résidu Met257 du domaine A PltF (Corpuz *et al.*, 2020). Le résidu Ile454 de celui-ci se loge au sein d'une cavité hydrophobe agencée par les résidus Ile19 et Trp37, localisés dans la boucle α_1 du partenaire (Corpuz *et al.*, 2020). De plus, le reste de l'interface, situé à proximité de l'hélice α_3 du domaine PCP PltL, met en jeu des liaisons hydrogènes formées entre le squelette et/ou les chaînes latérales des résidus Gly38/Ser62 de PltL et Lys457/Ser232 de PltF (Corpuz *et al.*, 2020).

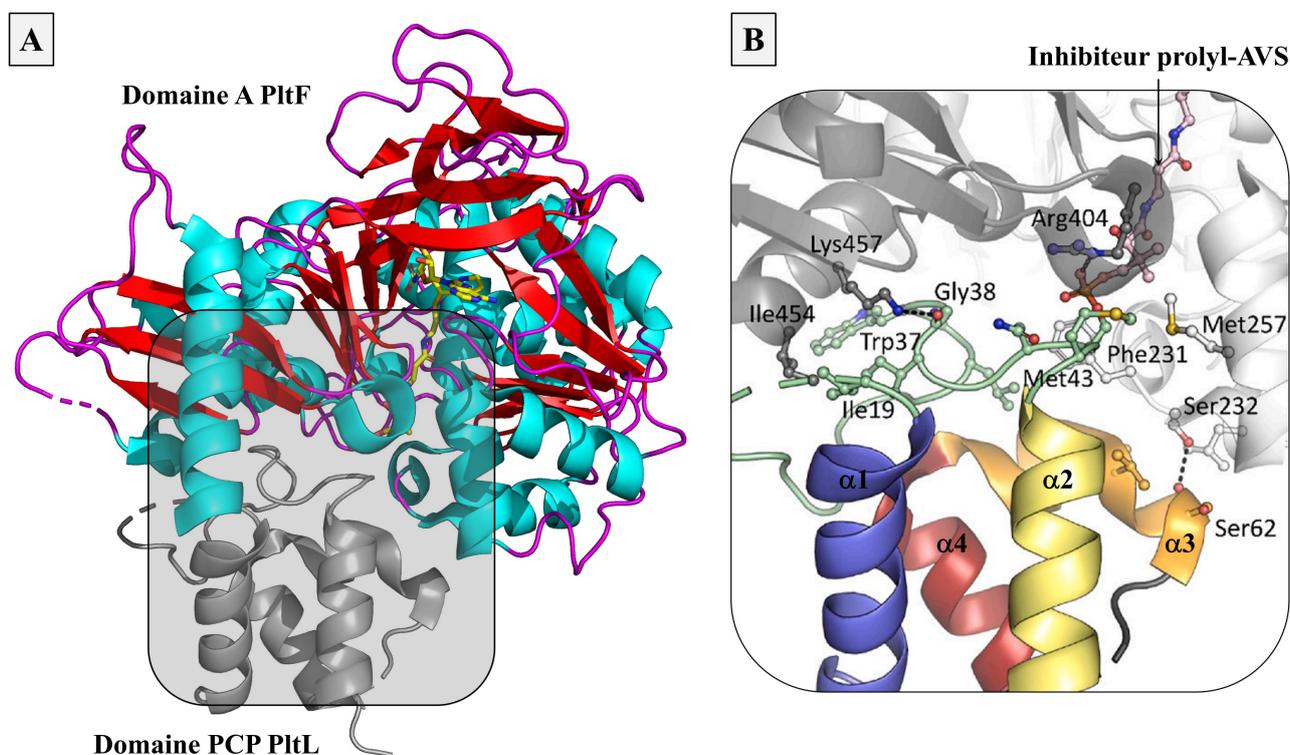


Figure 26. Représentation de la structure cristallographique du complexe autonome PltF-PltL (adaptée de Corpuz *et al.*, 2022). (A) La structure cristallographique du complexe de type II (PDB : 6O6E) a été déterminée en présence de l'inhibiteur prolyl-AVS, dont la fonction acyle correspond à la spécificité du domaine A, ce qui bloque donc la réaction catalytique. Les hélices α , les brins β , et les boucles du domaine A PltF sont représentés en cyan, rouge, et magenta, respectivement, et le domaine PCP partenaire en gris. (B) L'interface associée dépend à la fois de contacts hydrophobes, de liaisons hydrogènes ou encore d'une interaction électrostatique.

Les interfaces de domaines A et PCP analysées dans la littérature, y compris celle du complexe PltF-PltL considéré, dépendent principalement d'interactions hydrophobes et de liaisons hydrogènes médiées entre des résidus conservés (Corpuz *et al.*, 2022). Les mécanismes moléculaires à la base de l'interaction entre ces domaines et de l'incorporation du substrat restent en cours d'élucidation.

IV.4. Formation des liaisons peptidiques par le domaine C

IV.4.A. Analyse structurale et fonctionnelle

Le domaine C catalyse la formation d'une liaison amide entre le groupement carbonyle (thioester) de l'intermédiaire aminoacyle ou peptidyle lié au domaine PCP du module n-1 et la fonction amino du substrat attaché au module n (Felnagle *et al.*, 2008). Ces domaines appartiennent à la sous-classe des chloramphenicol acétyltransférases (CAT), et adoptent un repliement caractéristique en pseudo-dimère divisé en des lobes N- et C-terminaux, dont chacun possède un feuillet β central et des hélices α agencées contre les faces (Leslie *et al.*, 1988) (**Figure 27**). Les lobes maintiennent une interaction au moyen d'éléments structuraux conservés (*i.e.* jonction α -hélicoïdale et α/β -crossovers) (Miller et Gulick, 2016 ; Tan *et al.*, 2020) (**Figure 27**). Les sous-domaines présentent des poches distinctives de liaison du substrat, appelées site accepteur ou donneur, et forment à l'interface le site catalytique avec le motif C3 HHxxxDG conservé (boucle $\beta 6$ - $\alpha 4$) des enzymes CAT (Miller et Gulick, 2016).

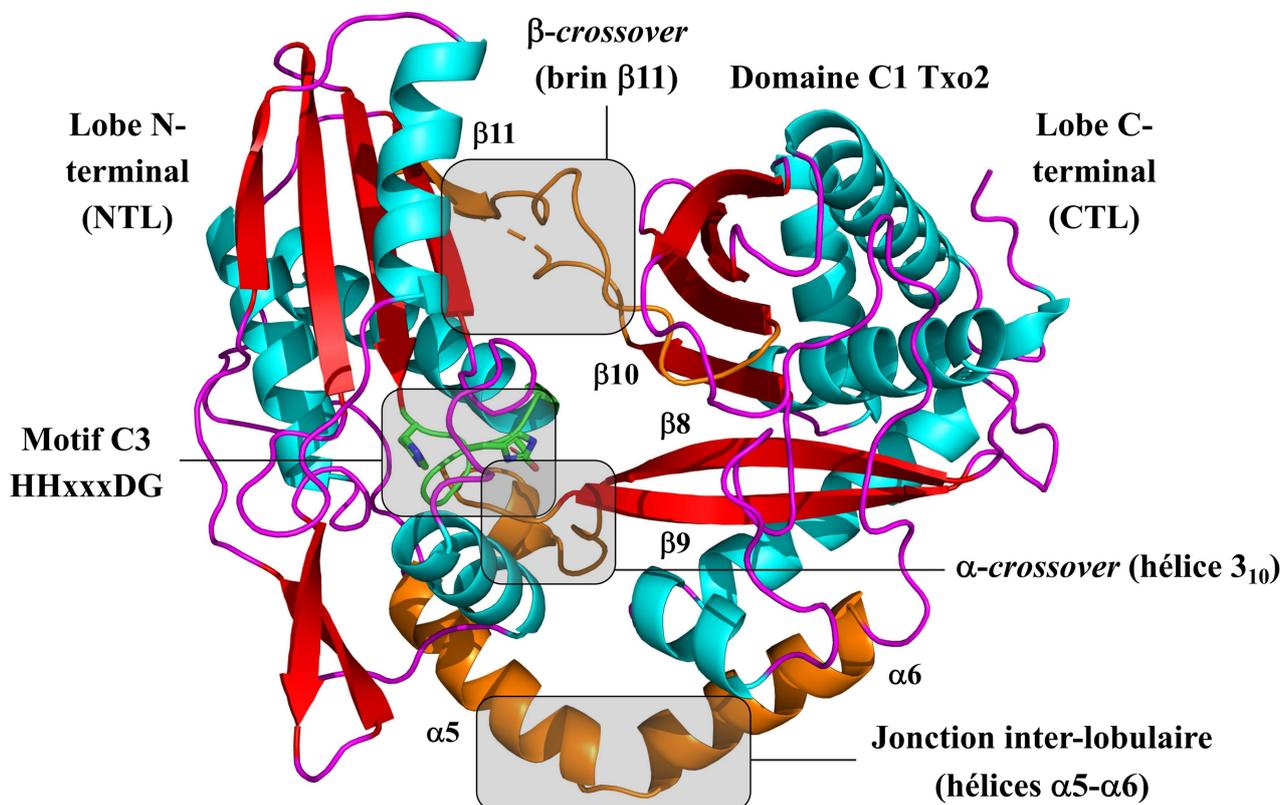


Figure 27. Représentation de la structure cristallographique du domaine C1 de la NRPS Txo2. Le domaine C1 intégratif du module Txo1 (PDB : 6P1J) est impliqué dans la biosynthèse de téixobactine (Tan *et al.*, 2020). La structure pseudo-dimérique présente deux lobes (NT et CT), composés chacun de deux feuillets β et cinq hélices α , assemblés de part et d'autre de l'ouverture centrale *via* les hélices $\alpha 5$ - $\alpha 6$, d'une hélice 3_{10} (α -crossover) et du brin $\beta 11$ (β -crossover). Le site actif contient le motif HHxxxDG conservé avec le second résidu histidine catalytique. Les hélices α , les brins β , et les boucles sont représentés en cyan, rouge, et magenta, respectivement. Les éléments de structure interlobulaires sont mis en évidence en orange.

Les domaines PCP interagissent simultanément au niveau de l'interface correspondant aux sites accepteur ou donneur du tunnel catalytique du domaine C et transportent les substrats corrects dans le site actif (Izoré et Cryle, 2018). Le site accepteur présente une sélectivité stricte envers la chaîne latérale et la configuration L- des substrats d'acides aminés, ce qui permet de contrôler le processus de sélection des monomères inhérent au domaine A (Dekimpe et Masschelein, 2021). Les domaines C démontrent en conséquence une stéréosélectivité différentielle : les domaines ^LC_L et ^DC_L couplent un substrat donneur d'isométrie L- ou D-, respectivement, avec un substrat accepteur d'isométrie L- strictement (Dekimpe et Masschelein, 2021).

Le mécanisme catalytique demeure en cours d'investigation, notamment à cause de l'absence de ligands piégés dans les structures cristallographiques déterminées. Des expériences de mutagenèse ont démontré que le second résidu His de la séquence consensus C3 HHxxxDG déprotone la fonction α -amino du peptidyl-PCP accepteur et que celle-ci réalise une attaque nucléophile sur le groupement carbonyle-thioester du peptidyl-PCP donneur (en amont) (Bloudoff et Schmeing, 2017) (**Figure 28**). À la lumière d'études sur plusieurs domaines C, le mécanisme proposé à été remis en question, vu que la base catalytique ne s'avère pas essentielle (*i.e.* mutation non-délétère, rôle structural, absence du résidu) (Izoré et Cryle, 2018). Ainsi, bien que le motif HHxxxDG soit hautement conservé, des domaines C semblent assurer des réactions de liaisons peptidiques d'une spécificité alternative, voire d'autres activités, comme cela sera détaillé à la section **IV.4.B**.

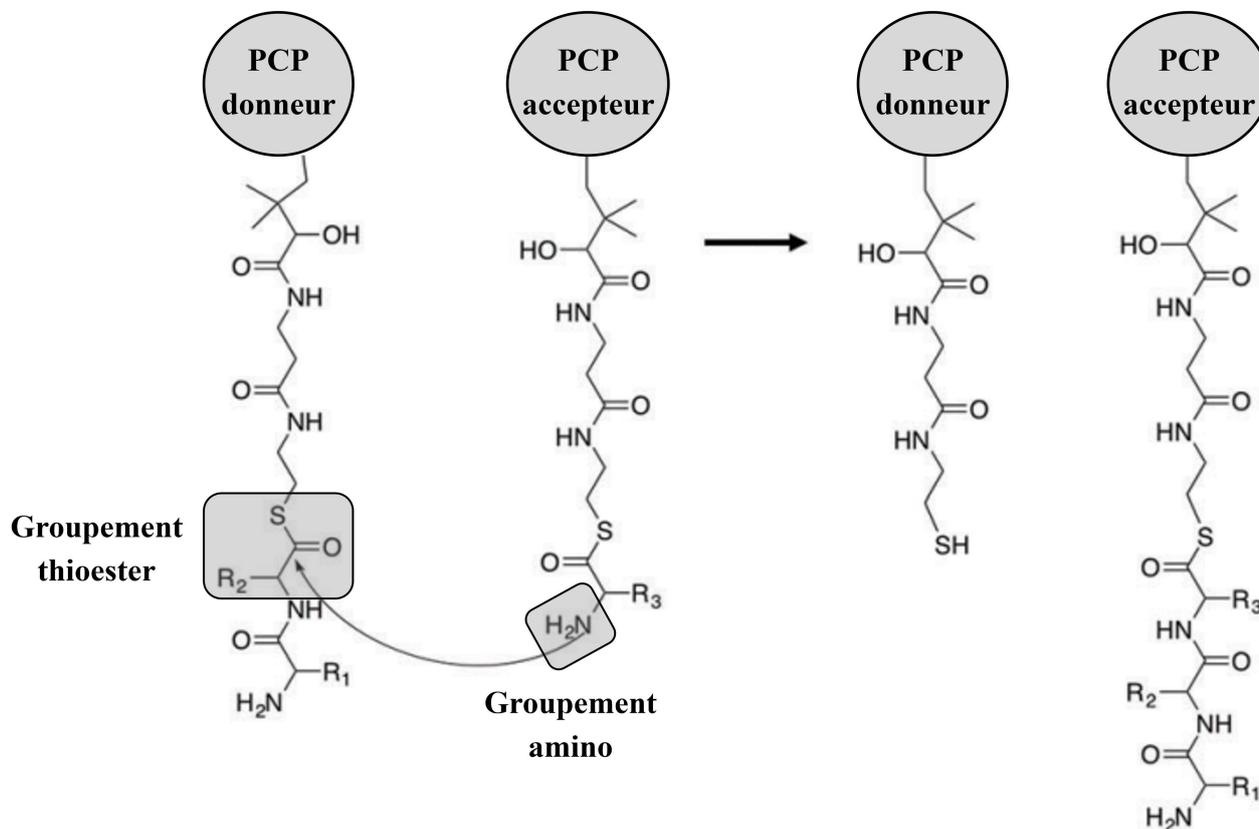


Figure 28. Réaction de condensation peptidique par le domaine C (adaptée de Payne *et al.*, 2017). (A) Au cours de la réaction, le groupement amino du résidu lié au domaine PCP accepteur attaque le groupement thioester du peptide lié au domaine PCP donneur, ce qui permet ainsi le transfert du peptide et l'élongation de la chaîne.

Les premières structures cristallographiques de domaines C résolues correspondent au domaine autonome VibH, au domaine intégratif CDA-C1, à un didomaine PCP-C (TyC5–6) et à un module NRPS (SrfA–C) (Keating *et al.*, 2002 ; Samel *et al.*, 2007 ; Tanovic *et al.*, 2008 ; Bloudoff *et al.*, 2013). Compte tenu de leurs natures pseudo-dimériques, les structures caractérisées jusqu'à ce jour adoptent une conformation davantage ouverte ou fermée en fonction du domaine C considéré, et une interaction avec le peptidyl-PCP pourrait provoquer des changements de cet état, bien que cela n'ait été observé que dans un unique cas (domaine C2 de la NRPS Txo1) (Izoré et Cryle, 2018 ; Tan *et al.*, 2020). Par exemple, la structure du domaine CDA-C1 présente une conformation fermée avec une inclinaison du lobe N-terminal supérieure de 15°–25° par rapport à des homologues (Bloudoff *et al.*, 2013) (**Figure 29**). De plus, divers essais enzymatiques ont démontré que l'accessibilité variable au site actif ne conditionne pas l'activité de la protéine *in vitro* (Bloudoff *et al.*, 2013). Par ailleurs, des analyses de diffusion des rayons X aux petits angles et de dynamique moléculaire ont confirmé que les conformations observées sont pertinentes sur le plan biologique (Samel *et al.*, 2007 ; Tanovic *et al.*, 2008 ; Bloudoff *et al.*, 2013).

De telles conformations revêtent une importance au cours de l'interaction avec les substrats liés au PCP afin de contrôler l'ordre de liaison des domaines donneur et accepteur (Izoré et Cryle, 2018). Les domaines C assurent de cette façon la stéréochimie du centre L- ou D- du peptidyl-PCP donneur, par le biais d'une compétition dynamique avec un éventuel domaine E adjacent, mais également la modification correcte du peptidyl-PCP accepteur au cours de la liaison avec des domaines auxiliaires *cis-/trans*-catalytiques (Izoré et Cryle, 2018). En raison des rôles majeurs des domaines C au sein des systèmes NRPS, de nombreux résultats cruciaux restent à obtenir grâce aux études biochimiques et structurales de complexes liés aux domaines PCP.

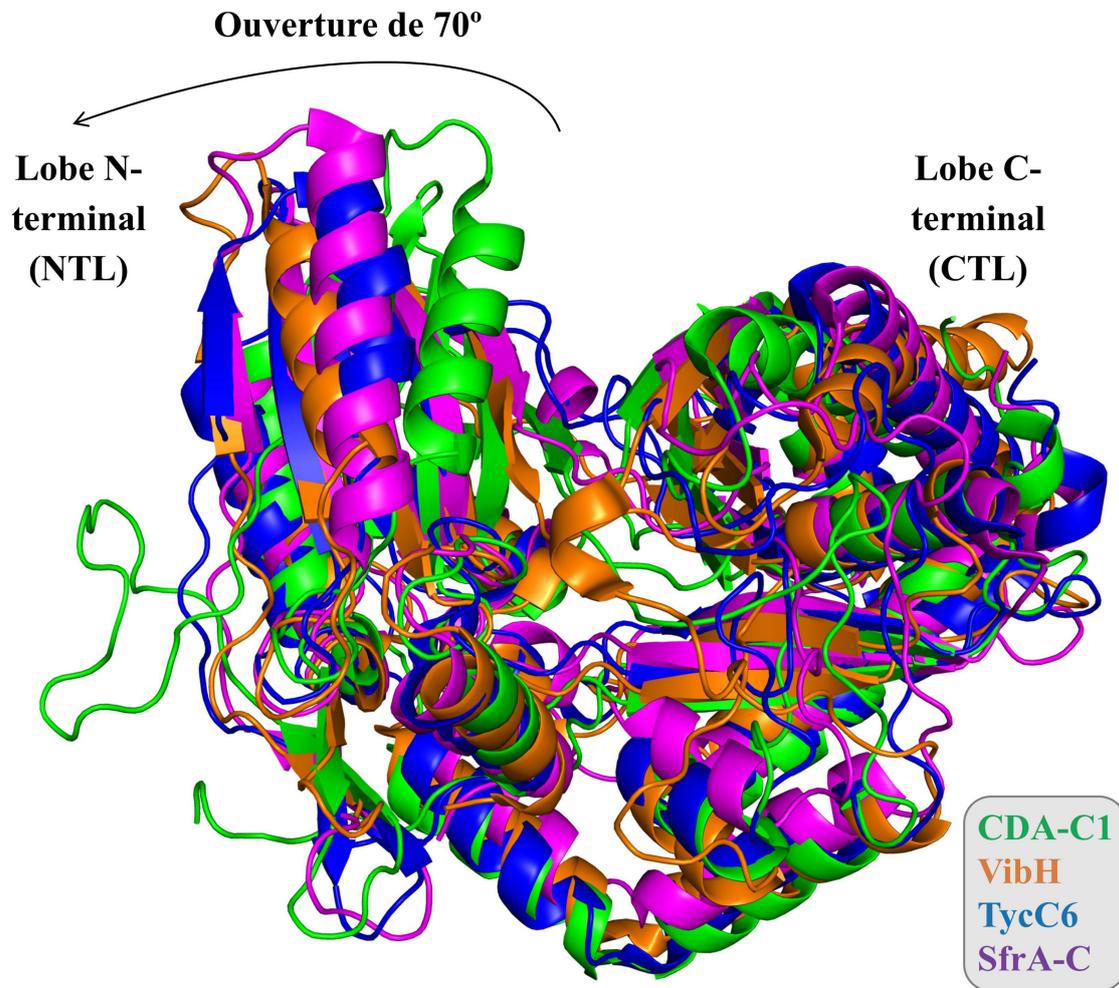


Figure 29. Comparaison des structures cristallographiques de domaines C homologues. Les structures du domaine C1 de la CDA synthase (CDA-C1, vert) (PDB : 4JN3) et de domaines homologues, en l'occurrence le domaine discret VibH (orange) (PDB : 1L5A), le domaine C6 de la NRPS TycA-C (bleu) (PDB : 2JGP) et le domaine C-terminal de la NRPS SrfA-C (violet) (PDB : 2VSQ) ont été superposées. Cette analyse a révélé que les sous-domaines N-terminaux se trouvent dans des positions différentes avec des variations des atomes de carbone en position α de 19 Å, 20 Å et 23 Å, respectivement.

Les domaines C interagissent à la fois avec les domaines A et PCP intramodulaires, mais aussi avec le domaine porteur du module précédent. Les études réalisées sur les didomains PCP-C ont impliqué l'usage de sondes chimiques pour promouvoir la cristallisation et obtenir des informations structurales à haute résolution sur les interactions médiées entre ces unités catalytiques (Reimer *et al.*, 2019 ; Corpuz *et al.*, 2022). Par exemple, la structure cristallographique du domaine C2 lié aux domaines PCP accepteur et donneur des modules de la NRPS LgrA a été déterminée en présence de l'analogue de substrat formyl-valine Ppant (Reimer *et al.*, 2019). Au niveau de l'interface formée avec le domaine PCP1, les interactions se produisent entre les hélices $\alpha2/\alpha3$ et le lobe C-terminal du domaine C2 par des contacts hydrophobes, tandis que sur l'interface opposée du complexe, les interactions se tiennent plutôt entre l'hélice $\alpha2$ du domaine PCP2 et les domaines N- et C-terminaux du domaine C2 (Reimer *et al.*, 2019) (**Figure 30**). En effet, les résidus Leu723, Phe752 ou Tyr743, situés dans la boucle $\alpha1-\alpha2$ ou dans l'hélice $\alpha3$, établissent des interactions hydrophobes avec les résidus Thr1013, Leu1085 ou Leu1088 localisés dans le site donneur du partenaire, respectivement. De plus, l'analogue du cofacteur se trouve au sein du canal composé par les deux lobes du domaine C2 et le groupement thiol à une distance d'environ 3 Å du résidu catalytique putatif His908 (Reimer *et al.*, 2019). En outre, l'analyse mutationnelle des interactions médiées au niveau des interfaces de complexes PCP-C a d'une part mis en évidence une diminution drastique de l'activité du domaine C, et d'autre part, l'implication de certains résidus dans le chargement du substrat au sein du site actif (Reimer *et al.*, 2019 ; Izoré *et al.*, 2021).

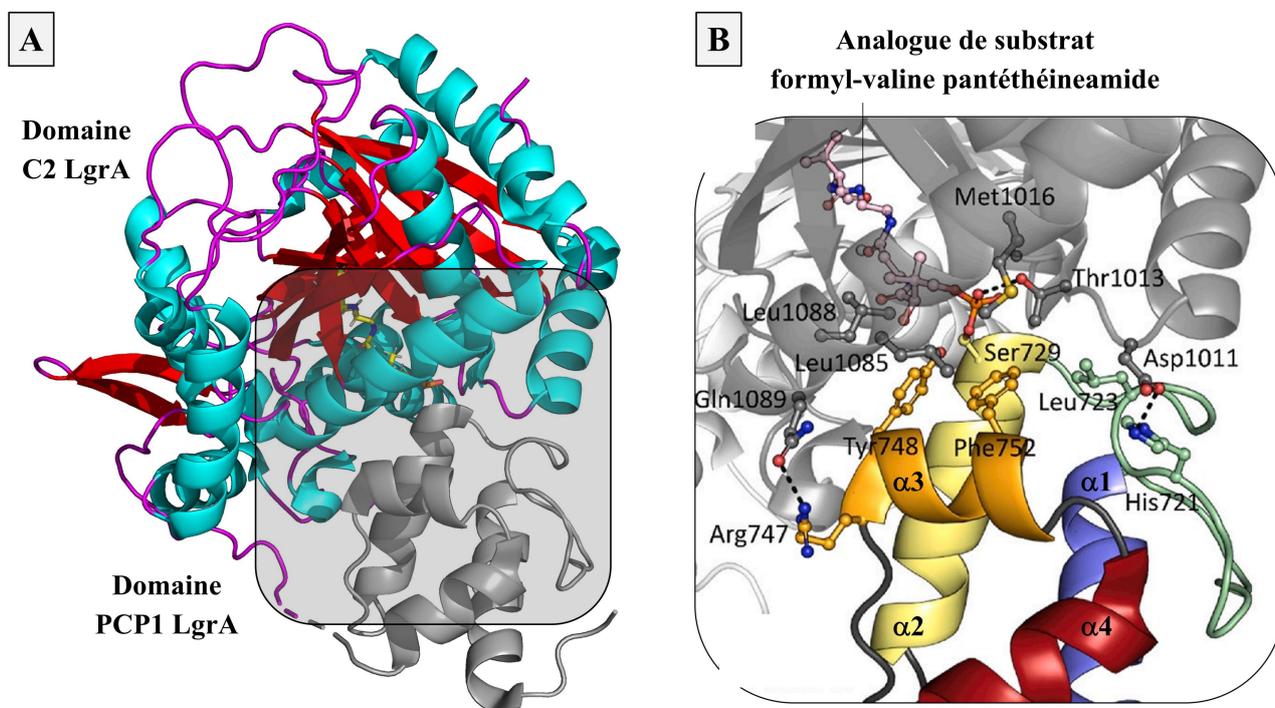


Figure 30. Représentation de la structure cristallographique du complexe PCP-C de la NRPS LgrA (adaptée de Corpuz *et al.*, 2022). (A) La structure cristallographique du didomaine PCP1-C2 de la NRPS dimodulaire LgrA (PDB : 6MFW) a été déterminée en présence de l'inhibiteur formyl-valine pantéthéinamide. (B) L'interface associée dépend exclusivement d'interactions hydrophobes.

IV.4.B. Évolution fonctionnelle

Dans les chaînes d'assemblage, des domaines C modifiés au niveau des éléments structuraux ou du motif catalytique C3 assurent des réactions d'estérification, d'épimérisation, de cyclisation et de déshydratation des NRP, voire du recrutement des enzymes partenaires (Dekimpe et Masschelein, 2021) (**Tableau III**). Les domaines C définissent la configuration énantiomérique des peptides, tel que cela a été décrit, ce qui reflète la relation avec les domaines d'épimérisation (E) apparentés sur le plan phylogénique. Les domaines E convertissent les résidus d'acides aminés à partir de la forme L- majoritairement activée par les domaines A (Chen *et al.*, 2016). Ces activités spécifiques sont combinées au sein des domaines ^DC_L/E hybrides du système ArfA–C à l'origine de la biosynthèse d'arthrofactine chez la souche *Pseudomonas* sp. MIS38 (Balibar *et al.*, 2005). En particulier, les 6 premiers résidus présentent l'énantiométrie dextrogyre, bien qu'aucun module des trois sous-unités catalytiques ne possède de domaines E propres (Balibar *et al.*, 2005).

Tableau III. Classification évolutive des domaines C (adapté de Süßmuth et Mainz, 2017).

Type	Motif catalytique C3	Fonction	PDB
C	xHHxxxDGxS	Condensation	1L5A
C/E	xHH [I / L] xxxxD	Condensation/épimérisation	Aucune structure disponible
C _{starter}	xHHxxxDGxS	N-acylation	Aucune structure disponible
E	xHHxxxDGxS	Épimérisation	2XHG
Cy	xDxxxxDxxxS	Hétérocyclisation	5T7Z
C _T	xSHxxxDxxxS	Libération/macrocyclisation	5EJD
PS	xH [R / Q] xxxxDxxxS	Cyclisation de Pictet-Spengler	Aucune structure disponible
βL	HHHxxxDxxx	Macrocyclisation de β-lactames	Aucune structure disponible
C _{mod}	xxHxxxDGxS	Déshydratation	7R9X
X	xHRxxxDDDS	Recrutement d'enzymes	4TX2

Au sein des domaines C, les résidus His2, His3, Asp7, Gly8 et Ser10 du motif catalytique C3 sont conservés et assurent les réactions de condensation et/ou d'épimérisation. Les résidus His1 ou Asp2 indiquent une propension à la formation d'hétérocycles simples. D'autres résidus substitués sont susceptibles de modifier la fonction de domaines C alternatifs.

Les domaines d'hétérocyclisation (Cy ou HC) et de déshydratation (C_{mod}) sont des domaines C spécialisés évolutifs, et possèdent les motifs C3 conservés DxxxxD ou xHxxxD, respectivement. Les domaines Cy catalysent la formation de liaisons peptidiques et la cyclodéshydratation subséquente de résidus L-Cys ou L-Ser et L-Thr en des noyaux thiazoline ou oxazoline (Dowling *et al.*, 2016). Les domaines C_{mod} assurent la déshydratation de résidus L-Ser/L-Thr en des motifs déshydro-alanine ou déshydrobutyrine (Patteson *et al.*, 2022). Les structures de domaines E et Cy seront décrites dans la section en lien avec les domaines auxiliaires. De même, la structure d'une enzyme post-NRPS ayant une nouvelle activité de déshydratation sera caractérisée dans la section **Résultats**.

IV.4.B.α. N-acylation par le domaine $C_{starter}$

Les modules d'initiation de NRPS sont généralement dépourvus de domaines C, aucune réaction de condensation n'étant requise pour permettre le chargement du premier monomère. Une exception notable concerne les systèmes impliqués dans la biogenèse de surfactine et de daptomycine, comme cela a brièvement été évoqué à la section III.2.E. dans ce cas. Ces lignes d'assemblage comportent chacune un domaine $C_{starter}$ N-terminal, d'une classe distincte sur le plan phylogénétique, mais dont le motif catalytique est identique aux domaines de condensation canoniques (Rausch *et al.*, 2007). Ces domaines catalysent la conjugaison d'un acide gras activé au groupement thioester de l'aminocyl-PCP du module initial, réaction de N-acylation désignée sous le terme de lipoinitiation (Dekimpe et Masschelein, 2021) (**Figure 31**). Par exemple, le domaine $C_{starter}$ du module de chargement SrfAA utilise directement le précurseur libre activé par la coenzyme A et en assure l'incorporation avec un groupement β -hydroxyle dans la structure du tensioactif surfactine (Kraas *et al.*, 2010). En revanche, ces domaines requièrent plus fréquemment l'activité préalable d'acyl-ligases d'acides gras (FAAL) responsables de l'activation et du transfert des chaînes acyles sur une protéine porteuse (Dekimpe et Masschelein, 2021). Dans tous les cas, la partie lipidique accroît l'hydrophobicité des NRP, ce qui améliore leur capacité d'interaction avec les membranes et donc leur bioactivité (Kraas *et al.*, 2010).

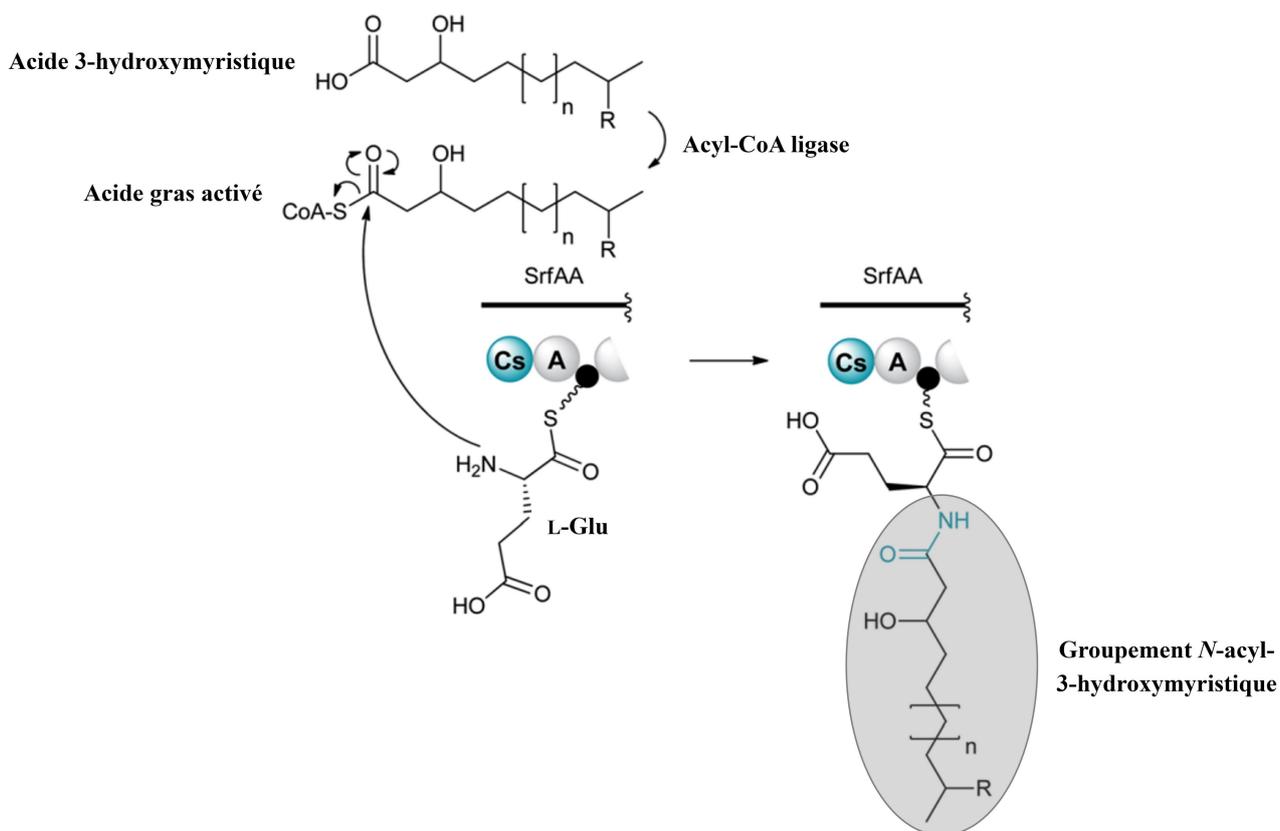


Figure 31. Lipoinitiation par le domaine $C_{starter}$ du module SrfAA (adaptée de Dekimpe et Masschelein, 2021). Une acyl-CoA ligase active l'acide 3-hydroxymyristique, puis le domaine $C_{starter}$ du module SrfAA initial le condense avec le glutamyl-PCP en un intermédiaire de la biosynthèse de la surfactine chez la bactérie *B. subtilis*. Cs : $C_{starter}$

IV.4.B.β. Formation de liaisons esters ou isopeptidiques par le domaine C

Certains domaines C catalysent la formation de liaisons esters, au lieu de liaisons amides, entre les substrats liés aux domaines PCP, ce qui forme alors des depsipeptides (Dekimpe et Masschelein, 2021). Dans les systèmes NRPS correspondants, les modules comportent un domaine cétoréductase (KR) supplémentaire impliqué dans la réduction stéréosélective de monomères α -cétoacyl-PCP en α -hydroxyacylthioester-PCP (Dekimpe et Masschelein, 2021). Ces intermédiaires deviennent ainsi les substrats des domaines spécialisés dans cette réaction d'estérification inhabituelle. Des exemples de tels modules se retrouvent dans les voies métaboliques d'anticancéreux (*e.g.* cryptophycine 1) et antifongiques (*e.g.* hectochlorine) potentiels, ou du piscicide antimycine produits par divers genres bactériens (Dekimpe et Masschelein, 2021) (**Figure 32**). Au terme de la biosynthèse du métabolite cryptophycine 1 chez la bactérie *Nostoc* sp. GSV224, le module terminal de la NRPS CrpD active et transfère l'acide α -cétoisocaproïque vers le domaine PCP accepteur, puis assure la formation d'une liaison ester-carboxylique avec le peptidyl-PCP adjacent (Ding *et al.*, 2011). De manière alternative, la fonction hydroxyle d'un résidu comme la sérine peut directement servir de nucléophile, tel est le cas de l'étape de dimérisation de l'échinomycine (voir section III.2.B.) (Felngle *et al.*, 2008).

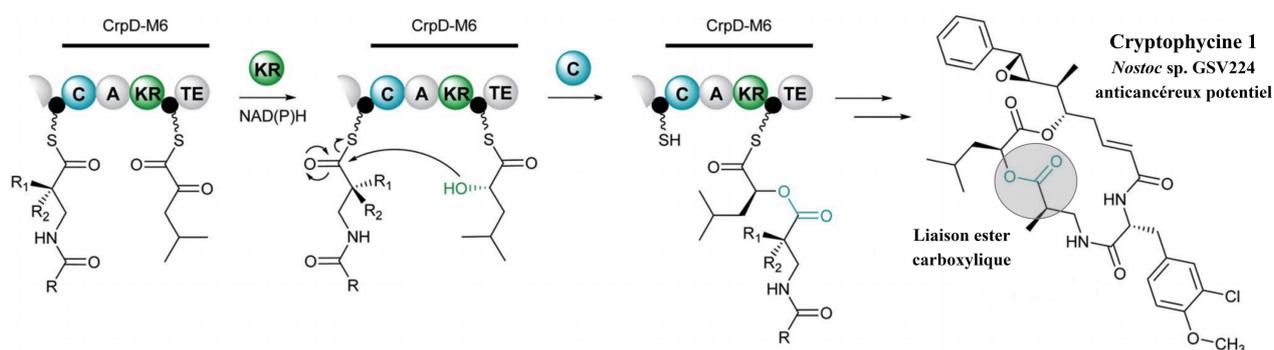


Figure 32. Formation d'une liaison ester par le domaine C6 du module terminal de la NRPS CrpD (adaptée de Dekimpe et Masschelein, 2021). Le domaine A du module terminal de la NRPS CrpD sélectionne et active l'acide α -cétoisocaproïque sur le domaine PCP6, puis le domaine KR le réduit et le domaine C exploite le groupement hydroxyle résultant afin de catalyser la formation d'une liaison ester dans la structure chimique de la cryptophycine. Les domaines C et KR sont représentés en bleu et vert, respectivement. A : domaine d'adénylation, KR : domaine cétoréductase, C : domaine de condensation, TE : domaine thioestérase, NADPH : nicotinamide adénine dinucléotide phosphate.

D'autres domaines C catalysent la formation de liaisons isopeptidiques entre le squelette et les chaînes latérales de résidus greffés aux domaines PCP (Dekimpe et Masschelein, 2021). Par exemple, lors de la biogenèse de toxines microcystines chez la bactérie *Microcystis aeruginosa*, les domaines A et A2 des modules McyE/McyB activent et thioestérifient les chaînes latérales des résidus D-Glu ou β -Me-D-Asp sur les domaines porteurs associés (Meyer *et al.*, 2017). Les modules suivants dans la chaîne NRPS génèrent chacun une liaison isopeptidique à partir de la fonction amino des résidus Me-Dha ou L-Arg du peptidyl-PCP accepteur correspondant (Meyer *et al.*, 2017) (**Figure 33A**). De même, au cours de la biosynthèse de l'érythrochéline chez la bactérie *Saccharopolyspora erythraea*, le domaine C3 de la NRPS EtcD catalyse la formation d'une liaison isopeptidique entre les fonctions δ -amino du résidu L-Orn et carbonyl-thioester du résidu attaché au peptidyl-PCP2 donneur (Lazos *et al.*, 2010) (**Figure 33B**). Néanmoins, dans ce cas, la chaîne latérale nucléophile n'est pas activée par le domaine A, contrairement à l'exemple précédent.

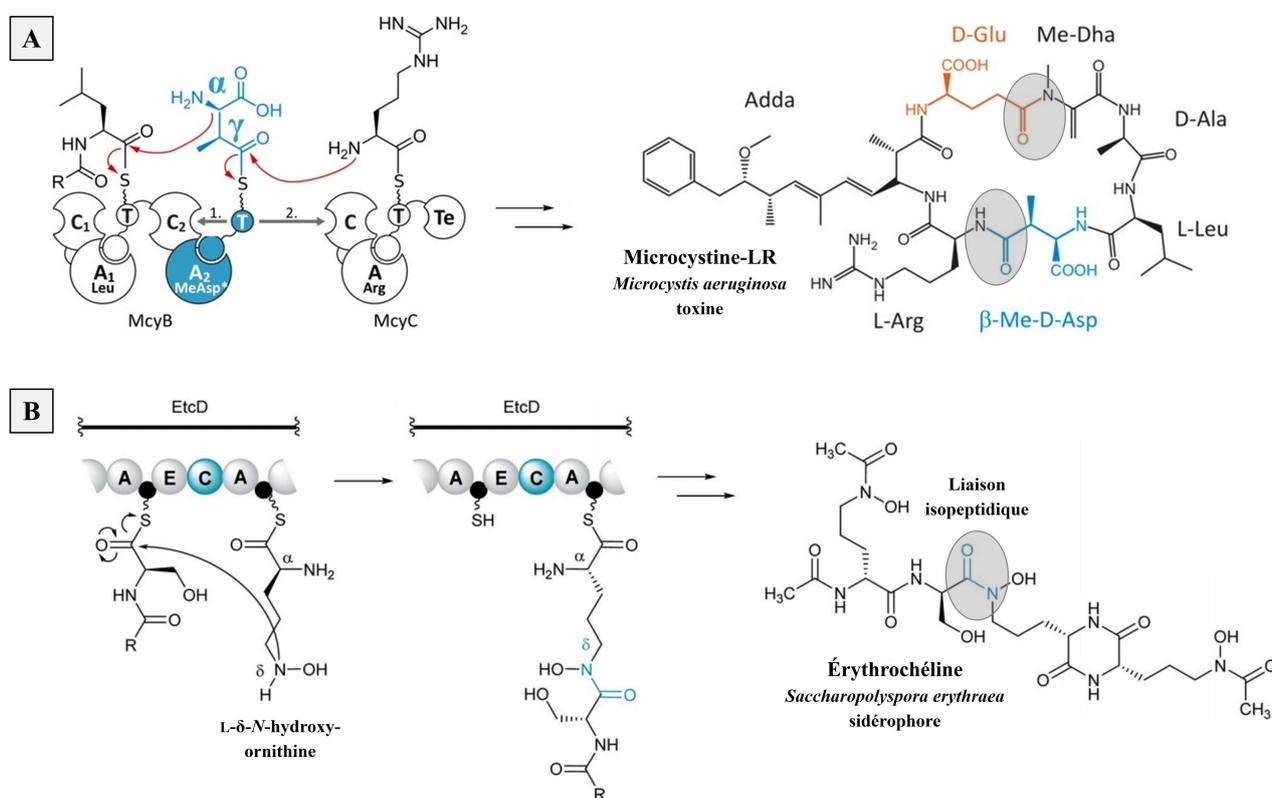


Figure 33. Formation de liaisons isopeptidiques par les domaines C de modules NRPS (adaptée de Meyer *et al.*, 2017 ; Dekimpe et Masschelein, 2021). Les domaines A des modules McyB (A) ou EtcD (B) activent et thioestérifient la chaîne latérale du résidu β -Me-D-Asp ou principale du résidu L- δ -N-hydroxyornithine, respectivement. Les domaines C du module suivant (A, McyC) ou correspondant (B, EtcD) catalysent ensuite la formation d'une liaison isopeptidique entre des groupements fonctionnels. A : domaine d'adénylation, C : domaine de condensation, T : domaine porteur de peptidyle, E : domaine d'épimérisation. Adda : acide 3-amino-9-méthoxy-2,6,8-triméthyl-10- phényldéca-4,6-diénoïque.

IV.4.B.γ. Formation d'un cycle β-lactame par le domaine C_{βL}

Les domaines C ont aussi été associés à des catalyses alternatives au sein des systèmes NRPS. Un exemple évocateur porte sur la formation d'un cycle β-lactame au cours de la biosynthèse de la norcardicine G chez la bactérie *Nocardia uniformis*, contrairement aux antibiotiques apparentés, où cette étape implique une isopénicilline N synthase (Gaudelli *et al.*, 2015) (**Figure 34**). En effet, le domaine C5 du module terminal de la sous-unité NocB possède un motif catalytique inhabituel avec un résidu histidine supplémentaire (His790) en amont de cette séquence (Gaudelli *et al.*, 2015). Des analyses mutationnelles et des essais enzymatiques *in vitro* ont révélé l'importance dudit résidu pour permettre la formation du noyau β-lactame (Gaudelli *et al.*, 2015). Un mécanisme en trois étapes a alors été proposé : la base His792 catalyse la réaction de β-élimination du groupement hydroxyle du résidu L-séryl attaché au domaine PCP4, ce qui génère un intermédiaire déhydroalanyl électrophile, puis le domaine A5 ajoute le monomère L-hydroxyphénylglycine, dont la fonction amine primaire attaque le centre sp² réactif (Gaudelli *et al.*, 2015). Lors de cette étape, le résidu His790 additionnel reprotrone l'espèce transitoire, ce qui forme une liaison amine secondaire réactive, avant que celle-ci n'attaque le groupement carbonyle-thioester du peptidyl-PCP4 et ne produise le métabolite final (Gaudelli *et al.*, 2015). Par ailleurs, cette suite de réactions particulières fait de ce mécanisme un moyen potentiel de piéger les domaines C en présence de leurs intermédiaires covalents (Payne *et al.*, 2017). En outre, un tel motif signature HHHxxxDG étendu de domaines C et un profil d'activation spécifique de précurseurs sérine ou thréonine pourraient être utilisés comme cibles de recherche afin d'identifier des NRPS impliquées dans l'assemblage de β-lactamines (Gaudelli *et al.*, 2015).

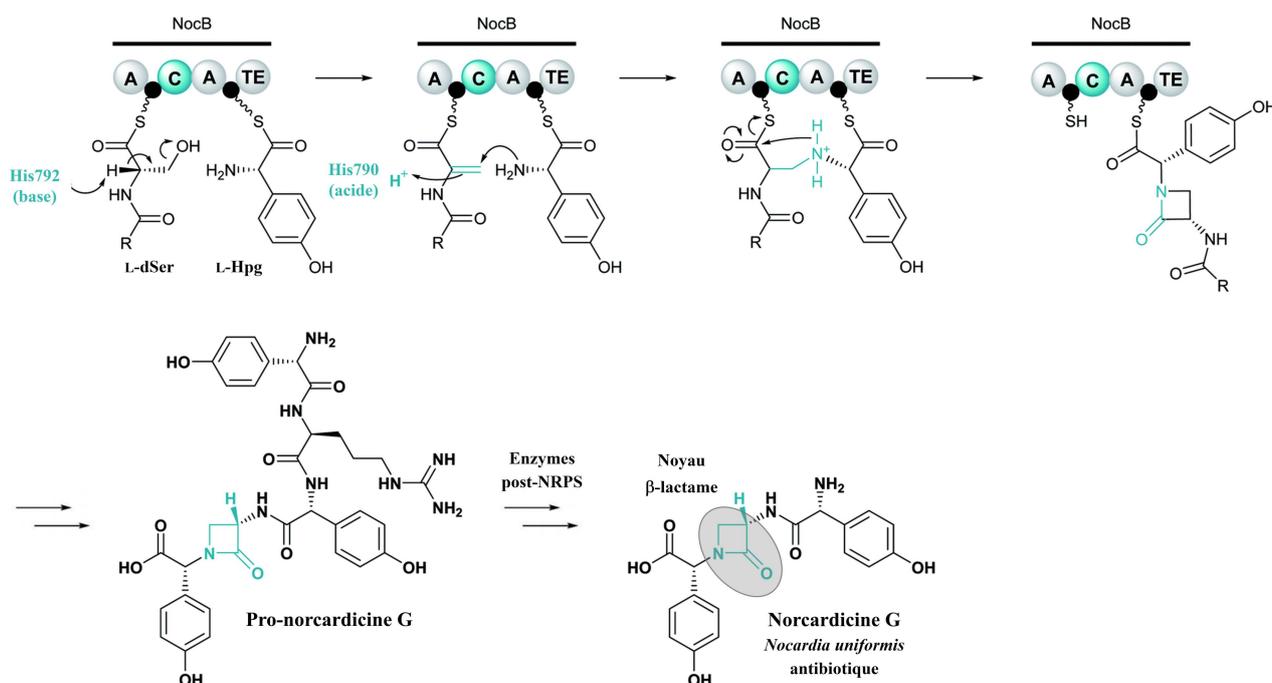


Figure 34. Déshydratation et cycloaddition par le domaine C de la NRPS NocB (adaptée de Dekimpe et Masschelein, 2021). Le domaine C5 du module terminal de la NRPS NocA–B catalyse la formation d'une liaison amide et du cycle β-lactame du précurseur de la norcardicine G. L-Hpg : hydroxyphénylglycine.

IV.4.B.δ. Cyclisation de Pictet-Spengler par le domaine PS

Les domaines C peuvent également catalyser une réaction de cyclisation de type Pictet-Spengler du squelette tétrahydroisoquinoléine de NRP, tel l'anticancéreux potentiel saframycine A produit du système SfmA–C chez la bactérie *Streptomyces lavendulae* (Koketsu *et al.*, 2010). En particulier, les motifs centraux C1, C3 et C5 montrent des modifications significatives dans ces domaines PS, par exemple, le résidu histidine catalytique du motif C3 est substitué par une arginine ou une glutamine (Koketsu *et al.*, 2010). Le mécanisme itératif implique le module terminal SfmC d'architecture PS-A-PCP-R où le domaine PS catalyse une première réaction, à partir d'une espèce aldéhyde libérée (9b) et d'un analogue de résidu tyrosine attaché au domaine porteur (2) (**Figure 35**). À ce stade, le domaine R réduit à nouveau l'intermédiaire formé en un composé aldéhyde tricyclique (7b), puis le domaine PS catalyse une deuxième réaction de condensation intramoléculaire et produit un peptide pentacyclique (Koketsu *et al.*, 2010). Le domaine terminal réducteur libère enfin celui-ci (10) de la sous unité (Koketsu *et al.*, 2010). Une cyclisation spontanée entre les groupements amine tertiaire et carbonyle génère un métabolite (5b) et des enzymes auxiliaires post-NRPS en assurent la maturation en saframycine A (Koketsu *et al.*, 2010). Cette biosynthèse singulière requiert en conséquence des étapes de recrutement d'intermédiaires carbonyles par le domaine PS et diverge donc du mécanisme fondamental des NRPS canoniques.

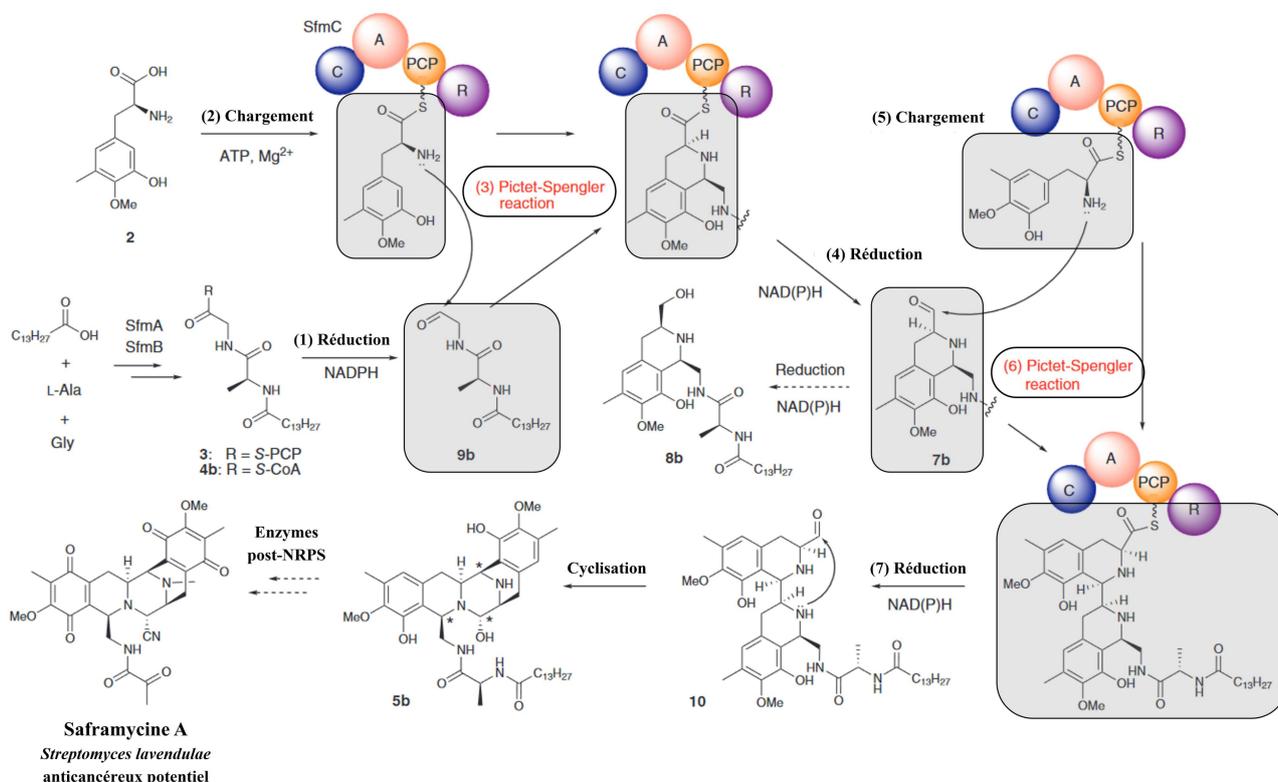


Figure 35. Cyclisation intramoléculaire de Pictet-Spengler par le domaine PS de la NRPS SfmC (adaptée de Koketsu *et al.*, 2010). Le domaine C de type Pictet-Spengler (PS) du module terminal SfmC catalyse deux réactions de cyclisation intramoléculaire sur des intermédiaires réduits, et forme ainsi le cycle isoquinoléine du produit saframycine A. A : domaine d'adénylation, PCP : domaine porteur de peptidyle, ATP : adénosine triphosphate, NADP : nicotinamide adénine dinucléotide phosphate.

IV.4.B.ε. Recrutement d'enzymes auxiliaires par le domaine X

Les modules de terminaison NRPS peuvent être pourvus d'un domaine C atypique, dit domaine X, dont un exemple démonstratif concerne les voies impliquées dans la biogenèse des antibiotiques glycopeptidiques, tels que la vancomycine ou la téicoplanine (Peschke *et al.*, 2016). Le domaine X possède un motif C3 au sein duquel le résidu histidine est remplacé par une arginine, ce qui obstrue le tunnel de liaison du site actif et empêche la catalyse (Haslinger *et al.*, 2015). Des expériences de modification génétique *in vivo* ont démontré que ce domaine inactif assure le recrutement séquentiel de cytochromes P450 (CYP450) vers le domaine PCP adjacent, et que ces enzymes, connues comme OxyB, E, A, C, catalysent des réactions de réticulations oxydatives du substrat heptapeptidique lié à la matrice (Peschke *et al.*, 2016) (**Figure 36**). Ces modifications sont introduites entre les chaînes latérales de résidus aromatiques du précurseur linéaire, ce qui octroie ainsi une importante rigidité à la structure d'aglycone essentielle à l'activité des produits (Peschke *et al.*, 2016). Des expériences cinétiques ont révélé que ces CYP450 sont en compétition pour accéder transitoirement à l'unique surface d'interaction du domaine X, et capables de reconnaître l'état de cyclisation de l'intermédiaire peptidyl-PCP en vue de leur action en *trans* à tour de rôle (Peschke *et al.*, 2016). Le changement conformationnel des complexes résultants en améliore l'affinité pour le substrat et initie la catalyse, ce qui fait du domaine X une plateforme intéressante afin de procéder à la modification chimique ponctuelle de peptides non-ribosomiques (Peschke *et al.*, 2016).

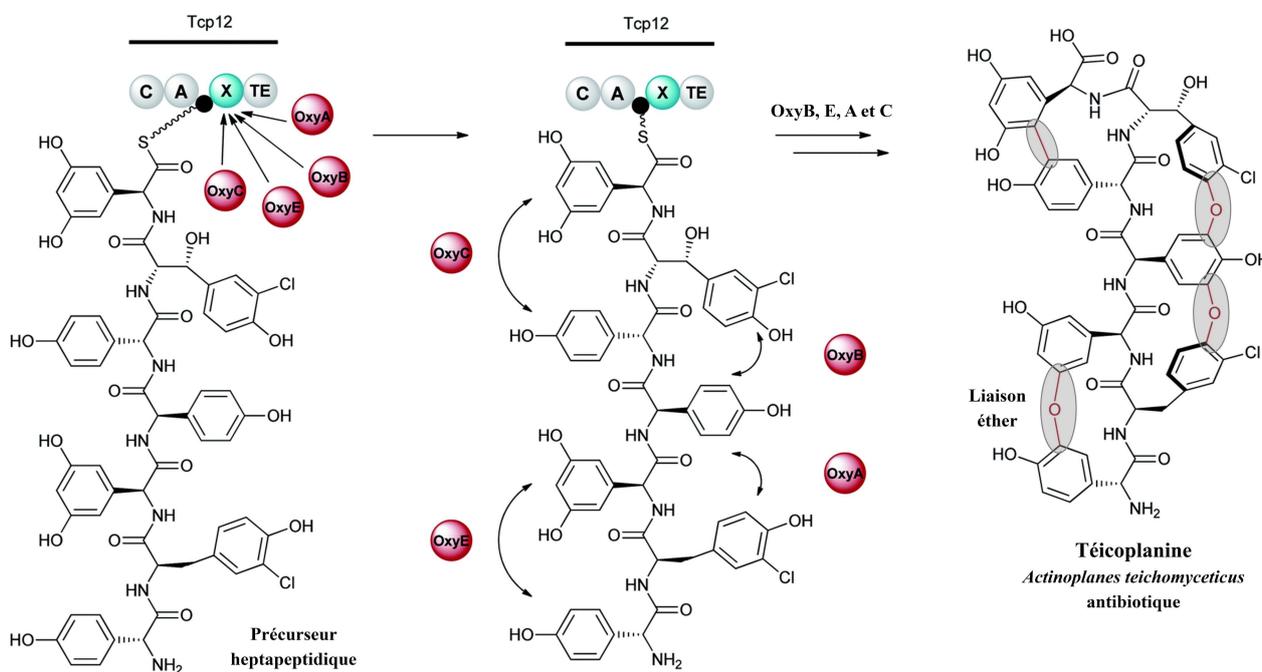


Figure 36. Cyclisation intramoléculaire par des CYP450 au sein du système NRPS Tcp9–12 (adaptée de Dekimpe et Masschelein, 2021). Le domaine X du module terminal de la sous-unité Tcp12 possède une activité *trans*-catalytique. En effet, il recrute quatre cytochromes P450 distincts, annotés OxyB, E, A, et C, responsables de l'oxydation des chaînes aromatiques du précurseur heptapeptidique lié au domaine PCP7, ce qui conduit à la tétracyclisation intramoléculaire en l'antibiotique téicoplanine. Le domaine X et les enzymes CYP450 sont représentées en cyan et magenta, respectivement. A : domaine d'adénylation, C : domaine de condensation, TE : domaine thioestérase.

IV.4.B.ζ. Libération de chaîne par le domaine C_T terminal

Au niveau du module terminal, certains domaines C_T médient la réaction de libération du produit biosynthétisé et supplantent alors le domaine thioestérase responsable de cette étape dans la plupart des lignes d'assemblage NRPS (Dekimpe et Masschelein, 2021) (**Figure 37**). Ces réactions incluent la formation de macrolactames et macrolactones, des cyclisations régiosélectives, des condensations intermoléculaires ou l'hydrolyse classique (Dekimpe et Masschelein, 2021).

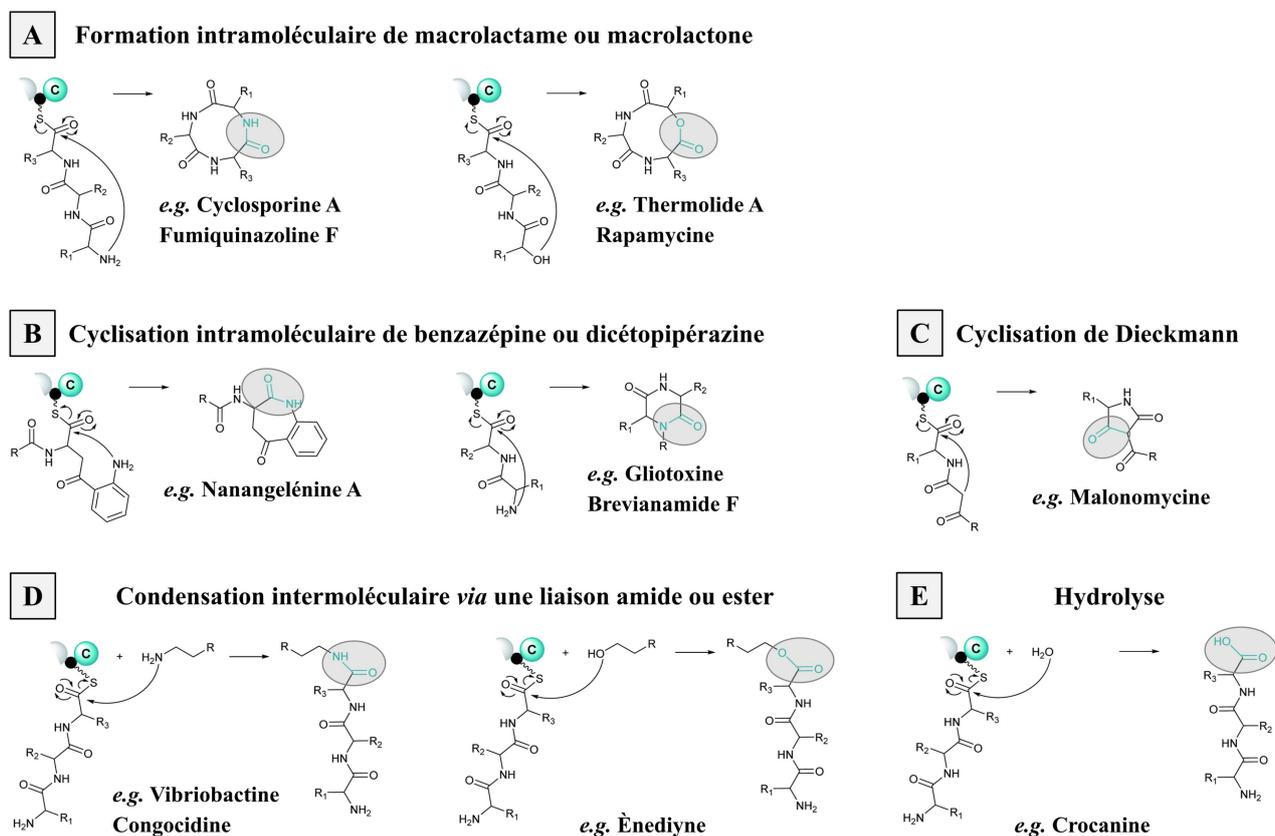


Figure 37. Libération de NRP par le domaine C_T terminal de systèmes NRPS (adaptée de Dekimpe et Masschelein, 2021). Le domaine C_T terminal assure des réactions de macrolactamisation (A, cyclosporine A) et de macrolactonisation (A, thermolide A), de cyclisations intramoléculaires régiosélectives de benzazépine-dione (B, nanangelénine A) de dicétopipérazine (B, gliotoxine), la cyclisation de Dieckmann (C, malonomycine), des condensation intermoléculaires *via* des liaisons amide (D, vibriobactine) ou ester (D, ènediïne), ou encore une simple hydrolyse (E, crocanine). Les domaines PCP et C sont représentés en noir et vert, respectivement.

Les domaines C présentent souvent de subtiles altérations au niveau du site actif et les fonctions résultantes constituent un point essentiel pour produire des structures peptidiques à haute complexité par ingénierie biosynthétique. Dans ce contexte, les complexes des domaines en interaction gagnent à être étudiés, notamment les régions *linkers* inter-domaines spécifiques au sein des NRPS.

IV.5. Communication et interaction entre les sous-unités par les domaines de *docking*

Les modules NRPS sont généralement distribués sur trois chaînes polypeptidiques, par exemple A, B, et C (systèmes de type I), et ces sous-unités sont ordonnées pour garantir la directionnalité de la chaîne d'assemblage et la fidélité de la biosynthèse (Smith *et al.*, 2021). Les domaines de *docking* sont de courtes régions de 20–30 résidus présentes aux extrémités N- et C-terminales des sous-unités multimodulaires, et jouent un rôle crucial dans les interactions intermoléculaires et la translocation du substrat au niveau des interfaces (Smith *et al.*, 2021) (**Figure 38**). Ces domaines d'ancrage, dits NDD et CDD , s'organisent en paires complémentaires capables d'assurer des interactions spécifiques de l'ordre du micromolaire (Kosol *et al.*, 2019). Pour cette raison, les *DD* ont été exploités dans des systèmes NRPS/PKS pour étendre le potentiel combinatoire de plateformes artificielles d'ingénierie, comme cela sera décrit ultérieurement dans cette section.

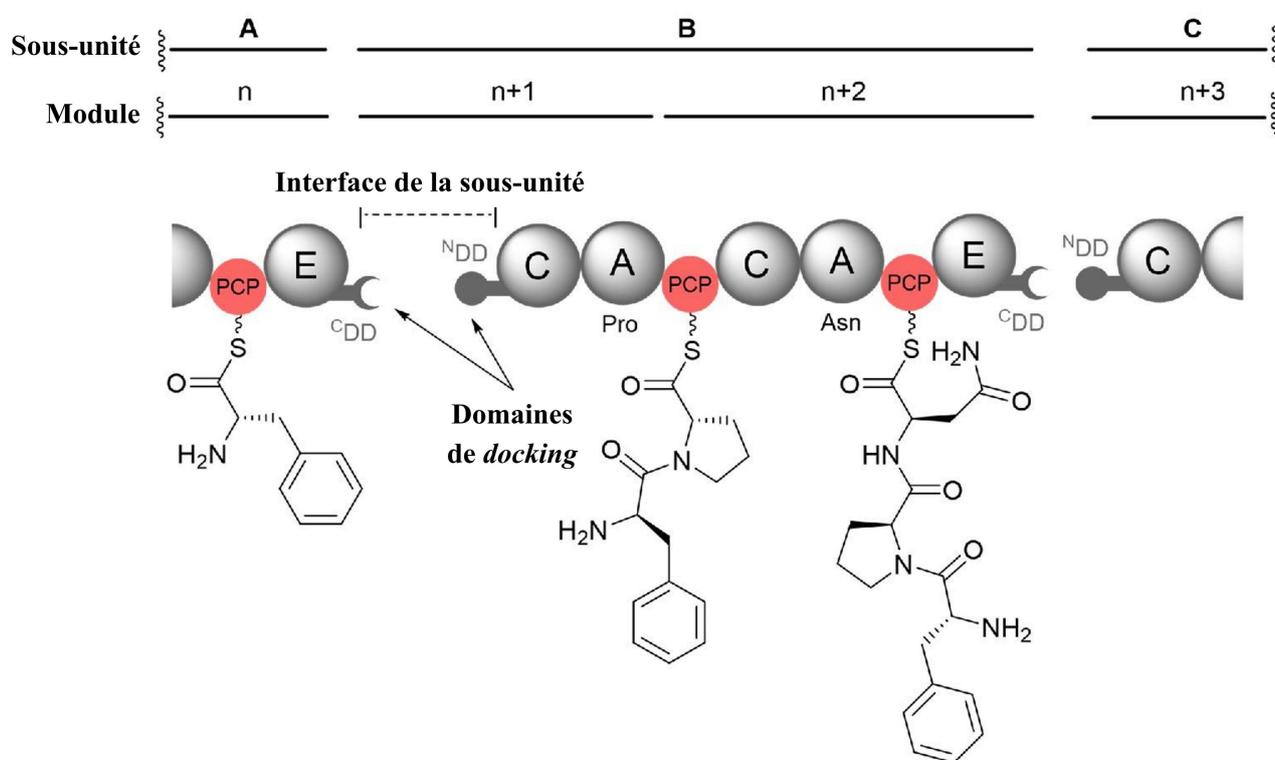


Figure 38. Organisation des domaines de *docking* au sein d'une NRPS hypothétique (adaptée de Smith *et al.*, 2021). Les domaines de *docking* situés aux extrémités C- et N-terminales des domaines E (module n) et C (module n+1), notés CDD et NDD , respectivement, assurent l'interaction entre les sous-unités A et B au niveau de l'interface, et contribuent également à la condensation du peptidyl-PCP. A : domaine d'adénylation, C : domaine de condensation, PCP : domaine porteur de peptidyle, E : domaine d'épimérisation, *DD* : domaine de *docking*.

Au cours des dernières années, les recherches ont été appuyées afin d'élucider les structures des domaines de *docking* et les mécanismes moléculaires sous-jacents à leurs fonctions biologiques. Les *DD* se regroupent en différentes classes définies : les domaines *COM* (*CO*munication-*ME*diating), les domaines β *hD* et les domaines *PAX* (*Peptide-Antimicrobial-Xenorhabdus*) (Hahn et Stachelhaus, 2004 ; Richter *et al.*, 2008 ; Dowling *et al.*, 2016 ; Watzel *et al.*, 2020).

IV.5.A. Communication intermodulaire par les domaines *COM*

Des alignements de séquences des extrémités N- et C- terminales des sous-unités TycA–C, ainsi que des analyses mutationnelles, ont initialement permis d'identifier de courts segments de 20 à 30 résidus importants dans la médiation de la reconnaissance interdomaine (Hahn et Stachelhaus, 2004). Des expériences d'échanges de ces domaines *COM* entre les modules donneurs TycA et TycB3, et accepteurs TycB1 et TycC1, ont mis en évidence que des paires correspondantes sont nécessaires et suffisantes pour établir la communication entre les NRPS partenaires (Hahn et Stachelhaus, 2004).

Les structures cristallographiques des domaines *COM* donneur (C-A-PCP-E-*COM*^D) et accepteur (*COM*^A-C-A-PCP-TE) entre les sous-unités SrfAB–C ont chacune été déterminées indirectement au sein des modules correspondants (Tanovic *et al.*, 2008). Le domaine *COM*^A, localisé à l'extrémité N-terminale du domaine C7, adopte une structure formée d'une hélice α d'un tour et de trois brins β non-contigus, dont le motif dit *hand-shaped* accueille le domaine partenaire *COM*^D (du domaine E6) (Tanovic *et al.*, 2008) (**Figure 39**). Des travaux de photoréticulation sous irradiation UV en présence de p-benzoylphénylalanine ont mis en lumière que les domaines *COM*^D du module GrsA et *COM*^A du module TycB1 présentent un repliement comparable (Dehling *et al.*, 2016). Des expériences de spectrométrie de masse complémentaires ont quant à elles suggéré que certains résidus du domaine C situé en aval font partie de l'interface protéique, contrairement au postulat initial.

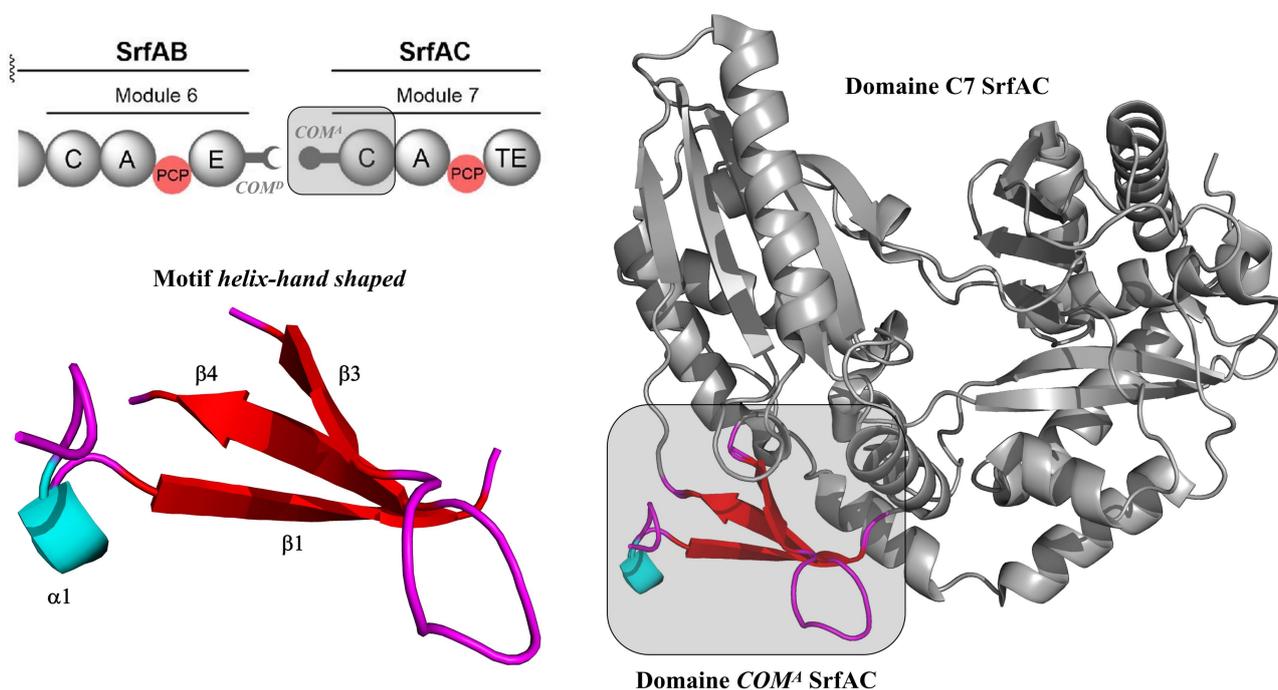


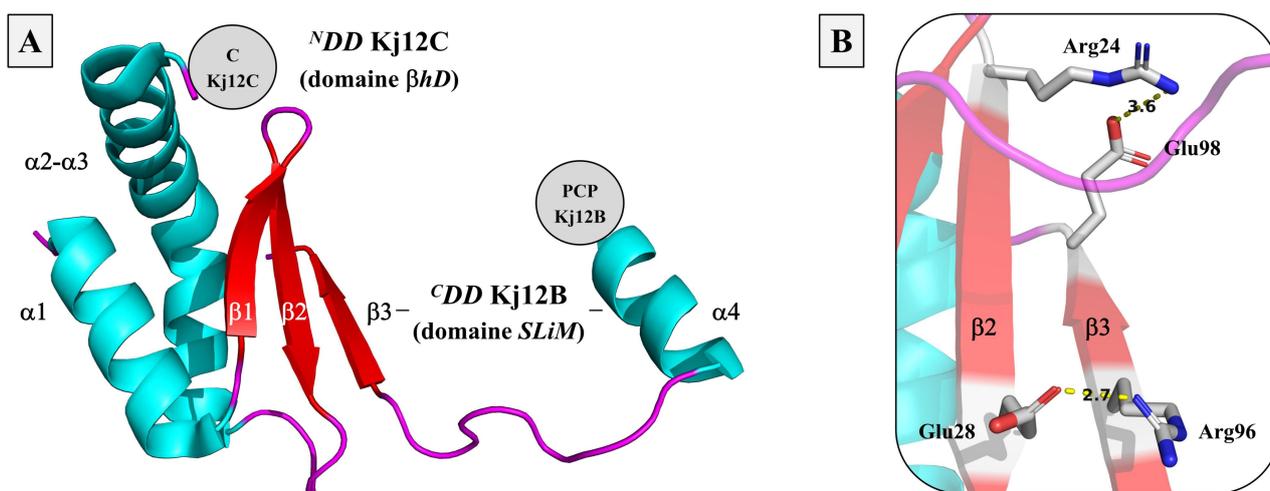
Figure 39. Représentation de la structure cristallographique du didomaine *COM*^A-C7 de la NRPS SrfAC. Le domaine *COM*^A, localisé à l'extrémité N-terminale du domaine C7 de la sous-unité SrfAC (PDB : 2VSQ), se structure en une hélice $\alpha 1$ et trois brins β non-consécutifs ($\beta 1$, $\beta 3$ et $\beta 4$). Ce domaine assure la communication avec l'hélice α du domaine *COM*^D situé à l'extrémité C-terminale du domaine E6 de la sous-unité SrfAB *via* des interactions hydrophobes (Tanovic *et al.*, 2008). *COM* : domaine *communication-mediating*.

Des expériences d'ingénierie biosynthétique ont exploité la promiscuité inhérente aux domaines COM compatibles pour induire des interactions productives entre des NRPS distinctes et également générer des dérivés originaux (Hahn et Stachelhaus, 2006). Par exemple, le complexe artificiel des modules TycA, BacB2 et SrfAC, assemblés dans cet ordre à partir des domaines *COM* adéquats, a conduit à la formation du tripeptide L-Phe-D-Orn-L-Leu (Hahn et Stachelhaus, 2006).

Toutefois, les domaines *COM* ne sont pas indépendants du domaine catalytique associé, vu que des structures secondaires proviennent des domaines E/C adjacents, alors cela soulève des questions quant aux possibilités d'utiliser directement ces régions dans l'ingénierie des voies NRPS.

IV.5.B. Interactions intermoléculaires par les domaines βhD et *PAX*

Des alignements de séquences de NRPS et systèmes hybrides ont permis d'identifier une famille d'éléments de liaison N- et C- terminaux aux interfaces des domaines PCP/C ou PCP/Cy (Richter *et al.*, 2008 ; Dowling *et al.*, 2016 ; Hacker *et al.*, 2018 ; Kosol *et al.*, 2019). Des études structurales ont montré que les domaines ^NDD, dits βhD , adoptent un repliement conservé de type $\alpha\beta\beta\alpha$, avec un motif en ruban β , et interagissent avec des ^CDD *SLiM* (*Short-Linear Motif*) en partie désordonnés (Hacker *et al.*, 2018 ; Kosol *et al.*, 2019). Par exemple, la structure résolue par RMN du complexe non-covalent ^CDD-^NDD (Kj12A-B) des sous-unités NRPS a révélé que le domaine βhD interagit uniquement avec cinq résidus (LRGEI) en brin β du motif *SLiM* (Hacker *et al.*, 2018) (**Figure 40A**). L'interface conservée se trouve entre les brins β des domaines de *docking*, et dépend d'interactions hydrophobes et de ponts salins (Hacker *et al.*, 2018) (**Figure 40B**). Les résidus chargés à la surface du motif β -*hairpin* déterminent la cinétique de l'association et conduisent ainsi à la liaison entre les partenaires selon un « code de *docking* » (Hacker *et al.*, 2018) (**Figure 41**).



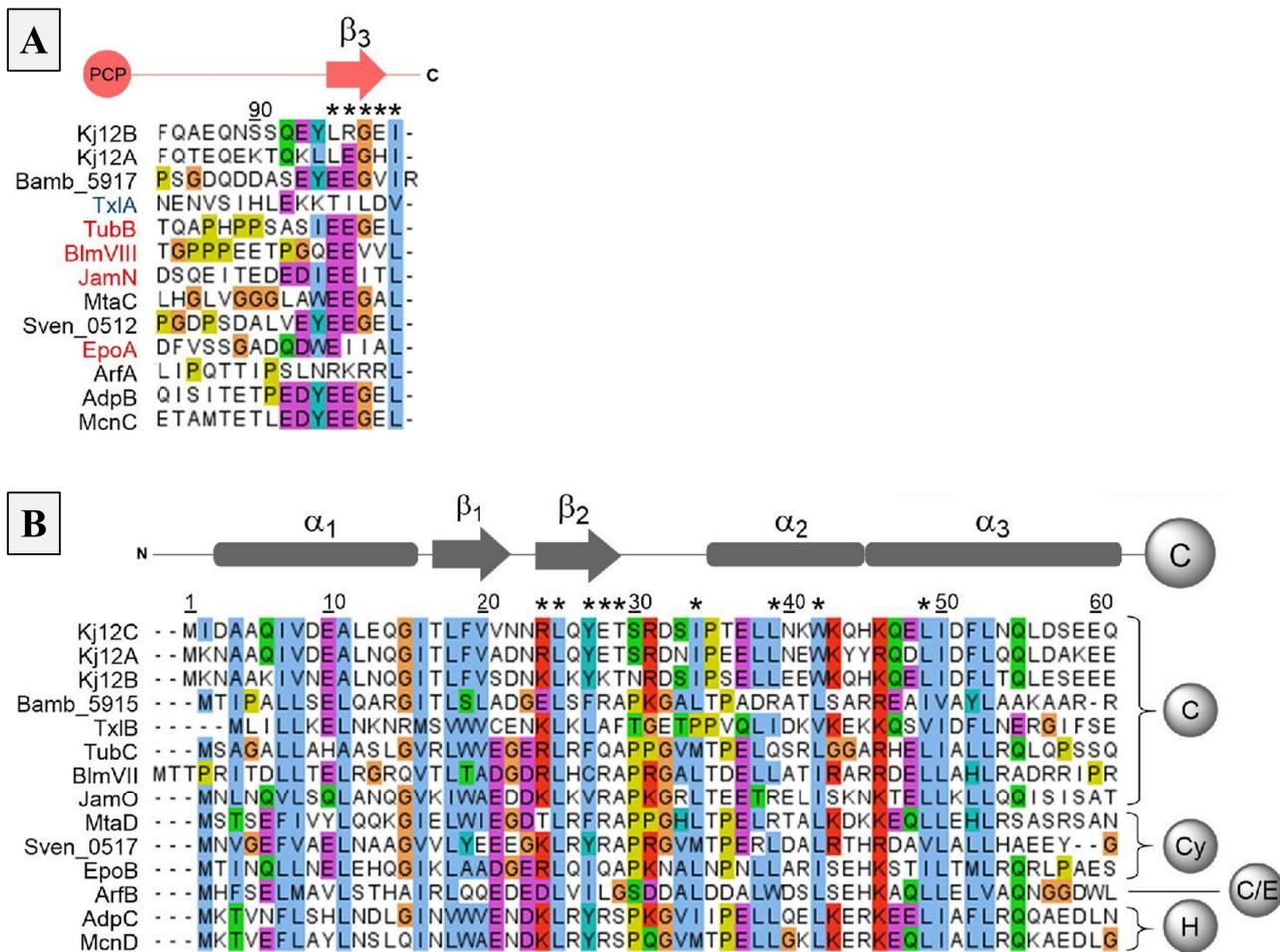


Figure 41. Caractérisation du code de docking des domaines *SLiM-βhD* (adaptée de Smith *et al.*, 2021). (A) La signature spécifique des domaines ^CDD *SLiM* (A) et ^NDD *βhD* (B) de NRPS (séquences en noir) ou PKS (séquences en rouge) est indiquée. Les résidus impliqués dans l'interaction au niveau des interfaces des complexes non-covalents sont mis en évidence avec des astérisques.

Les domaines *βhD* se connectent au domaine catalytique localisé en aval *via* des régions *linkers* flexibles d'environ 20 résidus (Dowling *et al.*, 2016 ; Kosol *et al.*, 2019). Par exemple, la structure cristallographique du domaine d'hétérocyclisation (Cy) EpoB expose le domaine ^NDD associé dans 3 conformations distinctes (Dowling *et al.*, 2016). Compte tenu de cela, les domaines *βhD* mobiles pourraient faciliter l'interaction avec le partenaire *SLiM* correspondant et la translocation du substrat au sein du site actif du domaine C adjacent (Kosol *et al.*, 2019 ; Smith *et al.*, 2021).

Des expériences d'ingénierie ont démontré que les ^CDD *SLiM* peuvent recruter divers domaines catalytiques à condition de les fusionner à des ^NDD *βhD* compatibles (Hacker *et al.*, 2018 ; Smith *et al.*, 2021). Cela indique un certain niveau de promiscuité des interfaces des domaines associés entre eux. Par exemple, chaque paire *SLiM-βhD* des sous-unités NRPS K_j12A–C permet une interaction productive, excepté une combinaison non-native parmi les 6 possibles testées (Hacker *et al.*, 2018). De cette manière, les systèmes NRPS reconstitués conduisent à la production de métabolites dérivés.

Dans tous les cas, cette classe de *DD* a le potentiel d'être exploitée dans l'ingénierie de NRPS artificielles, tel que cela a notamment déjà été entrepris avec succès au sein des voies de biosynthèse des rhabdopeptides et xénortides, ou xéfoampeptides (Cai *et al.*, 2019 ; Kegler et Bode, 2020).

L'analyse de la jonction des domaines PCP/C entre les sous-unités de la chaîne NRPS PaxA–C de la bactérie *Xenorhabus bovienii* a permis de découvrir une nouvelle classe (PAX) de domaines de *docking* (Watzel *et al.*, 2020). Les expériences de titration calorimétrique isotherme et de RMN des ^CDD (PaxB)–^NDD (PaxC) ont révélé que les domaines présentent des conformations α-hélicoïdales transitoires et que le complexe résultant se structure en un faisceau de trois hélices α avec une fente centrale (Watzel *et al.*, 2020) (**Figure 42A**). De plus, le complexe a une constante de dissociation $K_D = 25 \pm 8 \mu\text{M}$, ce qui indique que l'interaction est d'une force moyenne et est réversible (Watzel *et al.*, 2020). Des études de mutagenèse ont quant à elles défini les résidus impliqués dans la liaison entre ces domaines (Watzel *et al.*, 2020). En l'occurrence, les résidus Leu8, Leu11, Val15 et Leu16 du domaine ^NDD (C) établissent des contacts hydrophobes avec les résidus Val72, Leu76, Leu84 et Leu88 du domaine ^CDD (PCP), alors que des ponts salins (Lys19/Glu75 et Arg23/Asp68) confèrent une spécificité à l'interface (Watzel *et al.*, 2020) (**Figure 42B**). Par ailleurs, des analyses ont prédit d'autres exemples de cette classe de domaines de *docking* PAX liés à des domaines C, PCP et OX, ce qui montre leur utilité potentielle dans l'ingénierie de systèmes NRPS (Watzel *et al.*, 2020).

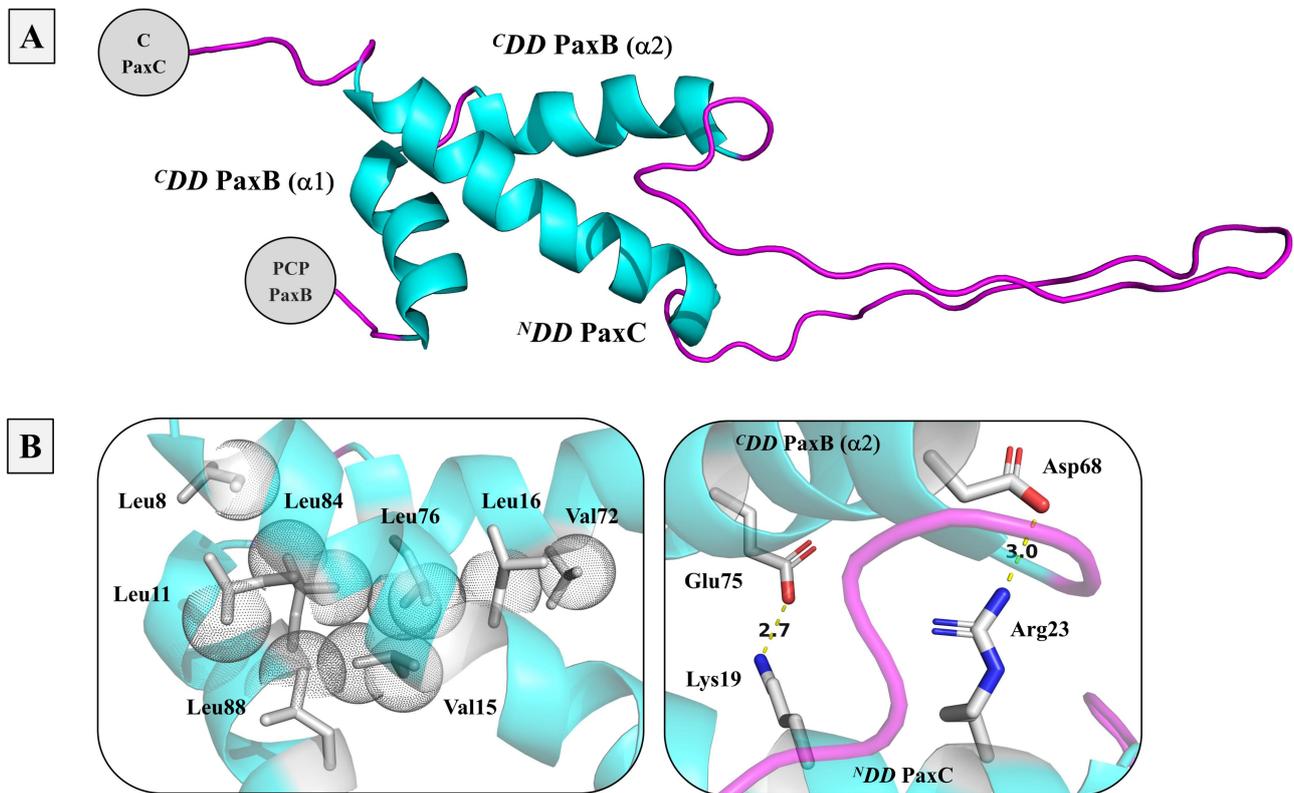


Figure 42. Représentation de la structure RMN du complexe des domaines de *docking* PaxB–C. (A) Les domaines de *docking* situés à l’interface des domaines PCP/C des sous-unités PxaB et PxaC forment un faisceau de trois hélices α (PDB : 6TRX), dont l’une (*NDD*) s’enfâsse au sein des deux autres (*CDD*). (B) Ces domaines interagissent par le biais d’interactions hydrophobes (cercles) et de ponts salins (traits pointillés). *DD* : domaine de *docking*.

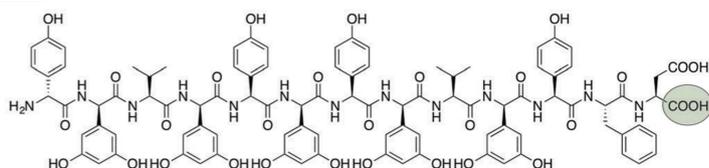
Actuellement, assez peu de données structurales détaillent les spécificités des différents types de complexes non-covalents, ni même la sélectivité des domaines accepteurs, ce qui fait de l’ingénierie des systèmes NRPS à travers ces domaines de *docking* un défi ardu avec un succès modeste, bien que des récentes avancées ouvrent la voie à de remarquables réussites (Cai *et al.*, 2019 ; Kegler et Bode, 2020). Des travaux réalisés en ce sens ont conduit au développement de nouveaux outils, tels qu’un ensemble de paires de bZIP (*basic-region leucine zipper*) synthétiques (*i.e.* SYNZIP), dont la structure en superhélice permet de médier des interactions de haute affinité au niveau des interfaces de modules NRPS ou PKS (Thompson *et al.*, 2012 ; Bozhüyük *et al.*, 2021).

IV.6. Fonctionnalisation terminale du NRP par le domaine TE

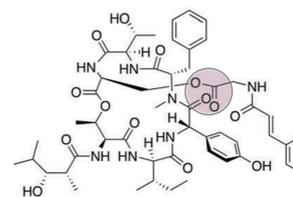
IV.6.A. Analyse fonctionnelle

Au sein des systèmes NRPS, deux types de domaines thioestérase (TE) peuvent être retrouvés, dont les domaines de type I responsables de la libération du produit nouvellement synthétisé, ce qui garantit l'intégrité du processus de biosynthèse (Izoré et Cryle, 2018). Des domaines TE intégratifs ou des enzymes autonomes de type II régénèrent quant à eux des domaines PCP actifs *via* le clivage des substrats inappropriés, ce qui rend à nouveau disponibles les domaines avec des fonctions thiols libres sur leur cofacteur 4'-Ppant (Kotowska et Pawlik, 2014). Par exemple, deux domaines TE en tandem, une disposition singulière de la chaîne d'assemblage de la lysobactine, assurent à la fois la macrocyclisation et la libération du produit (TE1), ainsi que la régénération du domaine porteur par hydrolyse du peptidyl-thioester (TE2) (Hou *et al.*, 2011).

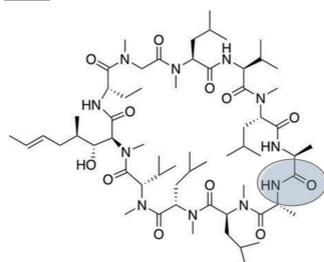
Les domaines thioestérase appartiennent à la super-famille des α/β hydrolases et possèdent une triade catalytique Ser-His-Asp conservée avec un résidu nucléophile (sérine) (Izoré et Cryle, 2018). Les domaines TE de type I catalysent la libération du produit de synthèse par une réaction en deux étapes : premièrement, le transfert du groupement peptidyle du PCP donneur sur le résidu catalytique activé, ce qui forme un intermédiaire *O*-acyl-enzyme (Izoré et Cryle, 2018). Ces domaines assurent ensuite l'hydrolyse, l'acylation, ou la cyclisation intra-/intermoléculaire *via* l'attaque nucléophile sur la liaison ester du NRP (Izoré et Cryle, 2018) (**Figure 43A–C**). Dans le cas des macrocyclisations, uniquement, les domaines TE présentent une sélectivité relative aux résidus liés à proximité du site actif, à la longueur du produit biosynthétisé, ainsi qu'aux structures secondaires formées (Kohli *et al.*, 2001). Par exemple, la réaction se produit par l'intermédiaire de la formation de lactames avec la fonction *N*-amino du premier résidu du NRP (*e.g.* tyrocidine A), d'un résidu de la chaîne linéaire (*e.g.* bacitracine) ou de l'atome de carbone β d'un lipide (*e.g.* bacillomycine) (Felnagle *et al.*, 2008). De manière comparable, la formation de lactones est également possible avec la fonction hydroxyle d'une chaîne latérale d'un résidu (*e.g.* daptomycine), d'un lipide (*e.g.* surfactine) ou d'un polycétide (produit hybride) (*e.g.* épothilone) (Felnagle *et al.*, 2008). Cette flexibilité de substrat a été exploitée pour générer des banques de NRP cycliques par des approches chémoenzymatiques (Heberlig *et al.*, 2018 ; Kobayashi *et al.*, 2023). Par contre, les domaines responsables des réactions d'hydrolyse des peptides linéaires n'ont jusqu'à présent pas été définis par une telle promiscuité (Payne *et al.*, 2017).

A Rupture intermoléculaire d'une liaison ester (hydrolyse)

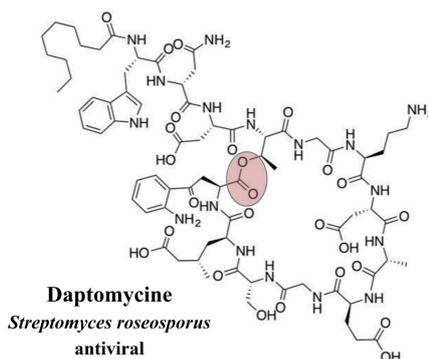
Feglymycine
Streptomyces sp. DSM 11171
antibiotique potentiel

B Acylation intermoléculaire

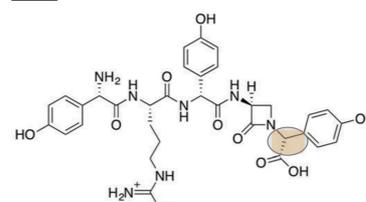
Déméthylsalinamide C
Streptomyces albus
antibiotique potentiel

C Formation intramoléculaire de macrolactame ou macrolactone

Cyclosporine A
Tolypocladium inflatum
immunosuppresseur



Daptomycine
Streptomyces roseosporus
antiviral

D Épimérisation

Norcardicine G
Nocardia uniformis
antibiotique

Figure 43. Libération de NRP par le domaine TE de systèmes NRPS (adaptée de Payne *et al.*, 2017). Le domaine TE chargé assure de simples réactions d'hydrolyse de NRP linéaires (A, feglymycine), d'acylation intermoléculaire (B, déméthylsalinamide C), de cyclisation intramoléculaire à partir de groupements amine ou alcool (C, cyclosporine A et daptomycine), ainsi que d'épimérisation d'un monomère (D, norcardicine G).

Par ailleurs, un autre exemple de cette polyvalence concerne la NRPS Sln, où les domaines TE assurent deux modifications consécutives : une cyclisation intramoléculaire (module Sln6), puis une transestérification intermoléculaire d'une unité (4-méthylhexa-2,4-diényl)-glycine (module Sln9), ce qui produit la déméthylsalinamide C à activité antibiotique (Ray *et al.*, 2016) (**Figure 43B**).

En outre, les domaines TE présentent une tolérance vis-à-vis du substrat chargé, mais certains s'avèrent plus stricts (Gaudelli *et al.*, 2015). Le domaine thioestérase de la NRPS NocAB requiert en effet la liaison *sine qua non* d'un substrat avec noyau β -lactame, et assure en plus l'épimérisation du résidu L-Hpg terminal, avant la libération du produit norcardicine G (Gaudelli et Townsend, 2014 ; Gaudelli *et al.*, 2015) (**Figure 43D**).

IV.6.B. Analyse structurale

Les approches structurales appliquées ont permis d'obtenir des données à haute résolution sur les interactions entre les domaines thioestérase et leurs substrats en fin du processus de biosynthèse. La structure cristallographique du domaine TE du module SrfAC a été résolue dans des conformations alternatives au sein de l'unité asymétrique (Bruner *et al.*, 2002) (**Figure 44**). En effet, ce domaine adopte fondamentalement un état fermé par l'intermédiaire d'hélices 3_{10} désordonnées qui, lors de l'interaction avec les hélices $\alpha 2$ et $\alpha 3$ du domaine PCP adjacent, sont sujettes à un repliement induit, ce qui rend accessible le site actif à un substrat conforme (*i.e.* acylpeptide) (Bruner *et al.*, 2002). La cavité hydrophobe accueille ce ligand et tolère son repliement en vue de la réaction de cyclisation en un macrolactone, la surfactine, aux propriétés tensioactives (Bruner *et al.*, 2002).

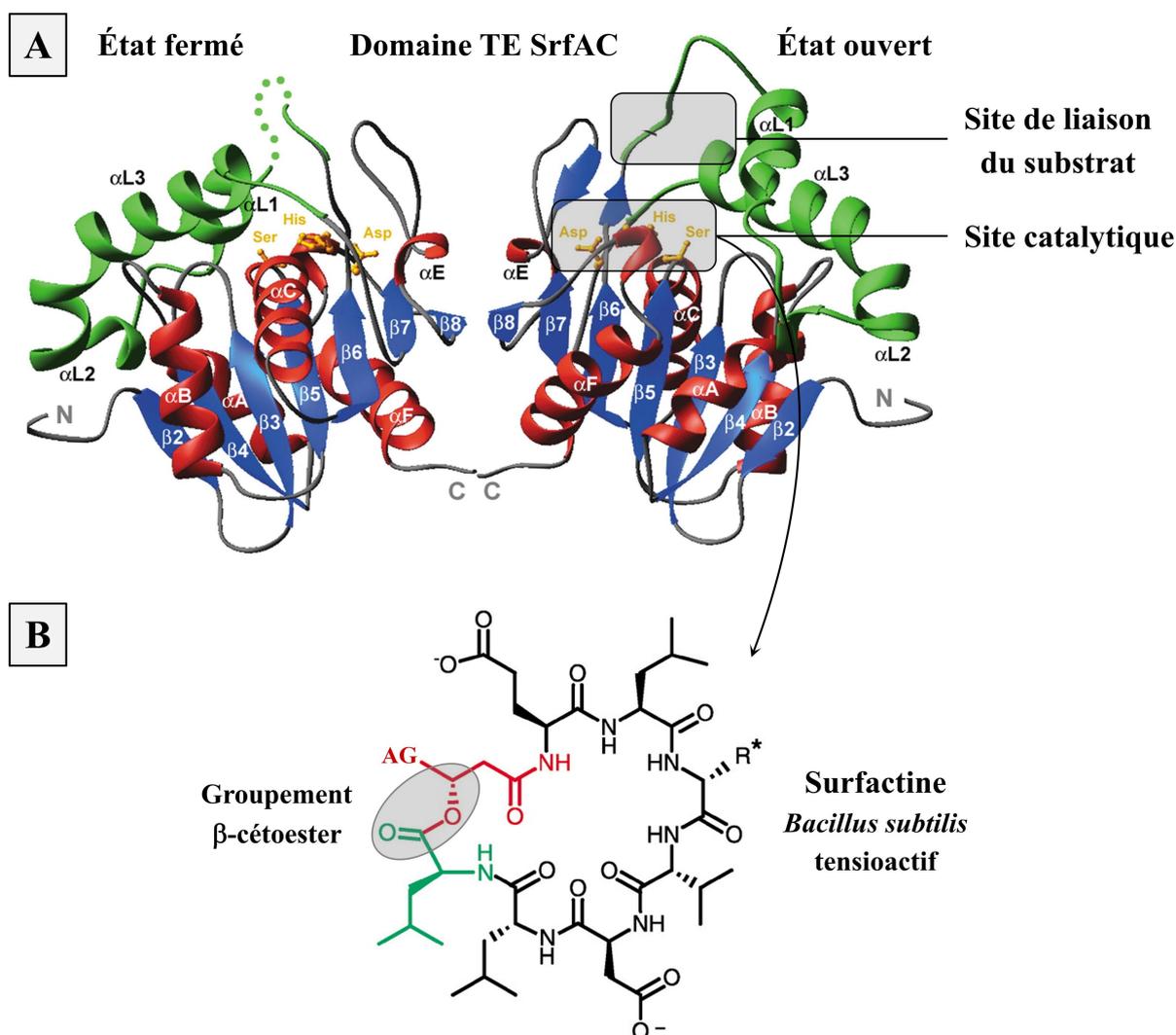


Figure 44. Représentation de la structure cristallographique du domaine TE de la NRPS SrfAC (adaptée de Bruner *et al.*, 2002). (A) Le domaine TE de la sous-unité SrfAC (PDB : 1JMK) est présent dans l'unité asymétrique en des états fermé et ouvert. Il adopte le repliement commun des α/β hydrolases, formé d'un feuillet β à sept brins flanqués d'hélices α , mais ne possède pas le brin N-terminal ($\beta 1$), ni l'hélice αD , mais trois hélices 3_{10} , dont l'interaction avec le substrat promeut la réaction de cyclisation intramoléculaire. Le site actif contient les résidus conservés Ser80, His207, et Asp107. (B) La structure chimique du tensioactif surfactine produit par la bactérie *B. subtilis* est représentée. AG : acide gras.

La structure cristallographique du domaine thioestérase du module FenB a été déterminée dans un état catalytiquement inactif en présence de l'inhibiteur fluorure de phénylméthylsulfonyle (Samel *et al.*, 2006). Celui-ci réagit avec le résidu nucléophile, ce qui génère un adduit covalent et inhibe la réaction (Samel *et al.*, 2006). Des expériences de *docking* et de dynamique moléculaire de l'enzyme avec le produit fengycine ont permis d'en identifier le mode de liaison au sein du site actif (Samel *et al.*, 2006) (Figure 45).

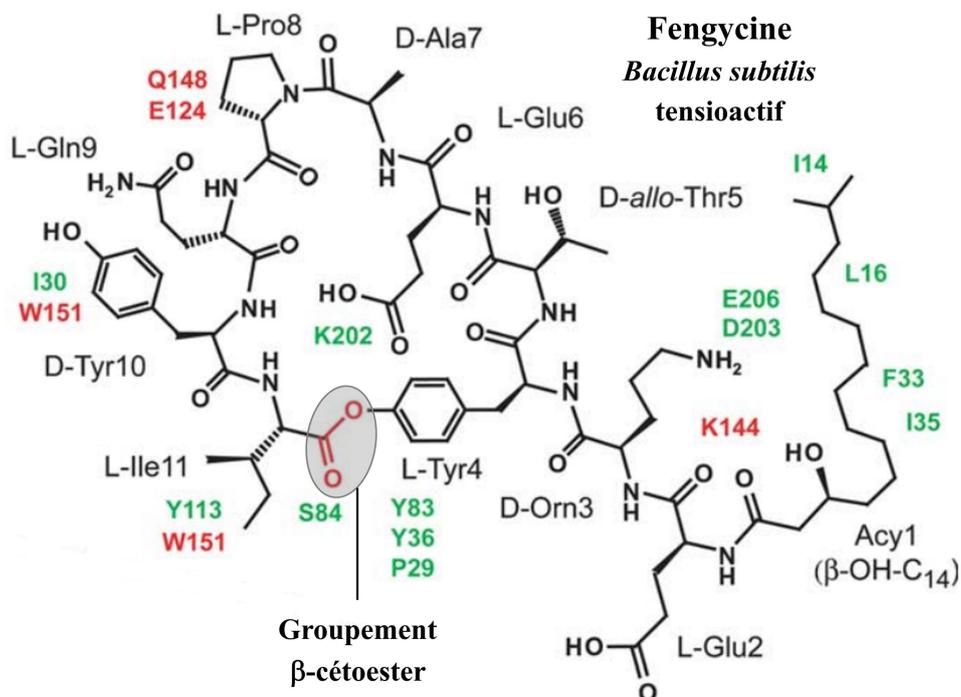


Figure 45. Liaison de la fengycine au sein du site actif du domaine TE de la NRPS FenB (adaptée de Samel *et al.*, 2006). Le domaine TE de la sous-unité FenB (PDB : 2CB9) catalyse la formation du noyau lactone entre les fonctions carboxylate et phénol d'un résidu tyrosine (L-Tyr4). Les résidus de l'enzyme sont indiqués en code à une lettre et ceux du lipopeptide en code à trois lettres. Les résidus hydrophobes I14, L16, Phe33 et I35 bordent le groupement β -hydroxy acétyl du NRP. Les résidus D203 et E206 forment des ponts salins avec la fonction amino du résidu non-protéino-gène D-Orn3. La région constituée des résidus hydrophobes P29, Y36 et Y83 porte le résidu peptidique L-Tyr4 correctement positionné en vue de la cyclisation intramoléculaire. L'hélice α de la coiffe participe à des liaisons hydrogènes avec le résidu peptidique L-Gln9 et des sous-sites établissent des contacts hydrophobes avec les résidus peptidique D-Tyr10/L-Ile11. Acy1 : β -hydroxyacyle, D-Orn3 : D-ornithine.

Au vu de la polyvalence des domaines TE, les activités correspondantes s'avèrent capitales pour générer des peptides macrocycliques ayant une haute rigidité structurale et stabilité protéolytique par des stratégies chémoenzymatiques (Süssmuth et Mainz, 2017).

IV.7. Intervention de domaines catalytiques subsidiaires

Des domaines supplémentaires sont directement intégrés à la chaîne d'assemblage ou possèdent une activité *trans*-catalytique (Miller et Gulick, 2016). Ces protéines modifient les précurseurs du substrat ou les peptides en cours d'élongation par de multiples réactions chimiques, ce qui augmente la complexité et la fonctionnalité des NRP produits (Miller et Gulick, 2016) (**Figure 46**).

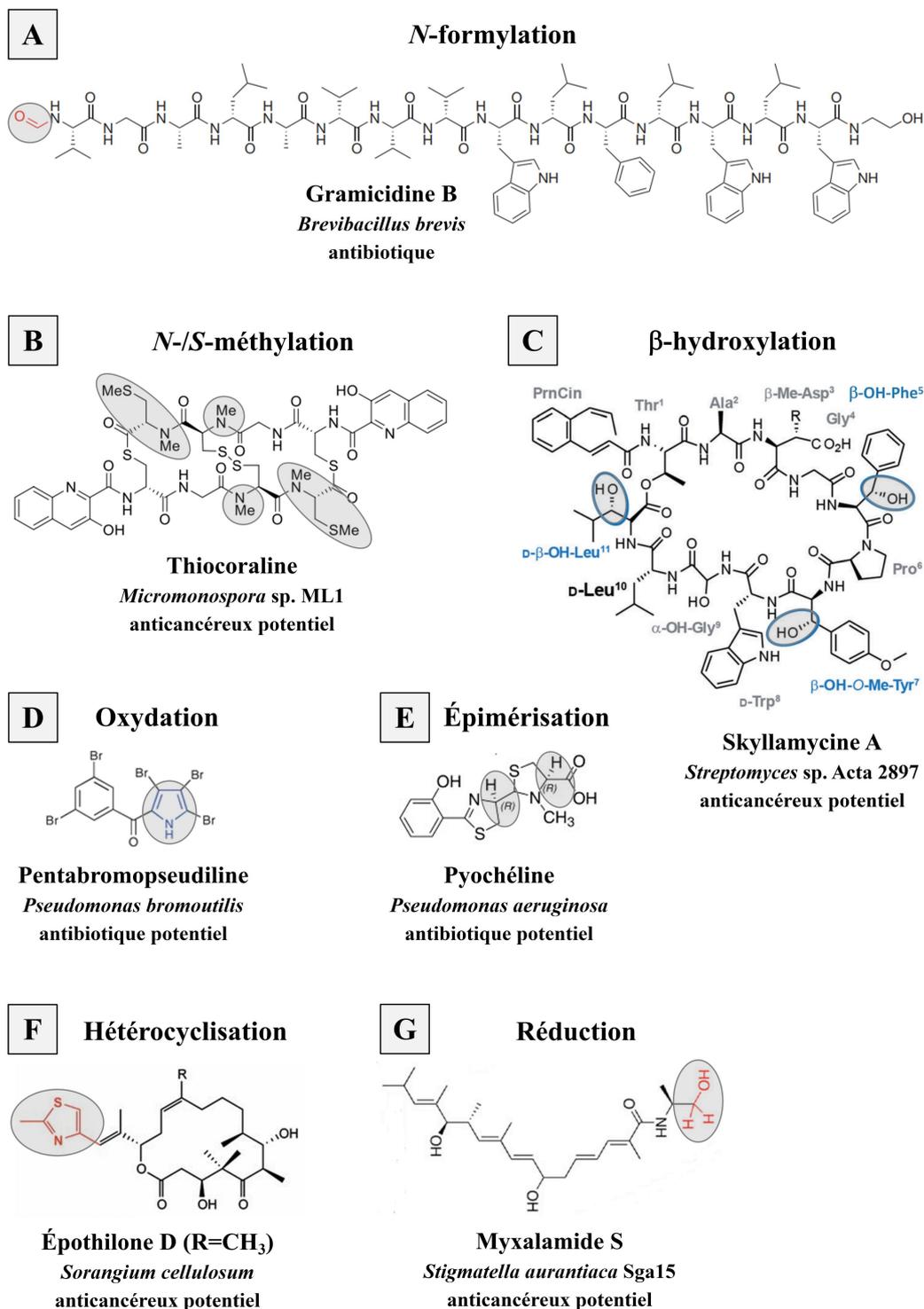


Figure 46. Modification de NRP par les domaines non-canoniques de systèmes NRPS. Les domaines additionnels assurent des réactions de formylation ou de réduction des extrémités terminales (A, gramicidine B ; G, myxalamide S), de méthylation (B, thiocoraline), de β-hydroxylation (C, skyllamycine A), d'oxydation (D, pentabromopseudiline), ou d'épimérisation (E, pyochéline) et d'hétérocyclisation (F, épothilone D) de résidus particuliers.

IV.7.A. N-acylation par les acyl-CoA ligases

La réaction de N-acylation est habituellement obtenue par couplage d'un acide gras linéaire ou ramifié, lié à un domaine PCP ou à la coenzyme A, à l'acide aminé initial de la ligne d'assemblage NRPS (Süssmuth et Mainz, 2017). Les acides gras varient d'une chaîne courte (*e.g.* époxomycine) à moyenne (*e.g.* éponémicine) ou longue (*e.g.* polymyxine B), et diffèrent aussi en fonction du degré d'insaturation (*e.g.* échinochandine) et de ramification (*e.g.* daptomycine) (Süssmuth et Mainz, 2017). Des groupements fonctionnels, tels que hydroxyles, carboxyles ou amines, situés à la position C β de la fraction lipidique, permettent des réactions de macrocyclisation, comme cela a été détaillé pour la surfactine et la fengycine (se référer à la section IV.6.) (Süssmuth et Mainz, 2017).

Différents mécanismes de fixation d'une unité acyle au monomère N-terminal du peptide lié à la matrice ont été rapportés (Süssmuth et Mainz, 2017) (**Figure 47**). Par exemple, une acyl-CoA ligase (AL) et un domaine ACP activent un précurseur sous forme d'acyl-adénylate, puis le transfèrent sur la chaîne d'assemblage DptA de l'antiviral daptomycine chez la bactérie *S. roseosporus* (Robbel et Marahiel, 2010) (**Figure 47A**). Ces unités catalytiques peuvent également agir (*en cis*) au sein du module d'initiation de systèmes hydrides, comme lors de la biosynthèse de l'antifongique potentiel mycosubtiline A produit chez la bactérie *B. subtilis* (Hansen *et al.*, 2007) (**Figure 47B**).

De plus, des protéines auxiliaires modifient les unités initiales, à l'image de l'époxydase HxcO responsable de l'insertion d'un atome d'oxygène au niveau d'une liaison C-C d'un hexanoïl-CoA lié à un domaine ACP, préalablement à l'incorporation du précurseur résultant (Kraas *et al.*, 2012) (**Figure 47C**). Par ailleurs, des domaines A activent des résidus acyles aromatiques, tels que l'acide dihydroxybenzoïque (Dhb) ou l'acide 4-méthyl-3-hydroxyanthranilique (4-Mha) en tant que motifs structuraux, notamment du sidérophore entérobactine ou de l'antibiotique actinomycine D produits par les ordres bactériens éponymes (Süssmuth et Mainz, 2017) (**Figure 47E**). En outre, des NRPS monomodulaires convertissent des acides aminés, comme le tryptophane en un noyau quinoxaline-2-carboxylique constitutif du composé antinéoplasique échinomycine chez la bactérie *S. echinatus* (Zhang *et al.*, 2013) (**Figure 47F**). Une réaction similaire a lieu avec la conversion de la proline en acide pyrrolique, structure cyclique centrale de l'antibiotique potentiel pyolutéorine produit par des espèces du genre *Pseudomonas*, y compris le pathogène humain *P. aeruginosa* (Thomas *et al.*, 2002).

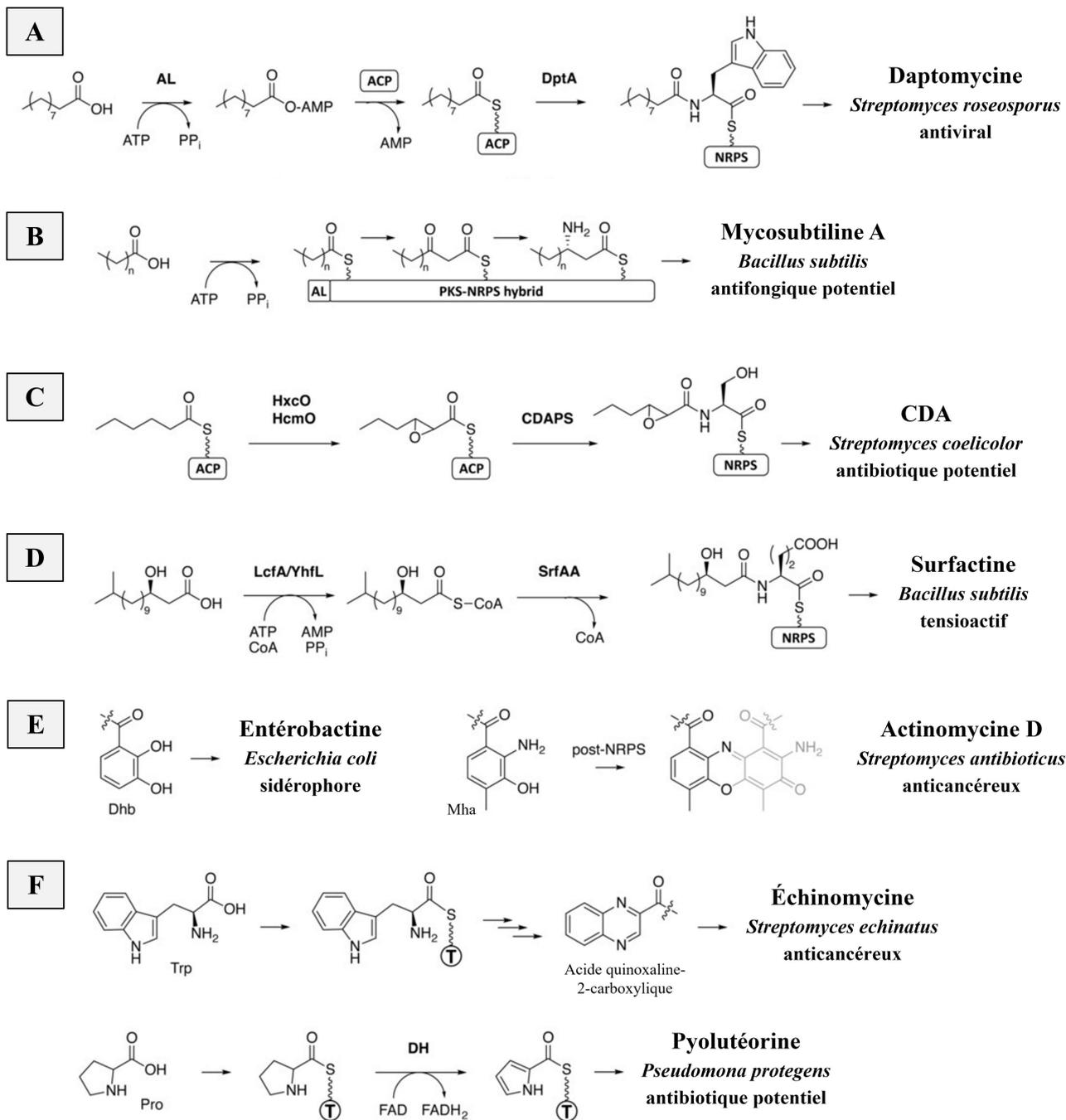


Figure 47. N-acylation par les acyl-CoA ligases (adaptée de Süßmuth et Mainz, 2017). (A) Une acyl-CoA ligase (AL) et un domaine porteur (ACP/PCP) catalysent l'activation et/ou le transfert d'un acide gras linéaire au sein de systèmes NRPS-PKS hybrides. (C) L'enzyme auxiliaire HxcO/HcmO catalyse l'époxydation d'un hexanoyl-CoA, puis un domaine d'adénylation autonome (CdAPS) active et lie ce précurseur à la NRPS associée. (D) Les acyl-CoA ligases autonomes LcfA/YhfL convertissent l'acide 3-hydroxymyristique (Hma) en l'acyl-CoA correspondant et un domaine C_{starter} couple celui-ci au résidu glutamate N-terminal incorporé dans le peptide. (E) Des domaines A de NRPS modulaires sélectionnent et activent des résidus aromatiques, tels que l'acide dihydroxybenzoïque (Dhb) ou l'acide 4-méthyl-3-hydroxyanthranilique (Mha). (F) Des NRPS monomodulaires modifient des acides aminés en unités acyles préalablement à la biosynthèse de produits naturels. CDA : antibiotique dépendant du calcium.

IV.7.B. *N*-formylation par le domaine F

Le domaine de formylation (F) est rarement trouvé au sein des systèmes, mais est par exemple intégré dans le module initial de la NRPS LgrA, dont la structure cristallographique a été résolue dans différents états conformationnels (Reimer *et al.*, 2016) (**Figure 48**). Dans l'état de formylation, le sous-domaine A_{sub} positionne le domaine valyl-PCP1 au niveau du site actif du domaine F. Celui-ci catalyse la *N*-formylation du résidu sur la fonction amino à partir du cofacteur N¹⁰-fH₄F (**Figure 49**). Le groupement *N*-formyl greffé stabilise la dimérisation des gramicidines cycliques, lesquelles forment des pores dans la membrane des bactéries à Gram + (Reimer *et al.*, 2016). Cela perturbe les gradients ioniques de cations monovalents et provoque donc la lyse des cellules (Corpuz *et al.*, 2022).

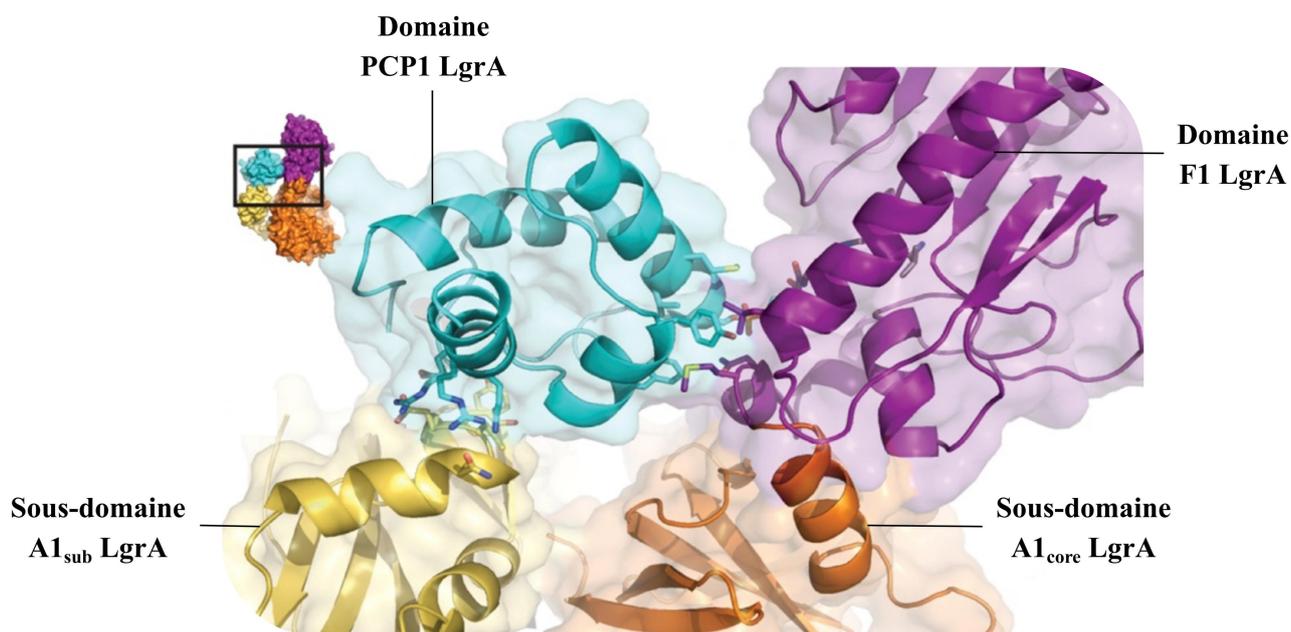


Figure 48. Représentation de l'interface de la NRPS LgrA (adaptée de Reimer *et al.*, 2016). La sous-unité LgrA (PDB : 5ES9) crée des interfaces spécifiques entre les domaines A, PCP et F. Le sous-domaine A1_{sub} place le domaine PCP1 chargé d'un résidu L-valine dans le site actif du domaine F, dont les résidus Arg170 et Asn177 interagissent avec la fonction acide phosphorique du monomère valyl-Ppant. Le domaine F1 catalyse ensuite la réaction de *N*-formylation à partir du cofacteur N¹⁰-formyltétrahydrofolate (voir **Figure 49**). A_{core} : sous-domaine N-terminal du domaine A, A_{sub} : sous-domaine C-terminal du domaine A, PCP : domaine porteur de peptidyle, F : domaine de formylation.

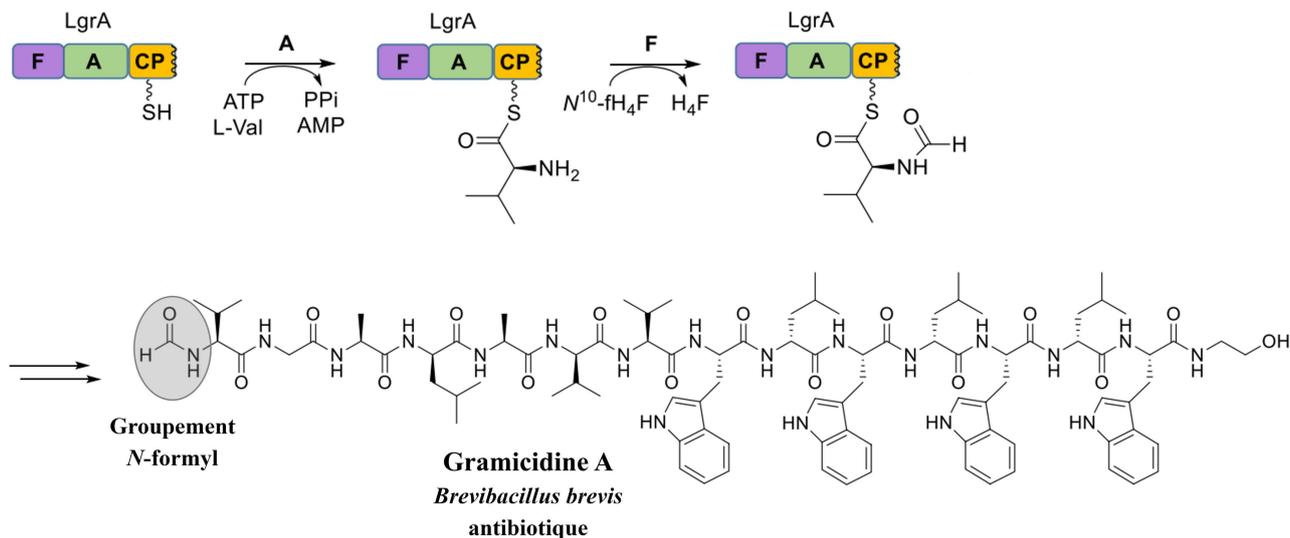


Figure 49. Réaction de *N*-formylation par le domaine F1 de la NRPS LgrA (adaptée de McErlean *et al.*, 2019). Le module initial de la NRPS LgrA adényle un résidu L-Val, puis y ajoute un groupement *N*-formyl. La structure chimique de l'antibiotique gramicidine A produit par la bactérie *Brevibacillus brevis* est représentée. AMP/ATP : adénosine mono-/triphosphate, N¹⁰-fH₄F : N¹⁰-formyltétrahydrofolate, PPI : pyrophosphate.

Un mécanisme alternatif de *N*-formylation de la marformycine A, cyclodepsipeptide produit des NRPS MfnC–L chez la bactérie *Streptomyces drozdowiczii*, a fait l'objet d'une étude approfondie (Liu *et al.*, 2015). L'analyse bio-informatique du BGC correspondant a d'une part révélé que la *N*-formyltransférase est une enzyme distincte (MfnA) et d'autre part que le module initial présente une architecture de domaines C-A-PCP, contrairement au module F-A-PCP de la NRPS LgrA (Liu *et al.*, 2015). À l'instar des gramicidines, la formylation de ce composé intervient dès les premières étapes de la biogenèse, bien que l'activité précise de la méthionyl-ARNt formyltransférase et du domaine C1 dans ce mécanisme réactionnel demeure à clarifier (Liu *et al.*, 2015). Ces unités pourraient le cas échéant catalyser la formation d'une liaison amide entre le domaine L-valyl-*S*-PCP et les cofacteurs N¹⁰-fH₄F ou fMet-ARNt^{Met} (Liu *et al.*, 2015).

Les domaines F présents à l'extrémité N-terminale des modules NRPS sont un moyen opportun de générer une coiffe avec une certaine fonctionnalité au sein de dispositifs synthétiques par le biais de l'ingénierie biosynthétique (Süssmuth et Mainz, 2017).

IV.7.C. Méthylation par le domaine MT

Le domaine méthyltransférase (MT) est généralement inclus entre les motifs A8 et A9 du sous-domaine C-terminal (A_{sub}) du domaine d'adénylation adjacent (Mori *et al.*, 2018). Les réactions de méthylation d'atomes d'azote, d'oxygène et de soufre des groupements ciblés sont catalysées par ces domaines à partir de précurseurs α -cétoacides et du cofacteur *S*-adénylméthionine (SAM) (Mori *et al.*, 2018). Ces modifications ont un impact significatif sur la polarité des liaisons, la résistance à la protéolyse, ainsi que sur la liberté conformationnelle des résidus (Süssmuth et Mainz, 2017).

Le principal système biosynthétique étudié à ce sujet contient deux NRPS dimodulaires (TioR-TioS), ainsi qu'une *S*-méthyltransférase discrète (TioN), et produit la thiocoraline, cyclodepsipeptide à activité anticancéreuse, initialement isolée chez la bactérie *Micromonospora* sp. ML1 (Mori *et al.*, 2018) (**Figure 50**). Une série d'expérimentations a démontré que le domaine chimérique A3-MT3 de la sous-unité TioS catalyse d'abord la *N*-méthylation du squelette du NRP au niveau de résidus L-Cys, puis que l'enzyme autonome TioN catalyse le *S*-méthylation de ces chaînes latérales, avant qu'une oxydation spontanée du résidu *N*-méthyl-L-Cys en un pont disulfure forme le produit cyclique final (Mori *et al.*, 2018) (**Figure 50**). Néanmoins, le domaine de fusion A-MT du module 4 (TioS) est prédit pour activer un résidu L-Val selon le code de spécificité NRPS, mais du fait de l'absence d'un tel résidu dans la molécule synthétisée, sa fonction exacte reste en suspens (Mori *et al.*, 2018).

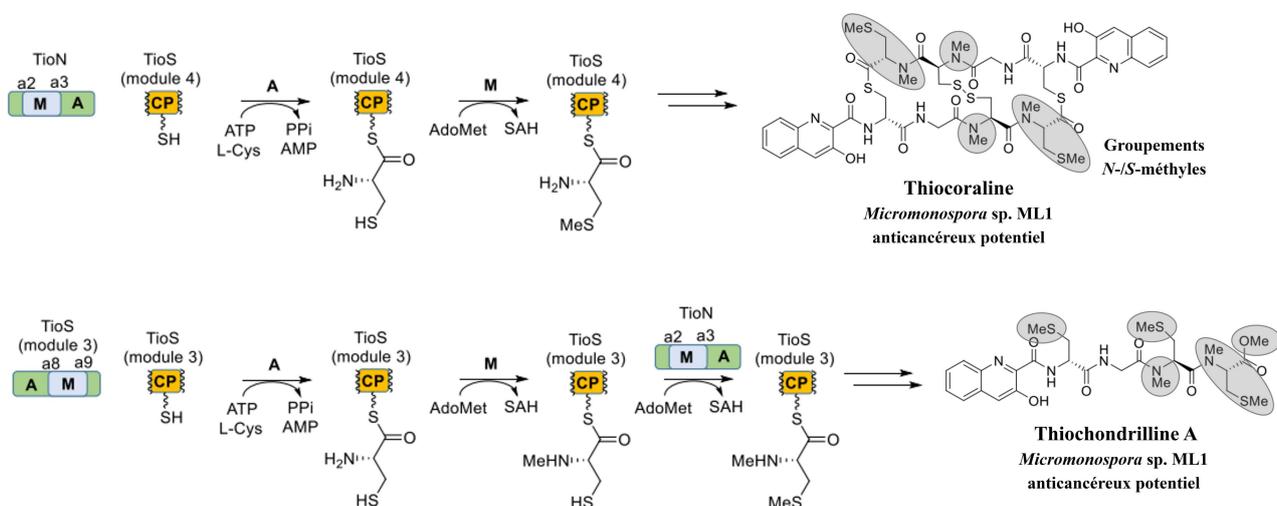


Figure 50. Réactions de méthylation par les domaines MT de la NRPS TioS et de l'enzyme TioN (adaptée de McErlean *et al.*, 2019). (A) Le domaine chimérique A3-MT3 de la sous-unité TioS et l'enzyme discrète TioN catalysent les *N*- et *S*-méthylations au niveau des intermédiaires réactionnels liés aux domaines porteurs. Les structures chimiques des anticancéreux potentiels thiocoraline et thiocondrilline A produits par la bactérie *Micromonospora* sp. ML1 sont représentées. A-M : domaine d'adénylation-méthyltransférase, AMP/ATP : adénosine mono-/triphosphate, PPi : pyrophosphate, AdoMet : *S*-adénylméthionine, SAH : *S*-adénylhomocystéine.

Dans ce cadre, la structure cristallographique du domaine chimérique A4-MT4 de la NRPS TioS a été résolue en complexe avec la protéine activatrice MLP TioT et en présence de l'inhibiteur SAH et du substrat L-Val (Mori *et al.*, 2018) (**Figure 51**). Le domaine MT4 s'intègre entre le brin β 24 et l'hélice C, chacun situé dans le sous-domaine A_{sub} des domaines A ininterrompus (Mori *et al.*, 2018). Contrairement aux fusions internes de protéines, flanquées de boucles flexibles, les liaisons entre les domaines A4 et MT4 se structurent à chaque extrémité sans modifier le repliement canonique du domaine A (Mori *et al.*, 2018). Par exemple, l'hélice α 24 se compose de deux fois plus de tours que la structure secondaire analogue (Mori *et al.*, 2018). La région de connexion entre ces domaines A4-MT4 comporte un feuillet β à trois brins et deux hélices α : les brins β 23- β 24 et l'hélice α 12 relient le domaine A à l'extrémité N-terminale du domaine MT ; l'hélice α 24 et le brin β 37 le connectent à l'autre borne (Mori *et al.*, 2018). Par ailleurs, l'hélice α 10 et le brin β 18, chacun localisé au sein du sous-domaine A_{core} , assurent une interaction native avec le partenaire MLP TioT (Mori *et al.*, 2018). Un feuillet β à deux brins perpendiculaires situé à l'extrémité distale de la structure permet la liaison du complexe avec le domaine PCP4 de la NRPS TioS (Mori *et al.*, 2018).

Le domaine MT présente une architecture unique, avec notamment un sous-domaine C-terminal atypique parmi les méthyltransférases caractérisées (Shi *et al.*, 2010 ; Mori *et al.*, 2018). En effet, la région N-terminale adopte le repliement α/β de Rossmann canonique, composé d'un feuillet β à six brins parallèles et lié à deux paires d'hélices α (Mori *et al.*, 2018). En plus des éléments impliqués dans la liaison du ligand, le domaine MT4 contient une structure secondaire originale, l'hélice α 12, ainsi que des résidus hydrophobes (Tyr484 et Tyr488), responsables de l'interaction avec l'adénine et le ribose du cofacteur SAM, respectivement (Mori *et al.*, 2018) (**Figure 51**). Par contre, la région C-terminale présente une certaine similitude avec le repliement des protéines SirA de transport du fer, formé en l'occurrence d'un feuillet β à quatre brins et de trois hélices α (Shi *et al.*, 2010 ; Mori *et al.*, 2018). Ce type de repliement particulier permet vraisemblablement la connectivité adéquate avec le domaine A4 au sein du didomaine (Mori *et al.*, 2018). Toutefois, la pertinence biologique de la réaction catalysée reste incertaine, possiblement à cause de l'existence d'un vestige évolutif non-fonctionnel *in vivo*, comme évoqué avec l'absence d'un résidu L-Val dans la structure chimique de la thiocoraline et de la thiochondrilline A (Mori *et al.*, 2018 ; McErlean *et al.*, 2019).

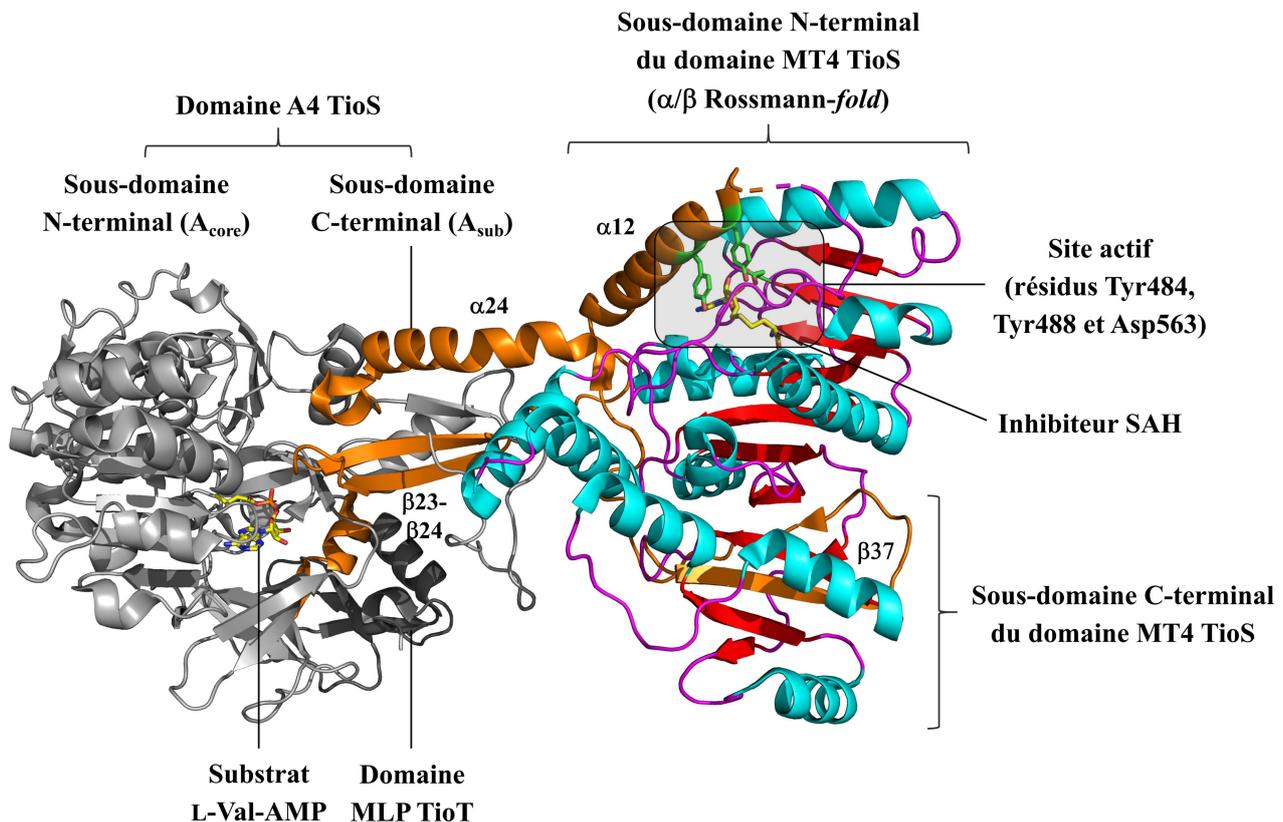


Figure 51. Représentation de la structure cristallographique du didomaine A-MT de la NRPS TioS. Le didomaine de fusion A-MT du module 4 de la NRPS TioS (PDB : 5WMM) adopte un repliement caractéristique α/β de Rossmann (région N-terminale) des méthyltransférases, et présente en plus un feuillet β à quatre brins et trois hélices α adjacentes. Cette enzyme bifonctionnelle se lie à son partenaire MLP TioT, et le complexe activé assure l'adénylation d'un résidu L-Val puis la *N*-méthylation de celui-ci. Le domaine A4 est illustré en gris, et le domaine MT4 en cyan et rouge selon les structures secondaires. Les régions *linker* N-MT (α 24, β 37) et C-MT (β 24, α 12) du domaine MT inclus entre les motifs A8 et A9 du domaine A sont représentées en orange. AMP : adénosine monophosphate, SAH : S-adénosylhomocystéine.

En plus des réactions de *N*- et *S*-méthylations de liaisons peptidiques et de résidus L-Cys, l'*O*-méthylation de fonctions hydroxyles est parfois observée dans les NRP (McErlean *et al.*, 2019). Par exemple, un domaine MT putatif, intégré dans le domaine A2 du module d'élongation de la NRPS KtzH, est responsable de la *O*-méthylation de la chaîne latérale d'un résidu L-Ser lors de la biogenèse des kutznerides chez la bactérie *Kutzneria* sp. 744 (Broberg *et al.*, 2006).

De par leur architecture singulière, des domaines A-MT artificiels ont déjà été exploités avec succès pour introduire des motifs de méthylation spécifiques dans des structures de NRP (McErlean *et al.*, 2019). Dans un cas concret, les domaines MT des NRPS TioS ou KtzH ont chacun été insérés entre les motifs A8 et A9 du didomaine A-T Ecm6, ce qui a permis de générer des dérivés bioactifs de l'échinomycine (Lundy *et al.*, 2018).

IV.7.D. β -Hydroxylation par les cytochromes P450

Les domaines oxygénase (Oxy) et TauD de NRPS, ainsi que les mono-oxygénases CYP450 ou les dioxygénases α -cétoglutarate dépendantes catalysent la β -hydroxylation des acides aminés avec une spécificité propre (*e.g.* L-Asp par TauD, ou les acides aminés aromatiques par les cytochromes P450) (Haslinger *et al.*, 2014 ; Duban *et al.*, 2022). Certains résidus sont de ce fait convertis en des structures chimiques atypiques, comme L-Tyr et L-His oxydés en cycles coumarine ou imidazolone lors de la biosynthèse de novobiocine ou nikkomycine, respectivement (Süssmuth et Mainz, 2017).

Des données indiquent que certaines oxydations ont lieu pendant l'assemblage du peptide sur la NRPS principale par divers CYP450 (Haslinger *et al.*, 2014). Ces métallo-enzymes sont reconnues pour leur capacité à catalyser nombre d'oxydations sur certains substrats, y compris l'hydroxylation de liaisons C-H (Haslinger *et al.*, 2014). Par exemple, l'isoenzyme CYP163B3 (P450sky) assure la modification stéréospécifique de 3 résidus hydrophobes de la skyllamycine A ((2*S*,3*S*)- β -OH-L-Phe, (2*S*,3*S*)- β -OH-L-Tyr-(OMe) et (2*R*,3*S*)- β -OH-D-Leu), produite chez la souche *Streptomyces sp. Acta 2897* (Haslinger *et al.*, 2014) (**Figure 52**). Ce métabolite inhibe la voie de signalisation du facteur de croissance des thrombocytes et bloque notamment la mitose de cellules tumorales, ce qui en fait un principe actif bénéfique contre les maladies cancéreuses (Trinidad-Calderón *et al.*, 2023).

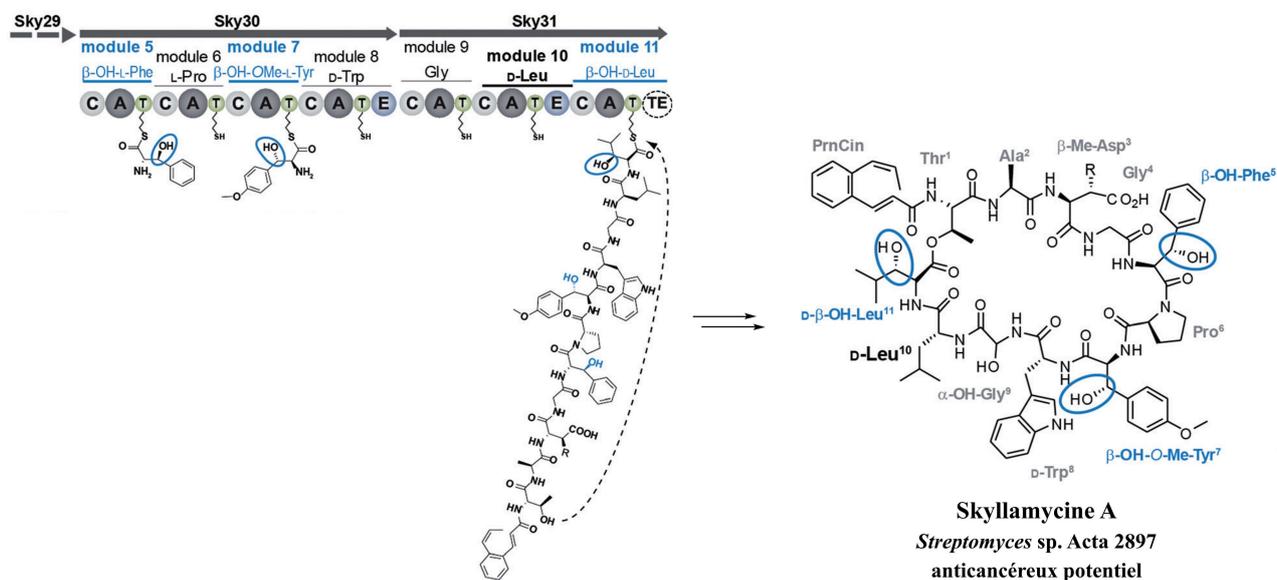


Figure 52. Représentation de la voie de biosynthèse de la skyllamycine A (adaptée de Haslinger *et al.*, 2014). Les NRPS Sky29–31 assemblent la skyllamycine A, tandis que le CYP450B3 catalyse la β -hydroxylation des résidus L-Phe, L-Tyr et D-Leu liés au domaine PCP du module correspondant. A : domaine d'adénylation, C : domaine de condensation, T : domaine porteur de pependyle, TE : domaine thioestérase.

La structure cristallographique du complexe CYP163B3-PCP7 a été déterminée en présence d'un inhibiteur imidazole-phosphopantéthéinyl (Haslinger *et al.*, 2014). Le CYP450 adopte un repliement canonique avec une région centrale hélicoïdale (14 hélices α), flanquée de trois feuillets β , dont les hélices α F et α G forment l'interface avec le domaine porteur (Haslinger *et al.*, 2014) (**Figure 53A**). Les partenaires ne subissent chacun que des ajustements mineurs de leur surface d'interaction, avant la formation du complexe, ce qui suggère que la structure tertiaire détermine la sélectivité de substrat (Haslinger *et al.*, 2014). L'interface en question présente un caractère hydrophobe prononcé : d'une part, les chaînes latérales des résidus Trp193 et Leu194 de l'hémoprotéine se logent au sein d'une poche du faisceau d'hélices α du domaine PCP7 ; d'autre part, les résidus aliphatiques Ala90, Met94, Leu200 et Leu239 forment une fente entre les hélices α G et α I où se positionnent le groupement *gem*-diméthyle du cofacteur 4'-Ppant (Haslinger *et al.*, 2014) (**Figure 53B**). En revanche, l'interface contient peu d'interactions ioniques : les résidus Asp191, Asn197 et Glu235, situés dans les hélices α G ou α I du cytochrome, contactent les résidus Arg63, Thr46 et Lys47 du partenaire, tandis qu'un unique couple de résidus polaires participe à une liaison hydrogène (**Figure 53B**).

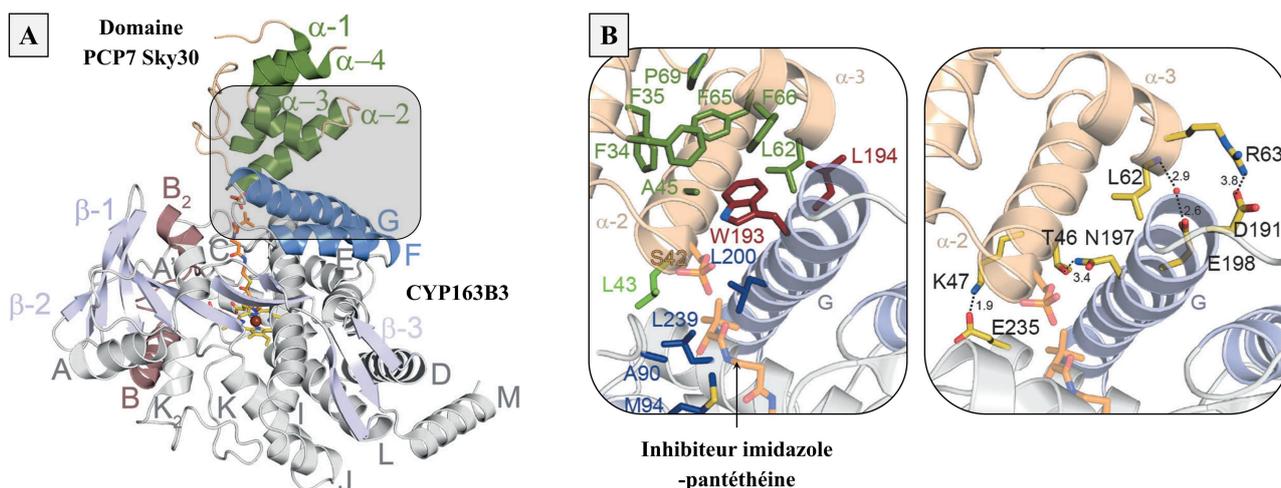


Figure 53. Représentation de la structure cristallographique du complexe CYP163B3-PCP7 au sein du système NRPS Sky29–31 (adaptée de Haslinger *et al.*, 2014). (A) Le cytochrome CYP163B3 présente le repliement typique de la super-famille des CYP450 avec un cœur α -hélicoïdal flanqué de trois feuillets β . (B) L'interface du complexe dépend principalement d'interactions hydrophobes et électrostatiques.

Au terme d'investigations approfondies sur ces interactions systémiques avec les domaines PCP, leur activité de recrutement de diverses enzymes, dont les CYP450, apparaît essentielle pour opérer des modifications *trans*-catalytiques et produire des NRP bioactifs, tant au sein de systèmes naturels, que de NRPS artificielles.

IV.7.E. Oxydation par les enzymes NAD⁺/FAD-dépendantes

Des oxydoréductases catalysent l'oxydation d'acides aminés thioestérifiés aux domaines PCP et requièrent les coenzymes NAD⁺ ou FAD en tant qu'accepteurs d'électrons lors de la réaction (Thapa *et al.*, 2019). Un système de type II caractérisé dans ce contexte inclut l'adényltransférase Bmp4, l'oxydase Bmp3, l'halogénase Bmp2, ainsi que le didomaine PCP-TE Bmp1, et synthétise la partie bromopyrrole de la pentabromopseudiline, métabolite aux propriétés antibiotiques initialement isolé chez la bactérie marine *Pseudomonas bromoutilis* (Thapa *et al.*, 2019) (**Figure 54**). Des expériences ont démontré que l'oxydase Bmp3 catalyse l'oxydation du prolyl-S-PCP en un cycle pyrrole avec un transfert d'électrons à l'aide de la coenzyme FAD vers l'accepteur final O₂, puis que l'halogénase Bmp2 catalyse l'addition régiosélective de quatre atomes de brome dans la structure (Thapa *et al.*, 2019). Ensuite, une déhalogénase thiorédoxine-like (Bmp8) catalyse l'élimination de l'halogène à la position C2 du tétrabromopyrrole et le domaine TE libère le monomère correspondant, avant que le CYP450 Bmp7 assure enfin le couplage de celui-ci avec une molécule de 2,4-dibromophénol dans le produit final (Thapa *et al.*, 2019) (**Figure 54**).

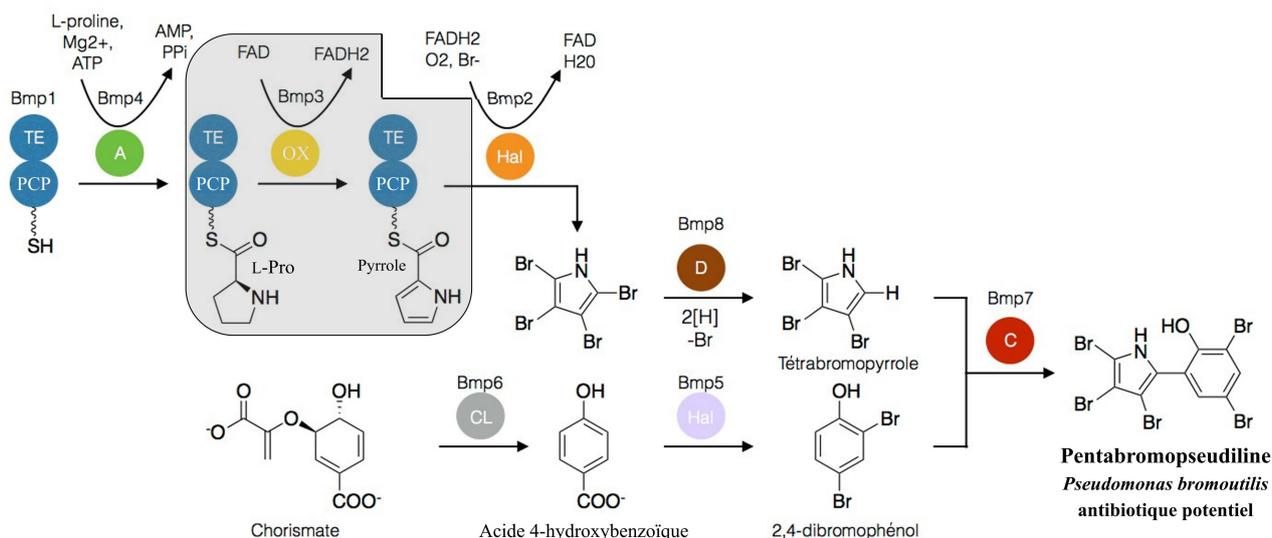


Figure 54. Représentation de la voie de biosynthèse de la pentabromopseudiline. Le système Bmp1–4 assemble la partie bromopyrrole, avec notamment l'oxydase Bmp3 responsable de la conversion du prolyl-PCP en pyrrolyl-PCP (encadré gris). En parallèle, la chorismate lyase Bmp6 produit l'acide 4-hydroxybenzoïque (4-HBA), puis l'halogénase Bmp5 catalyse l'addition de deux atomes de brome dans la structure. Le CYP450 Bmp7 condense les intermédiaires réactionnels en le métabolite final de la voie chez la bactérie *P. bromoutilis*. A : domaine d'adénylation, PCP : domaine porteur de peptidyle, OX : oxydase, Hal : halogénase, D : déhalogénase, TE : domaine thioestérase, CL : chorismate-pyruvate lyase, AMP/ATP : adénosine mono-/triphosphate, PPi : pyrophosphate, FAD : flavine adénine dinucléotide.

Dans ce contexte, la structure cristallographique de l'oxydase Bmp3 a été résolue en complexe avec le domaine porteur Bmp1, la coenzyme FAD et le produit pyrroly-PCP (Thapa *et al.*, 2019). L'enzyme Bmp3 présente des domaines α -hélicoïdaux N- et C-terminaux avec en plus un domaine central en feuillets β (Thapa *et al.*, 2019) (**Figure 55A**). L'interface établie avec le partenaire dépend de multiples résidus hydrophobes conservés, lesquels permettent de positionner le cofacteur 4'-Ppant, dans une orientation correcte au niveau du site actif, et le substrat L-Pro à proximité de la coenzyme FAD (Thapa *et al.*, 2019) (**Figure 55B**). Aucune chaîne latérale ne contacte directement la fonction phosphorique du bras 4'-Ppant, localisé dans une poche remplie de solvant, et seul le résidu Arg243 interagit avec les groupements carbonyles de celui-ci (**Figure 55B**). Les chaînes latérales des résidus Arg268 et Phe278 participent à une interaction ionique avec le groupement pyrophosphate ou à un empilement π - π avec la base adénine de la coenzyme FAD, respectivement (Thapa *et al.*, 2019). La fonction hydroxyle du résidu Thr158 et les groupements amides du squelette protéique stabilisent le noyau isoalloxazine de la coenzyme au moyen de liaisons hydrogènes (Thapa *et al.*, 2019). L'un des groupements hydroxyles de la chaîne ribityl du FAD et les chaînes latérales des résidus catalytiques (*i.e.* Glu242, Ser363) constituent le trou de l'oxyanion (Thapa *et al.*, 2019) (**Figure 55B**). Celui-ci est responsable de la stabilisation de l'espèce énolate générée lors de la réaction (Thapa *et al.*, 2019).

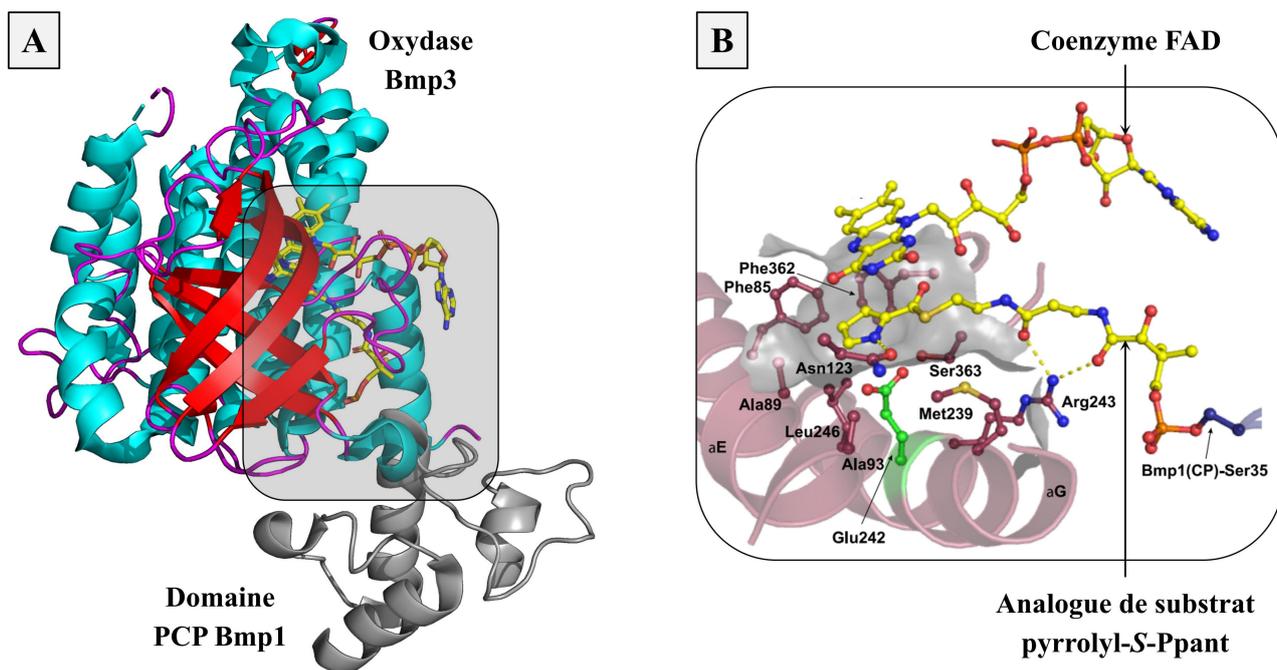


Figure 55. Représentation de la structure cristallographique du complexe autonome Bmp3-Bmp1 (*adaptée de Thapa et al.*, 2019). (A) Le complexe de type II Bmp3-Bmp1 (PDB : 6CXT) présente des régions N- et C-terminales α -hélicoïdales flanquées de sept feuillets β orthogonaux. (B) Le site actif assure l'oxydation du résidu L-prolyl-PCP en un cycle pyrrole. L'atome d'oxygène 2-carbonyl de celui-ci est à une distance de liaison de la partie ribityl de la coenzyme FAD et de l'atome d'azote du squelette du résidu Ser363. La base catalytique Glu242 (vert), positionnée au centre des résidus Met239 et Leu246 sur l'hélice α G, déprotone la liaison C-H de l'atome de carbone C2 du pyrrolyl-PCP, ce qui forme un intermédiaire énolate stabilisé *via* le trou de l'oxyanion.

IV.7.F. Épimérisation par le domaine E

Le domaine d'épimérisation est généralement compris entre les domaines PCP et ^DC_L de la ligne d'assemblage, et catalyse la conversion d'un acide aminé d'une configuration L- à D-, préalablement à la condensation stéréosélective avec l'unité d'extension en aval (Dekimpe et Masschelein, 2021). Ces modifications ont un effet notable sur la résistance aux protéases spécifiques des énantiomères L- répandus, et également sur l'activité des produits de synthèse (Chen *et al.*, 2016). Par exemple, la pyochéline contient des résidus D-cystéine et exerce un rôle de sidérophore, ce qui confère ainsi un avantage sélectif au pathogène *Pseudomonas aeruginosa* lorsque la disponibilité en fer est restreinte en réponse à l'infection (Patel *et al.*, 2003).

Les domaines E présentent une forte homologie structurale avec leur ancêtre évolutionnaire, le domaine C, à savoir une taille de 40–50 kDa et un repliement symétrique général en pseudo-dimère (Dekimpe et Masschelein, 2021) (**Figure 56**). Cependant, leur séquence diffère significativement, même si ces domaines partagent le motif catalytique E2/C3 HHxxxDG conservé (Samel *et al.*, 2014). Des études structurales réalisées sur le module initial du système NRPS TycA–C (tyrocidine A) ont mis en évidence que le deuxième résidu histidine (His146) de la séquence consensus E2 et un résidu glutamate (Glu285) exclusif aux domaines E (motif E4, EGHGRE) se positionnent au niveau des faces opposées du site actif (Samel *et al.*, 2014) (**Figure 56**). Le mécanisme catalytique demeure en cours d'élucidation, mais ces résidus ont été initialement proposés pour assurer une réaction à deux nucléophiles indépendante de cofacteurs exogènes (Samel *et al.*, 2014). Des analyses mutationnelles ont d'une part démontré que le résidu Glu285 soustrait le proton C α du substrat lié au domaine PCP donneur, ce qui forme un intermédiaire énolate (Samel *et al.*, 2014). D'autre part, cela a révélé que le résidu His146 et le moment dipolaire de l'hélice α 4 stabilisent l'espèce transitoire, et également que cette étape est suivie d'une reprotonation spontanée (Stachelhaus et Walsh, 2000 ; Samel *et al.*, 2014). Toutefois, des preuves biochimiques additionnelles semblent nécessaires pour confirmer les rôles présumés des résidus du site actif (Dekimpe et Masschelein, 2021).

Par ailleurs, les domaines C/E hybrides ne possèdent pas le fameux résidu glutamate catalytique des domaines E, mais portent un motif HH[I/L]xxxxGD au sein du site actif (Rausch *et al.*, 2007). Ceci suggère que ces catalyseurs bifonctionnels ont évolué avec un mécanisme réactionnel distinct, bien que celui-ci reste à étudier (Dekimpe et Masschelein, 2021).

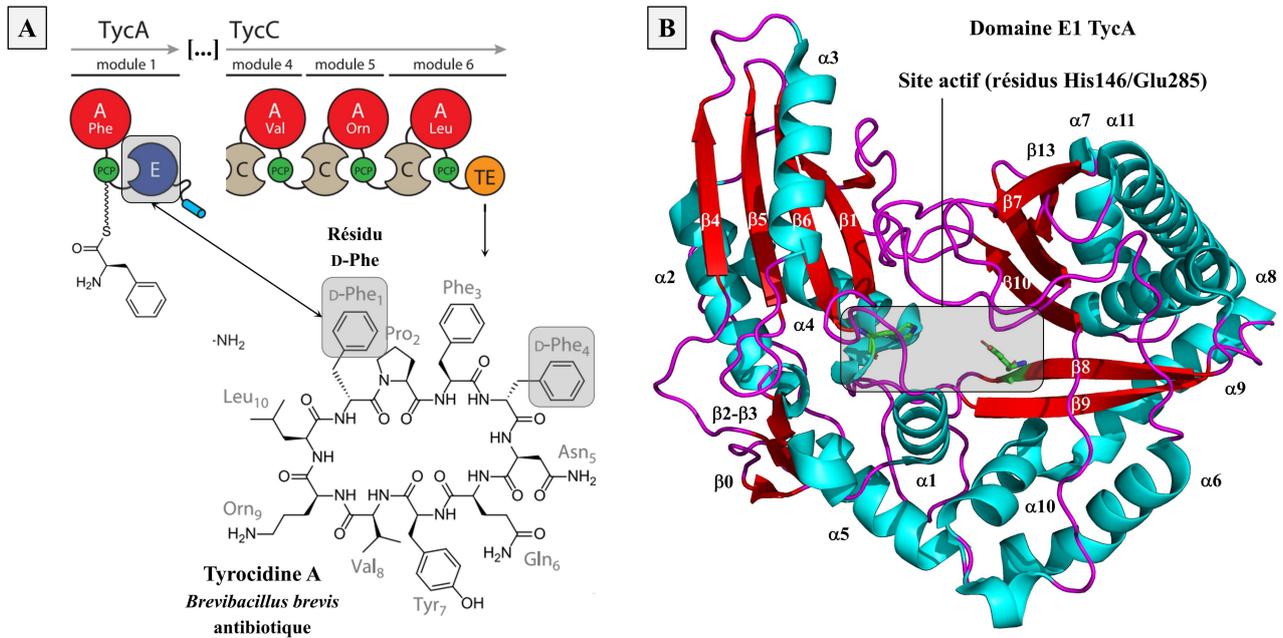


Figure 56. Représentation de la structure cristallographique du domaine E1 de la NRPS TycA (adaptée de Fage et al., 2021). (A) Le domaine E1 du module TycA (PDB : 2XHG) est impliqué dans la biosynthèse de tyrocidine A chez la bactérie *Brevibacillus brevis*. (B) La structure présente le repliement canonique des domaines C en deux sous-domaines avec une fente centrale. Le site actif contient les séquences consensus E2 et E4 avec les résidus catalytiques His146 et Glu285, respectivement. Les hélices α , les brins β , et les boucles sont représentés en cyan, rouge, et magenta.

IV.7.G. Hétérocyclisation par le domaine Cy

Le domaine Cy catalyse l'élargissement de la chaîne avec des résidus L-Cys ou L-Ser/L-Thr, suivie de l'hétérocyclisation des chaînes latérales sur le groupement carbonyle-amide du squelette et de la déshydratation des intermédiaires hémiaminaux correspondants en noyaux thiazoline ou oxazoline à cinq chaînons, respectivement (Süssmuth et Mainz, 2017). En plus, des domaines OX intégratifs ou des réductases à action *trans*-catalytique peuvent modifier ces hétérocycles, comme dans le cas de produits anticancéreux (*e.g.* bléomycine et épéthilone oxydés) ou de sidérophores (*e.g.* pyochéline et vibriobactine réduits), respectivement (Patel et Walsh, 2001 ; Schneider *et al.*, 2003). Par exemple, le domaine Cy de la sous-unité EpoB d'un système hybride assemble le cycle 2-méthylthiazoline de l'épéthilone, à partir des précurseurs acétyl et L-Ser, avant que le domaine OX adjacent oxyde cette structure cyclique essentielle à l'activité biologique (Dowling *et al.*, 2016) (**Figure 57**).

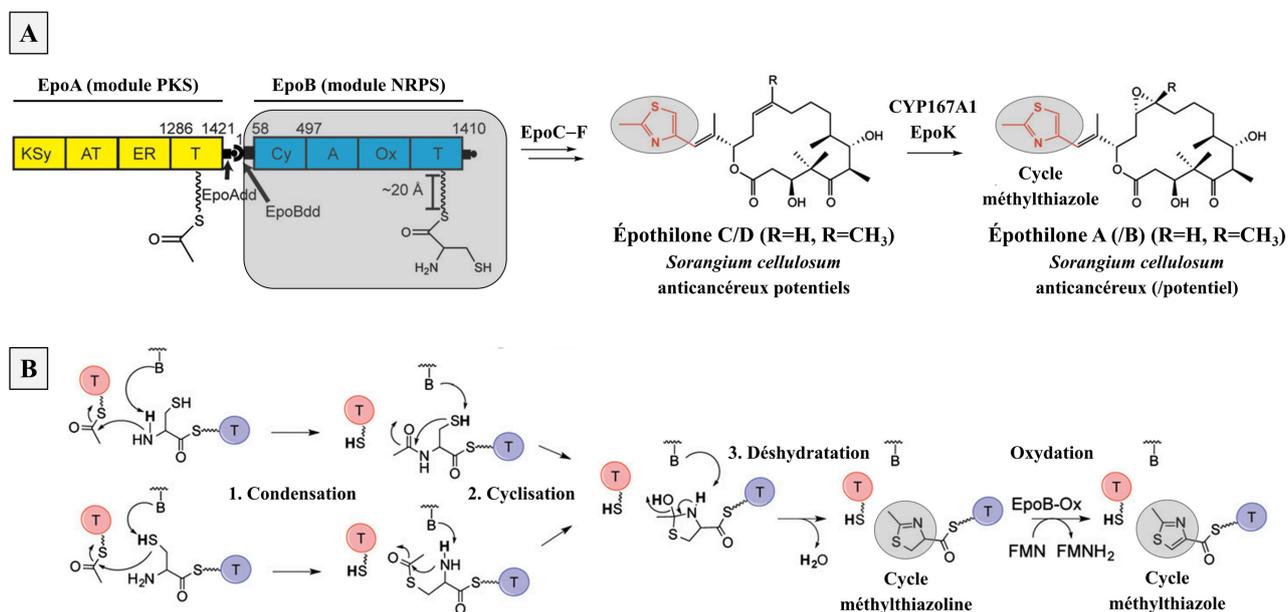


Figure 57. Représentation de la voie de biosynthèse de l'épéthilone (adaptée de Dowling *et al.*, 2016). La biogenèse de l'épéthilone implique notamment les modules PKS EpoA et NRPS EpoB, suivis de huit autres modules répartis sur quatre sous-unités protéiques. En particulier, le module EpoB comprend un domaine Cy responsable de l'incorporation d'un cycle thiazoline (oxydé en thiazole) dans la structure chimique des épéthilones C–D. Le CYP167A1 EpoK assure enfin l'époxydation d'une liaison alcène de ces produits en épéthilone A–B. (B) Le domaine Cy catalyse les réactions de condensation d'un résidu L-Thr, puis de cyclodéshydratation de l'intermédiaire en un cycle méthylthiazoline. Ensuite, le domaine OX intégré au module catalyse l'oxydation de ce noyau en méthylthiazole. Cy : domaine d'hétérocyclisation, A : domaine d'adénylation, T : domaine porteur de peptidyle, OX : domaine oxydase, FMN : flavine mononucléotide.

Les domaines Cy présentent également une homologie structurale importante avec les domaines C, comme le montre la structure cristallographique du domaine Cy EpoB résolue avec le domaine de *docking* N-terminal associé (Dowling *et al.*, 2016) (**Figure 58**). En effet, le domaine Cy adopte le repliement canonique en pseudo-dimère, dont chaque région possède un feuillet β à cinq brins et autant d'hélices α disposées contre celui-ci (Dowling *et al.*, 2016). De plus, le domaine de *docking* EpoBdd (55 résidus), connecté au moyen d'un *linker* flexible de 20 résidus, présente un repliement de type $\alpha\beta\beta\alpha$, dont le tour β et l'hélice α_3 permettent l'interaction avec le domaine partenaire du module PKS EpoA situé en amont (Dowling *et al.*, 2016). Lors de cette interaction, le *linker* dirige le domaine acétyl-PCP EpoA vers le site donneur du domaine Cy EpoB et contribue à transférer efficacement le substrat en vue de la réaction catalytique (Dowling *et al.*, 2016). Les domaines Cy se caractérisent par le motif DxxxxDxxS conservé, à la place du motif HHxxxDG des domaines C, dont les résidus, en l'occurrence Asp201 et Asp206, jouent un rôle dans le maintien de la structure des canaux de solvants (Dowling *et al.*, 2016). Cela suggère que le site actif se trouve à une position différente par rapport aux domaines ancestraux et que l'évolution favorise la modulation du spectre catalytique (Dowling *et al.*, 2016). Des expériences de mutagenèse dirigée et des tests d'activité ont démontré que les résidus Asn335, Gln445 et Asp449 contribuent aux réactions de condensation et de cyclodéshydratation, mais le mécanisme détaillé n'est pas encore compris (Dowling *et al.*, 2016).

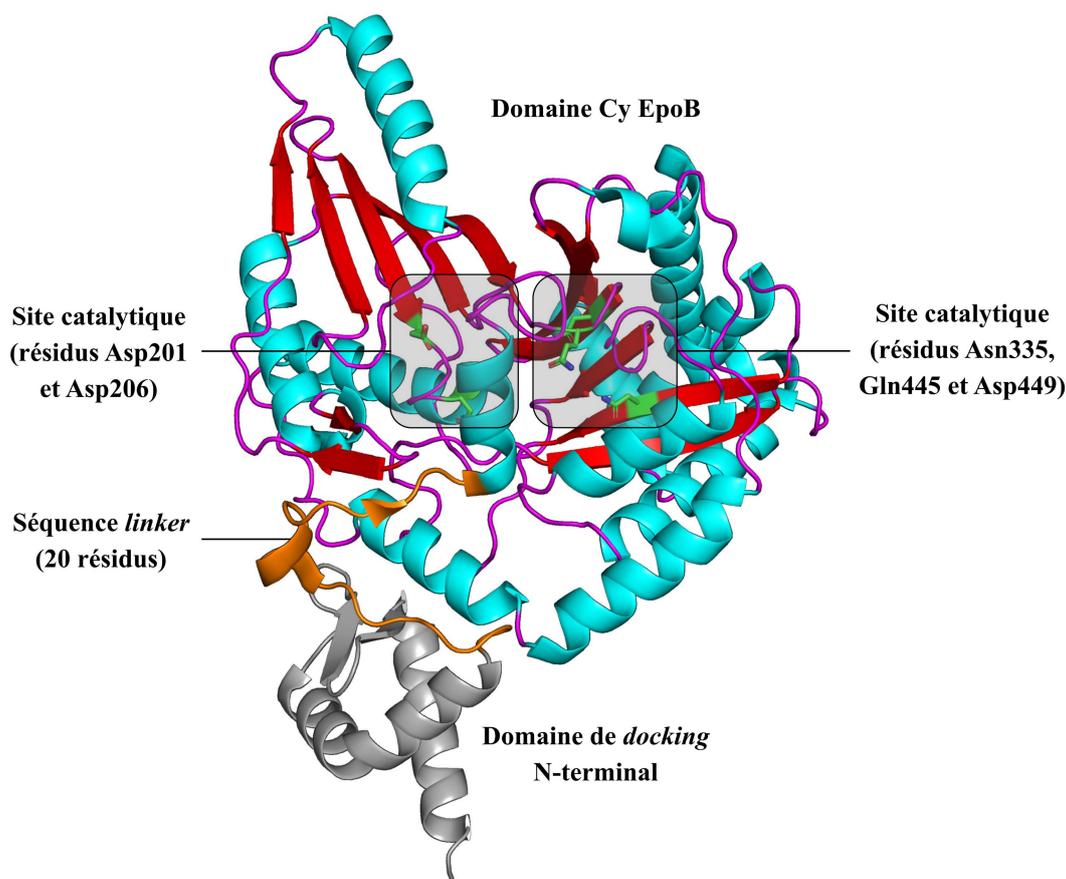


Figure 58. Représentation de la structure cristallographique du domaine Cy de la NRPS EpoB. Le domaine Cy de la sous-unité EpoB (PDB : 5T81) adopte la structure caractéristique des domaines C. Le site actif comporte les résidus conservés Asp201 et Asp206 à rôles structuraux, ainsi qu'une triade de résidus catalytiques putatifs.

IV.7.H. Réduction par le domaine R

Le domaine réductase (R) catalyse la réduction de la liaison thioester dans certains systèmes, à la place du domaine TE, et libère un NRP avec un groupement aldéhyde (saframycine A) ou alcool (myxalamide S) à l'extrémité C-terminale, au lieu d'une fonction carboxylate (Corpuz *et al.*, 2022). Le mécanisme catalytique implique la réduction cinétiquement limitante du substrat peptidyl-PCP en un intermédiaire aldéhyde avec un transfert de deux électrons, puis la dissociation du complexe, couplée à l'échange du cofacteur NADPH (Barajas *et al.*, 2015). Dans certaines voies, une seconde réduction assure la transformation de l'espèce libre en un produit alcool (Barajas *et al.*, 2015). Par exemple, le domaine R du module terminal MxaA d'une chaîne d'assemblage NRPS-PKS hybride catalyse une réduction non-processive de l'intermédiaire final en une fonction alcool primaire avec un transfert de quatre électrons et libère ainsi la myxalamide S (Barajas *et al.*, 2015) (**Figure 59**).

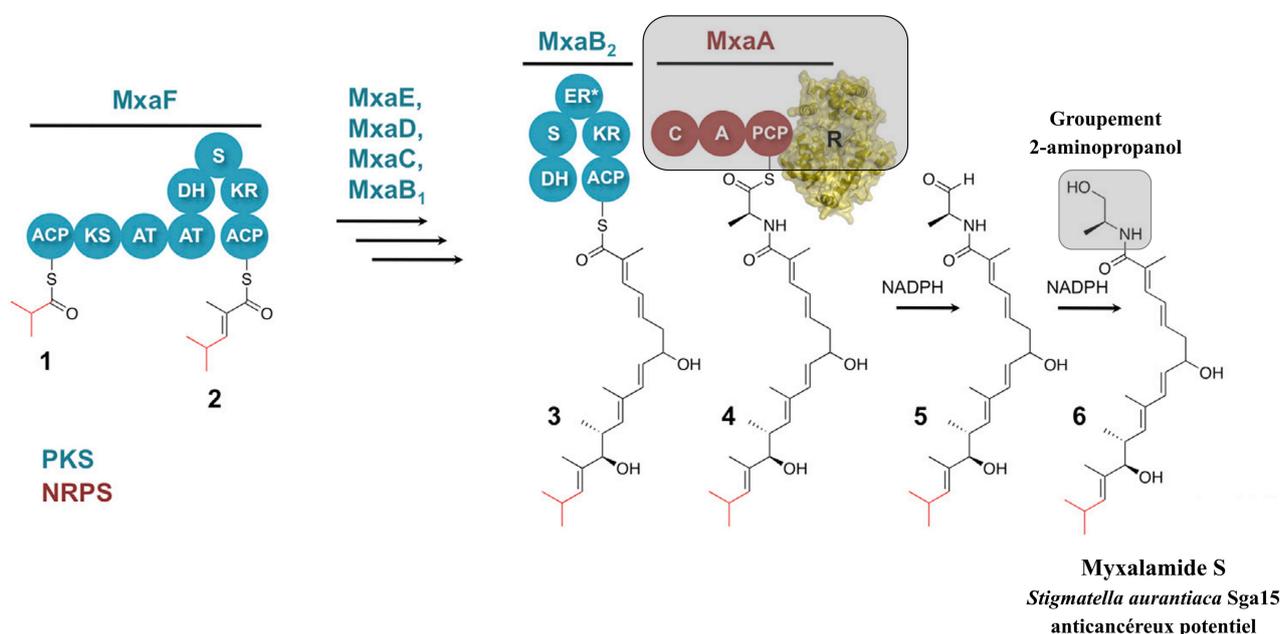


Figure 59. Représentation de la voie de biosynthèse de la myxalamide S (adaptée de Barajas *et al.*, 2015). La biogenèse de la myxalamide S implique notamment le modules PKS MxaF, E, D, C et B, ainsi que le module NRPS MxaA. Celui-ci condense un monomère L-Ala avec l'intermédiaire polyénique liés aux domaines porteurs associés, et le domaine R intégratif assure ensuite la libération réductrice du myxalamide S en deux étapes successives en présence du cofacteur NADPH. A : domaine d'adénylation, C : domaine de condensation, PCP : domaine porteur de peptidyle, R : domaine réductase, NADPH : nicotinamide adénine dinucléotide phosphate.

Par ailleurs, certains domaines R terminaux catalysent la cyclisation intramoléculaire de l'espèce carbonyle réactive libérée au niveau du groupement N-terminal du produit, comme cela est décrit dans le cas de l'aureusimine chez le pathogène *Staphylococcus aureus* (Süssmuth et Mainz, 2017). En outre, une enzyme autonome (SpsM) a été rapportée pour être responsable de la réduction de la sparsomycine produite chez la bactérie *Streptomyces sparsogenes* (Rui *et al.*, 2015).

Quelques études ont défini les bases structurales de domaines R excisés de modules terminaux (Süssmuth et Mainz, 2017). Par contre, aucune donnée expérimentale n'a encore pu être enregistrée sur l'interaction entre domaines R et PCP, certainement à cause de multiples états conformationnels. La structure cristallographique du domaine réductase du module MxaA a été déterminée en présence du cofacteur NADPH (Barajas *et al.*, 2015) (**Figure 60**). En effet, le domaine R présente une forte homologie avec les déshydrogénases/réductases à chaîne courte (*i.e.* SDR) (Barajas *et al.*, 2015). La région N-terminale est hautement conservée avec le repliement α/β de Rossmann étendu, composé d'un feuillet β à sept brins parallèles et flanqué de cinq hélices α (Barajas *et al.*, 2015) (**Figure 60**). Ce sous-domaine est responsable de la liaison au cofacteur et inclut la triade catalytique canonique, à savoir les résidus Thr1283, Tyr1311 et Lys1315 conservés, chacun localisé au niveau d'une hélice (3_{10} ou α_6), ainsi qu'un motif hélice-boucle-hélice caractéristique (Barajas *et al.*, 2015) (**Figure 60**). La région C-terminale est quant à elle impliquée dans la liaison du substrat et, de fait, elle présente une séquence variable par rapport aux homologues de la super-famille SDR (Barajas *et al.*, 2015). En particulier, elle comporte deux brins β parallèles et cinq hélices α , dont une autre paire d'hélices en tandem distinctive (α_{16} - α_{17}) (**Figure 60**). Ce motif structural contient des résidus hydrophobes souvent présents dans les domaines R terminaux de méga-synthases, et forme l'interface de liaison avec d'autres résidus non-polaires situés à une distance proche (Barajas *et al.*, 2015) (**Figure 60**).

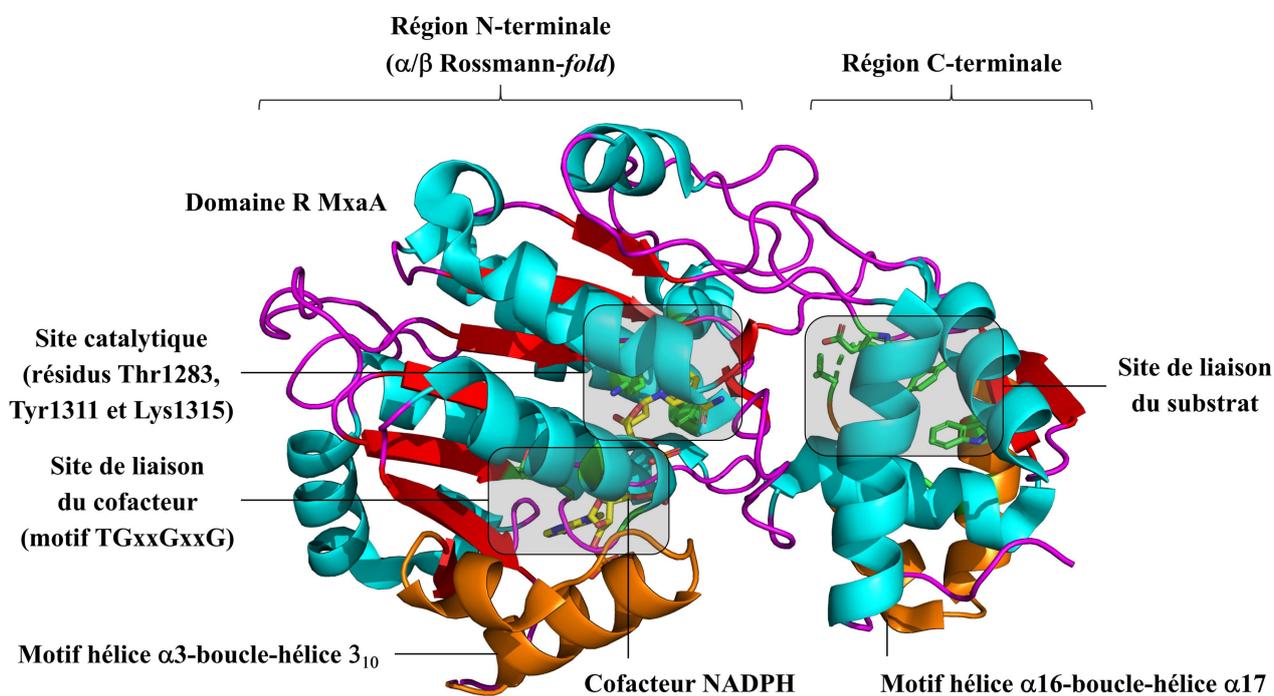


Figure 60. Représentation de la structure cristallographique du domaine R de la NRPS MxaA. Le domaine R du module MxaA (PDB : 4U7W) adopte un repliement canonique α/β de Rossmann, et comporte en plus une région C-terminale avec deux brins β parallèles et cinq hélices α . Le site actif possède le motif TGxxGxxG de liaison au cofacteur NADPH, ainsi qu'une triade de résidus catalytiques, tandis que le site de liaison du substrat se compose de multiples résidus hydrophobes au sein du motif hélice-boucle-hélice. NADPH : nicotinamide adénine dinucléotide phosphate.

Le sous-domaine α/β de Rossmann est impliqué dans la reconnaissance du cofacteur NADPH *via* le motif TGxxGxxG conservé, dont la fonction principale de certains résidus forme des liaisons hydrogènes avec la partie diphosphate (Barajas *et al.*, 2015) (**Figure 61A**). De plus, le groupement carbonyle du résidu Gly1155 établit aussi une liaison hydrogène avec la fonction hydroxyle (3') de l'adénosine, tandis que des résidus conservés (Thr1157, Arg1181 et Arg1191) interagissent avec le phosphate adjacent (Barajas *et al.*, 2015) (**Figure 61A**). De même, les chaînes latérales des résidus Tyr1311 et Lys1315 participent à des liaisons hydrogènes avec les groupements hydroxyles du ribose du noyau nicotinamide (Barajas *et al.*, 2015) (**Figure 61A**). Cet ensemble d'interactions spécifiques permet à la fois de positionner et d'orienter correctement le cofacteur au sein du site actif en vue de la réduction de l'intermédiaire réactionnel en alcool terminal (Barajas *et al.*, 2015). Pour cela, des molécules d'eau ordonnées pourraient participer à un système de relais de protons avec les fonctions hydroxyles du résidu Tyr1311, du ribose-nicotinamide, et amine du résidu Lys1315 (**Figure 61B**).

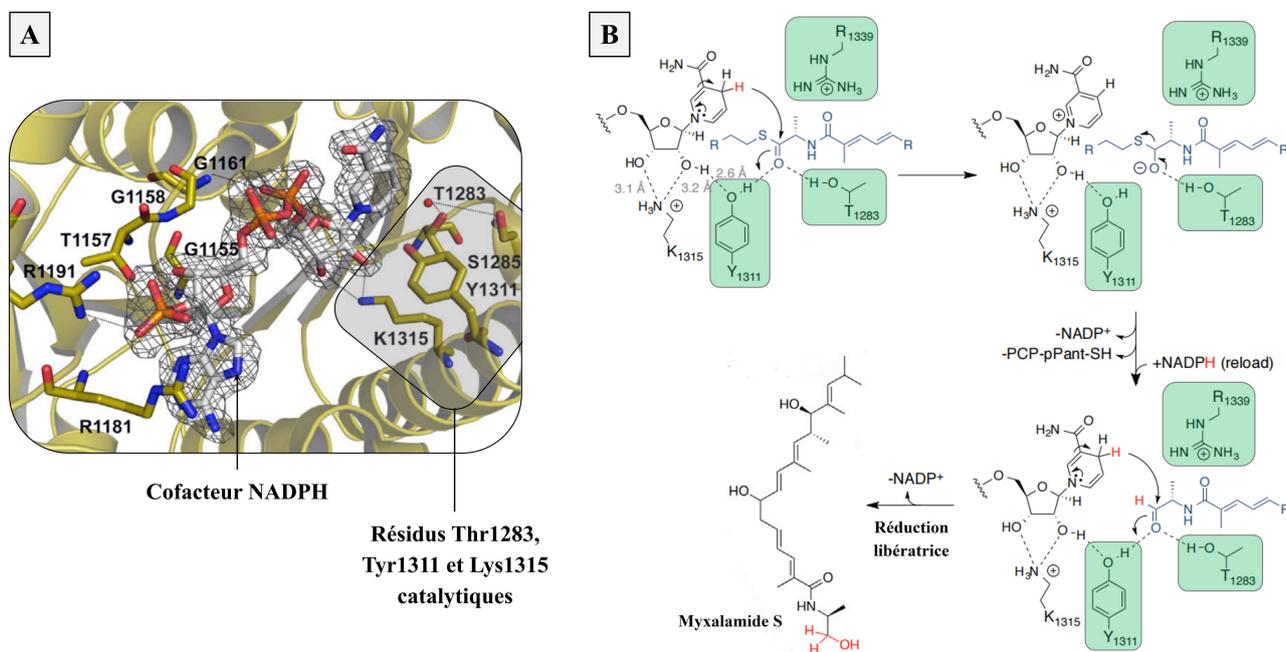


Figure 61. Liaison du cofacteur NADPH au sein du site actif du domaine R de la NRPS MxaA (adaptée de Barajas *et al.*, 2015). (A) Le domaine R du module MxaA (PDB : 4U7W) catalyse la réduction d'un intermédiaire réactionnel à l'aide du cofacteur NADPH, dont l'accommodation au sein du site actif requiert les résidus Gly1155, Thr1157, Gly1158, Leu1160 et Gly1161 du motif TGxxGxxG consensus, ainsi que la paire Arg1181 et Arg1191. Ces acides aminés forment des liaisons hydrogènes, *via* leur chaîne principale ou latérale, avec les groupements fonctionnels du cofacteur. (B). Le mécanisme catalytique implique la triade de résidus Thr1283, Tyr1311 et Lys1315 responsables des étapes de réduction en myxalamide S *via* un relais de protons et avec le transfert de quatre électrons.

Les domaines R terminaux permettent donc d'éviter les charges négatives (*e.g.* gramicidine D, myxalamide S), d'y greffer des groupements fonctionnels (*e.g.* glycolipopeptide) ou de générer des macrocycles (*e.g.* saframycine). Les activités correspondantes serviront sans aucun doute à l'avenir pour dicter le mécanisme de libération ou de cyclisation des NRP dans des systèmes rationnellement conçus (Süssmuth et Mainz, 2017).

IV.7.I. Modification par d'autres domaines rares

Les rares domaines FkbH retrouvés au sein de systèmes NRPS-PKS hydrides se substituent aux domaines A (Auerbach *et al.*, 2018). En effet, ces domaines atypiques catalysent le chargement du D-1,3-biphosphoglycérate sur le domaine PCP et la déphosphorylation du précurseur en une partie glycéryle, tel que cela est illustré dans le cas des vioprolides chez la myxobactérie *Cystobacter* sp. (Auerbach *et al.*, 2018) (**Figure 62**). Au cours de la biosynthèse de ces composés, une autre unité catalytique inhabituelle, un domaine ligase d'acide gras (FAAL) du module PKS en amont, recrute et active un monomère acétyle, tandis que le domaine C1 assure l'estérification du glycérate à partir du groupement α -hydroxyle de celui-ci (Auerbach *et al.*, 2018) (**Figure 62**). Suite à l'incorporation de huit autres résidus au sein de la chaîne d'assemblage VioA–D, le domaine TE catalyse ensuite la macrocyclisation de l'intermédiaire en pré-vioprolide B, dont la fraction lipidique a un impact sur la régiosélectivité de la réaction (Auerbach *et al.*, 2018). Une enzyme périplasmique inconnue catalyse enfin l'hydrolyse de la chaîne acyle, préalablement à l'export extracellulaire du métabolite bioactif (Auerbach *et al.*, 2018) (**Figure 62**). Des tests d'activité ont démontré que la vioprolide B possède un effet anticancéreux contre la lignée HCT-116 de cellules cancéreuses du côlon (Auerbach *et al.*, 2018). Ces essais ont aussi révélé que le produit pré-vioprolide montre des résultats comparables, ce qui suggère que l'acylation ne conditionne pas l'activité (Auerbach *et al.*, 2018).

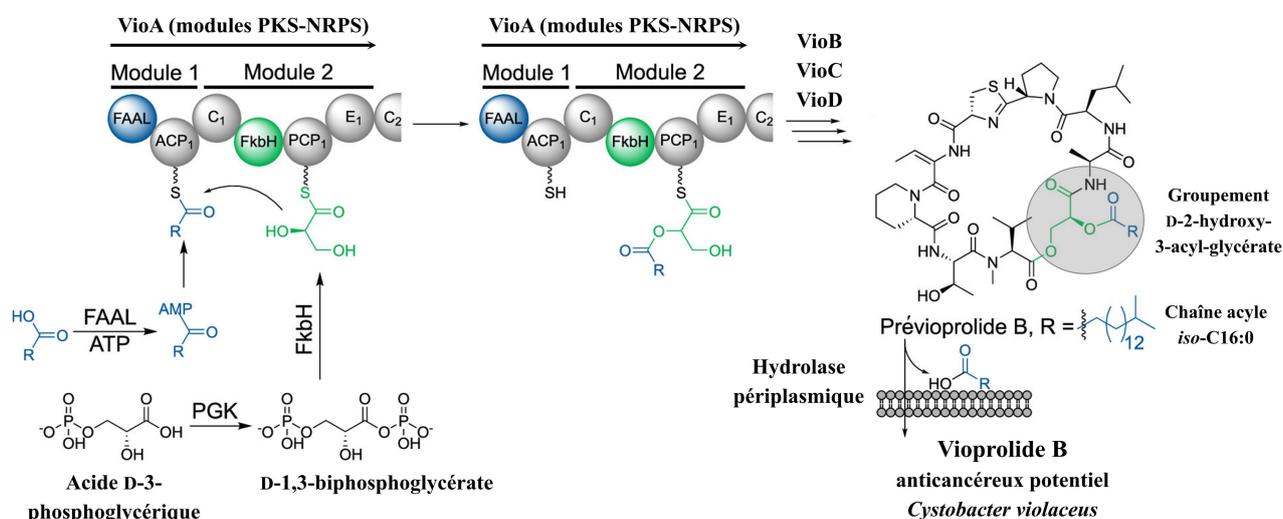


Figure 62. Représentation de la voie de biosynthèse de la vioprolide B (adaptée de Auerbach *et al.*, 2018). La genèse de la vioprolide B implique spécifiquement les sous-unités VioA–D. Au sein des deux premiers modules de VioA, des domaines FAAL, FkbH et C particuliers catalysent l'estérification d'un glycérate sur son groupement α -hydroxyle avec un acyle *iso*-C16:0. Suite à la libération du précurseur, une hydrolase cellulaire assure la rupture de la liaison ester, et le NRP mature gagne l'espace extracellulaire. PGK : phosphoglycérate kinase, FAAL : domaine ligase d'acide gras, FkbH : domaine FkbH, ATP : adénosine triphosphate, ACP : domaine porteur d'acyle, PCP : domaine porteur de peptidyle, C : domaine de condensation.

Autre exemple, le domaine BtrH terminal du système hydride PamA–H recrute une molécule de 4,3-spermidine et en catalyse la condensation nucléophile au niveau du résidu Gly du peptidyl-PCP (Müller *et al.*, 2014) (**Figure 63**). De manière alternative, l'enzyme post-NRPS PamI, homologue à une aminoglycoside *N*-acyltransférase, pourrait également assurer cette réaction (Müller *et al.*, 2014). Ces données indiquent qu'un mécanisme potentiellement nouveau pourrait médier la libération des NRP dans les lignes d'assemblage dépourvues de domaines TE (Müller *et al.*, 2014). Dans tous les cas, cette suite de réactions libère un précurseur, dont la peptidase périplasmique PamJ en catalyse la maturation en paenilamicines *via* le clivage du résidu acyl-D-Asn1 (Müller *et al.*, 2014) (**Figure 63**). Au sein du pathogène *Paenibacillus larvae*, quatre dérivés à activité antibiotique ont été décrits.

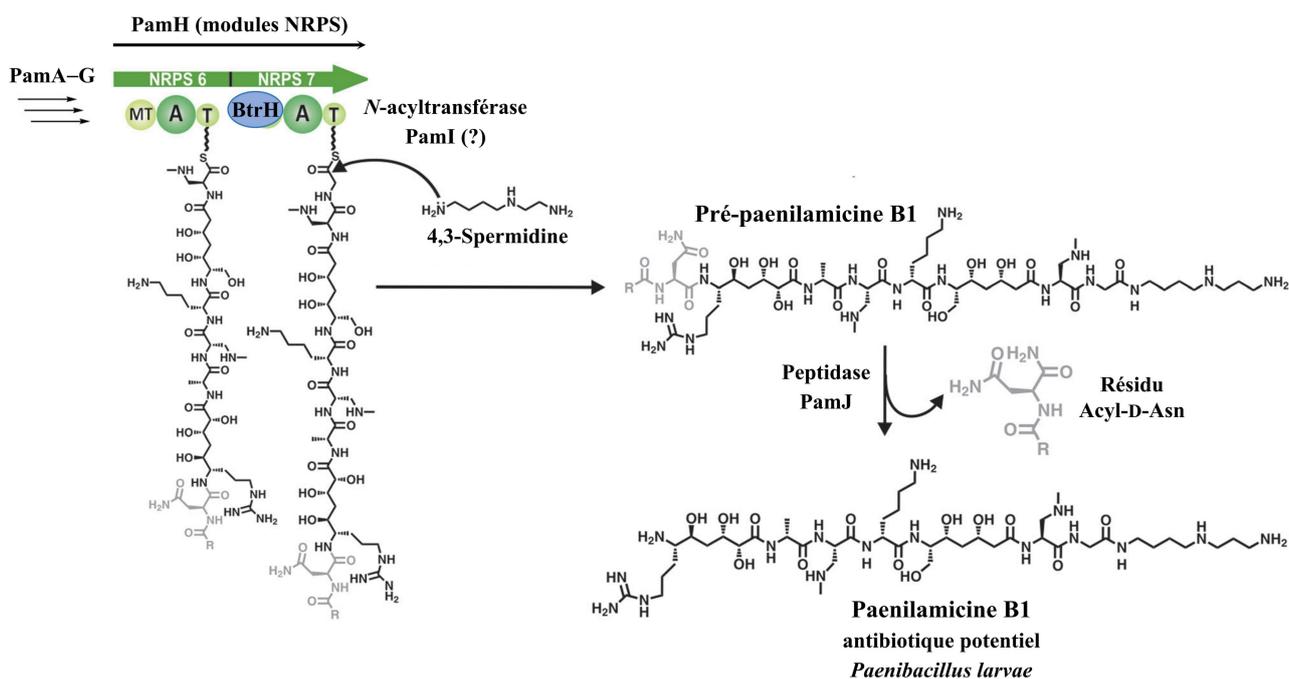


Figure 63. Représentation de la voie de biosynthèse de la paenilamicine B1 (adaptée de Müller *et al.*, 2014). La chaîne d'assemblage hybride des paenilamicines exploite un domaine BtrH responsable de la condensation nucléophile de la molécule de spermidine avec la fonction thioester-carbonyle du résidu glycine N-terminal du peptidyl-PCP, ce qui libère l'intermédiaire pré-paenilamicine B1. L'enzyme PamJ clive le résidu C-terminal de celui-ci et assure la sécrétion transmembranaire du produit. La paenilamicine B1 comprend une structure composée de blocs de construction uniques : les acides 2,3,5-trihydroxypentanoïque, *N*-méthyl-diaminopropionique, et galantinique, ainsi que la 4,3-spermidine. La partie N-terminale présente un résidu L-Arg décarboxylé, l'agmatine. A : domaine d'adénylation, T : domaine porteur de peptidyle, BtrH : domaine BtrH.

Dans l'ensemble, les domaines catalytiques auxiliaires, insérés dans les NRPS ou à activité *trans*, interagissent avec le peptidyl-PCP ou reconnaissent directement l'intermédiaire réactionnel afin de le modifier, ce qui augmente considérablement la diversité fonctionnelle des métabolites.

IV.8. Maturation post-NRPS des produits par des enzymes auxiliaires

Suite à la libération par les domaines TE, C_T ou R de produits de synthèse primaire, des enzymes auxiliaires en assurent la maturation finale, parfois nécessaire à leur bioactivité (Duban *et al.*, 2022). Ces enzymes autonomes post-NRPS regroupent des hydrolases, des oxydoréductases (*e.g.* CYP450, BVMO, SDR) et des transférases (*e.g.* glycosyltransférases), entre autres. Toutefois, ces enzymes de maturation restent relativement peu étudiées par rapport aux domaines NRPS, quelques dizaines de structures ayant été déposées à ce jour dans la banque de données protéiques (PDB).

Des processus pertinents de fonctionnalisation et de potentialisation concernent les antibiotiques glycopeptidiques (GPA), tels que la chloroéremomycine, la vancomycine ou la téicoplanine (Kahne *et al.*, 2005) (**Figure 64**). Consécutivement à la libération des heptapeptides réticulés des NRPS (se référer à la section **IV.4.B.ε.**), des *O*-glycosyltransférases catalysent des modifications spécifiques à partir d'hexoses et de nucléosides TDP (Kahne *et al.*, 2005). Par exemple, lors de la biosynthèse de la chloroéremomycine, GtfB transfère un monomère de D-glucose sur la chaîne latérale du résidu 4-hydroxyphényl glycine de l'aglycone AGV (Kahne *et al.*, 2005) (**Figure 65**). Ensuite, GtfA greffe le monomère L-4-*épi*-vancosamine sur le groupement β-hydroxyle du résidu L-Tyr du produit (Kahne *et al.*, 2005) (**Figure 65**). Enfin, GtfC catalyse une réaction identique, mais sur la fonction hydroxyle en C2 du fragment glycosyle de la chloroorienticine B (Kahne *et al.*, 2005) (**Figure 65**). Les enzymes orthologues GtfE et GtfD assurent les glycosylations de l'aglycone de la vancomycine, alors qu'une mannosyltransférase et une acyltransférase diversifient en plus l'aglycone de la téicoplanine (Kahne *et al.*, 2005) (**Figure 64**). Les hexoses présentent ainsi une nature et des positions différentes dans les GPA, ce qui augmente à la fois la solubilité et la capacité de liaison de ces antibiotiques avec la paroi bactérienne, et influe donc sur leur bioactivité respective (Mulichak *et al.*, 2004).

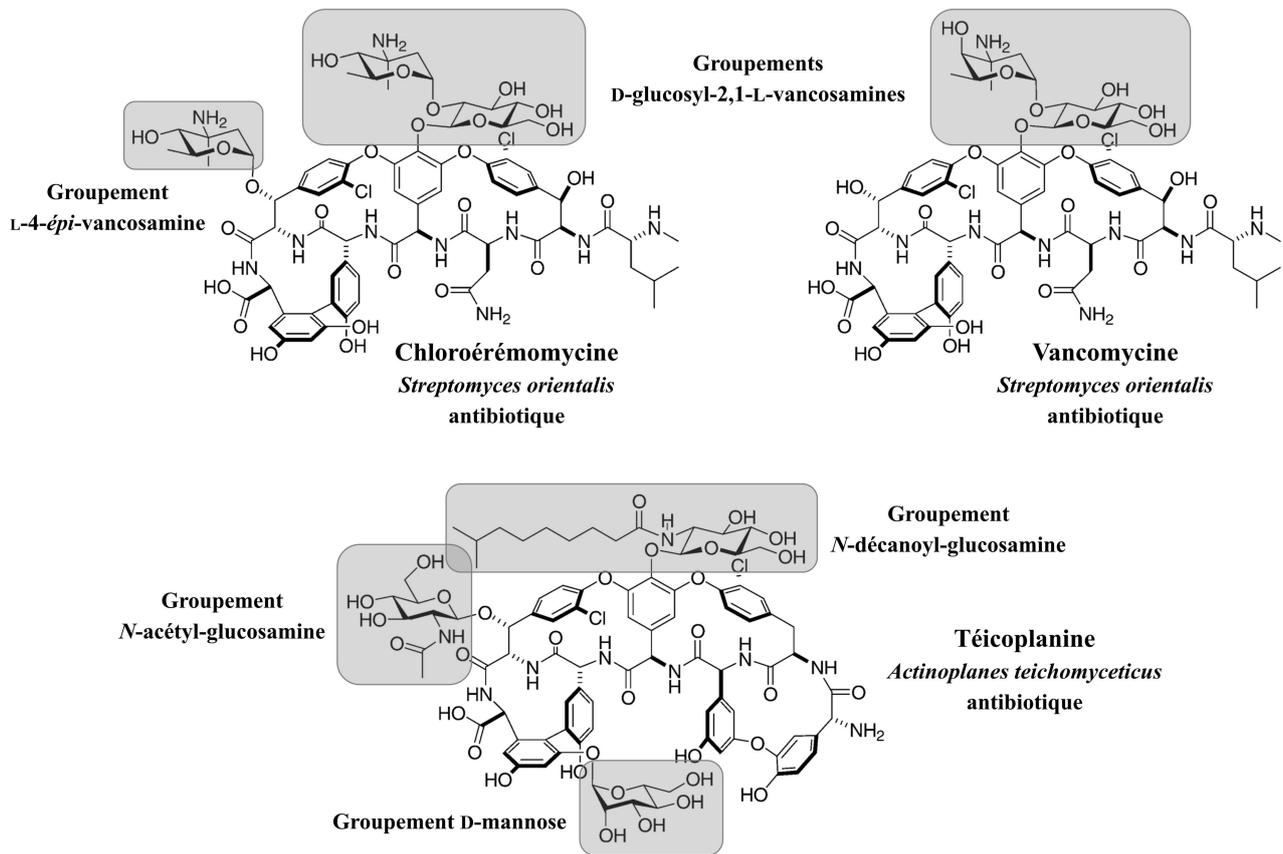


Figure 64. Représentation de structures chimiques de (lipo)-glycopeptides. Les antibiotiques glycopeptidiques (GPA) incluent notamment la chloroérymomycine et la vancomycine apparentées de la bactérie *Streptomyces orientalis*, dont le premier présente une fonction L-4-épi-vancosamine supplémentaire, et la téicoplanine avec des groupements lipidiques (*N*-acétyl/*N*-décanylglucosamine) additionnels.

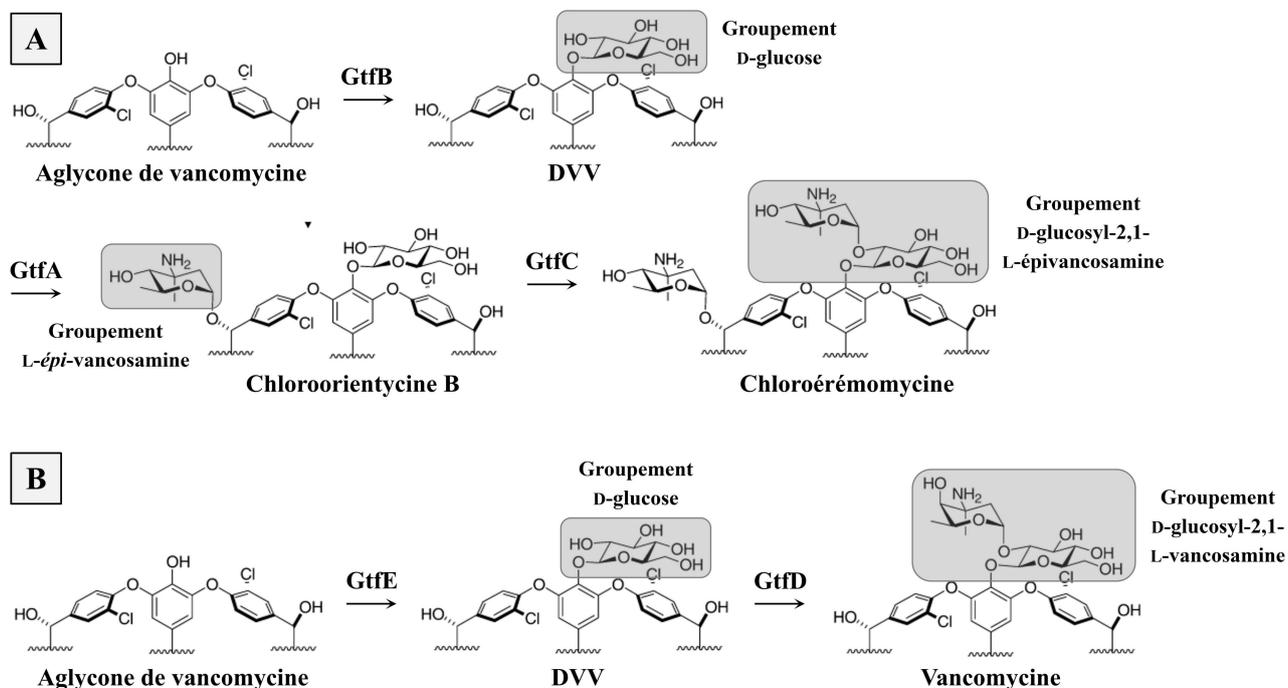


Figure 65. Maturation de la chloroérymomycine et de la vancomycine par les *O*-glycosyltransférases (adaptée de Kahne *et al.*, 2005). À partir de l'aglycone de la vancomycine et de dTDP-hexoses, les enzymes orthologues GtfB/GtfE et GtfC/GtfD catalysent l'addition de groupements D-glucose et L-épi-vancosamine. En plus, GtfA catalyse l'ajout d'un autre monomère sur le résidu β -hydroxytyrosine de la structure monoglycosylée (DVV).

La structure cristallographique de l'*O*-glycosyltransférase GtfD a été déterminée en présence des analogues du substrat donneur (TDP et D-glucose, à la place de TDP-L-vancosamine) et du substrat accepteur (DVV, dévancosaminy-vancomycine) (Mulichak *et al.*, 2004) (**Figure 66**). Ce complexe ternaire a fourni un aperçu détaillé de la spécificité de substrat et de la régiospécificité du transfert de l'hexose (Mulichak *et al.*, 2004). L'enzyme GtfD présente le repliement GT-B caractéristique en sous-domaines N- et C-terminaux, précédemment observé pour des homologues, y compris GtfB et GtfA (Mulichak *et al.*, 2004). Ces régions adoptent chacune un noyau de type Rossmann, composé de brins β parallèles et d'hélices α entrelacées, et se connectent au moyen d'une longue séquence *linker* (Mulichak *et al.*, 2004) (**Figure 66**). La fente profonde située entre les sous-domaines forme le site de liaison du substrat donneur, en l'occurrence le TDP-L-4-vancosamine, et des interactions inter-domaines (liaisons hydrogènes) stabilisent cette interface (Mulichak *et al.*, 2004) (**Figure 66**). Au niveau de la surface adjacente du sous-domaine N-terminal, les hélices α_3 , α_4 et α_7 , ainsi que les boucles intermédiaires hypervariables N3 et N5, créent le site de liaison du substrat accepteur, en particulier l'espèce monoglycosylée DVV (Mulichak *et al.*, 2004) (**Figure 66**). La structure de l'enzyme adopte un état dit en « fente fermée », à l'instar du complexe GtfA-TDP auparavant résolu, suite à la liaison du TDP-hexose et au repliement induit de la boucle β_7 - β_8 (Mulichak *et al.*, 2004). Cet ajustement conformationnel mineur permet le rapprochement des substrats donneur et accepteur, positionnés et orientés correctement au sein de leur poche de liaison respective en vue de la réaction catalytique (Mulichak *et al.*, 2004).

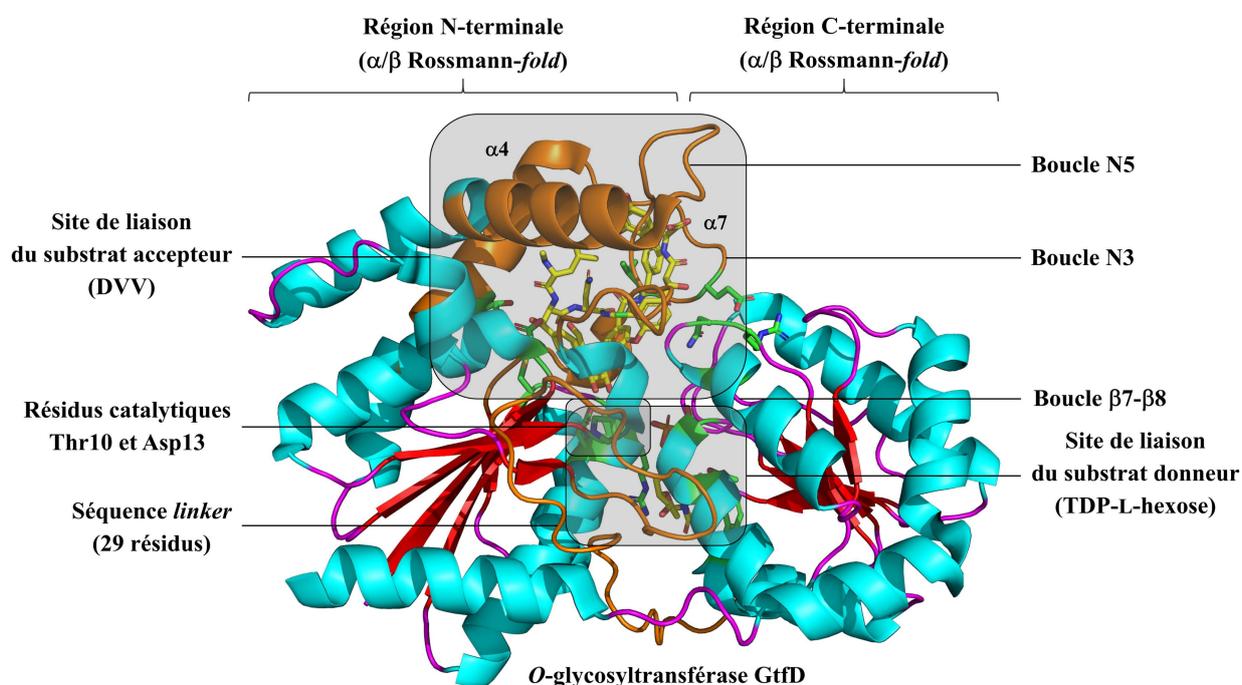


Figure 66. Représentation de la structure cristallographique de l'*O*-glycosyltransférase GtfD. L'enzyme GtfD (PDB : 1RRD) présente des sous-domaines N- et C-terminaux, dont chacun adopte le repliement caractéristique α/β de Rossmann. Ces régions forment à leur interface le site de liaison du substrat donneur d'hexose. Les hélices α_3 , α_4 et α_7 , et boucles N3 et N5, créent le site de liaison du substrat accepteur. Le site actif contient les résidus catalytiques Thr10 et Asp13. DVV : dévancosaminy-vancomycine, TDP : thymidine diphosphate.

En accord avec le mécanisme catalytique proposé pour des enzymes ayant un repliement GT-B, la base générale Asp13 soustrait dans un premier temps le proton de la fonction hydroxyle réactive de la partie glycosyle du substrat accepteur DVV (Mulichak *et al.*, 2004). Ensuite, l'atome d'oxygène O2' réalise une attaque nucléophile sur le centre C1 de l'hexose donneur, ce qui assure le transfert de celui-ci et ainsi la formation du groupement D-glycosyl-2,1-L-vancosamine du métabolite produit (Mulichak *et al.*, 2004). La chaîne latérale du résidu Thr10 lie à la fois l'aglycone et le ligand TDP, et stabilise les intermédiaires réactionnels (Mulichak *et al.*, 2004).

Compte tenu de la régiospécificité des *O*-glycosyltransférases GtfA–E, celles-ci ont été utilisées avec succès en présence de différents aglycones et NDP-hexoses afin de produire des antibiotiques glycopeptidiques dotés d'une puissance accrue ou encore d'une nouvelle activité (*e.g.* dalbavancine, oritavancine, *etc.*) (Marcone *et al.*, 2018). Ces approches de dérivation chimioenzymatique et de biosynthèse combinatoire contribueront à élargir la diversité chimique de cette classe de composés et faire face à l'évolution constante de la résistance aux antibiotiques (Marcone *et al.*, 2018).

MODÈLES D'ÉTUDE ET OBJECTIFS

La biosynthèse des alcaloïdes pyrrolizidiniques (AP) bactériens est un exemple remarquable de la diversité chimique des produits naturels bioactifs. L'assemblage de ces hétérocycles implique une NRPS bimodulaire de type I et une monooxygénase de Baeyer-Villiger (BVMO), ainsi que diverses enzymes auxiliaires et réactions chimiques spontanées. Des travaux récents ont montré que plusieurs de ces métabolites atténuent la virulence d'une bactérie productrice lors de l'interaction avec l'hôte. Cependant, les mécanismes moléculaires qui régissent la variété chimique des AP et déterminent leur activité biologique demeurent inconnus.

Cette section détaille l'identification des BGC des AP chez des bactéries (partie **I.1.**), ainsi que la caractérisation des structures chimiques et des activités biologiques de ces produits (partie **I.2.**). Elle expose ensuite les intérêts de la production d'analogues à travers l'exemple de la dérivatisation chimique d'une structure (bohémamine) (partie **I.3.**). Elle met enfin en évidence la biosynthèse des AP d'intérêts, avec un accent particulier sur la maturation post-assemblage (partie **II.**), et décrit les objectifs de la thèse en lien avec cela (partie **III.**).

I. Identification, caractérisation et dérivatisation des AP bactériens

I.1. Identification des BGC AP chez des γ -protéobactéries et actinobactéries

Dans le cadre des efforts continus des collaborateurs afin de découvrir la diversité chimique des NRP, des stratégies de *genome mining* ont permis d'identifier des clusters de gènes impliqués dans la production d'AP chez des γ -protéobactéries (*e.g. Xenorhabdus, Pseudomonas*) et actinobactéries (*e.g. Streptomyces*) (Bode *et al.*, 2019 ; Hong *et al.*, 2019) (**Figure 67**). Ces BGC divergent à la fois dans la composition et l'organisation génique, ce qui sous-tend la diversité importante des structures chimiques des produits biosynthétisés. Dans un premier temps, les clusters d'intérêt ont (1) soit été activés par couplage de l'inactivation du gène *hfq*, codant un régulateur du métabolisme spécialisé, et de la permutation de promoteurs au sein de l'hôte natif (*Xenorhabdus* spp.), (2) soit étudiés par inactivation génique (*Pseudomonas aeruginosa*), (3) ou alors exprimés dans un châssis hétérologue (*Streptomyces* spp.) (Bode *et al.*, 2019 ; Hong *et al.*, 2019).

Ensuite, les souches bactériennes modifiées ont été soumises à des profilages métaboliques (LC-HRMS) et à des analyses spectroscopiques (UV, RMN). Ces expériences ont permis de caractériser les composés produits, tels que la pyrrolizwilline ou les azétidomonamides A/B chez les pathogènes *X. hominickii* et *P. aeruginosa*, respectivement, dont la biosynthèse sera détaillée à la section **II.1.**

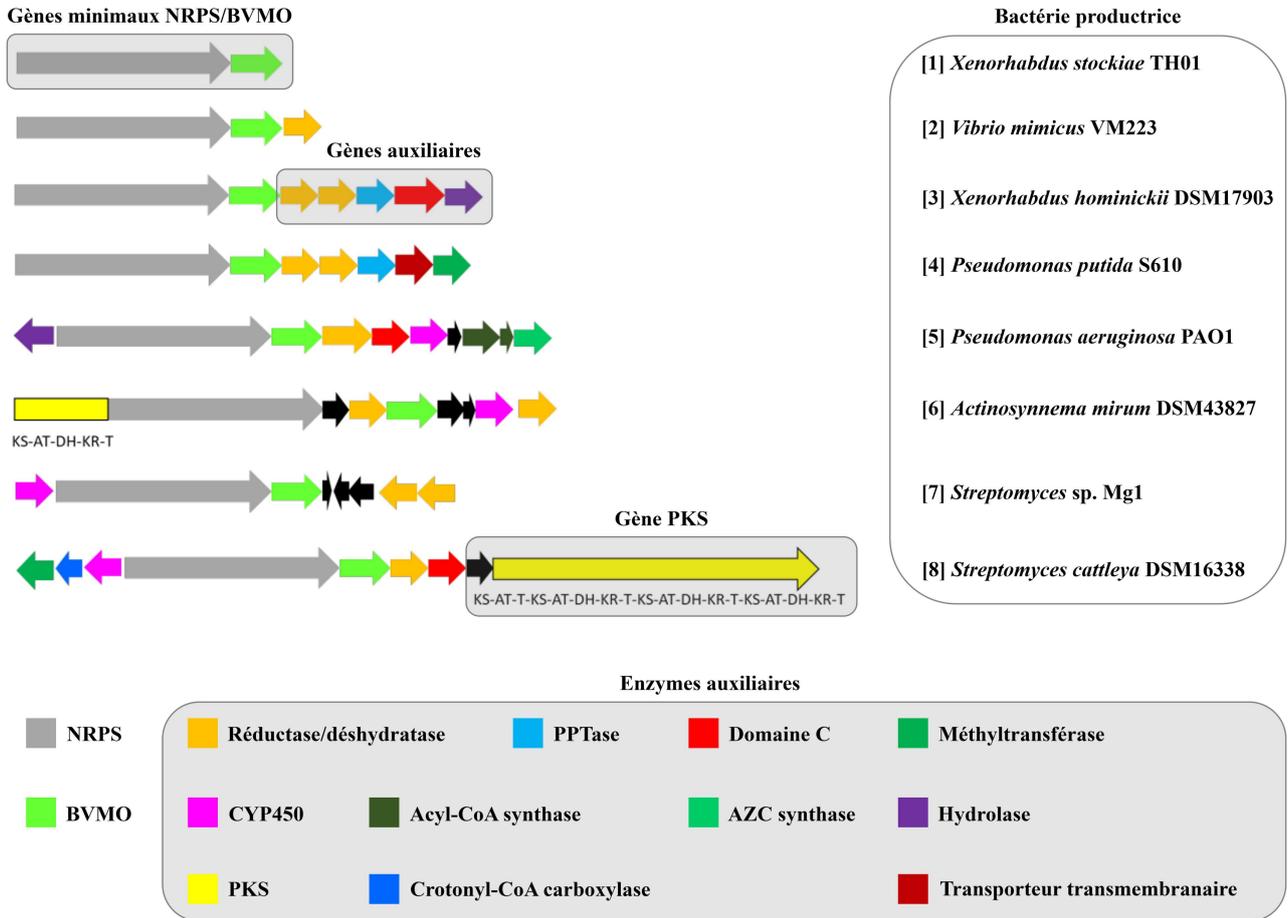


Figure 67. Organisation des BGC des AP chez des γ -protéobactéries et actinobactéries (adaptée du projet ANR-20-CE92-0038). Les BGC impliqués dans la biosynthèse des AP bactériens se retrouvent chez des γ -protéobactéries (*e.g.* *Xenorhabdus*, *Vibrio*, *Pseudomonas*, *Acinetobacter*, *etc.*) et des actinobactéries (*e.g.* *Streptomyces*, *Saccharothrix* *etc.*). Ils codent au minimum une NRPS bimodulaire de type I et une mono-oxygénase de Baeyer-Villiger (BVMO) (système [1]), responsables de l'assemblage du squelette bicyclique, et fréquemment des enzymes auxiliaires. Celles-ci regroupent des enzymes de synthèse de précurseurs, telles qu'une acyl-CoA synthase et une AZC synthase, dans le cas du système [5] du pathogène *Pseudomonas aeruginosa*. Elles incluent également des enzymes de maturation post-NRPS, y compris des oxydoréductases (*e.g.* CYP450, SDR), des transférases, comme une méthyltransférase (systèmes [4] et [8]), ou des hydrolases (système [3]). Par ailleurs, ces BGC peuvent exprimer des machineries NRPS-PKS hybrides ([6] et [8]).

I.2. Caractérisation de la structure chimique et de l'activité biologique des AP bactériens

Les AP bactériens, dont quelques dizaines ont été caractérisés, incluent les clazamycines, les jénomidines, les bohémamines, les spithionéines, les légonmycines et les pyrrolizixénamides, ainsi que des analogues naturels (cyclocarbamides, brabantamides, et spinoxazines) (Horiuchi *et al.*, 1979 ; Isogai *et al.*, 1986 ; Hu *et al.*, 2003 ; Schmidt *et al.*, 2014 ; Huang *et al.*, 2015 ; Schimming *et al.*, 2015 ; Fu *et al.*, 2016) (**Figure 68**). Ces produits présentent un noyau pyrrolizidine, deux cycles pyrroles accolés *via* l'atome d'azote, et divers groupements chimiques ajoutés sur un ou plusieurs atomes de carbone (Fang *et al.*, 2021). Cette diversité structurale importante est à l'origine de leurs rôles écophysiologiques variés, sous-jacents à leurs propriétés antibiotiques, antitumorales ou anti-inflammatoires prometteuses (Huang *et al.*, 2015). À cet égard, ces métabolites spécialisés suscitent un intérêt considérable de la part des laboratoires de recherche et industries pharmaceutiques, mais leur utilisation thérapeutique doit encore être évaluée (Fang *et al.*, 2021).

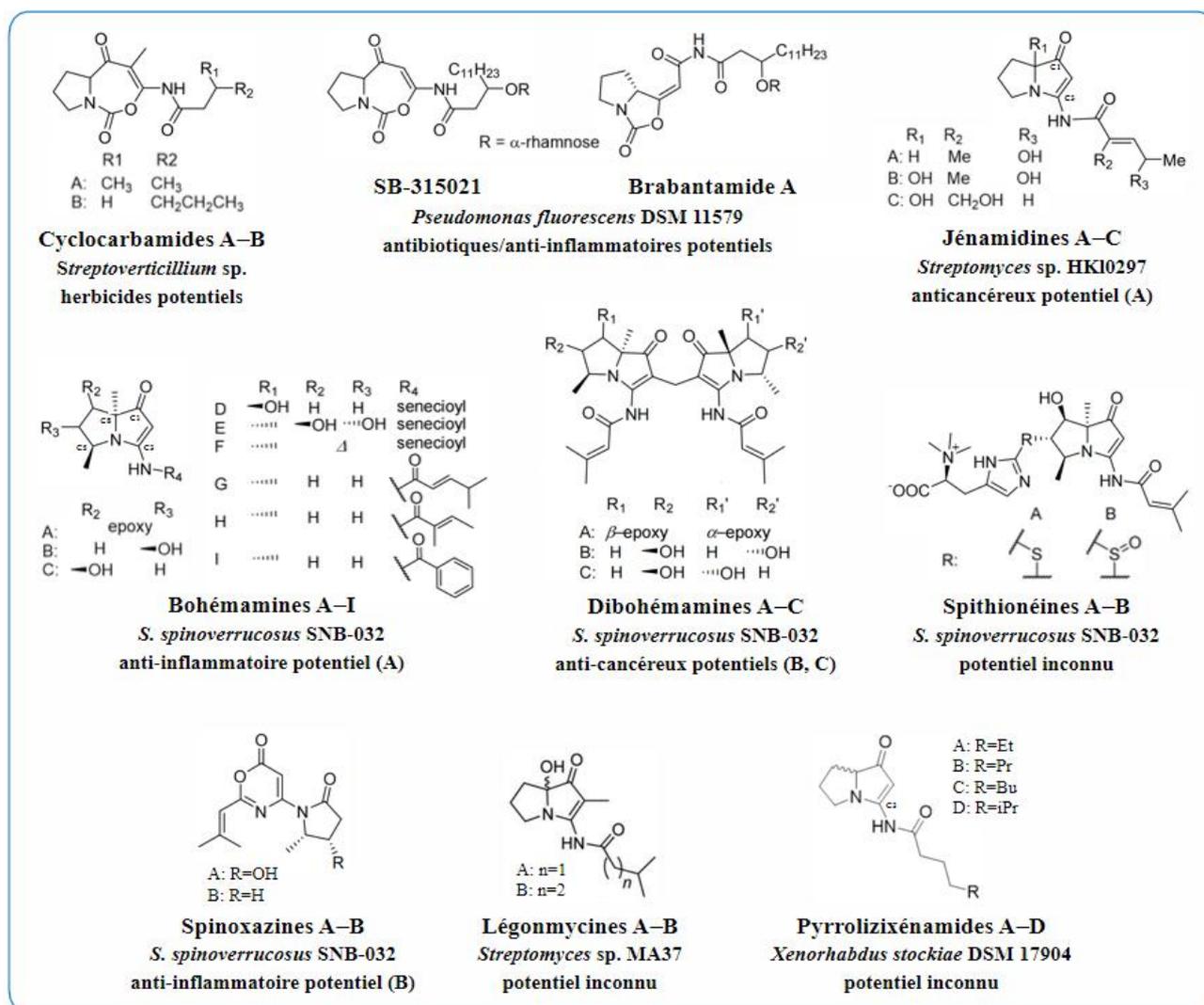


Figure 68. Représentation de structures chimiques et utilisation potentielle d'AP bactériens et analogues naturels.

Les AP sont essentiellement synthétisés par certains genres bactériens, et ont une structure chimique basée sur un noyau pyrrolizidine ramifié de groupements fonctionnels. Des dérivés naturels sont produits, comme les cyclocarbamides ou les oxazinones. Assez peu de voies de biosynthèse ont été dévoilées et, dans tous les cas, la bioactivité potentielle des métabolites nécessite d'être étudiée de manière approfondie.

Par exemple, les jénamidines A–C ont été découvertes à partir du filtrat de culture de la bactérie *Streptomyces* sp. HK10297 par une approche de criblage chimique (Hu *et al.*, 2003) (**Figure 68**). Ces composés possèdent un squelette ramifié avec des groupements aldéhyde (C1) et hydroxyle (radical R1), ainsi qu’avec une chaîne buténamide méthylée (Hu *et al.*, 2003). La jénamidine A montre des effets antiprolifératifs contre la lignée K-562 de la leucémie myéloïde chronique, contrairement aux isomères B et C dépourvus d’activité (Hu *et al.*, 2003).

Un autre métabolite, la bohémamine A, a été initialement isolé de la souche *Actinosporangium* sp. C36145 à partir d’un mélange complexe d’anticancéreux (anthracyclines) co-produits (Doyle *et al.*, 1980) (**Figure 68**). Ce composé présente une structure tricyclique singulière, avec un époxyde formé *via* les atomes de carbone C6 et C7. De plus, le noyau est ramifié de groupements méthyles aux positions C5 et C8, et également d’une chaîne *iso*-buténamide (Doyle *et al.*, 1980). Par ailleurs, le produit inhibe spécifiquement l’adhésion entre les cellules HL-60 et CHO-ICAM-1, ce qui laisse présager un usage potentiel en tant qu’anti-inflammatoire (Zhang *et al.*, 2003).

Des analogues naturels de ce métabolite ont été caractérisés avec les bohémamines B–I chez la souche *S. spinoverrucosus* SNB-032 (Fu *et al.*, 2016) (**Figure 68**). Aucun des dérivés ne possède plus le cycle époxyde avec, à la place, une ou deux fonctions hydroxyles (B, C et E), ou des atomes de carbone secondaires (D et G–I) ou tertiaires (F). Seule la bohémamine D présente un groupement hydroxyle à la position C8, au lieu d’une fonction méthyle commune. Les analogues G–I sont des isomères de constitution : de la longueur de la chaîne radicalaire (G, 4-méthylpent-2-énamide, H, 2-méthylbut-2-énamide) ou de la nature de celle-ci (I, phénylamide) (Fu *et al.*, 2016).

Des nouveaux AP, la spithionéine A et son isomère sulfoxide (B), ont aussi été identifiés dans les extraits de la bactérie productrice (Fu *et al.*, 2016) (**Figure 68**). Ces composés présentent un squelette identique à la bohémamine C avec, en plus, un fragment inhabituel L-ergothionéine à la position C6 (radical R2). Ils ne démontrent aucune cytotoxicité contre des lignées cancéreuses (du poumon), ni d’activité antibactérienne contre des pathogènes courants (Fu *et al.*, 2015).

D’autres produits apparentés issus de la souche, les spinoxazines, dérivent de la bohémamine D *via* une acylation intramoléculaire (dérivé B) entre les fonctions aldéhydes, suivie d’une oxydation (dérivé A) (Kim *et al.*, 2015 ; Fu *et al.*, 2016) (**Figure 68**). Ces réactions conduisent à la formation de structures conjuguées avec des cycles 2-méthylpropényle-1,3-oxazin-6-one et 1-oxopyrrolidinyle (Kim *et al.*, 2015). La spinoxazine B inhibe spécifiquement la production de monoxyde d’azote dans les cellules microgliales BV-2, ce qui laisse supposer des propriétés anti-inflammatoires, tandis que son isomère est inactif (Kim *et al.*, 2015).

Parallèlement, les légonmycines ont été isolées chez la bactérie *Streptomyces* sp. MA37 et leurs structures chimiques ont été élucidées (Huang *et al.*, 2015) (**Figure 68**). Celles-ci ont le squelette classique d'un AP, substitué en particulier avec des groupements aldéhyde, méthyle, et hydroxyle aux positions C1, C2, et C8, respectivement, ainsi qu'une chaîne *iso*-butane/propanamide (Huang *et al.*, 2015). Aucun test d'activité n'a été rapporté pour ces métabolites.

Les pyrrolizixénamides, quant à eux identifiés dans la souche *Xenorhabdus stockiae* DSM17904, possèdent un radical alkylamide d'une longueur variable à la position C3 (Schimming *et al.*, 2015) (**Figure 68**). Un des intermédiaires de leur biosynthèse partage une structure très similaire avec des cyclocarbamides, en l'occurrence la brabantamide A, isolée au sein de *Pseudomonas* spp. (Busby *et al.*, 2000 ; Schmidt *et al.*, 2014) (**Figure 68**). Ce produit bicyclique est doté d'une chaîne d'acide gras saturé à 14 atomes de carbone, dont l'un est greffé d'un fragment *O*-rhamnose (Schmidt *et al.*, 2014). La brabantamide A inhibe la lipoprotéine liée à la phospholipase A2 (Lp-PLA2), responsable du processus inflammatoire au sein des parois artérielles (Schmidt *et al.*, 2014). Le glycolipopeptide a donc été considéré comme un point de départ pour développer des traitements contre des maladies cardiovasculaires (*e.g.* athérosclérose), malgré que les essais cliniques soient désormais interrompus, faute de l'efficacité probante du médicament (darapladib) (Choi *et al.*, 2018).

La recherche réalisée sur les alcaloïdes pyrrolizidiniques bactériens met en lumière des progrès majeurs dans l'identification des BGC, ainsi que dans la caractérisation des structures chimiques et des activités biologiques. Cependant, l'émergence de bactéries multi-résistantes à des antibiotiques et la persistance de maladies chez certains sujets fait valoir le besoin d'améliorer la compréhension du rôle des dérivés d'AP. Dans ce contexte, certaines structures ont été dérivatisées par des stratégies d'hémi-synthèses chimiques pour développer le potentiel pharmaceutique de ces composés à haute valeur ajoutée (Fu *et al.*, 2016).

I.3. Dérivatisation chimique des AP bactériens

Un nombre important de produits naturels, y compris des AP bactériens, impliquent des étapes non-enzymatiques au cours de leur biosynthèse (Fu *et al.*, 2016). Dans ce cas, il est plutôt judicieux d'exploiter la chimie non-canonique *via* l'apport des précurseurs correspondants dans le milieu afin de dérivatiser la structure originelle et potentiellement générer des dérivés bioactifs (Fu *et al.*, 2016). Les voies métaboliques forment généralement un intermédiaire électrophile, tel qu'un aldéhyde ou une imine, puis une molécule endogène ou exogène réalise une attaque nucléophile sur l'espèce en question, ce qui donne un composé hybride (Fu *et al.*, 2016).

Un exemple pertinent concerne les dimères de bohémamine (nucléophile), découverts au sein de la bactérie *S. spinoverrucosus* SNB-032, avec un pont méthylène entre les noyaux pyrrolizidines (Fu *et al.*, 2016). Le mécanisme non-enzymatique correspond à une réaction de Baylis-Hillman, lors de laquelle l'atome de carbone C2 de l'intermédiaire A réalise une attaque nucléophile sur le méthanal, ce qui forme une espèce avec un groupement alcool primaire (Fu *et al.*, 2016) (**Figure 69**). Ensuite, cet intermédiaire est susceptible d'être protoné dans les conditions du milieu de fermentation, avant qu'une bohémamine équivalente à l'espèce initiale effectue une substitution de type S_N1 au niveau du carbocation, ce qui produit le dimère d'AP (Fu *et al.*, 2016) (**Figure 69**). Cette hypothèse a été confirmée à travers la dérivatisation semi-synthétique de la bohémamine A en présence de méthanal dans le milieu réactionnel (Fu *et al.*, 2016). Les dibohéamines B et C diffèrent du composé A *via* la réduction du groupement époxyde en diol et montrent une activité antitumorale contre la lignée NSCLC A549 des cellules cancéreuses du poumon, contrairement aux monomères correspondants (Fu *et al.*, 2016). Cela indique que le processus de dimérisation conditionne en partie la bioactivité de ces métabolites inhabituels.

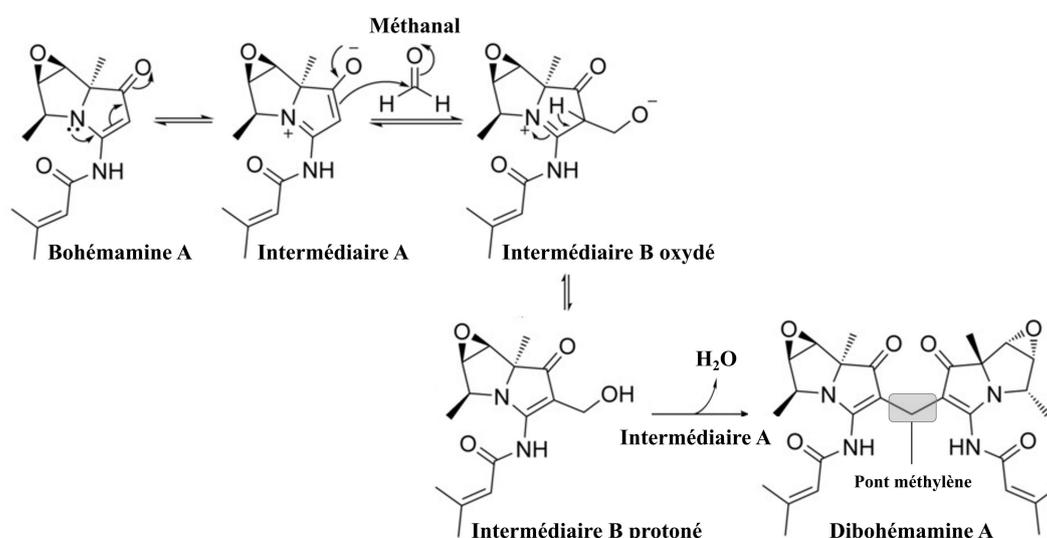


Figure 69. Dimérisation de la bohémamine A en présence de méthanal endogène (adaptée de Fu *et al.*, 2016). La réaction implique une attaque nucléophile de l'intermédiaire A sur le méthanal, puis une substitution de type S_N1 entre l'intermédiaire B protoné et un nouvel intermédiaire A, ce qui produit la dibohémamine A.

Ainsi, des banques d'analogues alkylés et arylés de dibohémamines A–C ont été générées afin de maximiser leur cytotoxicité, voire développer d'autres bioactivités (Fu *et al.*, 2016) (**Figure 70**). Un certain nombre de ces dérivés sont au moins aussi puissants que les molécules originelles et tous les analogues actifs contiennent des groupements hydroxyles au lieu du cycle époxyde (Fu *et al.*, 2016).

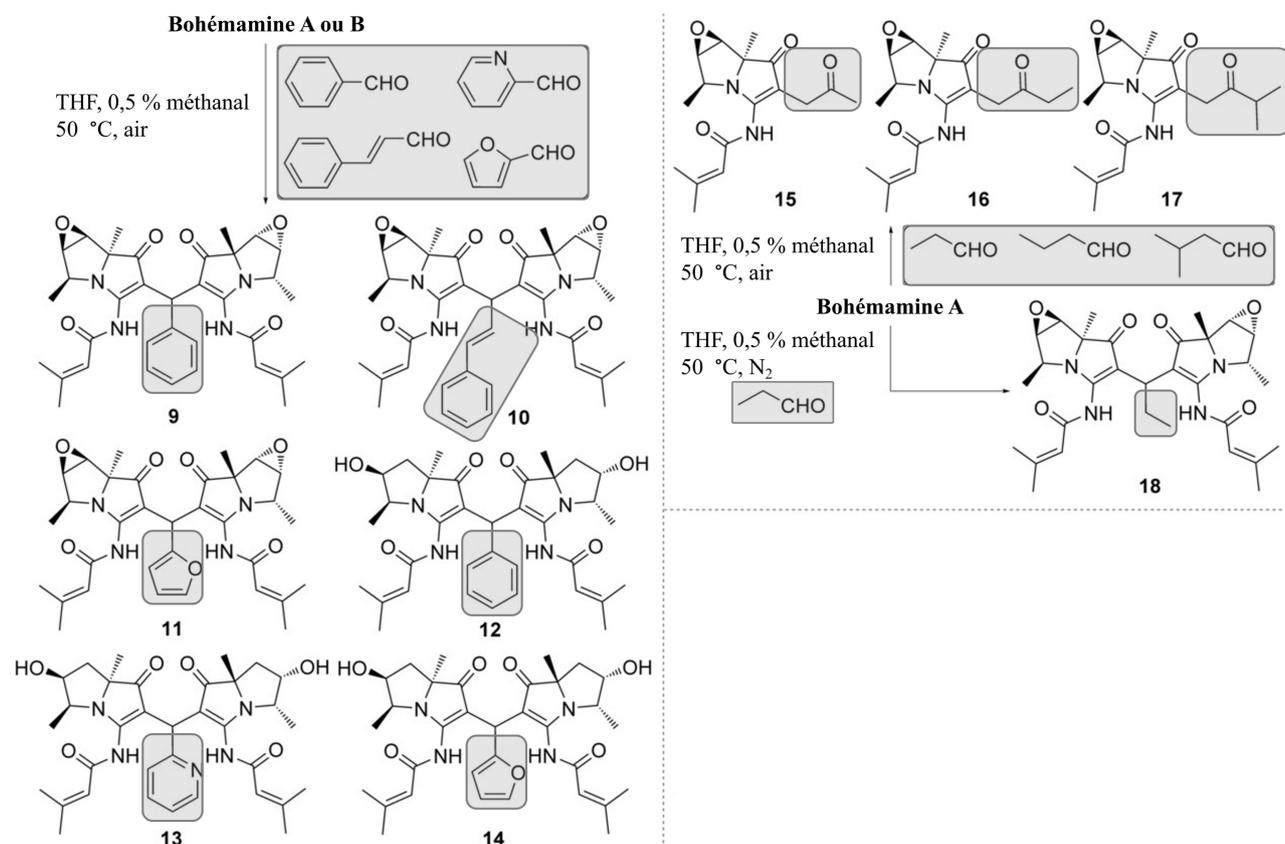


Figure 70. Création d'une banque de bohémamines en présence de radicaux linéaires ou cycliques (adaptée de Fu *et al.*, 2016). Des arylaldéhydes comme le benzaldéhyde, le 2-pyridinecarbaldéhyde, le 3-phénylpropénal ou encore le 2-furancarbaldéhyde ont été incorporés dans les bohémamines A ou B et cela a généré les dimères 9–14. Par contre, des alkylaldéhydes, tels que le propional, le butanal, ou l'*iso*-pental, ont empêché la dimérisation et ceci a donné lieu à une série de monomères oxydés (15–17). Cette réaction répétée sous une atmosphère d'azote en présence de méthanal et de propional est à l'origine de la dibohémamine A attendue (18). THF : tétrahydrofurane.

En raison de leurs structures complexes, l'immense majorité des NRP bioactifs a été créée par semi-synthèse chimique (Fu *et al.*, 2016). Parmi les AP bactériens précisément, un nombre croissant d'étapes non-enzymatiques ont été rapportées pour leur biogenèse (Fu *et al.*, 2016). Dans ce contexte, la biologie synthétique des systèmes NRPS apparaît comme une puissante méthode pour tirer parti de la réactivité inhérente des intermédiaires et ainsi modifier la capacité biosynthétique des micro-organismes. Or, les tentatives d'ingénierie aboutissent généralement à des machineries enzymatiques dépourvues de fonctionnalité ou avec une modeste efficacité. Par conséquent, les études génétiques, biochimiques et structurales doivent être approfondies pour mettre au point des outils et stratégies à même d'assurer une diversification moléculaire.

II. Caractérisation des systèmes biosynthétiques modèles chez des γ -protéobactéries

II.1. Organisation de la voie de biosynthèse commune aux AP bactériens

Les voies de biosynthèse orientées vers la genèse de multiples produits chez les protéobactéries présentent un intérêt fondamental en raison de réactions atypiques et d'une diversification chimique, et partagent les étapes de l'assemblage de la chaîne des AP (Hong *et al.*, 2019) (**Figure 71**). En effet, le domaine A1 de la NRPS bimodulaire sélectionne et active un acide aminé L-Ser/L-Thr, tandis que le domaine A2 est spécifique de L-AZC/L-Pro, selon la bactérie productrice (Hong *et al.*, 2019). Le domaine C2 catalyse l'extension du produit, mais également la déshydratation du résidu L-Ser/L-Thr, une réaction non-canonique (Hong *et al.*, 2019) (**Figure 71**). Le domaine TE assure la cyclisation et la libération concomitante du dipeptide en un intermédiaire azétidopyridin-2-one ou indolizidin-2-one (si L-AZC ou L-Pro, respectivement) (données non publiées) (**Figure 71**).

L'enzyme BVMO discrète catalyse ensuite l'insertion d'un atome d'oxygène dans l'hétérocycle azoté (Hong *et al.*, 2019). À partir de cette étape, le produit de synthèse primaire atteint un point de jonction, mais les facteurs moléculaires sous-jacents aux voies alternatives possibles restent inconnus (Hong *et al.*, 2019). Dans tous les cas, diverses enzymes de modification post-NRPS et des réactions chimiques spontanées mènent à la maturation de pyrrolizidines ou cyclocarbamates apparentés (Hong *et al.*, 2019) (**Figure 71**).

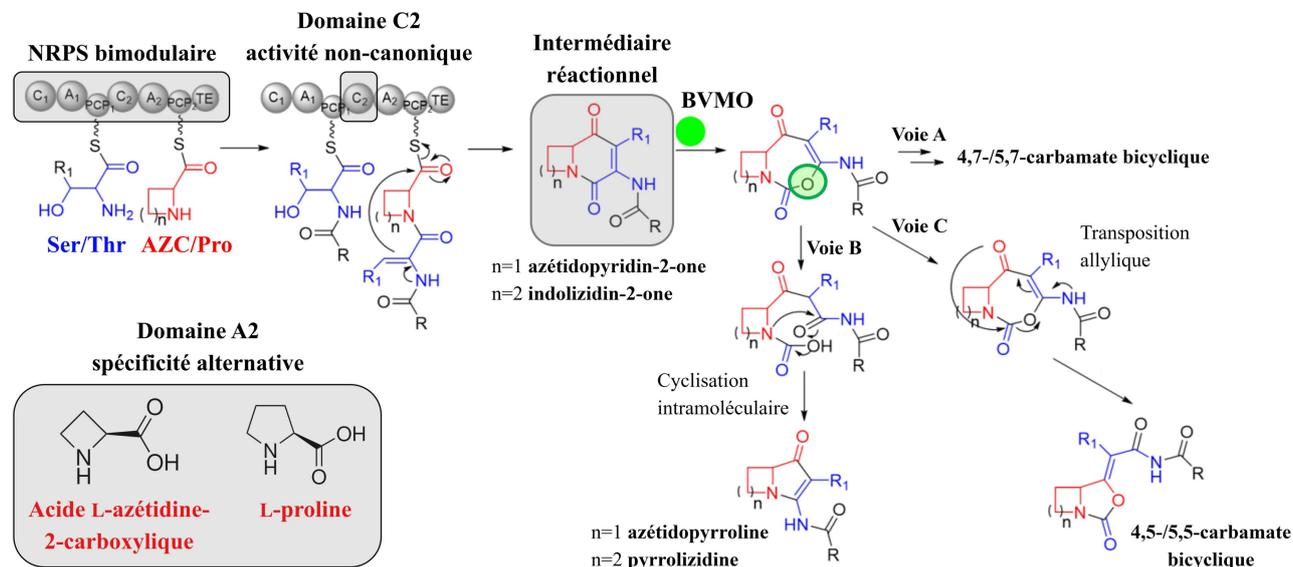


Figure 71. Représentation de la voie de biosynthèse des AP bactériens (adaptée du projet ANR-20-CE92-0038). Une NRPS bimodulaire, telle que la NRPS AzeB du pathogène *P. aeruginosa*, assemble un dipeptide avec un motif d'acide déshydroamino, avant que le domaine TE assure la libération d'un intermédiaire azétidopyridinone. Une BVMO (AzeC) catalyse ensuite l'insertion d'un atome d'oxygène dans l'hétérocycle (voie A), puis l'hydrolyse de cet intermédiaire et la cyclisation en azétidopyrroline (*i.e.* azétidomonamide B) (voie B). Au sein de la voie A, des enzymes post-assemblage modifient le produit en un 4,7-cyclocarbamate, tandis que dans la voie C, une transposition allylique spontanée génère un 4,5-cyclocarbamate. AZC : acide azétidine-carboxylique, BVMO : mono-oxygénase de Baeyer-Villiger, A : domaine d'adénylation, C : domaine de condensation, PCP : domaine porteur de peptidyle, TE : domaine thioestérase.

Des analyses *in silico* de la séquence du domaine A2 de la NRPS AzeB et de multiples domaines homologues ayant une spécificité pour le précurseur L-AZC ont démontré que ceux-ci présentent le code de 10 résidus conservés DMQLVSQQVK au sein de la poche de liaison du substrat (données non publiées). En comparaison, les domaines A2 avec une préférence pour l'acide aminé L-Pro, tels que celui de la NRPS XhpA chez la bactérie *Xenorhabdus hominickii*, intègrent un motif différent (DVQ(F/Y)(S/I)(A/Q/G)(H/V)V(V/I)K). Cette spécificité alternative entraîne à terme la production de métabolites variés, dont les cycles sont fusionnés *via* des atomes de carbone particuliers (*e.g.* 4,7-/5,7-cyclocarbamate, 4,5-/5,5-cyclocarbamate).

Des expériences de reconstitution de la NRPS AzeB *in vitro* ont confirmé que la seule activité de celle-ci est suffisante afin de catalyser les réactions nécessaires à la formation de l'intermédiaire azétidopyridin-2-one (Patteson *et al.*, 2019). Ce résultat indique que l'activité de cyclodéshydratation est intégrée dans la ligne d'assemblage, et n'est pas assurée par des enzymes discrètes, comme cela aurait pu être envisageable (Patteson *et al.*, 2019). Des analyses *in silico* de la séquence du domaine C2 l'ont phylogénétiquement regroupé avec des domaines de déshydratation, tels que le domaine C intégratif AmbE impliqué dans la biosynthèse de 4-méthoxyvinylglycine (Patteson *et al.*, 2022). En l'occurrence, le résidu histidine catalytique putatif de ce domaine aide à positionner le substrat dans le site actif ou alors en élimine le proton de la liaison C-H α (Patteson *et al.*, 2022). Les domaines C2 des NRPS responsables de l'assemblage des AP pourraient exercer un rôle identique. Concernant les domaines TE présents en aval dans chaque système, ils semblent distincts sur le plan évolutif de par leur activité de cyclisation (données non publiées). Un mécanisme réactionnel crédible impliquerait la délocalisation de la paire d'électrons de l'atome d'azote énaminique, avant l'attaque nucléophile de l'atome C β de l'acide déshydroamino sur le groupement carbonyle-thioester (Hong *et al.*, 2019).

Suite à la cyclisation libératrice des intermédiaires azétidopyridinone ou indolizidinone, ceux-ci subiraient une tautomérisation cétone-énol nécessaire à l'obtention du substrat d'une configuration *R* adéquate (Hong *et al.*, 2019). Des études d'expression de la BVMO orthologue LgnC, au sein d'un système NRPS de la bactérie *Xenorhabdus stockiae*, ont mis en évidence que cela ne modifie pas le produit natif (pyrrolizixénamide A) (Schimming *et al.*, 2015). Ceci suggère donc que le substrat ou des interactions protéines-protéines conditionnent la réaction catalytique.

II.2. Maturation post-assemblage des produits et réactions non-enzymatiques

La NRPS bimodulaire assure l'assemblage du squelette azabicyclique des AP, puis des enzymes auxiliaires y greffent des groupements chimiques fonctionnels au cours de la maturation finale des produits. Les NRP issus de ces systèmes acquièrent ainsi une complexité structurale à l'origine de leur activité biologique. Par exemple, l'équipe du Pr. H. Bode (Max Planck Institute, Allemagne) a identifié un dimère dérivé de pyrrolizixénamide, nommé pyrrolizwilline, chez l'entomopathogène *Xenorhabdus hominickii* (**Figure 72A**). La fonctionnalisation de ce composé implique en particulier l'hydrolase XhpG qui catalyse le clivage de la partie acétamide, ainsi qu'une étape non-enzymatique avec un glyoxylate-CoA (**Figure 72A**). Parallèlement, l'équipe de la Dr. Y. Li (UMR 7245 CNRS MCAM, Paris) a caractérisé un métabolite nommé PA496, constitué d'un motif azétidomonamide A et d'un groupement heptyl-4-quinolone *N*-oxyde, au sein du pathogène humain *P. aeruginosa* (Hong *et al.*, 2019) (**Figure 72B**). La modification du NRP primaire nécessite notamment la déshydratase AzeD, homologue à un domaine de condensation, qui catalyse l'élimination de la fonction hydroxyle du cycle central, préalablement à une réaction d'hydroxylation (CYP450 AzeF) (**Figure 72B**).

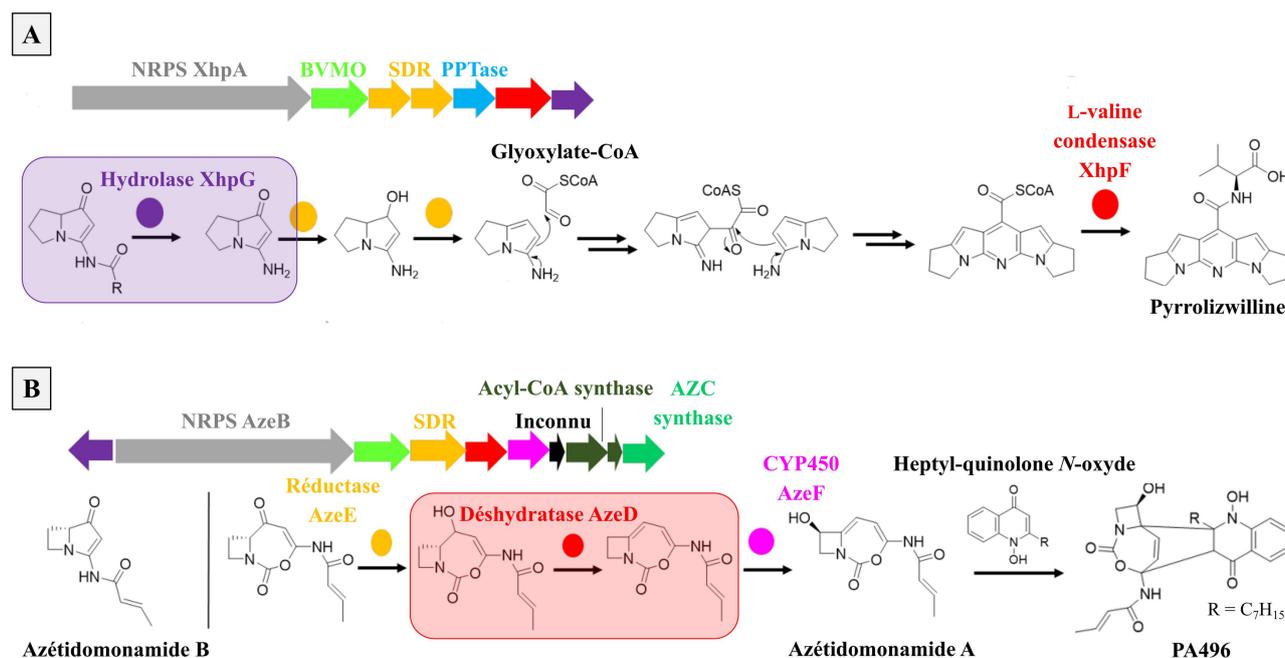


Figure 72. Maturation post-assemblage de la pyrrolizwilline et du PA496 (adaptée du projet ANR-20-CE92-0038).

Les méga-complexes NRPS et les enzymes de modification qui conduisent à la biosynthèse de pyrrolizwilline (A) et de PA496 (B) ont été étudiés dans ce projet de thèse, en particulier l'hydrolase XhpG et la déshydratase AzeD qui assurent une étape de leur maturation respective. AZC : acide azétidine-2-carboxylique, BVMO : mono-oxygénase de Baeyer-Villiger, SDR : déshydrogénase à chaîne courte, PPTase : phosphopanthéthéinyltransférase.

Ainsi, l'ensemble des données obtenues jusqu'à présent démontre que les voies de biogenèse des AP bactériens s'avèrent d'une certaine plasticité moléculaire, qui va à l'encontre de la production de métabolites uniques au sein des systèmes bactériens.

II.3. Fonction écophysiological de la pyrrolizidine et des azétidomonamides

Les BGC des alcaloïdes pyrrolizidiniques sont assez largement distribués chez les symbiotes de nématodes et les souches entomopathogènes du genre *Xenorhabdus*, et conservés chez le pathogène humain *P. aeruginosa*. Cette observation souligne leurs fonctions biologiques dans les interactions avec l'hôte et la virulence. En effet, des tests d'activité ont mis en évidence que la voie métabolique de l'azétidomonamide atténue la virulence de la bactérie productrice dans un modèle *in vivo* de la chenille *Galleria mellonella*, conformément aux études antérieures menées sur la souris (Chen *et al.*, 2018 ; Hong *et al.*, 2019). Cet effet est ainsi supposé bénéfique au développement d'une infection chronique avec l'hôte dans une relation symbiotique-parasitaire. Par contre, le rôle possible des AP isolés de la bactérie *X. homonickii* n'a pour lors pas été déterminé.

Dans l'ensemble, ces constatations suscitent l'engouement pour étudier l'écologie chimique des AP et les mécanismes d'action sous-jacents à leurs activités biologiques. Ces résultats seraient le cas échéant instructifs afin de générer des analogues structuraux aux propriétés thérapeutiques optimales. En effet, les dérivés de l'azétidomonamide pourraient être des agents d'antivirulence opportuns pour lutter contre le pathogène *P. aeruginosa*, multi-résistant à une large gamme de produits antibiotiques et responsable d'infections nosocomiales. De plus, la pyrrolizidine et les azacycles associés ont déjà été reconnus comme des motifs structuraux adaptés à la découverte de médicaments (Blass, 2018). Plusieurs AP et cyclocarbamides apparentés possèdent des propriétés antibactériennes, cytotoxiques, ou anti-inflammatoires prometteuses, comme cela a été vu, mais seule une connaissance approfondie de leur biosynthèse permettra d'en exploiter au mieux le vaste potentiel pharmaceutique. Toutefois, les enzymes des voies métaboliques correspondantes n'ont jamais fait l'objet d'une étude structurale détaillée, ce qui limite les applications de l'ingénierie génétique au sein de ces systèmes.

III. Éluclation des relations structure-fonction des enzymes biosynthétiques

Dans ce contexte, les objectifs de ma thèse étaient de caractériser les enzymes impliquées dans la maturation d'AP issus des voies NRPS modèles par la méthode de cristallographie aux rayons X afin d'élucider les relations structure-fonction. Pour ce faire, les expérimentations ont tout d'abord consisté à purifier l'hydrolase XhpG et la déshydratase AzeD recombinantes par chromatographie en phase liquide, dans leur forme native ou mutante, et à les analyser par diffusion de la lumière (DLS). Suite à ces expériences, les conditions étaient réunies pour entreprendre l'étude structurale, à savoir cristalliser les enzymes par diffusion de vapeur, enregistrer les spectres de diffraction au synchrotron, et résoudre les structures tridimensionnelles par remplacement moléculaire à l'aide d'un modèle.

L'objectif initial de la thèse était néanmoins d'étudier les domaines A2 et C2 de la NRPS AzeB, mais, étant donné qu'ils ne se prêtaient pas à une analyse structurale, les travaux ont été redéfinis en conséquence afin de les caractériser par des techniques biophysiques et/ou par modélisation *in silico*.

MATÉRIELS ET MÉTHODES

SECTION I. Matériels biologiques

I.1. Souches bactériennes

La souche DH5 α d'*E. coli* (Thermo Fisher Scientific) présente le génotype *supE44*, *lacZAM15*, *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*. Elle est utilisée afin de procéder à des transformations classiques. Elle prévient notamment les événements de recombinaison homologue et de dégradation endonucléolytique des plasmides (mutations *recA1* et *endA1*), ce qui assure ainsi une efficacité de transformation et de préparation plasmidique importante (Chan *et al.*, 2013).

La souche BL21(DE3) d'*E. coli* présente le génotype *F*⁻, *ompT*, *gal*, *dcm*, *lon*, *hsdSB*, et est utilisée pour surproduire des protéines recombinantes. En effet, le terme DE3 signifie que la souche a été infectée par le bactériophage λ DE3, dont le gène codant l'ARN polymérase T7 (sous contrôle du promoteur *lacUV5*, inductible par l'IPTG), s'est intégré au chromosome bactérien (Briand *et al.*, 2016). En conséquence, lorsque la souche est transformée par des vecteurs dont l'expression génique est contrôlée par le promoteur fort T7, cela résulte en une traduction accrue et en une sur-production des protéines recombinantes. De plus, la souche BL21 est dépourvue des gènes *ompT* et *lon* codant des protéases, ce qui empêche la dégradation des protéines produites (Ratelade *et al.*, 2009).

I.2. Plasmides d'expression

Le plasmide pBG102, dérivé du plasmide pET28a (Novagen), permet d'exprimer une protéine recombinante fusionnée à des étiquettes de solubilisation et de purification à l'extrémité N-terminale (**Figure 73**). Les sites multiples de clonage (MCS) assurent la linéarisation du plasmide et, par suite, un clonage avec des endonucléases de restriction spécifiques. L'origine de réplication *colE1* permet une propagation à haut nombre de copies et le maintien des plasmides dans les souches d'*E. coli*. Le marqueur de résistance à la kanamycine (KanR) assure la sélection des clones recombinants sur un milieu sélectif. Le gène *lacI* et la séquence opératrice *lacO*, ainsi que les séquences régulatrices du gène de l'ARN polymérase T7, permettent l'expression génique en présence d'IPTG. Les étiquettes protéiques de SUMOylation et 6 \times His assurent une purification de la protéine par chromatographie d'affinité sur ions métalliques immobilisés. Le plasmide pET28a ne présente pas l'étiquette SUMO.

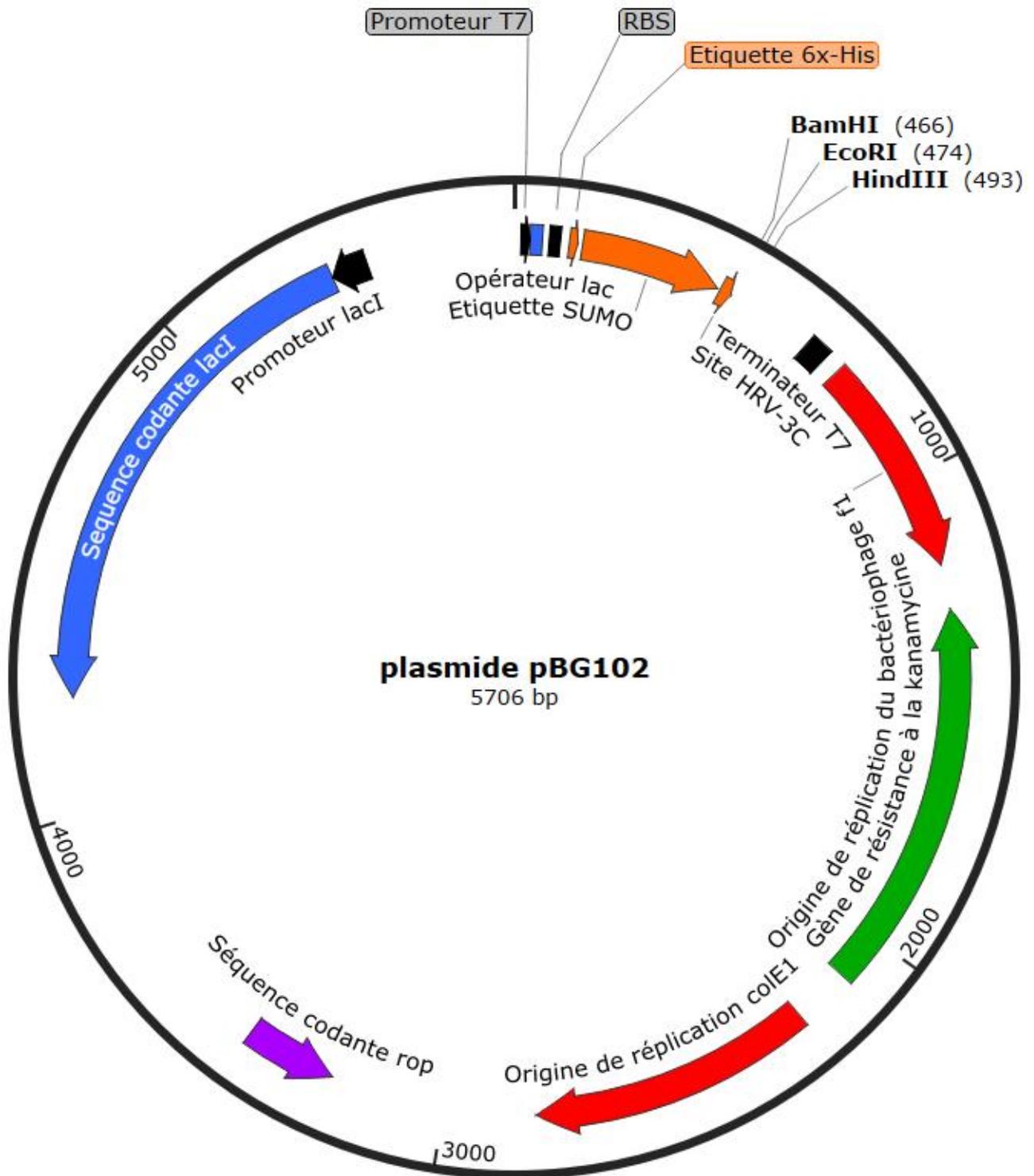


Figure 73. Représentation de la carte du plasmide d'expression pBG102. Le vecteur pBG102 porte un site multiple de clonage, présent en aval du site de clivage de la protéase 3C du rhinovirus humain (noté HRV-3C), des étiquettes de purification 6×His et de solubilisation SUMO, ainsi que les séquences régulatrices *lacI* et *lacO* de l'opéron lactose, pour permettre l'expression de protéines recombinantes. Il comprend l'origine de répllication *colE1* pour assurer la propagation à haut nombre de copies des plasmides et leur maintien stable dans les souches d'*E. coli*, et le gène *kan* de résistance à la kanamycine pour sélectionner les clones recombinants sur un milieu sélectif.

SECTION II. Méthodes d'étude des enzymes biosynthétiques XhpG et AzeD

II.1. Clonage des séquences géniques dans les plasmides d'expression

II.1.A. Amplification des séquences par PCR et ligation des inserts dans les plasmides

Les séquences des gènes de l'hydrolase XhpG et de la déshydratase AzeD ont été amplifiées par PCR (**Tableau IV et Figure 74**). Les amorces ont été conçues afin de s'apparier aux extrémités des régions ciblées, présenter une différence de température de demi-dénaturation (T_M) minimale, ainsi que pour introduire les sites de restriction des enzymes *Bam*HI et *Hind*III (**Tableau IV**).

Tableau IV. Composition du milieu réactionnel de PCR et séquences des amorces.

Réactifs	Concentrations	Volumes
<i>High GC enhancer</i> (5 ×)	1 ×	10 µL
Matrice réactionnelle (100 ng.µL ⁻¹)	100 ng	1 µL
Amorce sens (10 mM)	200 nM	1,2 µL
Amorce anti-sens (10 mM)	200 nM	1,2 µL
dNTP (5 mM)	1,2 mM	12 µL
Phusion HSII (2 U.µL ⁻¹)	1 U	0,5 µL
Eau ultra-pure	-	qsp 25 µL

Amorces	Séquences 5'-3'
XhpG-F	TTTCC <u>GGATCC</u> ATGAATAAAGTTAAAGTTGGCG
XhpG-R	TTTCCA <u>AAGCTT</u> TTTAAGGCACATCCTCCGAC
XhpG-S98A-F	TGTTGGTTTCGCACTTGGATCAGCGGTTGTGATAG
XhpG-S98A-R	CTGATCCAAGTGCGAAACCAACAAGATCAAAAGG
AzeD-F	TTTCC <u>GGATCC</u> ATGGCCGAGATACGACGC
AzeD-R	TTTCCA <u>AAGCTT</u> TTTATTGCTCCTCGAGGGCGGATC

L'enzyme recombinante Phusion résulte de la fusion d'un domaine de liaison à l'ADN hautement processif et de l'ADN polymérase Pfu dotée d'une activité correctrice, et présente une efficacité maximale dans le tampon *high GC enhancer*. Les amorces possèdent chacune les sites de restriction *Bam*HI/*Hind*III et assurent l'amplification en chaîne des gènes d'intérêt en présence d'un excès de dNTP. Les oligonucléotides S98A-F/R permettent quant à eux la mutagenèse site-spécifique du gène de l'hydrolase XhpG.

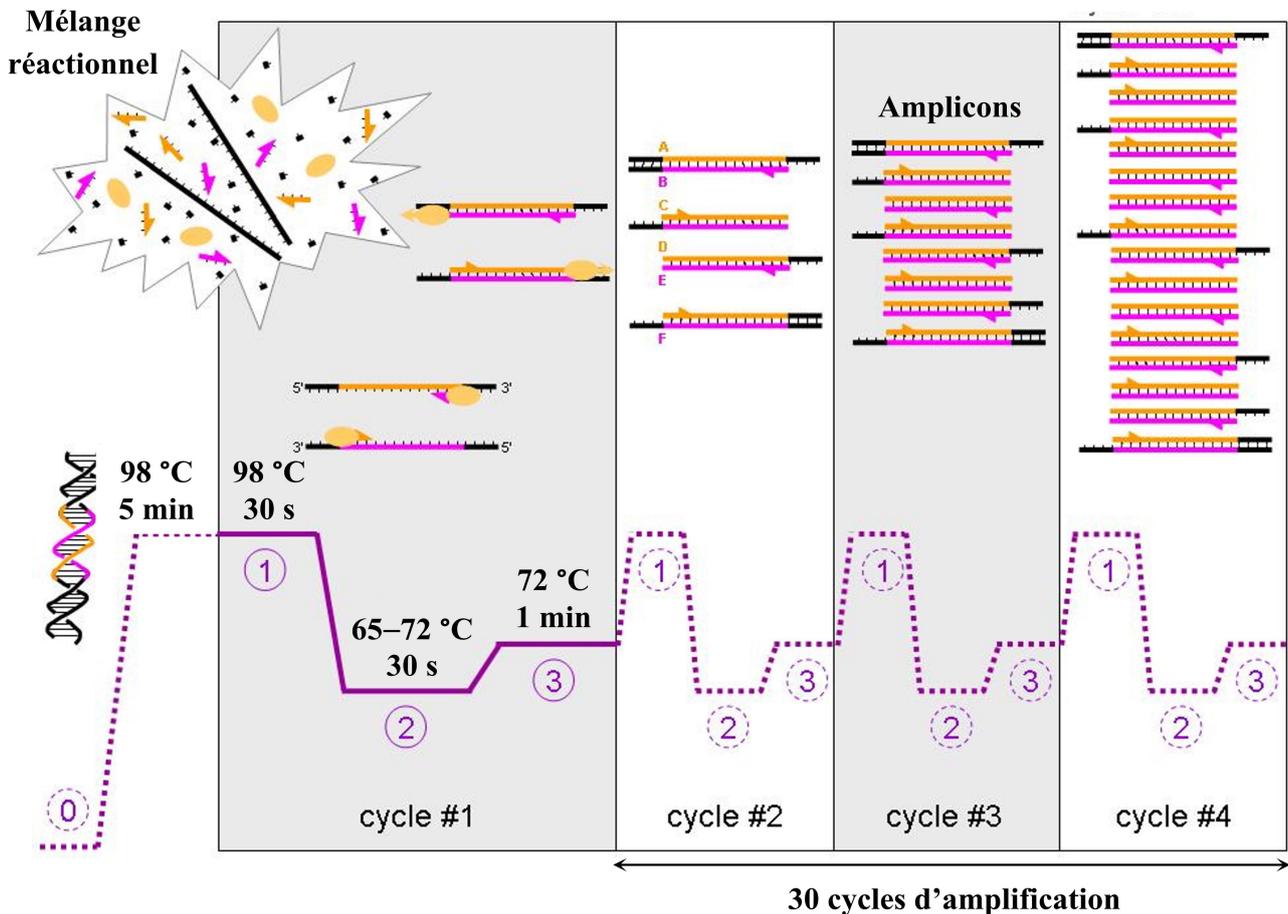


Figure 74. Conditions expérimentales d'amplification en chaîne par polymérase (adaptée de Wikipedia « PCR »). Les conditions expérimentales de PCR consistent en l'étape de dénaturation initiale, suivie de 30 cycles d'amplification, à savoir des étapes de dénaturation (1), hybridation (2) et élongation (3). L'étape de dénaturation initiale (98 °C, 5 min) contribue à homogénéiser le milieu réactionnel, déshydrider les duplex d'amorces et activer la polymérase Hot Start. La phase d'appariement des amorces sur l'ADN matriciel se produit à une température thermodynamiquement favorable, avant que la polymérase Phusion assure l'élongation du brin complémentaire à partir des dNTP en excès. Ces réactions en chaîne augmentent donc de manière exponentielle la quantité d'amplicons.

Les amplicons ont été purifiés par électrophorèse en gel d'agarose (0,8 %) et extraits à partir de 100 µL du mélange réactionnel avec le kit NucleoSpin® PCR Clean-Up Gel Extraction (Macherey-Nagel). Les produits d'extraction (300–500 ng) ont été digérés en présence de 5 unités des enzymes de restriction *Bam*HI (5'-G↓GATCC-3') et *Hind*III (5'-A↓AGCTT-3') dans le tampon Fast Digest (Thermo Fisher Scientific) à 37 °C pendant 1 h.

Les inserts (50 ng) ont été assemblés avec les plasmides pBG102 (XhpG) ou pET28a (AzeD), préalablement linéarisés, en présence de 10 unités de l'ADN ligase T4 dans un tampon spécifique (500 mM Tris-HCl, pH 7,6, 100 mM MgCl₂, 10 mM ATP, 10 mM DTT, 25 % PEG-8000) pour reformer les liaisons phosphodiester entre les extrémités 5'-P et 3'-OH. Le mélange réactionnel a été incubé à température ambiante pendant 20 h.

II.1.B. Préparation et transformation de bactéries *E. coli* compétentes par choc thermique

La préparation de bactéries compétentes a été opérée pour transformer la souche DH5 α d'*E. coli* avec le plasmide contenu dans le mélange réactionnel. À cette fin, une culture de 100 mL de milieu LB a étéensemencée avec 2 mL d'une pré-culture et incubée à 37 °C sous agitation jusqu'à atteindre une concentration cellulaire de 0,5–0,6 unité d'absorbance cellulaire à 600 nm ($A_{600\text{nm}}$). La culture a alors été incubée dans la glace pendant 30 min, centrifugée à 4000 g pendant 15 min à 4 °C, et le culot a été lavé dans 30 mL de tampon TfbI glacé (30 mM acétate de potassium, 100 mM KCl, 10 mM CaCl₂, 15 % glycérol, 50 mM MnCl₂). La suspension bactérienne a ensuite été incubée dans la glace pendant 10 min. La centrifugation a été réitérée et le culot a été lavé dans 4 mL de tampon TfbII glacé (10 mM MOPS, 75 mM CaCl₂, 10 mM KCl, 15 % glycérol). Le volume final a enfin été aliquoté par fraction de 100 μL et les échantillons ont été conservés à –80 °C.

La transformation de bactéries de la souche DH5 α d'*E. coli* a été opérée pour isoler le plasmide recombinant et l'amplifier en vue de sa préparation. Pour cela, 50 μL de bactéries compétentes ont été incubés avec 100 ng du mélange réactionnel dans la glace pendant 20 min. Les bactéries ont été transformées par choc thermique à 42 °C pendant 1 min, puis incubées dans la glace pendant 5 min et régénérées dans 1 mL de milieu LB à 37 °C sous agitation pendant 1 h. Les clones ayant intégré un plasmide, recombinant ou non, ont été isolés par étalement des bactéries sur boîtes de gélose LB, supplémentées avec la kanamycine (Kan) (50 $\mu\text{g}\cdot\text{mL}^{-1}$) et placées à 37 °C pendant 16 h.

II.1.C. Préparation d'ADN plasmidique par lyse alcaline

Les préparations d'ADN plasmidique ont été réalisées à l'aide du kit NucleoSpin[®] Plasmid DNA (Macherey-Nagel). Dans un premier temps, une pré-culture bactérienne de la souche DH5 α d'*E. coli* a été effectuée dans 5 mL de milieu LB additionné de l'antibiotique Kan, puis incubée à 37 °C sous agitation pendant au moins 16 h, et les bactéries ont été concentrées par centrifugation. Le culot a été repris dans le tampon de resuspension et les bactéries ont été lysées par ajout d'une solution de NaOH/SDS. Les lysats ont été neutralisés et ajustés à une haute concentration en sels chaotropiques par ajout d'une solution tampon, puis le surnageant a été isolé par centrifugation et ensuite appliqué sur une colonne ayant une membrane de silice. Les impuretés ont alors été éliminées par passage à travers la membrane et l'ADN adsorbant celle-ci a été élué dans 50 μL d'eau purifiée (Milli-Q). Les plasmides élués ont été dosés à 260 nm avec le spectrophotomètre Nanodrop 3000C (Thermo Fisher Scientific) et les échantillons ont été conservés à –20 °C.

II.1.D. Analyse des plasmides recombinants par restriction et séquençage

L'analyse des vecteurs a ensuite été effectuée pour vérifier la présence d'un insert et la séquence de celui-ci. Pour ce faire, les plasmides préparés ont été digérés avec 3 U d'enzymes de restriction *Bam*HI/*Hind*III à 37 °C pendant 1 h et les fragments produits ont été visualisés aux tailles attendues par électrophorèse en gel d'agarose. Les séquences des inserts ont alors été vérifiées par séquençage (Eurofins Genomics). Les vecteurs ont été renommés selon la nature du gène cloné (pBG102:*xhpG* /S98A ou pET28a:*azeD*).

II.1.E. Modification de séquences géniques par PCR mutagène

Des homologues des enzymes étudiées ont été identifiés à partir de la base de données PDB et du serveur de comparaison structurale Dali (Berman *et al.*, 2000 ; Holm *et al.*, 2023). Les séquences respectives ont été analysées par alignements multiples avec le programme ClustalW dans l'objectif d'identifier des résidus conservés importants (Larkin *et al.*, 2007). Une mutation site-spécifique au niveau du codon du résidu Ser98 catalytique de l'hydrolase XhpG a été opérée par PCR mutagène avec les oligonucléotides adéquats (se référer au **Tableau IV** précédent) et la polymérase Phusion de haute fidélité (**Figure 75**). L'ADN parental méthylé a été sélectivement digéré avec 10 unités de l'enzyme de restriction *Dpn*I Fast Digest (Thermo Fischer Scientific) à 37 °C pendant 1 h. Suite aux étapes de clonage bactérien et de préparation plasmidique, la séquence du mutant XhpG S98A a été confirmée par séquençage (Eurofins Genomics).

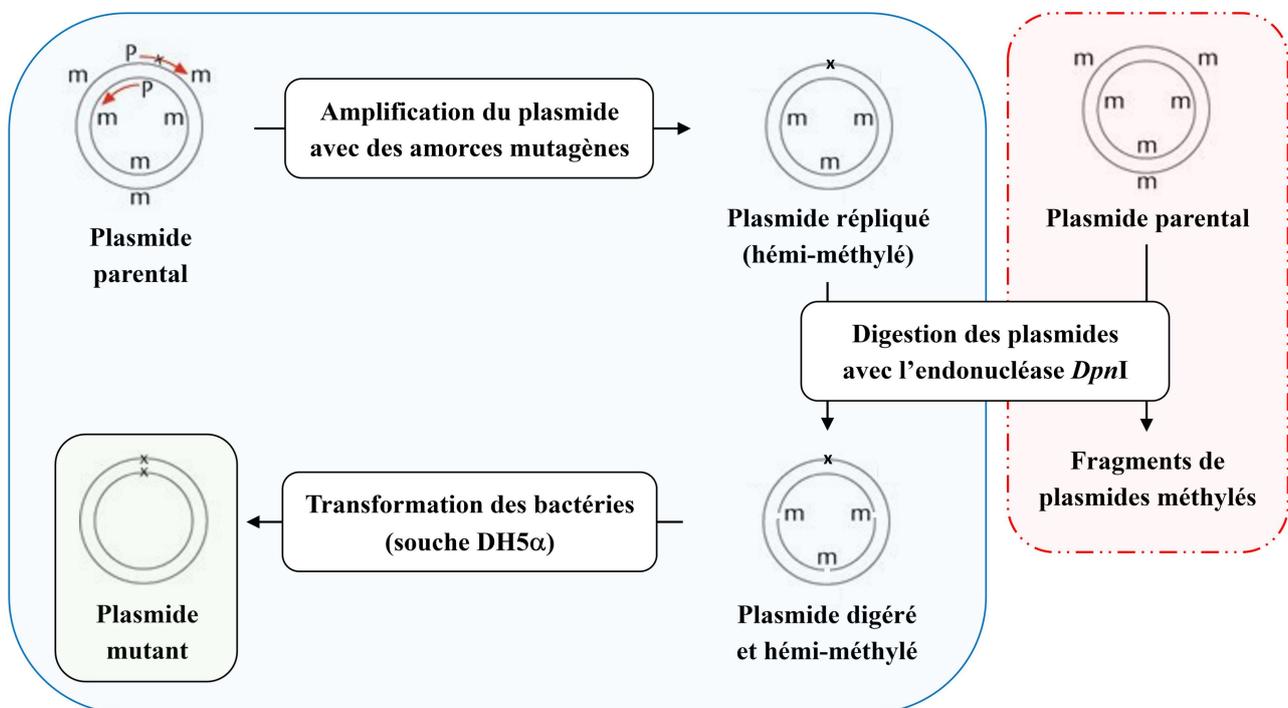


Figure 75. Mutagenèse site-spécifique d'un plasmide. Le plasmide parental a été amplifié avec des amorces mutagènes et les amplicons résultants ont été traités avec l'enzyme de restriction *Dpn*I, ce qui produit des plasmides héli-méthylés. Ceux-ci ont été transformés dans la souche *E. coli* DH5α et le processus de répllication bactérien a rétabli la mutation attendue sur chaque brin des plasmides.

II.2. Production et purification des protéines recombinantes

II.2.A. Expression des protéines chez la bactérie *E.coli* par un réactif inducteur

Les bactéries *E. coli* BL21(DE3) recombinantes transformées avec les plasmides recombinants pBG102:*xhpG*/S98A ou pET28a:*azeD* ont été inoculées dans 0,1 L de milieu LB supplémenté en kanamycine (50 $\mu\text{g.mL}^{-1}$) et les cultures ont alors été incubées à 37 °C sous agitation pendant 18 h. Plusieurs cultures de 2 L de milieu LB sélectif ont étéensemencées avec 0,15 unité d'absorbance à 600 nm ($A_{600\text{nm}}$) jusqu'à une $A_{600\text{nm}}$ de 0,7, et la production des protéines recombinantes a été induite par l'ajout d'IPTG (0,5 mM). Les cultures ont été incubées à 20 °C sous agitation pendant 24 h, puis les cellules ont été concentrées par centrifugation (4000 g, 4 °C, 30 min).

II.2.B. Lyse de la biomasse bactérienne par sonication

Les culots cellulaires ont été remis en suspension dans 30 mL de tampon de lyse (50 mM Tris-HCl, pH 7,8, 300 mM NaCl, 20 mM imidazole, 10 % glycérol), puis les bactéries ont été lysées par 4 cycles de sonication (15 kpsi, 4 °C, 2 min) (Sonifier S-250D) entrecoupés de pauses dans la glace. Les lysats ont alors été concentrés par centrifugation (48 000 g, 4 °C, 30 min) et le surnageant a été filtré au travers d'une membrane stérile (0,22 μm) à l'aide d'une seringue (Merck).

II.2.C. Purification des protéines par chromatographie en phase liquide

II.2.C.α. Purification par chromatographie d'affinité et protéolyse de l'étiquette protéique

Le surnageant de lyse a ainsi été injecté sur une colonne de chromatographie d'affinité HisTrap (Ni^{2+} -sépharose, 10 mL) (GE Healthcare), appareillée au système ÄKTA avant (Cytiva). La colonne a été préalablement équilibrée avec le tampon A (50 mM Tris-HCl pH 7,8, 300 mM NaCl, 10 % glycérol), supplémenté en imidazole (20 mM) pour limiter la fixation non-spécifique. La colonne a été lavée avec 10 volumes-colonnes (CV) de tampon à 5 mL.min^{-1} , puis la protéine a été éluée avec un tampon B à 0,3 M d'imidazole à un débit identique. Dans le cas de l'hydrolase XhpG, exprimée avec des étiquettes 6×His-SUMO à son extrémité N-terminale, les fractions ont été dialysées contre 5 L de tampon en présence de la protéase HRV-3C (10 $\mu\text{g.mL}^{-1}$, 4 °C) afin d'éliminer l'imidazole de l'échantillon et cliver l'étiquette protéique. L'échantillon a été à nouveau chargé sur la colonne HisTrap et la protéine d'intérêt a été éluée par lavage avec le tampon A à un débit constant.

II.2.C.β. Purification par chromatographie d'exclusion stérique

Les fractions de la protéine ont été concentrées à l'aide d'un concentrateur Amicon Ultracel-10 (4000 g, 4 °C). L'échantillon a été injecté *via* une boucle de 10 mL sur une colonne préparative de chromatographie d'exclusion stérique Superdex-75 (130 mL) (GE Healthcare), fixée sur le système ÄKTA pure (Cytiva) et équilibrée dans un tampon de purification (20 mM Tris-HCl, pH 7,8, 300 mM NaCl, 5 mM EDTA, 5 % glycérol). Cette étape de filtration sur gel a été effectuée pour séparer les macromolécules présentes dans l'échantillon selon leur volume hydrodynamique et ainsi éliminer les éventuels agrégats. La protéine d'intérêt a été éluée dans le tampon à 1 mL.min⁻¹ et concentrée jusqu'à un certain volume, puis congelée dans de l'azote liquide et conservée à -80 °C.

Les échantillons de protéines ont été quantifiés à 280 nm avec le spectrophotomètre Nanodrop 3000C à partir des coefficients ϵ d'extinction molaire, calculés à l'aide de l'outil ExPASy ProtParam (11 mg.mL⁻¹ et 20 mg.mL⁻¹ pour l'hydrolase XhpG et la déshydratase AzeD) (Gasteiger *et al.*, 2005). La pureté de ces protéines a été analysée par SDS-PAGE (10 %).

II.3. Analyse des échantillons par diffusion dynamique de la lumière

Les échantillons ont également été analysés par diffusion dynamique de la lumière (DLS) avec l'appareil Zetasizer Nano-ZS (Malvern Panalytical) afin de déterminer le diamètre hydrodynamique moyen des particules et leur indice de polydispersité en vue de essais de cristallisation subséquents. Dans ces expériences, la solution de protéines a été irradiée par un laser, responsable de la diffusion quasi-élastique des ondes électromagnétiques dans toutes les directions de l'espace en fonction de la taille des macromolécules et de la température/viscosité du tampon (Stetefeld *et al.*, 2016) (**Figure 76A**). Les fluctuations d'intensités résultant du mouvement brownien des particules ont été mesurées à un angle de 90° et le signal correspondant a été autocorrélé avec un temps de latence (ms) (**Figure 76B**). L'analyse de la fonction de corrélation, dont a été extrait un temps τ de relaxation, conduit aux coefficients de diffusion des protéines et, à partir de la relation de Stokes, elle fournit les diamètres hydrodynamiques (Stetefeld *et al.*, 2016) (**Figure 76B**).

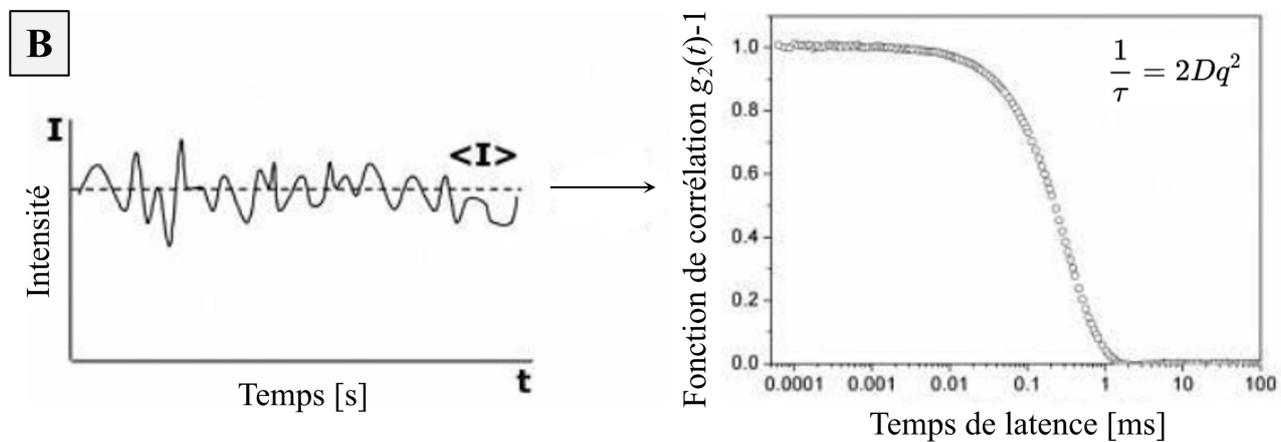
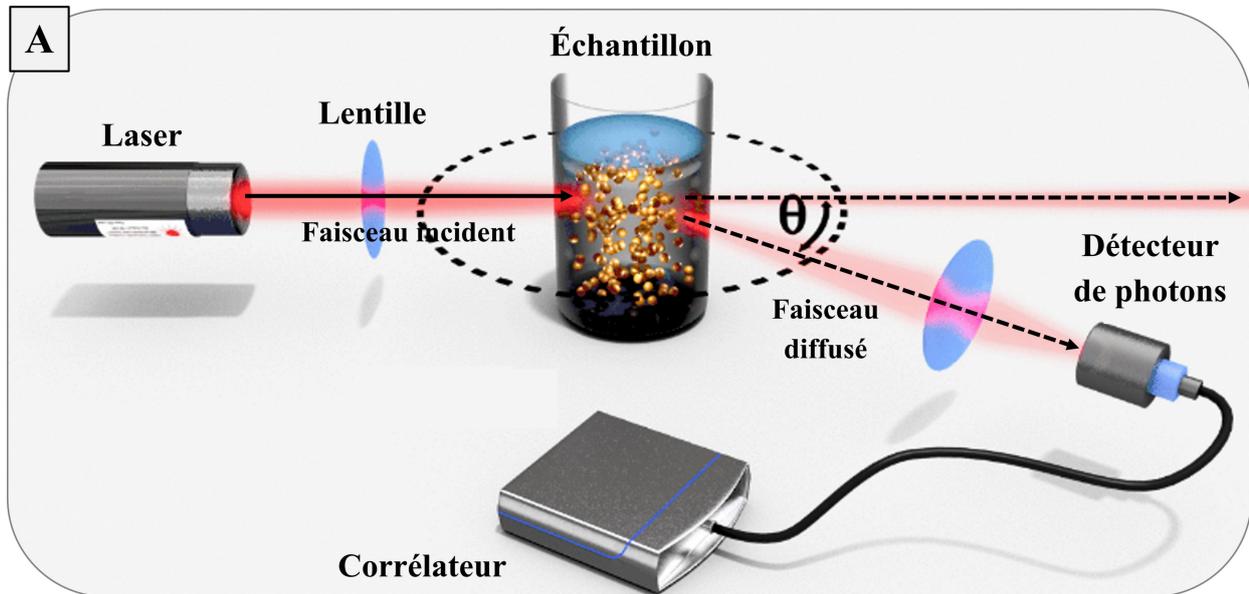


Figure 76. Analyse d'un échantillon par DLS (adaptée de LS Instruments). (A) Un faisceau de lumière traverse une solution protéique, ce qui provoque la diffusion (de Rayleigh) des ondes électromagnétiques dans toutes les directions de l'espace en fonction de la nature de l'échantillon (*e.g.* taille et forme des macromolécules, température et viscosité du tampon) et sans variation d'énergie. Les particules analysées étant soumises à une agitation thermique, des interférences constructives en résultent et produisent un signal mesurable. Un détecteur analyse l'ensemble des fluctuations d'intensités de la lumière à un angle optimal pré-défini, et un corrélateur numérique corrèle le signal mesuré selon le temps (ns– μ s). (B) La fonction de corrélation d'intensité $g_2(t)$ (de second ordre) décrit le mouvement brownien des macromolécules étudiées. La modélisation graphique utilise un temps τ de relaxation, relatif au coefficient de diffusion des particules *via* la relation suivante : $1/\tau = 2 \cdot D \cdot q^2$ avec q le vecteur de diffusion et D le coefficient de diffusion. Ce paramètre rentre en compte dans le calcul du rayon hydrodynamique (R_h) à partir de la relation suivante : $R_h = k_B \cdot T / 6 \cdot \pi \cdot \eta \cdot D$ avec k_B la constante de Boltzmann, T la température de l'expérience et η la viscosité du tampon.

II.4. Résolution des structures tridimensionnelles par cristallographie aux rayons X

La cristallographie aux rayons X apparaît comme la méthode de choix pour résoudre les structures de macromolécules à l'échelle atomique (0,8–3 Å). Schématiquement, ce processus se décompose en différentes étapes : (1) cristallisation de la protéine, (2) diffraction des cristaux, enregistrement et traitement des données, (3) détermination des phases, (4) construction et affinement du modèle et (5) validation de la structure. Les points fondamentaux de ce processus sont brièvement décrits en guise de préambule aux résultats expérimentaux.

II.4.A. Cristallisation des protéines recombinantes par diffusion de vapeur

La cristallisation correspond à la transition de la protéine d'un état libre en solution à un état solide cristallin. Une structure cristalline se caractérise par la répétition (translation) périodique d'un arrangement ordonné d'atomes dans une maille cristalline au sein d'un réseau direct. Ce changement de phase des macromolécules, chacune étant positionnée et orientée de manière identique dans toutes les cellules unitaires, permet d'amplifier le signal de diffraction, mais nécessite alors d'identifier au préalable les conditions favorables aux interactions intermoléculaires.

Bien que les structures de plus de 100 000 macromolécules aient ainsi été résolues, la plupart des facteurs responsables du processus de cristallisation restent mal maîtrisés, si bien que le criblage des conditions optimales repose majoritairement sur la méthode d'essai-erreur (McPherson, 2017). Les lois thermodynamiques à l'origine des cristaux sont néanmoins bien définies : la formation d'un état de sursaturation métastable de la liqueur mère, où la concentration de la protéine est supérieure à sa solubilité, est essentielle à la nucléation des intermédiaires paracristallins en noyaux critiques et à la croissance des germes cristallins (McPherson et Gavira, 2014) (**Figure 77**). Afin d'atteindre cet état de sursaturation critique, il peut être judicieux de : (1) augmenter la concentration de la protéine et modifier l'état d'ionisation des résidus de surface, (2) faire varier les multiples paramètres physico-chimiques (*e.g.* pH, température, nature et force ionique du tampon, *etc.*), ainsi que d'utiliser divers agents précipitants (*e.g.* sels, solvants organiques, polymères), additifs (*e.g.* ions divalents, alcools, monosaccharides, *etc.*) ou ligands (cofacteurs, inhibiteurs) (McPherson et Gavira, 2014). Plusieurs méthodes de cristallisation à l'échelle nanométrique, kits de criblage et outils de robotique ont donc été combinés pour échantillonner en premier lieu l'espace complet des combinaisons possibles.

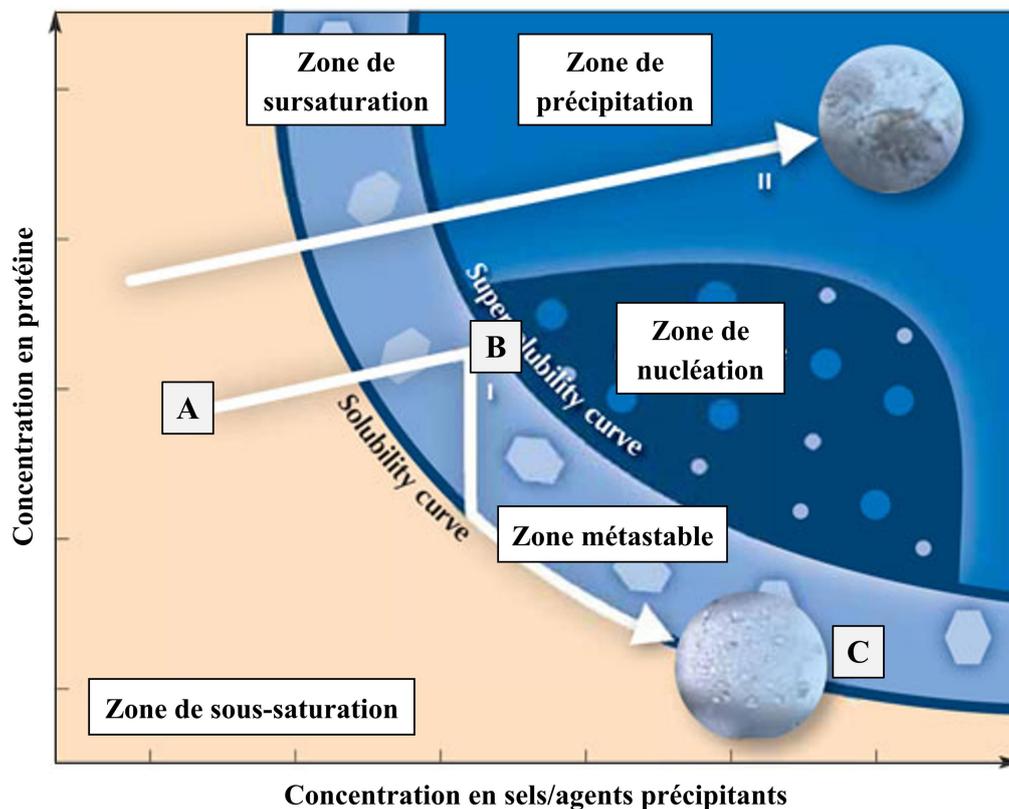


Figure 77. Sursaturation, nucléation et croissance des cristaux protéiques (adaptée de EMBL). Le diagramme de phase du processus de cristallo-genèse se divise en des zones de sous-saturation et de sursaturation. Celle-ci se scinde elle-même en des régions métastables, ou de nucléation et de précipitation. La flèche et les points A, B et C décrivent le phénomène de cristallisation au sein d'un système à l'équilibre par diffusion de vapeur (voir la **Figure 78**). En effet, au fur et à mesure que la protéine et les sels/précipitants se concentrent dans la solution, celle-ci passe d'un état sous-saturé (A) à la zone de nucléation (B), et ce changement de phase se traduit avec la formation des premiers germes cristallins. En conséquence du passage à l'état solide, la macromolécule en solution retourne à un état métastable, tandis que les sels/précipitants se concentrent davantage. Cela ralentit la nucléation et assure progressivement la croissance des cristaux à partir des noyaux critiques (C).

Les conditions de cristallisation des protéines recombinantes ont été criblées par la méthode de diffusion de vapeur en nano-gouttes assises dans des plaques MRC3 (96 puits) (Greiner Bio-One) au moyen d'un robot Mosquito (SPT Labtech) (**Figure 78**). Les résultats (*hits*) initiaux ont été obtenus avec les kits Structure Screen 1 (Molecular Dimensions) pour l'hydrolase XhpG ou Index (Hampton Research) pour la déshydratase AzeD. Les conditions optimales sont les suivantes pour l'hydrolase : 200 mM d'acétate de magnésium, 100 mM de cacodylate de sodium à pH 6,5, et 20 % de PEG-8000, et pour la déshydratase : 100 mM de bis-Tris méthane à pH 6,5 et 25 % de PEG-3350 (18 °C, 1–2 jours d'incubation). Les cristaux ont alors été manuellement reproduits par la technique de diffusion de vapeur en micro-gouttes suspendues dans des plaques Linbro® (24 puits) (Crystalgen) avec des mélanges de 1 µL de solution protéique et 2 µL de tampon de cristallisation (ou avec un ratio de 2:4) (**Figure 78**). Cette étape d'affinement a été réalisée à partir de gammes de solutions tampons et de plans factoriels incomplets, dont les variables supplémentaires multiplient les conditions propices à la genèse de cristaux de meilleure qualité.

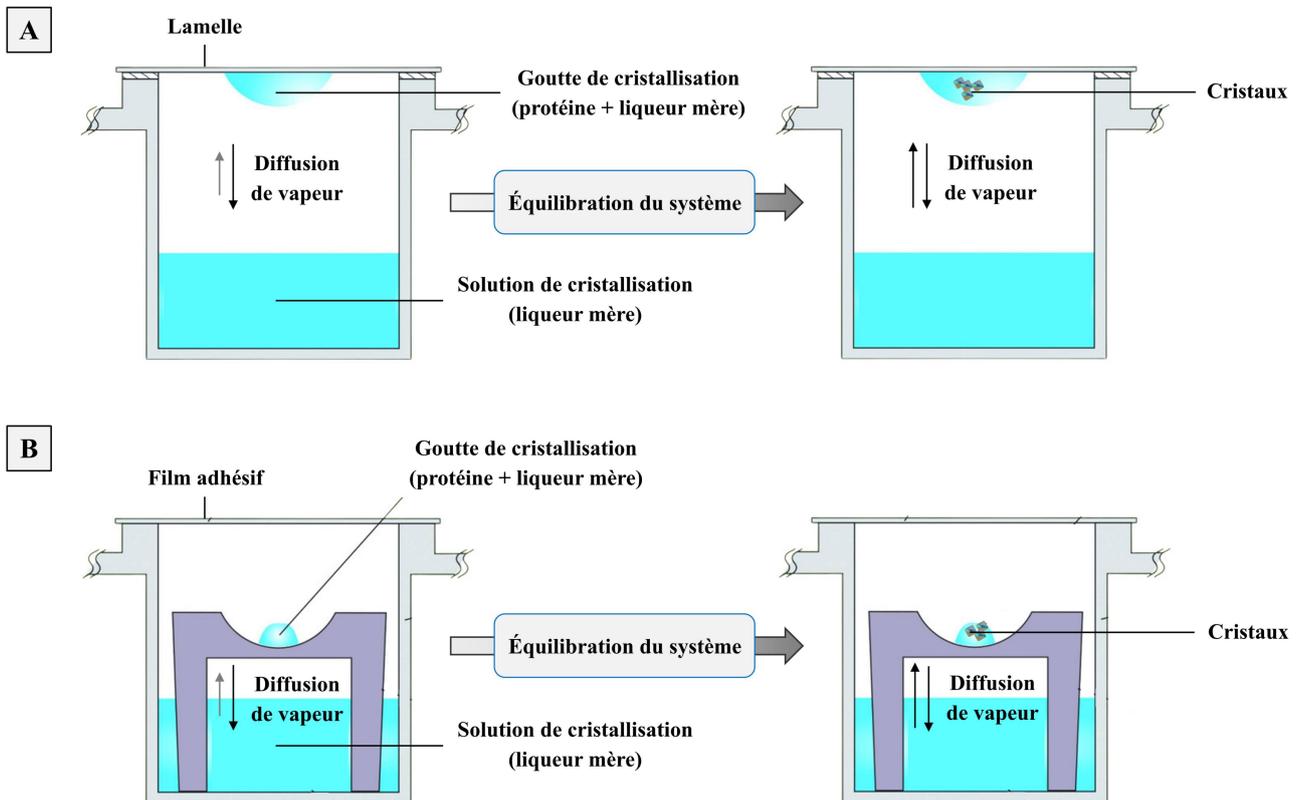


Figure 78. Cristallisation d'une protéine par diffusion de vapeur en système hermétique (adaptée de McPherson et Gavira, 2014). La méthode de diffusion de vapeur en gouttes suspendues (A) ou assises (B) consiste à promouvoir le processus d'évaporation d'espèces volatiles (*e.g.* molécules d'eau) entre un réservoir de la solution de cristallisation et une goutte de la solution de protéines au sein d'une enceinte close. En effet, le réservoir est initialement préparé à une concentration donnée en agent précipitant/cristallisant. La solution de protéines est quant à elle disposée sur une lamelle en verre (A) ou sur une plateforme (B) et diluée avec un certain volume de la liqueur mère prélevée directement du puits. Le compartiment est ensuite scellé hermétiquement et les deux solutions ayant des pressions partielles différentes, l'eau diffuse jusqu'à l'équilibre thermodynamique du système. En conséquence, les composants de la goutte se concentrent, ce qui initie le processus de cristallogenèse, comme décrit **Figure 77**.

II.4.B. Diffraction des cristaux par un rayonnement synchrotron

La diffraction d'un cristal suivi du traitement préliminaire des données brutes obtenues permet de définir les paramètres de maille et groupes d'espace, ainsi que d'accéder aux intensités I_{hkl} et erreurs associées $\sigma(I_{hkl})$ de chaque réflexion d'indice hkl possible, afin de calculer le module du facteur de structure $|F_{hkl}|$ (voir section C.). Une expérience d'enregistrement d'un jeu de données de diffraction nécessite : (1) un dispositif de préparation des cristaux, (2) une source de rayons X, (3) un système optique de focalisation, (4) une méthode de collecte et (5) un détecteur de technologie HPC (*Hybrid Photon Counting*), avant le traitement des clichés obtenus (Cavarelli, 2020).

Au laboratoire, les cristaux de meilleure qualité ont été prélevés à l'aide de cryo-boucles en nylon de tailles appropriées (50–500 nm) (CrystalCap SPINE HT) (Hampton Research), puis trempés dans une solution cryoprotectrice donnée (liqueur mère avec ou sans substrat (10 mM) et 30 % de glycérol ou d'éthylène glycol) pendant 0,1–10 minutes, avant d'être refroidis un par un dans un flux d'azote gazeux à -173 °C. Les données de diffraction des rayons X ont été collectées au synchrotron SOLEIL (St-Aubin, France) sur la ligne de lumière PROXIMA-2A. En particulier, les cristaux ont tour à tour été placés sur un goniomètre κ de haute performance, puis irradiés avec un faisceau de rayons X de haute énergie (6–18 keV) selon la méthode d'oscillation classique sur un secteur angulaire standard avec une amplitude de $0,1^\circ$. Les images de diffraction ont été capturées grâce à un détecteur EIGER X 9M à comptage de photons uniques ($750 \text{ images.s}^{-1}$) (Dectris) (**Figure 79**). Les jeux de données ont été indexés et intégrés avec le logiciel XDS, puis mis à l'échelle et réduits à l'unité asymétrique avec le programme aimless (CCP4) (Kabsch, 2010 ; Winn *et al.*, 2011 ; Evans et Murshudov, 2013). Leur qualité a été évaluée à partir de paramètres tels que la résolution maximale, le rapport signal-sur-bruit ($I/\sigma I$), la complétude, la redondance, et le coefficient de corrélation $CC_{1/2}$ (**Tableau V**).

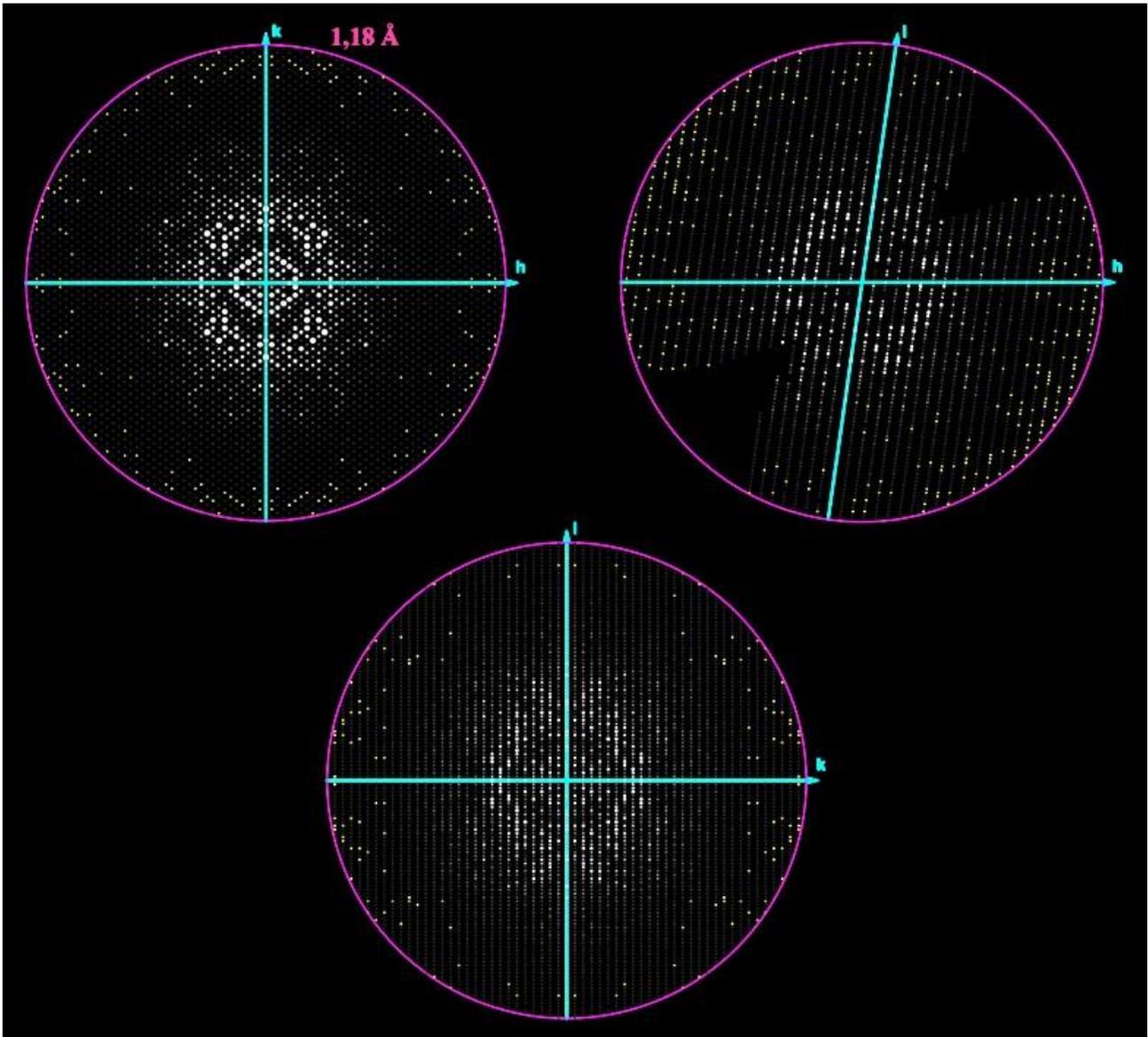


Figure 79. Traitement d'un patron de diffraction avec les programmes XDS, pointless et aimless. Un cliché traduit la distribution spatiale de l'intensité des pics de diffraction des rayons X par les plans hkl du réseau réciproque du cristal. Les patrons sont associés aux plans réticulaires $hk0$, $h0l$, et $0hk$ perpendiculaires à l'axe z , y , ou x , respectivement. Pour chaque enzyme étudiée, un jeu de données a été indexé et intégré avec le programme XDS (Kabsch, 2010). Cette étape consiste à attribuer des indices de Miller (h, k, l) (*i.e.* coordonnées) aux tâches de diffraction dans l'espace réciproque, à déterminer les paramètres (*i.e.* longueurs, angles) de la maille primitive du cristal, et à mesurer l'intensité des réflexions. Une fois ces informations connues, le groupe d'espace du cristal (*i.e.* ensemble des isométries affines) a été identifié avec le programme pointless (Evans, 2011). Les données ont ensuite été mises à l'échelle avec le programme aimless afin de diminuer les erreurs de mesure de l'intensité dues à l'absorbance anisotrope du cristal, à la dégradation de celui-ci, et à la décroissance du faisceau des rayons X incidents (Evan et Murshudov, 2013). L'intensité de la plupart des réflexions uniques étant mesurée plusieurs fois en raison du secteur angulaire expérimental et/ou du groupe de symétrie du cristal, les données ont enfin été moyennées en des valeurs d'intensité I_{hkl} avec l'erreur $\sigma(I_{hkl})$ associée.

Tableau V. Enregistrement des données cristallographiques des enzymes XhpG et AzeD.

Enregistrement des données	Hydrolase XhpG S98A	Déshydratase AzeD
Groupe d'espace	C2	P2 ₁ 2 ₁ 2 ₁
Dimensions de la maille		
a, b, c [Å]	48,83, 68,54, 75,48	56,46, 67,59, 116,68
α , β , γ [°]	90,00, 98,11, 90,00	90,00, 90,00, 90,00
Longueur d'onde [Å]	0,826561	0,980112
Résolution [Å]	39,50–1,18 (1,21–1,18)	44,20–1,55 (1,59–1,55)
<i>R</i> merge [%]	4,1 (148,0)	6,4 (107,9)
<i>R</i> pim [%]	1,7 (62,5)	1,9 (30,3)
<i>I</i> / σ <i>I</i>	17,9 (1,2)	20,0 (1,9)
Nombre d'observations totales	541379	857500
Nombre de réflexions uniques	78423	65206
Complétude [%]	98,0 (92,7)	99,9 (98,7)
Redondance/multiplicité	6,9 (6,4)	13,1 (13,6)
Facteur <i>B</i> de Wilson [Å ²]	21,248	31,535
Coefficient de corrélation <i>CC</i> _{1/2}	99,9 (67,4)	99,9 (76,9)

Un groupe d'espace est un ensemble de symétries géométriques et décrit la disposition périodique d'un motif unitaire dans le cristal. Par exemple, le cristal de l'hydrolase XhpG appartient au groupe d'espace C2 (système monoclinique), dont le réseau possède pour éléments de symétrie un axe binaire et un miroir plan perpendiculaire, avec les paramètres de maille $a = 48,83 \text{ \AA}$, $b = 68,54 \text{ \AA}$, $c = 75,48 \text{ \AA}$, $\alpha = \gamma = 90^\circ$, $\beta = 98,11^\circ$. Les rayons X incidents de haute énergie (15 keV (0,83 Å) et 12,7 keV (0,98 Å) pour XhpG et AzeD, respectivement) permettent d'atteindre une résolution atomique, en l'occurrence 1,21–1,18 Å (XhpG) et 1,59–1,55 Å (AzeD) dans la dernière coquille (*i.e.* plage spécifique). Le facteur de désaccord *R*merge entre réflexions équivalentes évalue la cohérence interne des données, c'est-à-dire l'accord entre les intensités reliées par la symétrie et la redondance, et doit être inférieur à 10 %, en particulier 4,1 % (XhpG) et 6,4 % (AzeD). Bien que très employé, ce facteur dépend de la multiplicité (*i.e.* redondance), c'est-à-dire le nombre de mesures d'un angle de diffraction spécifique pour un pic donné sur le détecteur. Le facteur *R*pim (*Precision-indicating merging R-factor*) reflète quant à lui la qualité des données relatives aux réflexions uniques. Le rapport signal/bruit *I*/ σ *I* dépend également de la redondance et doit être supérieur à 3 pour un jeu de données exploitable. La complétude de celui-ci fait référence à la proportion de mesures d'intensités d'ondes par rapport à l'ensemble théorique des données attendues dans l'espace de diffraction, ici 98,0 % (XhpG) et 99,9 % (AzeD). Un autre indicateur pertinent, le coefficient *CC*_{1/2}, traduit la corrélation entre les données recueillies à partir de deux demi-ensembles indépendants de mesures, calculé à 99,9 % dans les deux cas. Globalement, les jeux de données décrits sont d'une excellente qualité et ont chacun été exploités en vue de déterminer les structures tridimensionnelles des enzymes biosynthétiques.

Parallèlement, des données de diffusion des rayons X aux petits angles (SAXS) des enzymes en solution ont été enregistrées sur la ligne de lumière SWING et traitées avec la suite de programmes ATSAS (Manalastas-Cantos *et al.*, 2021). Les études structurales ont été appuyées par ces résultats en termes d'analyse des états oligomériques et conformationnels plausibles.

II.4.C. Détermination des phases ϕ_{hkl} par remplacement moléculaire

La détermination d'une structure de protéine requiert en premier lieu de calculer une fonction de densité électronique pour chaque point de coordonnées relatives (x, y, z) d'une maille cristalline à partir des modules et des phases du facteur de structure F_{hkl} de chaque réflexion (**Figure 80AB**). Une expérience de diffraction des rayons X permet uniquement de déterminer l'amplitude (module $|F_{hkl}|$) des ondes diffusées à partir de l'intensité des pics, l'information primordiale sur les phases ϕ_{hkl} étant perdue (**Figure 80C**). Les ondes diffractées ne peuvent être refocalisées qu'au moyen d'équipements et méthodes expérimentales spécifiques (*e.g.* monocapteurs, focalisation de Bragg), lesquels restent complexes à mettre en œuvre à l'heure actuelle lors du flux de travail habituel. Donc, il est essentiel de disposer de phases initiales les plus proches possibles des phases exactes pour poursuivre l'étude.

$$[\text{A}] \quad \rho(x, y, z) = \left(\frac{1}{V}\right) \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F_{hkl} \exp[-2i\pi(hx + ky + lz)]$$

F_{hkl} facteur de structure pour la réflexion hkl ,
 V volume de la maille cristalline.

$$[\text{B}] \quad F_{hkl} = |F_{hkl}| \exp(i\phi_{hkl})$$

$|F_{hkl}|$ module,
 ϕ_{hkl} phase du facteur de structure.

$$[\text{C}] \quad I_{hkl} \propto (|F_{hkl}|)^2 = F_{hkl} F_{hkl}^*$$

F_{hkl}^* complexe conjugué de F_{hkl} .

Figure 80. Fonction de densité électronique et équation complexe du facteur de structure F_{hkl} . L'intensité des pics de diffraction d'un cristal de protéine par les rayons X permet de calculer le module $|F_{hkl}|$ du facteur de structure à partir de la relation [C]. Cependant, compte tenu du fait que les ondes diffusées ne peuvent pas être refocalisées à travers une lentille optique conventionnelle, alors la connaissance des phases ϕ_{hkl} n'est possible pour aucune réflexion (relation [B]). Afin de calculer la fonction de densité électronique $\rho(x, y, z)$, c'est-à-dire la transformée de Fourier inverse des facteurs de structure des ondes diffractées, et d'interpréter la distribution continue d'électrons dans la maille cristalline (relation [A]), il est indispensable de résoudre le problème des phases *via* une méthode indirecte.

Actuellement, trois types de méthodes indirectes sont employés pour déterminer les phases ϕ_{hkl} inconnues : le remplacement isomorphe multiple (MIR) et la diffusion anormale (MAD) tirent profit de la présence de dérivés d'atomes lourds et des différences d'intensités résultantes entre les facteurs de structure, tandis que le remplacement moléculaire exploite la structure d'une protéine homologue déposée ou d'un modèle bio-informatique généré à partir de la séquence primaire (Cavarelli, 2020).

Cette méthode de phasage, initiée dans les années 1960, constitue un point de départ judicieux dès lors qu'une structure parente est accessible, et ne nécessite qu'un espace complet de diffraction de haute qualité de la protéine d'intérêt (Rossman et Blow, 1962). Elle consiste à échantillonner les orientations et les positions de l'homologue structural dans l'unité asymétrique de la macromolécule cible, et à optimiser une fonction de score objective entre les données expérimentales et les données calculées (Cavarelli, 2020). À titre indicatif, les algorithmes de remplacement moléculaire comparent la fonction d'autocorrélation de la densité électronique, dite fonction de Patterson, observée pour la protéine d'intérêt avec celle calculée pour la structure parente (Cavarelli, 2020). Ainsi, les phases ϕ_{hkl} déterminées à partir de celle-ci sont associées aux modules $|F_{hkl}|$ observés expérimentalement afin d'afficher une carte de densité électronique initiale. En raison des progrès significatifs réalisés dans la modélisation structurale, le remplacement moléculaire s'est progressivement imposé comme une méthode de référence en matière de phasage, si bien qu'elle est à l'origine du dépôt d'environ 60 % des macromolécules dans la PDB jusqu'à maintenant.

Dans cette étude, les phases initiales ont été attribuées avec le programme MOLREP à partir d'un modèle généré avec le logiciel d'apprentissage automatique AlphaFold2 (Vagin et Teplyakov, 1997 ; Jumper *et al.*, 2021). À l'aube de cette thèse, AlphaFold2 est apparu comme une avancée scientifique révolutionnaire en bio-informatique pour prédire la structure tridimensionnelle de protéines avec une extrême précision à l'échelle atomique, même dans le cas où aucun homologue n'est connu (Jumper *et al.*, 2021). Grâce à une approche inédite de réseaux neuronaux en graphiques (GNN), le système s'est vu couronné de succès lors de la 14^{ème} édition (2020) du concours CASP (*Critical Assessment of Structure Prediction*), et a ouvert la voie à de vastes perspectives quant à la compréhension des mécanismes biologiques et à la conception rationnelle de médicaments (Jumper *et al.*, 2021).

II.4.D. Construction et amélioration des modèles par des cycles itératifs d'affinement

L'affinement d'une structure est un processus cyclique d'optimisation non-linéaire d'une fonction de score, dont l'objectif est d'améliorer les cartes de densité électronique, c'est-à-dire diminuer les biais issus du modèle initial, pour en permettre l'interprétation en termes de positions et mouvements atomiques. En principe, cette étape repose sur des observations, comme les paramètres cristallins, les modules des facteurs de structure, ou la stéréochimie de la macromolécule, et dépend de paramètres relatifs au désordre statique et dynamique des atomes (*i.e.* facteurs q_j d'occupation et B_j d'agitation thermique) (Cavarelli, 2020) (**Figure 81**). La fonction de score à optimiser s'exprime soit sous forme de résiduels des moindres carrés ou d'énergie empirique associée (*i.e.* dynamique moléculaire), soit comme maximum de vraisemblance (Cavarelli, 2020). Pour ce faire, les méthodes correspondantes tirent parti d'algorithmes stochastiques afin de minimiser l'erreur entre les modules des facteurs de structure observés (F_{obs}) et calculés (F_{calc}) (Cavarelli, 2020).

$$F_{hkl}^{calc} = \sum_{j=1}^N q_j f_j w_j^{hkl} \exp(2i\pi(hx_j + ky_j + lz_j))$$

f_j facteur de diffusion de l'atome j de coordonnées relatives (x_j, y_j, z_j) ,

(q_j, w_j^{hkl}) paramètres de déplacements atomiques,

N nombre d'atomes de la maille,

q_j facteur d'occupation,

w_j^{hkl} facteur de Debye-Waller pour l'atome j

$$w_j^{hkl} = \exp\left(-B_j \frac{\sin^2 \theta}{\lambda^2}\right) \quad B_j = \frac{8\pi^2}{3} \overline{U_j^2}$$

B_j facteur d'agitation thermique,

U_j^2 déplacement quadratique moyen de l'atome j

Figure 81. Équation complexe du facteur de structure F_{hkl} (adaptée de Cavarelli, 2020). La nature, la position et le mouvement des atomes dans la maille cristalline permettent de calculer le facteur de structure à partir de ladite relation. Le facteur q_j reflète l'occupation d'un atome j donné à une position particulière dans la carte de densité électronique et le facteur w_j en traduit la distribution statistique dans des positions équivalentes. Le paramètre w_j se décompose en un facteur d'agitation thermique B_j lié au déplacement quadratique moyen U_j^2 dudit atome autour de sa position d'équilibre.

En pratique, les structures ont été construites et affinées par des cycles itératifs de reconstruction manuelle dans les cartes de densité électronique avec le programme WinCoot, suivis d'un affinement restreint à 1,18 Å (XhpG) ou 1,55 Å (AzeD) avec le programme REFMAC5 (Emsley *et al.*, 2010 ; Murshudov *et al.*, 2011) (**Figure 82**). Ces étapes d'affinements restreints à une stéréochimie idéale ont été réalisées afin d'augmenter le rayon de convergence de la méthode de minimisation, c'est-à-dire à identifier le minimum global de la fonction d'ajustement du modèle, et non pas un minimum local. La qualité des modèles construits a été évaluée au terme de chaque cycle à travers les facteurs de désaccord cristallographique R_{crist} et R_{libre} , ainsi que les valeurs de déviation quadratique moyenne (RMSD) des longueurs et angles de liaison (**Tableau VI**).

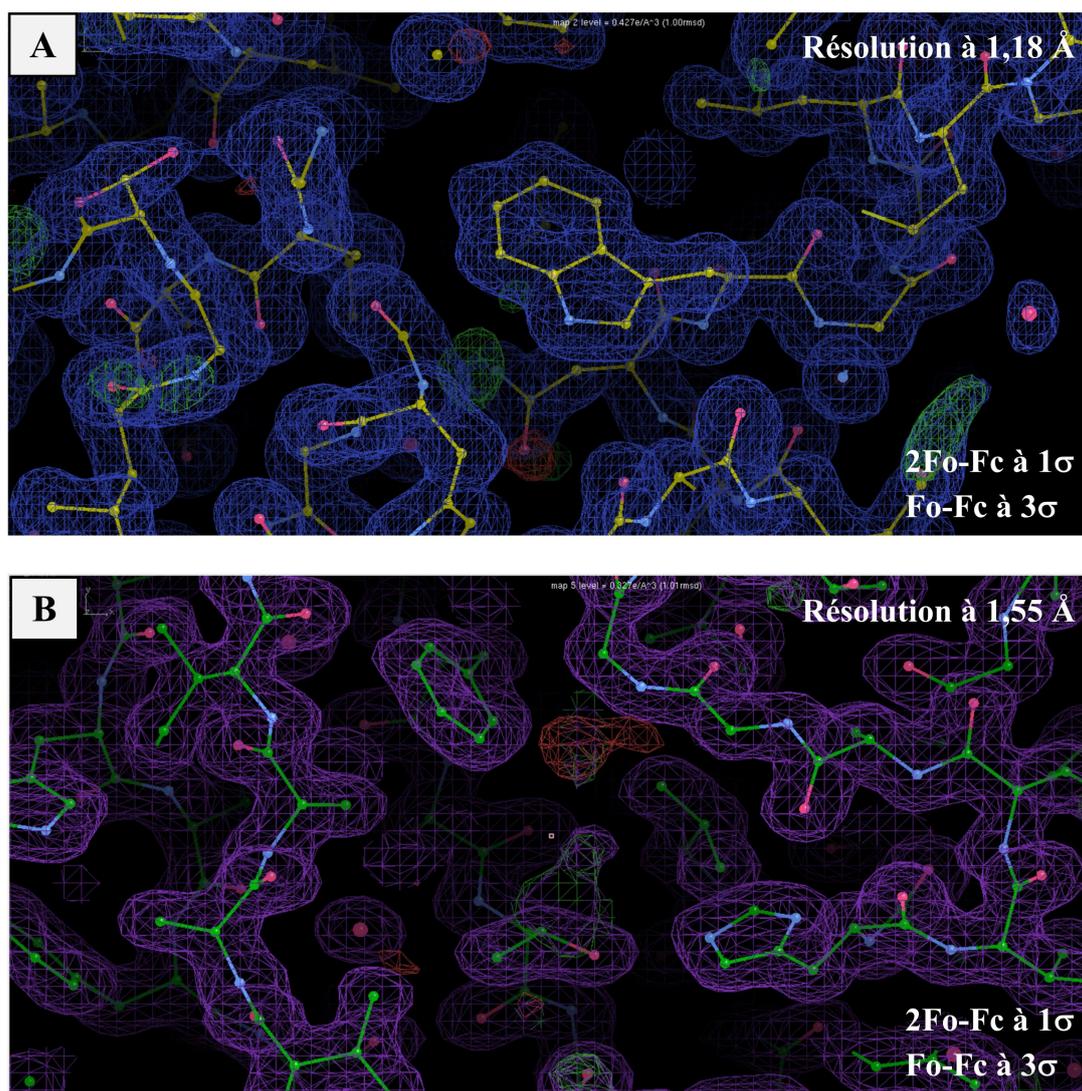


Figure 82. Observation de cartes de densité électronique de structures protéiques avec le programme WinCoot. Les cartes de densité électronique illustrent la distribution spatiale des électrons dans la structure cristalline de XhpG (A) ou de AzeD (B) avec un écart-type uniforme associé à l'ensemble des valeurs. Les cartes de type différences de Fourier (notées $2F_o-F_c$), calculées avec des amplitudes $2F_{\text{obs}}-F_{\text{calc}}$ et des phases ϕ_{calc} , permettent de construire le squelette de la protéine, avant d'en orienter au mieux les chaînes latérales dans la densité (contourée à 1σ). Les cartes de type « OMIT map » (notées F_o-F_c), calculées entre les données expérimentales et théoriques, s'avèrent complémentaires et permettent de mettre en évidence des densités négatives (zones rouges) ou positives (zones vertes) à 3σ . Celles-ci correspondent à des désaccords structuraux importants du modèle atomique ou à des artefacts expérimentaux.

Tableau VI. Affinement des données cristallographiques des enzymes XhpG et AzeD.

Affinement des données	Hydrolase XhpG S98A	Déshydratase AzeD
Résolution limite [Å]	20,00–1,18	44,20–1,55
Nombre de réflexions	74471	61945
Facteurs d'affinement R , R_{libre}	0,1663, 0,1963	0,1457, 0,1945
Nombre d'atomes	2286	3763
Protéine	2051	3411
Autres ligands/ions	4	12
Molécules d'eau	231	340
Facteur B moyen [Å ²]	19,582	24,346
Déviations $RM.S$		
Longueur de liaison [Å]	0,0123	0,0080
Angle de liaison [°]	1,7371	1,4670
PDB	8QZU	8S6L

Dans cette étude, la résolution limite permet d'individualiser les atomes dans les cartes de densité électronique, voire de distinguer les hydrogènes (1,18 Å dans le cas de XhpG), et ainsi de construire une structure tridimensionnelle complète avec ligands et molécules d'eau associés. Le facteur de désaccord cristallographique R_{crist} (ou R) relatif à l'ensemble des réflexions reflète l'accord entre les modules $|F_{hkl}|$ des facteurs de structure observés et calculés, et doit tendre vers 0,1–0,2 (*i.e.* 1/10^{ème} de la résolution effective) au terme du processus d'affinement. Le facteur R_{libre} prend uniquement en compte un sous-ensemble de 5–10 % de réflexions aléatoires dans les calculs d'optimisation et évalue la précision de la structure à partir de données expérimentales prédites. Ces paramètres doivent converger vers les valeurs attendues au fur et à mesure de l'amélioration du modèle, en termes de géométrie et de stéréochimie, et permettent par conséquent d'éviter d'éventuels problèmes de suraffinement (*i.e.* surinterpréter les cartes de densité électronique).

II.4.E. Validation des structures cristallographiques et soumission dans la PDB

La validation et le contrôle-qualité de la structure finale des macromolécules ont été réalisés au moyen du programme MolProbity pour parvenir au meilleur accord avec les données expérimentales (Chen *et al.*, 2010) (**Tableau VII**). Une fois que les structures résolues ont atteint un score optimal au niveau des contacts interatomiques et paramètres géométriques, cette étape a été réitérée à l'aide du serveur wwPDB (Young *et al.*, 2017). Les structures de l'hydrolase XhpG S98A mutante et de la déshydratase AzeD ont été déposées dans la base de données PDB sous les codes d'accèsion 8QZU et 8S6L, respectivement (**Annexes PDB**).

Tableau VII. Validation des structures des enzymes XhpG et AzeD avec le programme MolProbity.

A	All-Atom Contacts	Clashscore, all atoms:	0.74		99 th percentile* (N=256, 1.18Å ± 0.25Å)	
		Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.				
	Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%	
		Favored rotamers	221	99.55%	Goal: >98%	
		Ramachandran outliers	0	0.00%	Goal: <0.05%	
		Ramachandran favored	259	98.48%	Goal: >98%	
		Rama distribution Z-score	-0.63 ± 0.46		Goal: abs(Z score) < 2	
		MolProbity score [^]	0.74		100 th percentile* (N=1299, 1.18Å ± 0.25Å)	
		Cβ deviations >0.25Å	0	0.00%	Goal: 0	
		Bad bonds:	0 / 2092	0.00%	Goal: 0%	
Bad angles:	0 / 2843	0.00%	Goal: <0.1%			
B	All-Atom Contacts	Clashscore, all atoms:	1.18		99 th percentile* (N=730, 1.55Å ± 0.25Å)	
		Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.				
	Protein Geometry	Poor rotamers	1	0.28%	Goal: <0.3%	
		Favored rotamers	344	98.01%	Goal: >98%	
		Ramachandran outliers	0	0.00%	Goal: <0.05%	
		Ramachandran favored	423	98.14%	Goal: >98%	
		Rama distribution Z-score	-0.10 ± 0.37		Goal: abs(Z score) < 2	
		MolProbity score [^]	0.83		100 th percentile* (N=6779, 1.55Å ± 0.25Å)	
		Cβ deviations >0.25Å	0	0.00%	Goal: 0	
		Bad bonds:	0 / 3494	0.00%	Goal: 0%	
Bad angles:	0 / 4766	0.00%	Goal: <0.1%			

Dans l'ensemble, 98,48 % (259) des résidus de l'hydrolase XhpG S98A (A) se situent dans les régions favorisées du diagramme de Ramachandran et 99,55% (221) présentent des rotamères privilégiés. Plusieurs chaînes latérales adoptent des conformations alternatives et l'un des 9 résidus proline (Pro91) présente une conformation *cis* dans la structure. Les longueurs et angles de liaison sont idéaux. Compte tenu de ces éléments, la géométrie globale de la structure atteint le 100^{ème} percentile parmi N = 1299 de résolution semblable (1,18 Å ± 0,25 Å). Aucun contact interatomique défavorable n'est observé (99^{ème} percentile). Des remarques comparables s'appliquent à la structure de la déshydratase AzeD (B).

II.4.F. Modélisation des complexes binaires par *docking* moléculaire

Les structures des complexes avec le substrat ont ensuite été modélisées par *docking* moléculaire avec le programme CB-Dock2 (Liu *et al.*, 2022). La modélisation *in silico* permet d'identifier le site de liaison au sein de la cavité du site actif à partir de calculs de courbure des surfaces accessibles au solvant, et d'y prédire la position et l'orientation du ligand avec l'affinité théorique correspondante via l'algorithme d'homologie AutoDock Vina (Liu *et al.*, 2022). Cette étape a ainsi été réalisée pour comprendre les bases moléculaires de la liaison du substrat dans une conformation spatiale propice à la réaction. Les figures ont été préparées à l'aide du logiciel PyMOL (Lilkova *et al.*, 2015).

II.5. Analyse des échantillons par diffusion des rayons X aux petits angles

Les échantillons de protéines ont été analysés par SAXS au synchrotron SOLEIL sur la ligne de lumière SWING afin d'obtenir des informations structurales à basse résolution ($> 10 \text{ \AA}$) quant aux états oligomériques et conformationnels en solution (**Figures 83 à 87**).

Tout d'abord, les échantillons ($50 \mu\text{L}$, $5\text{--}10 \text{ mg}\cdot\text{mL}^{-1}$) ont chacun été injectés sur une colonne de chromatographie d'exclusion stérique Bio SEC-3 (100 \AA) (Agilent), pré-équilibrée dans le tampon de purification (20 mM Tris-HCl , $\text{pH } 7,8$, 300 mM NaCl , 5% glycérol) et connectée à une cellule capillaire en quartz dans une chambre sous vide du système BioSAXS. Au cours de leur séparation, les espèces en solution ont été irradiées *via* un faisceau de rayons X de $570 \times 200 \mu\text{m}^2$ avec un flux de $10^{12} \text{ photons}\cdot\text{s}^{-1}$ ($0,77 \text{ \AA}$, 16 keV). L'intensité diffusée a été mesurée par une diode Schottky ($2,2 \times 1,2 \text{ mm}^2$) en diamant et les trames de diffusion ont été enregistrées grâce à un détecteur EIGER X 4M (Dectris) à un angle de $0,1\text{--}3^\circ$ par rapport au faisceau incident (**Figures 83 et 84**). Ce paramètre angulaire rentre en compte dans le calcul du vecteur de diffusion q , dont dépend l'intensité diffusée $I(q)$. Celle-ci donne accès à des indications à la fois sur la masse moléculaire, la dimension, la forme et la conformation de la macromolécule selon la gamme q considérée, comme décrit ultérieurement.

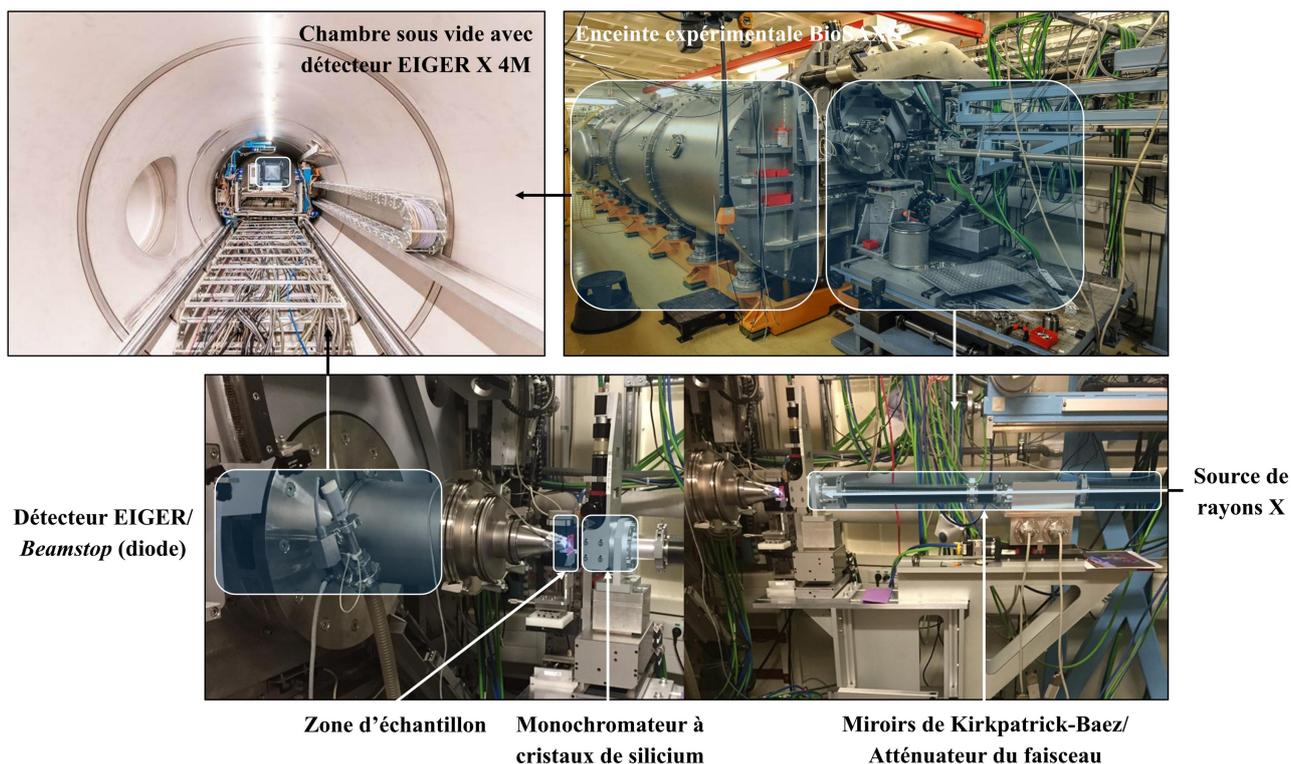


Figure 83. Configuration de la ligne de lumière SWING au synchrotron SOLEIL. La ligne synchrotron SWING comprend plusieurs sections avec un 'undulateur', source des rayons X, un atténuateur à rôle d'ajustement de l'intensité du faisceau incident ($6\text{--}16 \text{ keV}$), un monochromateur à double-cristaux de silicium et des miroirs de Kirkpatrick-Baez responsables de la focalisation à une longueur d'onde spécifique. Un autre compartiment présent en aval contient la zone de l'échantillon, la fente de définition, ainsi que le *beamstop* à diode de diamant et le détecteur EIGER X à comptage de photons uniques.

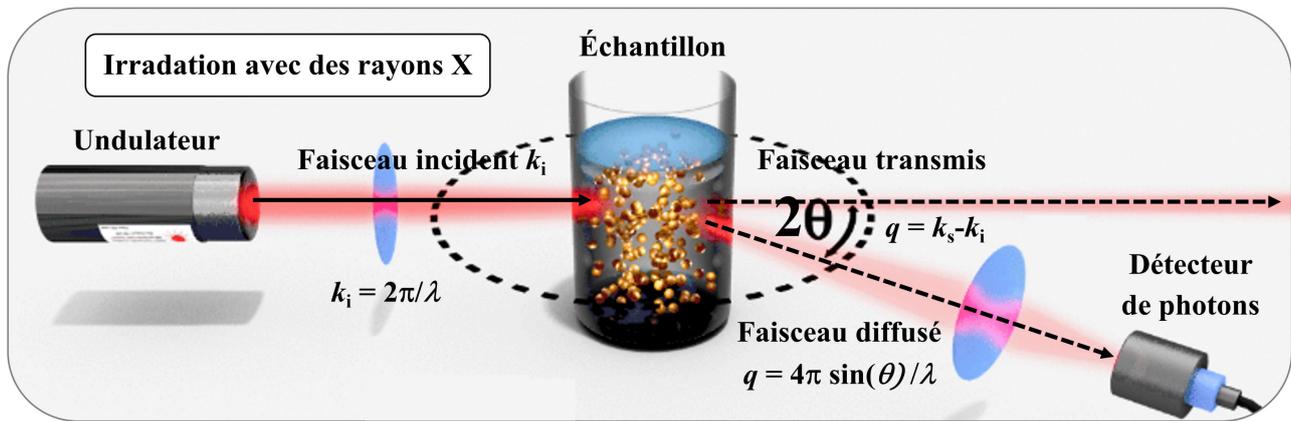


Figure 84. Analyse d'un échantillon par SAXS (adaptée de LS Instruments). Dans les expériences, un rayonnement monochromatique incident k_i ($\lambda = 0,77 \text{ \AA}$) traverse l'échantillon de protéines, ce qui provoque l'interaction élastique des photons avec les nuages électroniques des atomes et leur diffusion en un faisceau k_s avec un transfert d'énergie nul. Les ondes diffusées dépendent du paramètre q via la relation suivante : $q = |k_s - k_i| = 4 \cdot \pi \cdot \sin(\theta) / \lambda$ avec λ la longueur d'onde du faisceau incident k_i et 2θ l'angle de diffusion (Da Vela et Svergun, 2020). Un système de détection mesure alors les modulations d'intensités diffusées $I(q)$ à de faibles angles ($0,1-3^\circ$) par rapport au faisceau dans une gamme pré-définie ($q_{\min} = 3,65 \cdot 10^{-3} \text{ \AA}^{-1}$, $q_{\max} = 5,54 \cdot 10^{-1} \text{ \AA}^{-1}$) et collectent les trames correspondantes.

Les trames brutes de diffusion ont été réduites par normalisation de l'intensité $I(q)$ à une échelle absolue et soustraction du bruit de fond (*i.e.* tampon) avec le programme FOXTROT (Thureau *et al.*, 2021). Les valeurs de l'intensité $I(0)$ et du rayon de giration R_g de chacune des images individuelles ont été reportées sous forme d'un profil SAXS en fonction du temps d'éluion de la macromolécule (**Figure 85**). Les trames homogènes (*i.e.* $I(0)$ maximale, 88–113) ont été intégrées par moyennation radiale en une courbe de diffusion (**Figure 85**). L'intensité $I(q)$ étant proportionnelle au carré de la transformée de Fourier de la densité électronique excédentaire de la macromolécule par rapport au tampon, la courbe de diffusion en reflète une structure tertiaire/quaternaire moyenne en solution (Da Vela et Svergun, 2020). Dans le cas idéal, c'est-à-dire des échantillons dilués à l'infini et injectés en série, les modulations d'intensités dues à des corrélations intermoléculaires (*i.e.* agrégats ou petites molécules interférentes) sont négligeables (Da Vela et Svergun, 2020).

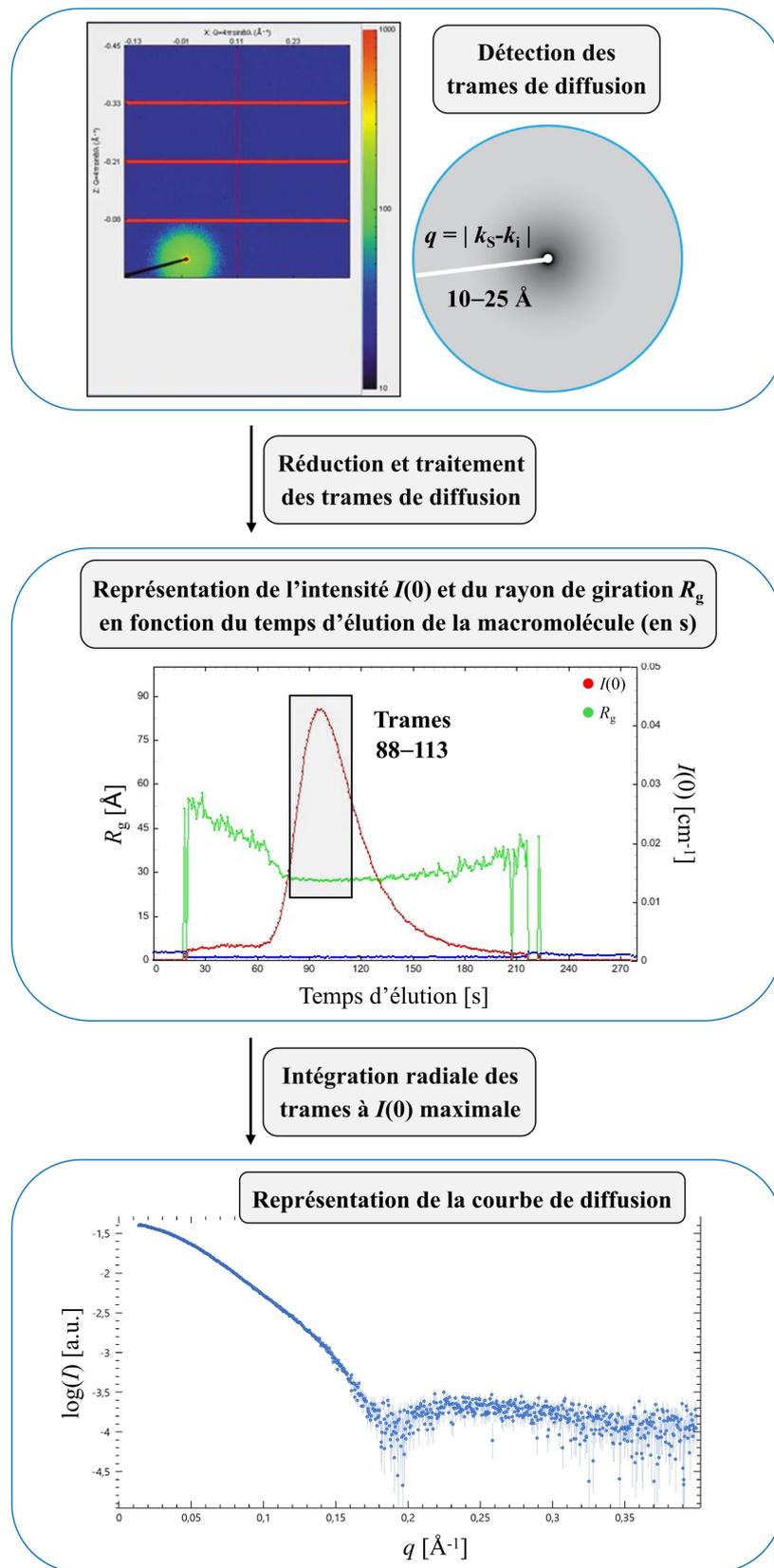


Figure 85. Réduction des trames et moyennation en une courbe SAXS avec le programme FOXTROT. Les trames de diffusion traduisent la distribution angulaire de l'intensité des rayons X diffusés en fonction du vecteur d'onde q . Au nombre total de 600 dans une expérience de collecte sur la ligne SWING, celles-ci ont été traitées et représentées en un profil 1D avec le programme FOXTROT (Thureau *et al.*, 2011). Suite à cela, les courbes associées à des valeurs de R_g similaires en fonction de $I(0)$ ont été intégrées en une courbe de diffusion radiale moyenne. Celle-ci exprime l'intensité de diffusion $\log(I)$ de la protéine en fonction du vecteur d'onde q dans une gamme caractéristique ($3,6 \cdot 10^{-3}$ – $5,5 \cdot 10^{-1} \text{ \AA}^{-1}$) de l'énergie incidente ($0,77 \text{ \AA}$, 16 keV) et de la distance échantillon-détecteur (3 m).

Pour chaque domaine NRPS ou protéine ayant été étudié, la courbe de diffusion obtenue a été dans un premier temps analysée avec le programme PRIMUSQT (ATSAS 3.0) (Manalastas-Cantos *et al.*, 2021) (**Figure 86**). Dans le domaine de q à faibles valeurs (*i.e.* $q < 5 \cdot 10^{-2} \text{ \AA}^{-1}$), l'analyse des données fournit la masse moléculaire ($q = 0$) et le rayon de giration R_g (approximation de Guinier) de la protéine (Da Vela et Svergun, 2020) (**Figure 86A**). Au sein d'une gamme de valeurs étendues (*i.e.* $q < 3-4 \cdot 10^{-1} \text{ \AA}^{-1}$), l'analyse de la courbe indique son état d'oligomérisation et sa conformation (courbe de Kratky), ainsi que la distribution des distances interatomiques (fonction $P(r)$), y compris la dimension maximale (D_{\max}) (Glatter et Kratky, 1982 ; Da Vela et Svergun, 2020) (**Figure 86BC**).

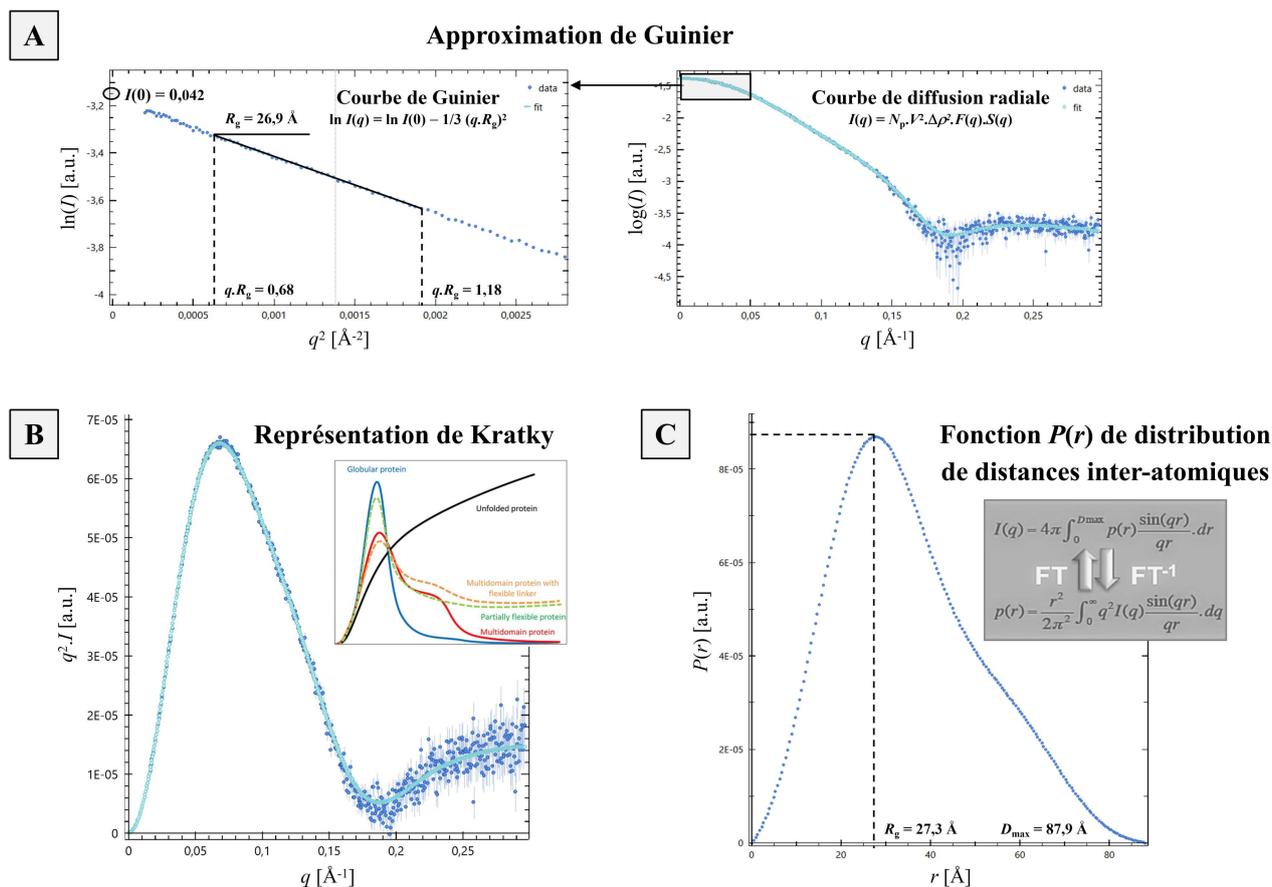


Figure 86. Analyse d'une courbe de diffusion radiale avec le programme PRIMUS. (A) La courbe obtenue décrit la distribution d'intensité isotrope proportionnelle à la diffusion d'une seule particule moyennée sur toutes les positions et orientations possibles dans un système idéal à partir de la relation suivante : $I(q) = N_p \cdot V^2 \cdot \Delta\rho^2 \cdot F(q) \cdot S(q)$. L'intensité de diffusion est liée au nombre de macromolécules (N_p), à leur volume et contraste de densité électronique ($V^2 \cdot \Delta\rho^2$), à leur forme ($F(q)$), ainsi qu'aux interférences ($S(q)$). Les informations structurales et paramètres associés ont ainsi été extraits du motif de diffusion radial. En premier lieu, l'analyse des données linéaires dans la plage de faibles q permet d'estimer l'intensité à un angle nul ($I(0)$), ainsi que le rayon de giration R_g (AUTORG, courbe de Guinier), moyenne quadratique des distances intramoléculaires par rapport au centre de masse de la densité électronique de la protéine. (B) Les données SAXS ont été converties en une courbe de Kratky ($q^2 \cdot I$ en fonction de q), en lien avec le repliement et la flexibilité locale de la macromolécule, mais aussi avec l'état d'oligomérisation de celle-ci. L'aire sous la courbe de Kratky est intégrée en un invariant de Porod (Q), proportionnel à la masse moléculaire M_w à travers le rapport surface/volume à une intensité nulle, et spécifique de la compacité globale à l'échelle des grands q . (C) Une transformée inverse de Fourier des données donne accès à la fonction $P(r)$ de distribution des distances interatomiques au sein de la molécule, dont l'analyse permet de déterminer à la fois les paramètres $I(0)$, R_g et D_{\max} (distance maximale) révélateurs de la conformation en solution.

Suite au traitement préliminaire des données SAXS avec le programme PRIMUS, les courbes de diffusion ont été analysées avec de multiples programmes de la suite ATSAS afin de modéliser des enveloppes *ab initio* à l'aide de modèles d'atomes factices (*e.g.* DAMs (DAMMIN, DAMAVER)), de comparer les courbes expérimentales et théoriques à partir d'une structure protéique résolue (*e.g.* CRY SOL), de générer des complexes moléculaires au moyen de corps rigides (*e.g.* SASREF, CORAL) ou d'en déterminer les conformations plausibles en solution (*e.g.* SREFLEX, EOM) (Figure 87). La qualité des modèles ainsi construits a été évaluée avec le paramètre d'ajustement χ^2 entre les données expérimentales et les courbes théoriques obtenues.

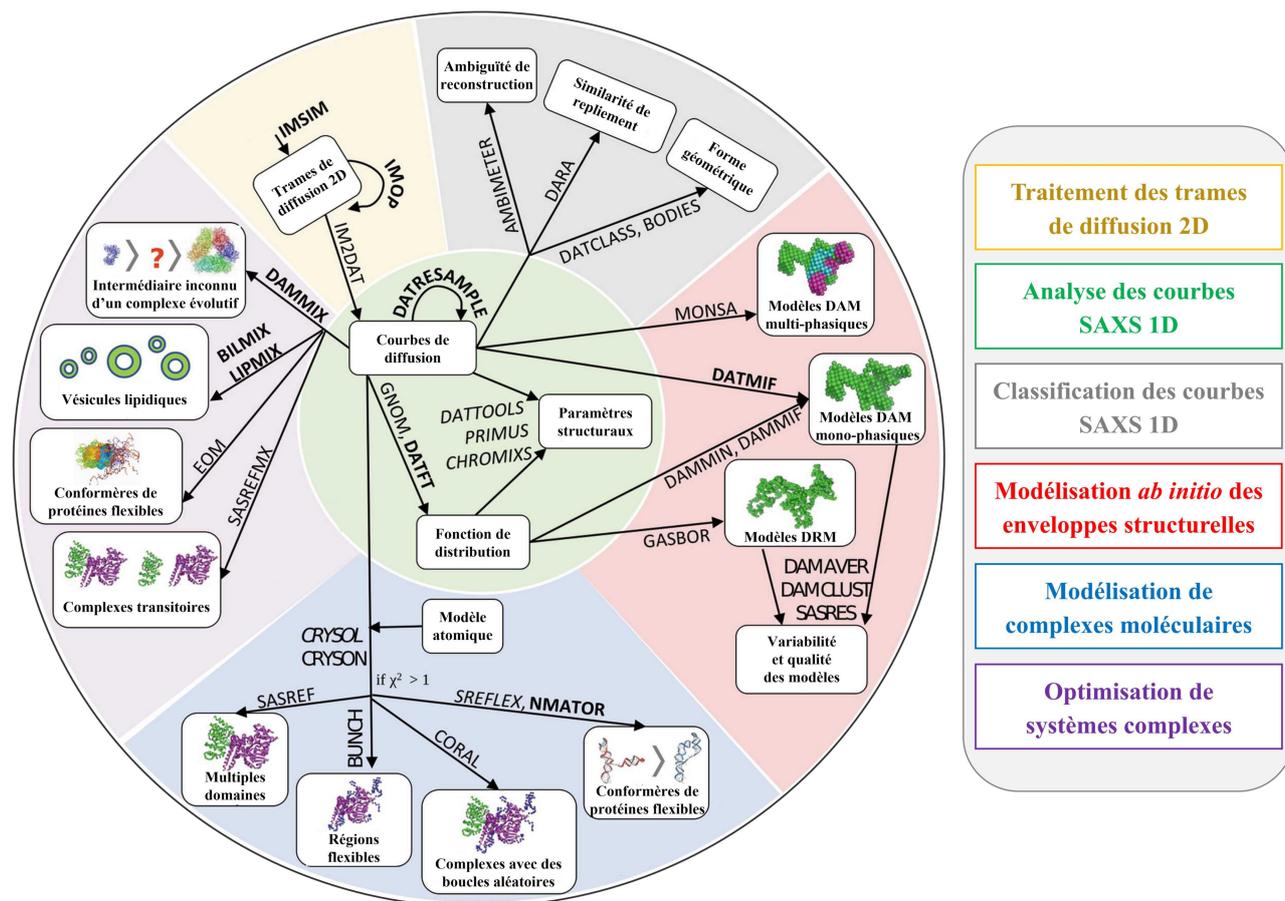


Figure 87. Utilisation des programmes de la suite ATSAS 3.0 (adaptée de Manalastas-Cantos *et al.*, 2021). La suite ATSAS comprend divers programmes utilisés pour traiter, analyser, modéliser et optimiser des données SAXS. Dans la version 3.0 du logiciel, de nouveaux programmes ont été développés. Dans un premier temps, le programme IMSIM permet par exemple de simuler des motifs de diffusion isotrope 2D, et IMOP d'effectuer des opérations sur les images et masques correspondants pour obtenir la courbe de diffusion radiale. À partir de celle-ci, le programme DATMIF permet de modéliser *ab initio* des enveloppes structurales à l'aide de modèles d'atomes factices (DAM) monophasiques. Le programme NMATOR permet quant à lui de modéliser les changements conformationnels de complexes protéines/ARN *via* une analyse des modes normaux dans l'espace des angles de torsion. Le programme DAMMIX sert enfin à générer la forme d'un complexe intermédiaire non-résolu au sein d'un système évolutif. L'ensemble des outils implémentés dans ATSAS est accessible depuis l'interface en ligne (<https://www.embl-hamburg.de/biosaxs/software.html>).

SECTION III. Méthodes d'étude du domaine C2 de la NRPS AzeB

Les domaines d'intérêt de la NRPS AzeB du pathogène *P. aeruginosa* ont chacun fait l'objet d'expérimentations avancées, au même titre que les enzymes XhpG/AzeD, du clonage des séquences géniques chez l'hôte *E. coli* à l'analyse biophysique (*i.e.* DLS et SAXS) des protéines recombinantes (se référer aux sections **II.1–3.** et **II.5.** précédentes). En dépit d'un travail conséquent, seul un cristal prismatique non-reproductible d'un didomaine PCP1-C2 a pu être analysé par diffraction des rayons X et, de surcroît, les données obtenues étaient d'une qualité insuffisante à une étude structurale. Dans ce contexte, une stratégie alternative a été développée, de la construction de plasmides d'expression originaux de protéines chimériques à l'analyse subséquente de complexes PCP1-C2-PCP2, en vue de stabiliser le domaine C2 et d'en élucider à terme les relations structure-fonction.

Pour ce faire, les expérimentations ont d'abord consisté à concevoir et construire des plasmides spécifiques à l'expression d'une protéine fusionnée à un domaine bZIP synthétique (SYNZIP), puis à cloner les séquences PCP1-C2 et PCP2 (**Figure 88A**). À l'origine, les SYNZIP ont été développés à partir des facteurs de transcription de type *basic-region leucine zipper* (bZIP), et leurs structures en superhélice ont été rationnellement optimisées pour procéder à l'ingénierie des interactions protéine-protéine avec une haute affinité (10–50 nM) (Reinke *et al.*, 2010 ; Thompson *et al.*, 2012 ; Klaus *et al.*, 2019). Les paires de SYNZIP hétérodimériques se caractérisent chacune par 5–7 motifs répétés de sept résidus [abcdefg]_n, avec des acides aminés hydrophobes (a, d), chargés (e, g), et variables (b, c et f) (Reinke *et al.*, 2010) (**Figure 88BC**). Dès lors que les séquences correspondantes s'enroulent en hélices α avec une configuration particulière, les chaînes latérales hydrophobes (Val, Leu et Ile) de chaque structure secondaire se font face et établissent des interactions responsables du repliement en superhélice (Reinke *et al.*, 2010 ; Thompson *et al.*, 2012) (**Figure 88BC**). Grâce à ces propriétés géométriques, 27 paires de SYNZIP définies peuvent être utilisées afin de moduler la spécificité et l'affinité de l'interaction entre partenaires protéiques au sein de réseaux moléculaires (Thompson *et al.*, 2012). Parmi celles-ci, la paire SZ17:SZ18 a été choisie en raison d'une forte affinité de liaison (10 nM) et d'un faible encombrement stérique (49 Å).

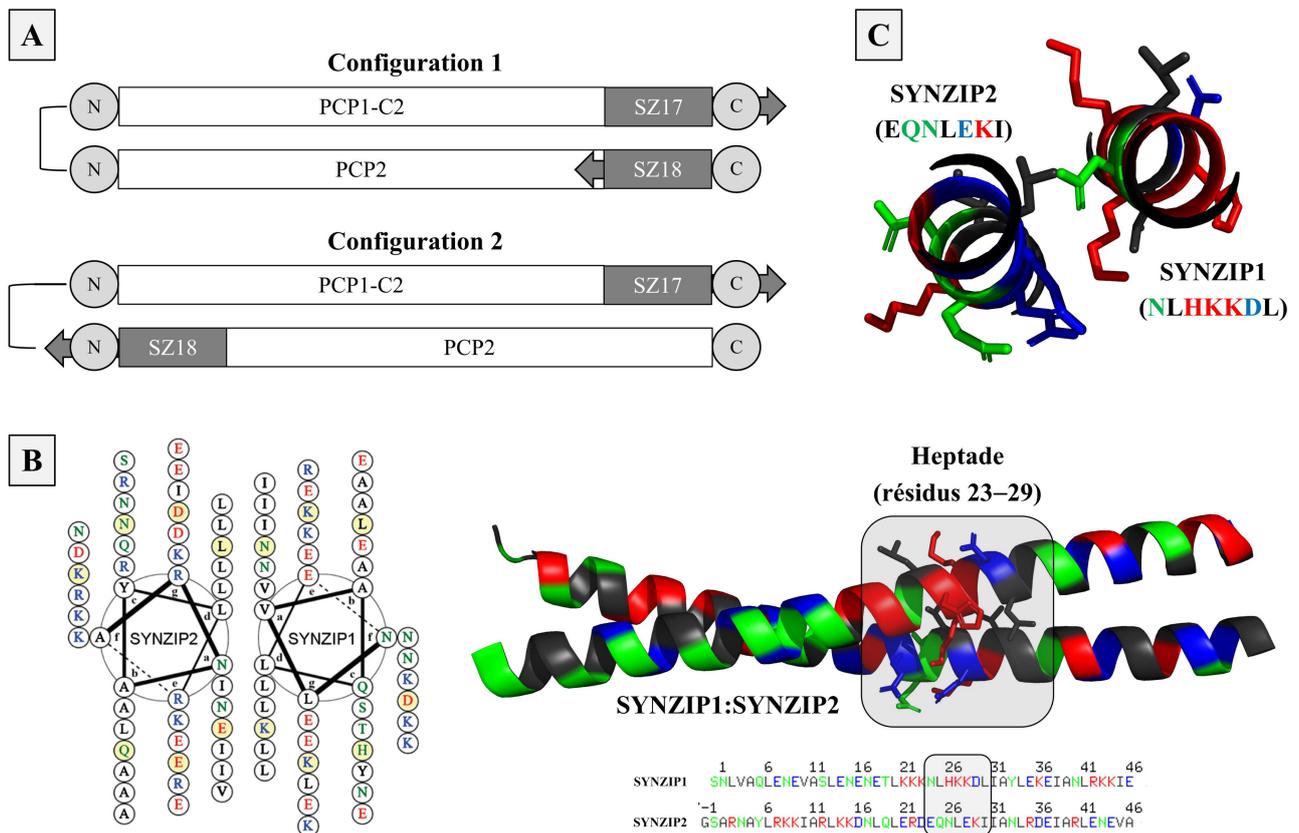


Figure 88. Configuration expérimentale des domaines chimériques et structuration heptamérique d'une paire de SYNZIP. (A) Les séquences du didomaine PCP1-C2 et du domaine PCP2 de la NRPS AzeB ont été clonées dans des plasmides d'expression originaux (voir section III.1.), et exprimées sous forme de protéines chimériques fusionnées aux SYNZIP17 ou SYNZIP18 aux bornes N- ou C-terminales, comme décrit sur le schéma. La paire de SZ17:SZ18 utilisée assure en particulier une interaction antiparallèle, contrairement aux autres SYNZIP disponibles (Reinke *et al.*, 2010). Cela permet théoriquement de placer les partenaires dans deux orientations différentes, chacune possiblement favorable à une interaction productive et à la formation d'un complexe PCP1-C2-PCP2 stabilisateur du domaine C2 d'intérêt. (B) Le diagramme en hélice du complexe SZ1:SZ2 (PDB : 3HE5) indique la composition en résidus à chaque position des séquences heptamériques [abcdefg]₇ (Reinke *et al.*, 2010). La structure adopte une superhélice gauche amphiphile de deux hélices α . Les 47 résidus hydrophobes, polaires, ou chargés +/- sont montrés en noir, vert, bleu et rouge. (C) Les sept séquences heptamériques contribuent chacune à la dimérisation du complexe *via* des interactions hydrophobes, ici les résidus 23–29 (NLHKKDL (SZ1) et EQNLEKI (SZ2)) inclus dans le quatrième motif sont représentés.

Les étapes suivantes de la stratégie établie ont impliqué d'exprimer les protéines chimériques en système bactérien hétérologue et de les purifier par chromatographie liquide, avant de caractériser des interactions potentielles par titration calorimétrique isotherme (ITC) et d'analyser les complexes résultants par DLS et SAXS. Dans ce cas, les SYNZIP présents aux bornes des domaines devraient réduire la flexibilité intrinsèque du domaine C2 et donc permettre d'obtenir des cristaux réguliers, ce qui a jusqu'alors fait défaut.

III.1. Conception et construction des plasmides d'expression originaux

III.1.A. Conception *in silico* des plasmides pSZ avec le logiciel SnapGene

Différents plasmides pSZ ont été conçus à partir d'un squelette pBG102 et de la séquence native des SZ17/SZ18 au moyen du logiciel SnapGene (**Figure 89**). Les plasmides ont été nommés comme suit : pSZ17WLC, pSZ18WLC, et pSZ18WLN, c'est-à-dire avec le site de clonage *Bam*HI-*Hind*III en phase avec la séquence SZ17 ou SZ18 présente à l'extrémité N- ou C-terminale, et sans séquence *linker* introduite entre ces éléments, exceptés des sites de restriction de type IIS résiduels.

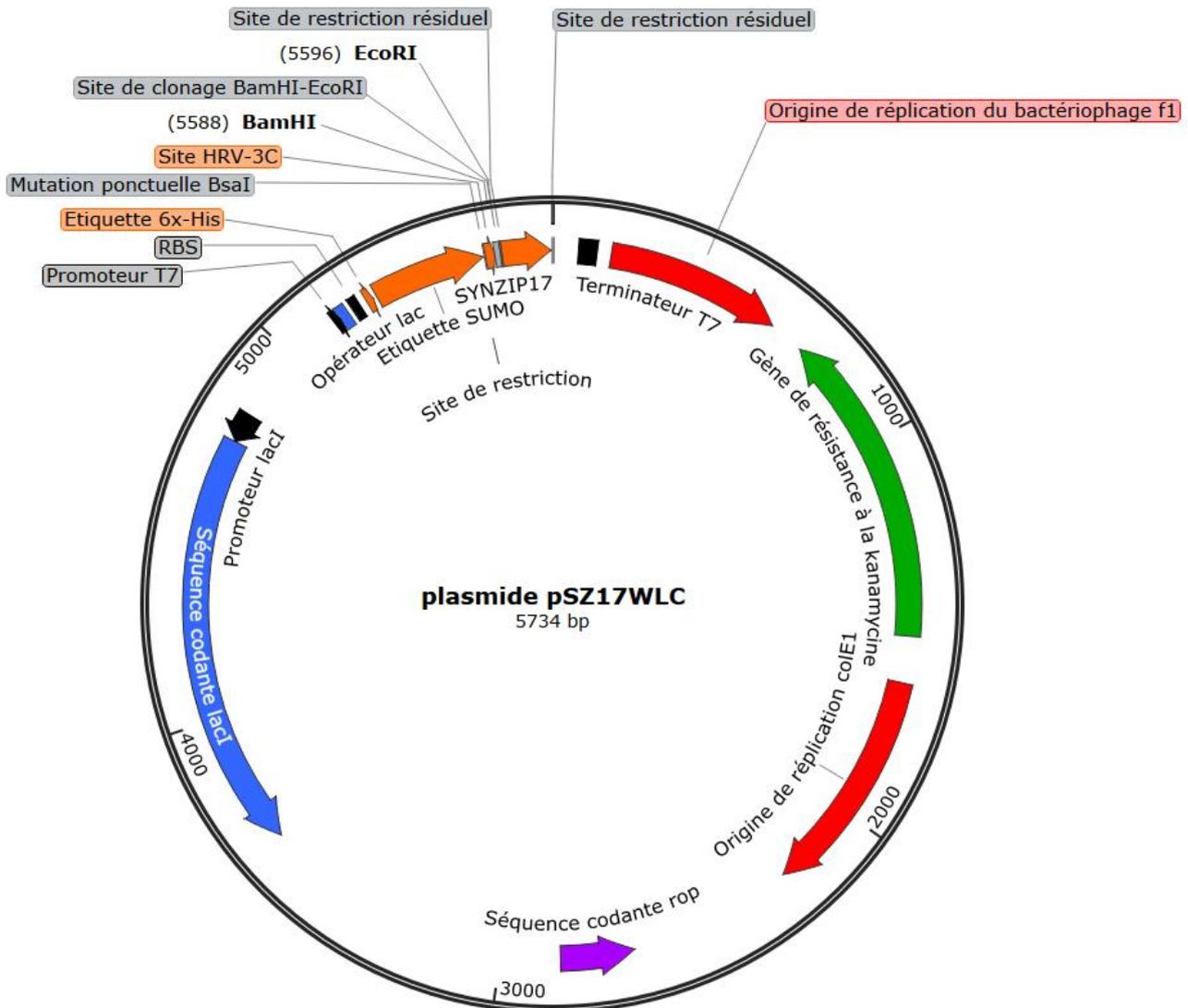


Figure 89. Représentation de la carte du plasmide d'expression pSZ17WLC. Le vecteur pSZ17WLC partage des éléments de séquences invariables avec le plasmide pBG102 (section I.2.), à savoir un site de clonage, en l'occurrence *Bam*HI-*Hind*III, des étiquettes de purification (6×His, SUMO), les séquences régulatrices de l'opéron lactose, l'origine de réplication *colE1* et le gène de résistance Kan^R. Il porte en plus, juste en aval du site de clonage, la séquence codante du SYNZIP17 et la séquence résiduelle de l'assemblage *in vitro*.

III.1.B. Construction des plasmides pSZ par un assemblage de type Golden Gate

Les séquences d'assemblage des plasmides, de 24 pb à plus de 5000 pb, ont été amplifiées par PCR sur matrice plasmidique (e.g. pBG102) ou directement appariées par chauffage-refroidissement des oligonucléotides complémentaires (**Tableaux VIII et IX**). Les amorces ont été conçues afin de s'hybrider au niveau de régions R1/R2 du plasmide pBG102 ou de séquences SYNZIP de plasmides pENTR-SZ17/SZ18 (Addgene) (**Tableau VIII**). En plus, les extensions des amorces ont été conçues pour intégrer la séquence de reconnaissance de l'enzyme de type IIS *BsaI* (5'-GGTCTC(N)₁↓-3') et un site de restriction pré-défini (TGTA, GTGA, TCAC, TCCT, AGGA, ACAT) (**Tableau VIII**).

Tableau VIII. Composition du milieu réactionnel de PCR et séquences des amorces/oligonucléotides.

Réactifs	Concentrations	Volumes
<i>High GC enhancer</i> (5 ×)	1×	10 µL
Matrice plasmidique (100 ng.µL ⁻¹)	100 ng	1 µL
Amorce sens (10 mM)	200 nM	1,2 µL
Amorce anti-sens (10 mM)	200 nM	1,2 µL
dNTP (5 mM)	1,2 mM	12 µL
Phusion HSII (2 U.µL ⁻¹)	1 U	0,5 µL
Eau ultra-pure	-	qsp 25 µL

Amorces	Séquences 5'-3'
R1C-pBG102-F	caGGTCTCATGTAgatccggctgctaacaagc
R1C-pBG102-R	gtGGTCTCTGTGAgcctcaataatcgttatcc
R1N-pBG102-F	caGGTCTCATGTAgatccggctgctaacaagccccgaaaggaagctgagttg
R1N-pBG102-R	gtGGTCTCTTCCTgagggcccctggaacagaactccagcccAagaccaccaatctg
R2WLC-pBG102-F	caGGTCTCATCACagagaacagattggtggtctTgggctggaagtctg
R2WLC-pBG102-R	gtGGTCTCTTCCTGAATTCggGGATCCgggcccctggaacag
SZ17C-pENTR-F	caGGTCTCAAGGAAACGAAAAAGAAGAAGACTGAAATC
SZ17C-pENTR-R	gtGGTCTCTTACATCATTGTGAAGCTTCGATTTC
SZ18C-pENTR-F	caGGTCTCAAGGAAGCATCGCGGCGACCCTG
SZ18C-pENTR-R	gtGGTCTCTTACATCAGAAGTACGCTTCTTCACGCTC
SZ17N-pENTR-F	caGGTCTCAAGGAAACGAAAAAGAAGAAGACTGAAATC
SZ17N-pENTR-R	gtGGTCTCTTCACTTTGTGAAGCTTCGATTTCTTTAC
SZ18N-pENTR-F	caGGTCTCAAGGAAGCATCGCGGCGACCCTG
SZ18N-pENTR-R	gtGGTCTCTTACGAAGTACGCTTCTTCACGCTCC
Oligonucléotides	Séquences 5'-3'
R2WLN-pBG102-F	caGGTCTCAGTGAgcGGATCCccGAATTC
R2WLN-pBG102-R	gtGGTCTCTTACATCAaGAATTCggGGATCC

Tableau IX. Conditions expérimentales d'une réaction de PCR des séquences d'assemblage.

Phases de la réaction		Temps	Températures
Dénaturation initiale		5 min	98 °C
Dénaturation des matrices double-brin		30 s	98 °C
Hybridation des amorces	30 cycles	30 s	55–72 °C
Élongation des brins		30 s –5 min	72 °C
Élongation terminale		10 min	72 °C

Les fragments ont été purifiés par électrophorèse en gel d'agarose (0,8–2 %) et extraits à partir de 50 µL du mélange réactionnel avec le kit NucleoSpin® PCR Clean-Up Gel Extraction (Macherey-Nagel). Les plasmides ont été construits avec une quantité équimolaire des fragments (50 fmol) par un assemblage de type Golden Gate (50 cycles) en présence de 6 U de l'enzyme de restriction *BsaI* (5'–GGTCTC(N)₁↓–3') et 40 U de l'ADN ligase T4 dans un tampon spécifique (500 mM Tris-HCl, pH 7,6, 100 mM MgCl₂, 10 mM ATP, 10 mM DTT, 25 % PEG-8000) (Thermo Fischer Scientific) (**Figure 90, Tableaux X et XI**). La méthode met à profit les propriétés des endonucléases de type IIS, lesquelles reconnaissent leur séquence spécifique (asymétrique) à chaque extrémité des fragments et clivent l'ADN à un site de 4 nucléotides pré-définis non-palindromiques situés en aval (Engler *et al.*, 2008). Étant donné que les fragments générés ne contiennent plus aucune séquence de reconnaissance de l'enzyme de restriction, la ligase T4 catalyse simultanément l'assemblage directionnel du produit, au niveau des régions simple-brin complémentaires, au cours d'une réaction monotope irréversible.

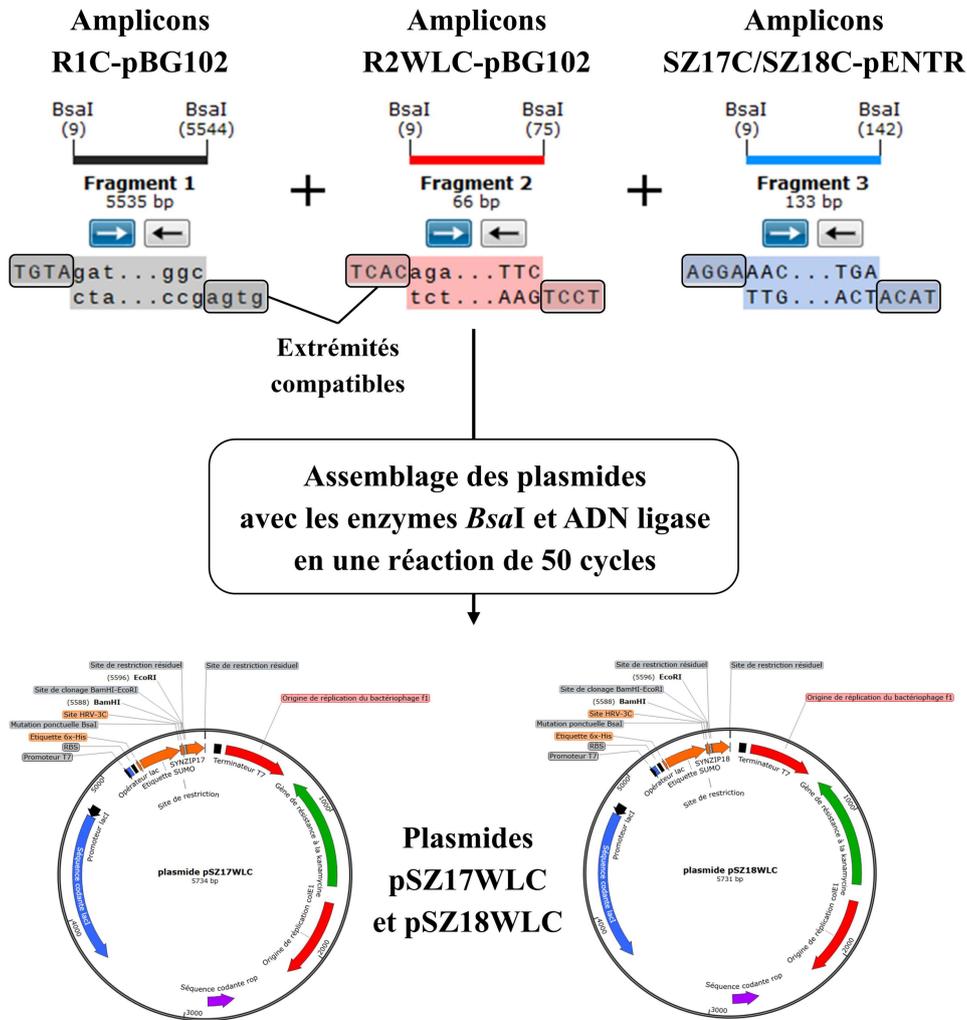


Figure 90. Construction des plasmides pSZ par assemblage de type Golden Gate. Dans le cas de l'assemblage des plasmides pSZ17WLC et pSZ18WLC, les fragments R1C, R2WLC et SZ17C/SZ18C, ayant préalablement été amplifiés avec des sites de restriction conformes (TGTA, GTGA, TCAC, TCCT, AGGA, ACAT), ont été digérés par l'enzyme de type IIS *BsaI* et ligaturés dans un ordre défini par l'ADN ligase T4 lors d'une réaction *one-pot* de 50 cycles thermiques (Tableaux X et XI). De manière analogue, les fragments R1N, SZ18N et R2WLN, ici non-représentés, ont donné lieu à l'assemblage du plasmide pSZ18WLN.

Tableau X. Composition du milieu réactionnel d'une réaction de Golden Gate.

Réactifs	Concentrations	Volumes
Tampon de l'ADN ligase T4 (10 ×)	1 ×	2 µL
R1C/R1N-pBG102 (300 ng.µL ⁻¹)		0,3 µL
R2WLC/R2WLN-pBG102 (30 ng.µL ⁻¹)	50 fmol	0,3 µL
SZ17/SZ18-pENTR (30 ng.µL ⁻¹)		0,3 µL
Enzyme de restriction <i>BsaI</i> (20 U.µL ⁻¹)	6 U	0,3 µL
ADN ligase T4 (80 U.µL ⁻¹)	40 U	0,5 µL
Eau ultra-pure	-	qsp 20 µL

Tableau XI. Conditions expérimentales d'une réaction de Golden Gate.

Phases de la réaction	Temps	Températures
Activation de l'endonucléase <i>BsaI</i>	50 cycles	37 °C
Activation de l'ADN ligase T4		16 °C (optimale)
Activation conjointe des enzymes	15 min	37 °C
Inactivation de l'endonucléase <i>BsaI</i>	20 min	65 °C
Inactivation de l'ADN ligase T4 (optionnelle)	5 min	80 °C

Les mélanges réactionnels de Golden Gate ont été traités avec 10 U de l'endonucléase *DpnI* dans le tampon FastDigest (Thermo Fischer Scientific) à 37 °C pendant 1 h. Cette étape a été réalisée afin d'éliminer sélectivement les plasmides originels pBG102 et pENTR-SZ17/SZ18 méthylés ayant été utilisés comme matrice de PCR. Les bactéries *E. coli* DH5 α ont alors été transformées par 5 μ L des mélanges réactionnels, puis les produits attendus ont chacun été détectés par PCR sur colonies avec les amorces ayant été utilisées dans l'assemblage des fragments. À partir de clones bactériens positifs, les plasmides chimériques ont été préparés avec le kit NucleoSpin Plasmid DNA (Macherey-Nagel) et linéarisés pour procéder à l'étape de clonage des inserts PCP1-C2 et PCP2.

III.2. Clonage des séquences géniques dans les plasmides d'expression

Les séquences du didomaine PCP1-C2 et du domaine PCP2 ont été amplifiées par PCR (**Tableaux XII et XIII**). Les amorces ont été conçues afin de s'hybrider aux bornes des séquences délimitées, et d'incorporer les sites de restriction des enzymes *Bam*HI et *Eco*RI (**Tableau XII**).

Tableau XII. Composition du milieu réactionnel de PCR et séquences des amorces.

Réactifs	Concentrations	Volumes
<i>High GC enhancer</i> (5 \times)	1 \times	10 μ L
Matrice réactionnelle (100 ng. μ L ⁻¹)	100 ng	1 μ L
Amorce sens (10 mM)	200 nM	1,2 μ L
Amorce anti-sens (10 mM)	200 nM	1,2 μ L
dNTP (5 mM)	1,2 mM	12 μ L
Phusion HSII (2 U. μ L ⁻¹)	1 U	0,5 μ L
Eau ultra-pure	-	qsp 25 μ L

Amorces	Séquences 5'-3'
PCP1-C2-F	TTTCCGGATCCATGGAACAGGTGGCGGCGGTTGC
PCP1-C2-R	TTTCCGAATTCccGTCGCGGCGAGGGGCGAG
PCP2-F	TTTCCGGATCCATGGAGCACGGGACGAACATCG
PCP2-R	TTTCCGAATTCccACGGATCGGTACCAGCGG

Tableau XIII. Conditions expérimentales d'une réaction de PCR des séquences PCP1-C2 et PCP2.

Phases de la réaction		Temps	Températures
Dénaturation initiale		5 min	98 °C
Dénaturation des matrices double-brin		30 s	98 °C
Hybridation des amorces	30 cycles	30 s	58–63 °C
Élongation des brins		1 min 30 s	72 °C
Élongation terminale		10 min	72 °C

Les amplicons ont été purifiés par électrophorèse en gel d'agarose (0,8 %) et extraits à partir de 100 µL du mélange réactionnel avec le kit NucleoSpin® PCR Clean-Up Gel Extraction (Macherey-Nagel). Les produits d'extraction (300–500 ng) ont été digérés en présence de 5 unités des enzymes de restriction *Bam*HI (5'-G↓GATCC-3') et *Eco*RI (5'-G↓AATTC-3') dans le tampon Fast Digest (Thermo Fisher Scientific) à 37 °C pendant 1 h.

Les inserts (50 ng) ont été assemblés avec les plasmides pSZ17WLC (PCP1-C2), pSZ18WLC (PCP2) et pSZ18WLN (PCP2) en présence de 10 U de l'ADN ligase T4 dans un tampon spécifique (500 mM Tris-HCl, pH 7,6, 100 mM MgCl₂, 10 mM ATP, 10 mM DTT, 25 % PEG-8000) et le mélange réactionnel a été incubé à température ambiante pendant 20 h.

Les bactéries *E. coli* DH5α ont alors été transformées par 5 µL des mélanges réactionnels. Les plasmides recombinants ont été préparés avec le kit NucleoSpin® Plasmid DNA (Macherey-Nagel), et analysés par digestion et séquençage (Eurofins Genomics). Cette étape a été opérée pour vérifier la fonctionnalité des plasmides pSZ nouvellement construits et l'intégrité de la séquence des inserts.

III.3. Production et purification des protéines recombinantes

Les protéines chimériques PCP1-C2-SZ17C, PCP2-SZ18C et SZ18N-PCP2 ont été exprimées à partir des plasmides recombinants chez l'hôte *E. coli* BL21 (DE3) par l'addition d'IPTG (0,5 mM). La biomasse cellulaire a été lysée par sonication (15 kpsi, 4 °C, 2 min) et le surnageant obtenu a été filtré au travers d'une membrane stérile (0,22 µm). Les échantillons de protéines ont été purifiés par chromatographie en phase liquide, en l'occurrence des étapes d'IMAC (colonne Ni²⁺-sépharose), de dialyse intermédiaire et protéolyse des étiquettes, et de filtration sur gel (colonnes Superdex-75/200) dans le tampon A (20 mM Tris-HCl pH 7,8, 300 mM NaCl, 5 % glycérol) avec les appareils ÄKTA. Les protéines d'intérêt ont été dosées à 280 nm par spectrophotométrie, analysées par SDS-PAGE et conservées à -80 °C. L'ensemble de ces étapes a davantage été détaillé dans la section II.2.

III.4. Caractérisation des interactions protéine-protéine par titration calorimétrique isotherme

Les échantillons ont été analysés par titration calorimétrique isotherme avec l'appareil MicroCal iTC200 (Malvern Panalytical) pour déterminer les paramètres thermodynamiques relatifs à la liaison des protéines chimériques, tels que la stœchiométrie réactionnelle et la variation enthalpique (ΔH), ainsi que les constantes cinétiques d'association/dissociation (K_a/K_d) des complexes résultants. Dans ces expériences, le didomaine PCP1-C2-SZ17C (300 μL , 65 μM) a d'abord été injecté dans la cellule d'échantillonnage du calorimètre, pré-calibré avec le tampon, puis le partenaire (soit PCP2-SZ18C, soit SZ18N-PCP2) (50 μL , 650 μM) a été progressivement ajouté en excès ($\times 10$) (20 injections de 2 μL , une par période de 120 s) (**Figure 91A**). En temps réel, chaque variation de chaleur ($\mu\text{cal}\cdot\text{s}^{-1}$) a été détectée, mesurée et compensée par rapport à la cellule de référence, jusqu'à l'épuisement du réactif titrant (**Figure 91B**). Les aires des pics correspondants ont été corrigées du bruit de fond (*i.e.* chaleur de dilution) et intégrées en une courbe isotherme avec le programme NITPIC (Scheuermann et Brautigam, 2015) (**Figure 91B**). Le rapport molaire pris au centre du modèle de liaison indique la stœchiométrie N et, à partir de la relation de van't Hoff, le coefficient directeur de la tangente donne l'affinité K_d de liaison des protéines en solution (Scheuermann et Brautigam, 2015) (**Figure 91B**).

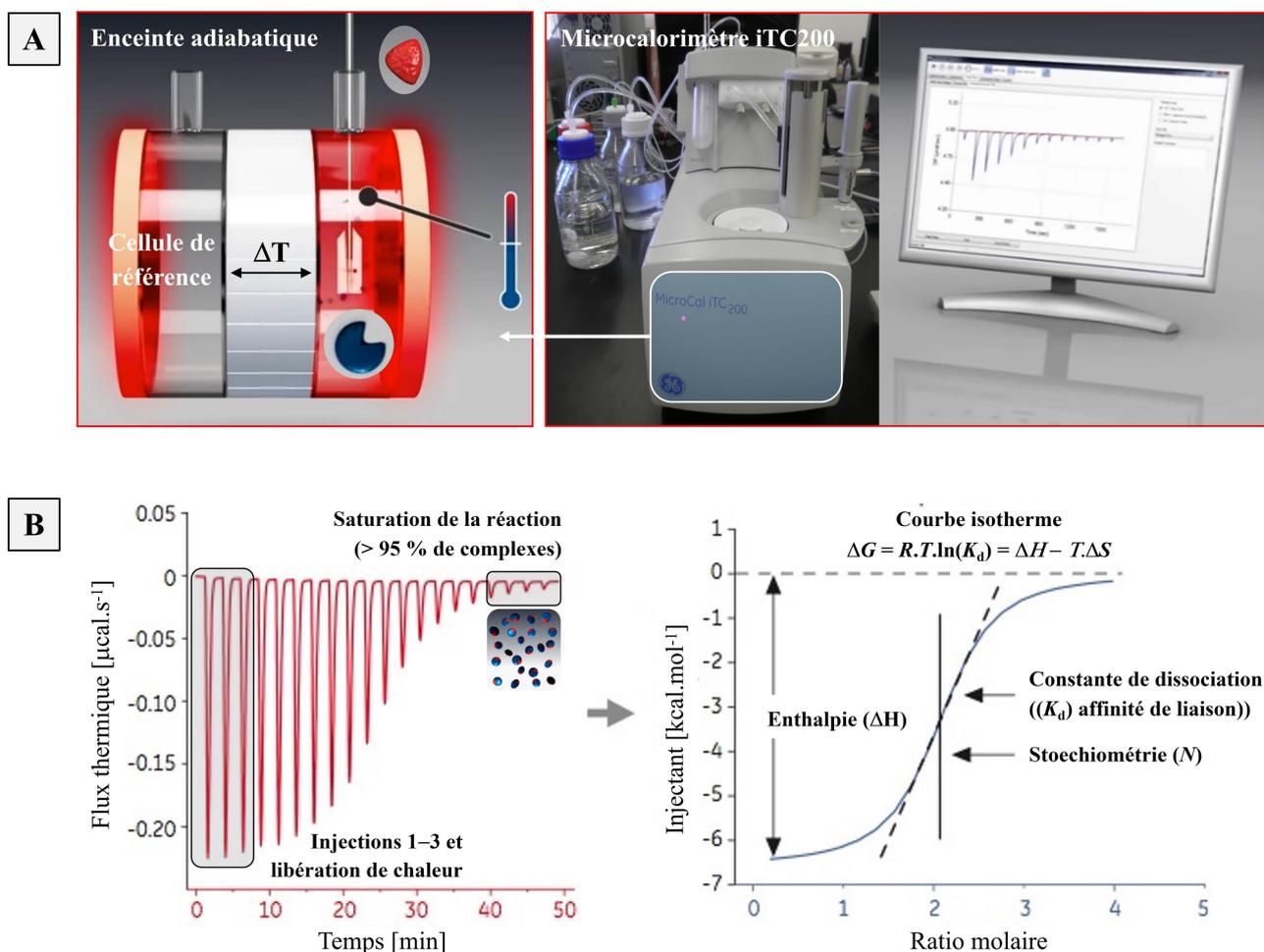


Figure 91. Analyse d'un échantillon par ITC (adaptée de Malvern Panalytical). (A) Dans une expérience d'ITC, les protéines interagissent dans une cellule d'échantillonnage, et le microcalorimètre mesure et compense les flux de chaleur résultants. (B) L'aire de chaque pic a été intégrée en une courbe de titrage et celle-ci a été ajustée à un modèle de liaison.

III.5. Analyse des échantillons par diffusion d'ondes électromagnétiques

Les échantillons des protéines ont été un à un analysés par DLS et les complexes correspondants ont été étudiés par SAXS afin d'obtenir des informations structurales pertinentes, comme cela a été décrit aux sections **II.3.** et **II.5.**

III.6. Cristallisation des complexes intermoléculaires par diffusion de vapeur

Les protéines chimériques ont été diluées à des concentrations équimolaires (100 μM ou 150 μM) et les conditions de cristallisation des complexes non-covalents (*i.e.* PCP1-C2-SZ17C:PCP2-SZ18C et PCP1-C2-SZ17C:SZ18N-PCP2) ont été criblées par diffusion de vapeur en nano-gouttes dans des plaques MRC3 avec un robot Mosquito. Les essais sont voués à être poursuivis.

RÉSULTATS ET DISCUSSION

SECTION I. Élucidation des relations structure-fonction de l'hydrolase XhpG

En vue d'étudier le mécanisme catalytique et la spécificité de substrat de l'hydrolase XhpG, une stratégie intégrative a été appliquée de l'analyse de la séquence d'ADN à la résolution de la structure cristallographique. En préambule, la séquence primaire de l'enzyme a été alignée avec des séquences homologues et, de cette façon, une triade catalytique putative a été identifiée (**Figure 92**). Dès lors, la séquence génique a été clonée dans le plasmide pBG102 et le codon du résidu nucléophile Ser98 a été modifié par PCR mutagène, avant que la construction correspondante (S98A) ne soit exprimée chez un hôte hétérologue (*E. coli* BL21 (DE3)). La protéine recombinante a ensuite été purifiée par chromatographie en phase liquide, et analysée par DLS et SAXS (**Figures 93 et 94**). Ces étapes ont permis de confirmer l'homogénéité et la monodispersité des échantillons, des conditions nécessaires à l'étude structurale approfondie de l'hydrolase XhpG S98A.

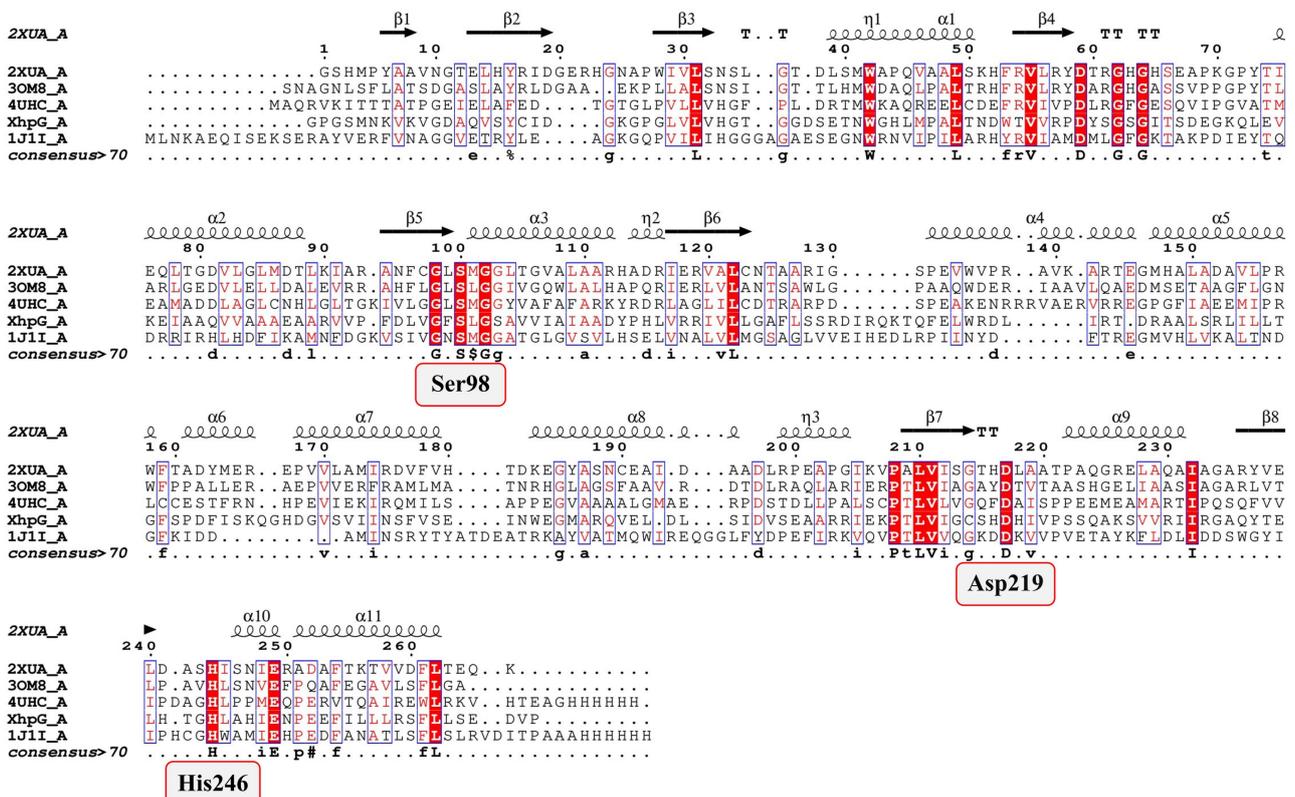


Figure 92. Alignement séquentiel des homologues de l'hydrolase XhpG. La séquence primaire de l'hydrolase XhpG a été analysée par alignement avec des séquences d'enzymes homologues, y compris une énoïl-lactonase (PDB : 2XUA), une hydrolase putative (PDB : 3OM8), une estérase (PDB : 4UHC), et une sérine-hydrolase (PDB : 1J11), identifiées à l'aide de la banque PDB et du serveur de comparaison structurale Dali (Berman *et al.*, 2000 ; Holm *et al.*, 2023). Les enzymes présentent une identité séquentielle supérieure à 25 %, à l'exception de la sérine-hydrolase CarC (PDB : 1J11, 17 %), et partagent des déterminants responsables de la liaison du substrat et de la réaction catalytique. En l'occurrence, l'hydrolase XhpG contient les résidus Ser98, Asp219, et His246 constitutifs de la triade catalytique conservée. De plus, elles montrent un Z-score de similarité de repliement important, compris entre 30 % et 50 %. Étant donné que le résidu Ser98 du motif GxSxG consensus est essentiel à la catalyse, le codon en question a été muté afin d'inactiver l'enzyme.

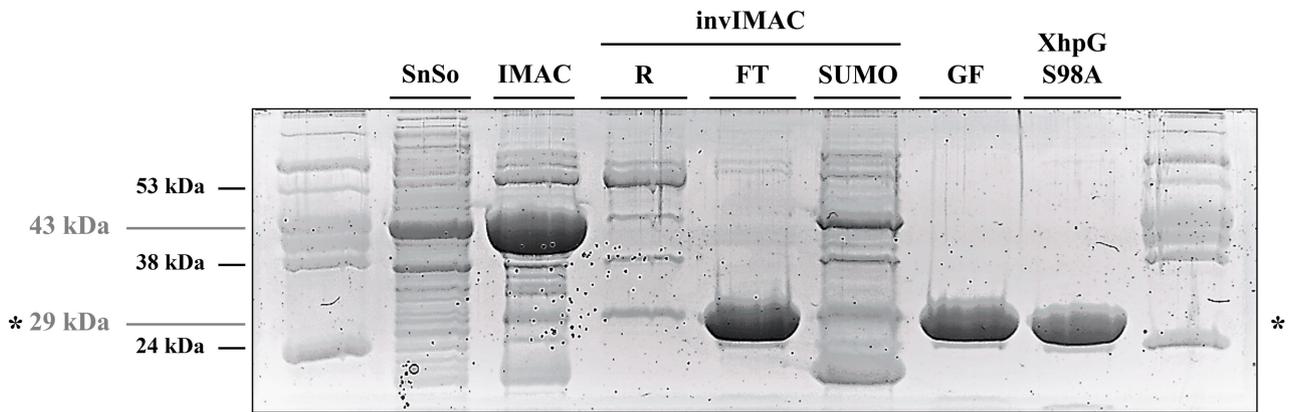


Figure 93. Purification de l'hydrolase XhpG S98A mutante par chromatographie liquide. L'enzyme XhpG S98A mutante a été purifiée par une chromatographie d'affinité sur ions nickel immobilisés (notée IMAC), suivie d'une dialyse en présence de la protéase HRV-3C, afin de cliver les étiquettes SUMO et 6×His. Suite à une seconde étape d'IMAC (notée invIMAC), où les protéines dépourvues d'affinité ont été éluées par lavage (noté FT), l'enzyme a été purifiée par chromatographie d'exclusion stérique (notée GF). Le rendement final de la purification a été estimé à 7 mg.L^{-1} .

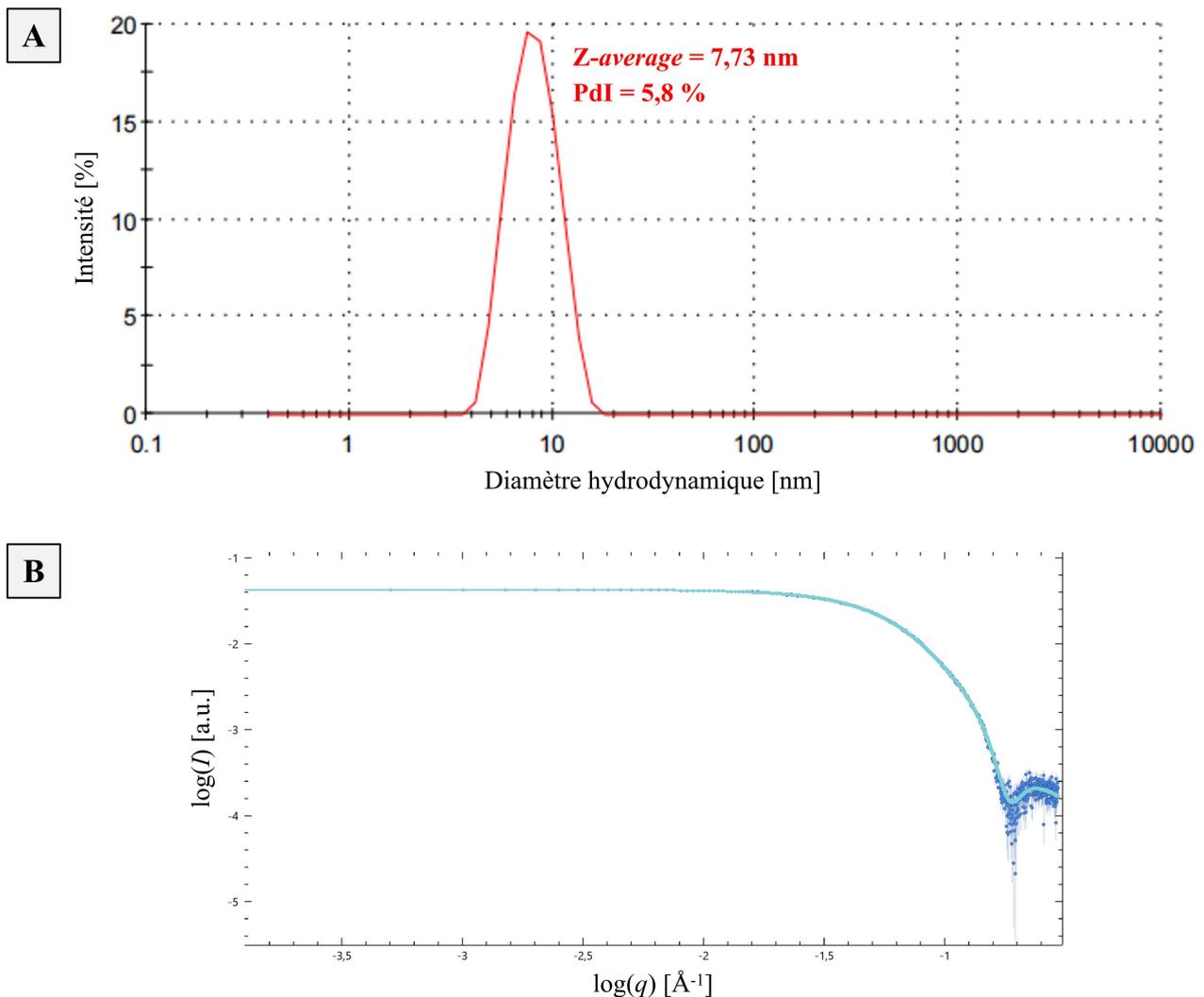


Figure 94. Analyse de l'hydrolase XhpG S98A mutante par DLS et SAXS. (A) Les résultats de DLS indiquent que l'enzyme a un diamètre hydrodynamique moyen (*Z-average*) de $8 \pm 2 \text{ nm}$ et que l'échantillon est homogène avec un faible indice de polydispersité ($\text{PDI} = 5,8 \%$). (B) La courbe SAXS a été convertie en un graphique avec une double-échelle logarithmique, où les données à faibles valeurs de q présentent un plateau sans artéfact, ce qui est donc cohérent avec un échantillon de protéines monodispersé.

I.1. Résolution de la structure cristallographique

L'hydrolase XhpG S98A mutante a été cristallisée sous forme d'une *apo*-protéine par diffusion de vapeur et les conditions optimales ont été affinées (**Figure 95**). Cette étape a permis d'améliorer la reproductibilité des cristaux de protéines, avant d'en collecter un jeu de données de haute qualité par diffraction des rayons X au synchrotron. Parallèlement, les expériences de co-cristallisation du complexe binaire et de diffusion du substrat ont échoué, à cause de la grande instabilité de celui-ci, de la faible teneur en solvant (31 %) et de l'inaccessibilité du site actif dans l'empilement cristallin.

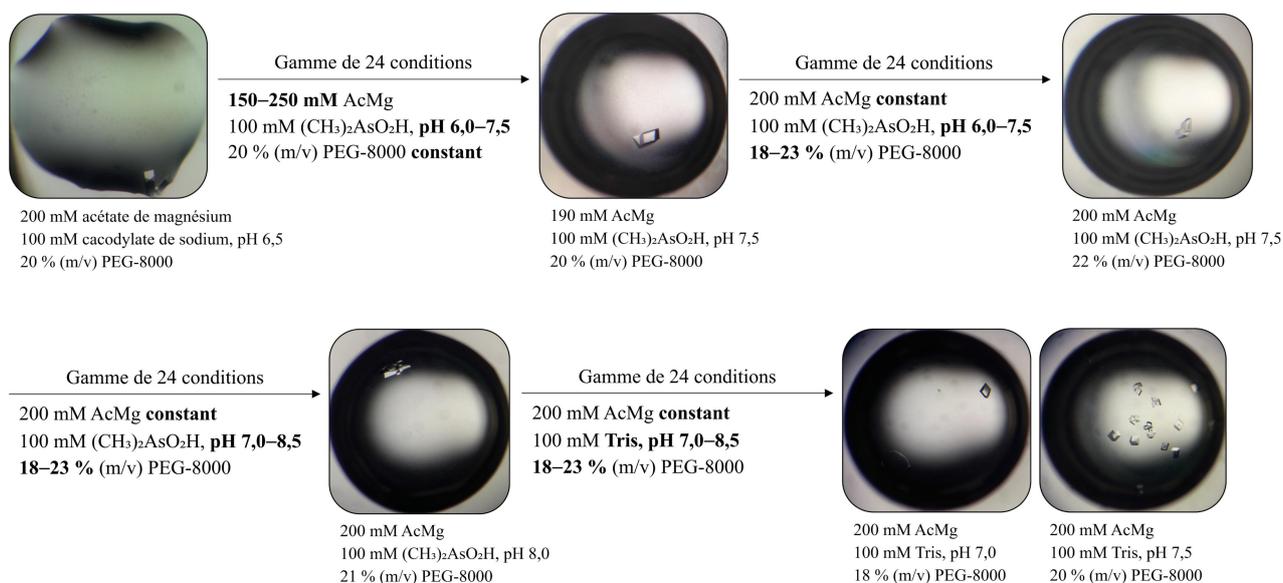


Figure 95. Reproduction des cristaux de l'hydrolase XhpG S98A mutante par diffusion de vapeur. Les cristaux de macromolécules obtenus à 11 mg.mL⁻¹ en présence de 200 mM d'acétate de magnésium (AcMg), 100 mM de cacodylate de sodium ((CH₃)₂AsO₂H) à pH 6,5, et 20 % (m/v) de PEG-8000, ont alors été affinés par diffusion de vapeur en microgouttes suspendues dans des plaques à 24 puits à partir de gammes de solutions tampons spécifiques et de volumes avec des ratios différents (1:2, 2:4, *etc.*). Les variables chimiques comme la concentration en agents précipitants ou le pH du tampon ont été ajustées à des valeurs adéquates en vue d'optimiser la qualité des cristaux produits. Des plans factoriels incomplets, non-représentés ici, ont également été réalisés dans cet objectif. PEG : polyéthylène glycol.

La structure tridimensionnelle de l'hydrolase XhpG S98A de *X. hominickii* a été déterminée par remplacement moléculaire (AlphaFold2) à une résolution de 1,18 Å (PDB : 8QZU) (**Annexe PDB**). L'unité asymétrique du cristal contient un monomère de 29 kDa. Les résidus désordonnés sont rares dans la structure : Gly1 et Pro2 de la séquence résiduelle, ainsi que Asp268, Val269 et Pro270 de la boucle C-terminale sont absents des cartes de densité électronique.

I.1.A. Description du repliement α/β hydrolase

L'enzyme XhpG appartient à la super-famille des α/β hydrolases qui catalysent un large éventail de réactions et partagent un repliement caractéristique, composé d'un cœur catalytique et d'un sous-domaine α -hélicoïdal (Ollis *et al.*, 1992). Le noyau central de l'hydrolase consiste en un feuillet α/β , formé de huit brins β protégés du solvant par quatre hélices α d'un côté et par deux de l'autre (Bugg, 2004 ; Carr et Ollis, 2009) (**Figure 96**). Le feuillet β présente la topologie canonique 12435678 avec le brin $\beta 2$ antiparallèle aux autres, et une structure torsadée, au point que le premier ($\beta 1$) et le dernier brin ($\beta 8$) soient presque orthogonaux (Carr et Ollis, 2009). Le sous-domaine α -hélicoïdal constitué de cinq hélices ($\alpha 4$ – $\alpha 8$) et connecté au cœur catalytique par une région charnière (boucle $\beta 6$ - $\alpha 4$) de huit résidus (GAFLSSRD), fait saillie de la structure et agit comme une « coiffe » au niveau du site actif, comme cela est décrit pour les homologues (Bains *et al.*, 2011) (**Figure 96**).

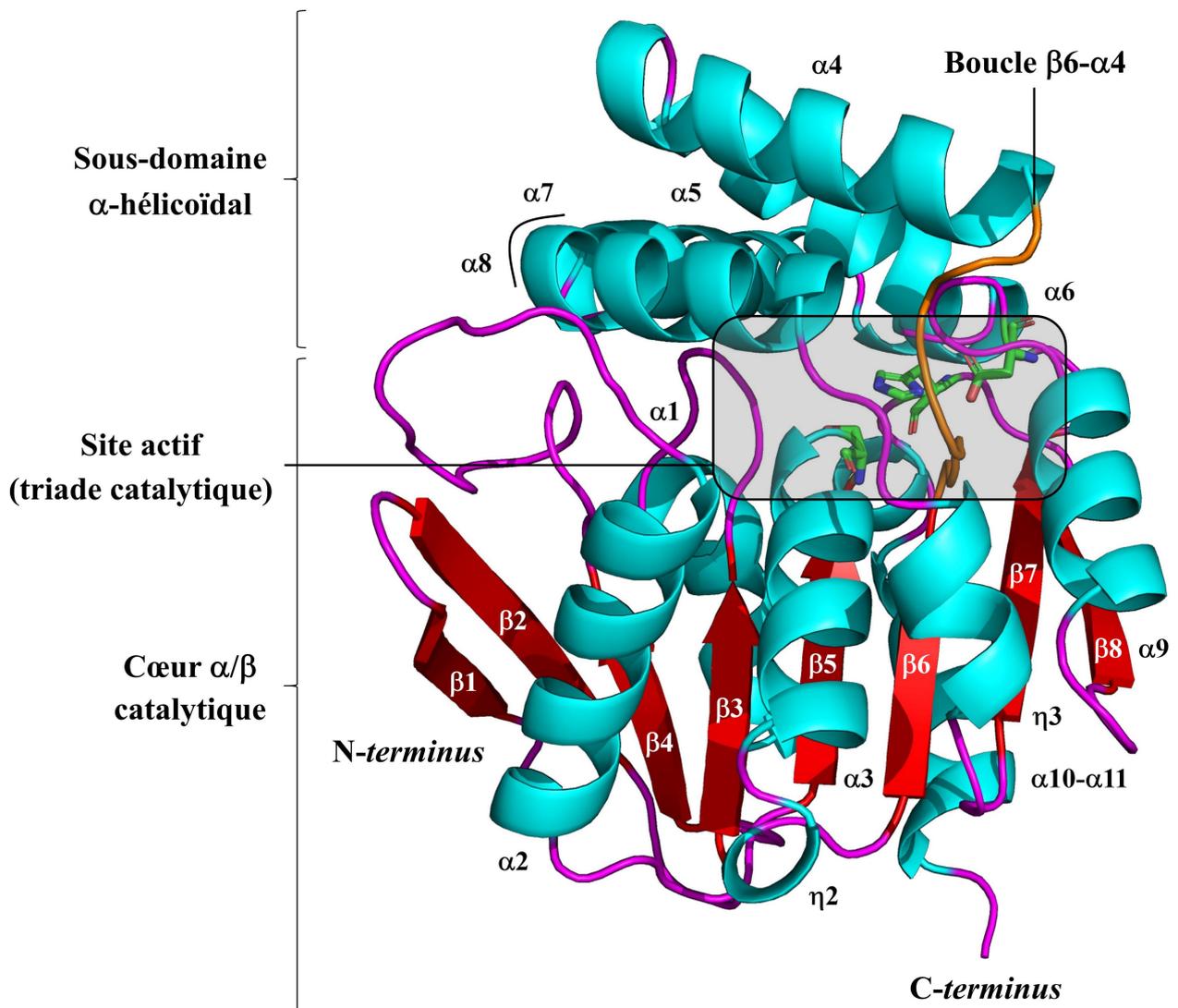


Figure 96. Représentation de la structure cristallographique de l'hydrolase XhpG. Le monomère de l'hydrolase XhpG contient un cœur catalytique, structuré en huit brins β ($\beta 1$ - $\beta 2$ - $\beta 4$ - $\beta 3$ - $\beta 5$ - $\beta 6$ - $\beta 7$ - $\beta 8$) flanqués de six hélices α , et un sous-domaine α -hélicoïdal ($\alpha 4$ – $\alpha 8$), connecté par une séquence *linker* (boucle $\beta 6$ - $\alpha 4$), responsable de la reconnaissance du substrat. Les hélices α et les hélices 3_{10} (η) sont représentées en cyan, et les brins β en rouge. PDB : 8QZU.

I.1.B. Organisation structurale du site actif et liaison du substrat

Le site actif de l'hydrolase XhpG se situe au niveau de l'interface entre les sous-domaines central et α -hélicoïdal et, dans une forme *apo* de l'enzyme, le canal hydrophobe est relativement accessible (8 Å de diamètre) au substrat attendu. Au contraire des lipases, par exemple, dont le sous-domaine hélicoïdal en obstrue le tunnel d'entrée et où le substrat y induit des changements conformationnels requis pour l'activation de l'interface catalytique (Ericsson *et al.*, 2008 ; Sayer *et al.*, 2015). Dans la structure de XhpG, la triade catalytique correspond aux résidus Ser98, Asp219 et His246, chacun étant situé dans une région non-structurée distincte : un coude γ (boucle $\beta 5$ - $\alpha 3$), connu sous le terme de « coude nucléophile » avec la sérine positionnée au centre, ainsi que les boucles $\beta 7$ - $\alpha 9$ (Asp) et $\beta 8$ - $\alpha 10$ (His) (Carr et Ollis, 2009). Le mécanisme réactionnel correspond incontestablement à celui établi pour les protéases à sérine, au cours duquel la triade Ser-His-Asp et un « trou de l'oxyanion » stabilisent les états de transition et autres intermédiaires céto-énoliques (Hedström, 2002) (**Annexe 1**). Au sein de l'hydrolase XhpG, le résidu Ser98 est inclus dans le motif GFSLG, dont les résidus Gly y imposent une courbure marquée, ce qui permet audit nucléophile de s'approcher à proximité (2,8 Å) du substrat oxopyrrolizidinacétamide. Le « trou de l'oxyanion » comporte les groupements amides des résidus Leu99 et Thr32, localisés au sommet de l'hélice $\alpha 3$ ou en aval du brin $\beta 3$, respectivement, comme l'illustrent les alignements structuraux réalisés avec plusieurs enzymes homologues.

Pour comprendre les bases moléculaires de la liaison du substrat, des expériences de *docking in silico* ont été effectuées avec le programme CB-Dock2 (Liu *et al.*, 2022) (**Figure 97**). Ce modèle a révélé que la fixation du substrat dans une orientation correcte par rapport à la machinerie catalytique résulte à la fois d'interactions hydrophobes et de liaisons hydrogènes. En effet, les chaînes latérales de résidus apolaires contactent les atomes de carbone C1 (Trp139, Ile221), C4 (Phe97, Leu247) et C8 (Ile154, Phe179 et Ile183). De plus, les fonctions du résidu Asn39 et du nucléophile établissent chacune une liaison hydrogène forte (30 kcal.mol⁻¹) avec un groupement carbonyle du ligand. Donc, le mode de fixation du substrat pourrait stabiliser la conformation tridimensionnelle du site actif et créer l'environnement chimique favorable à la réaction catalytique. Toutefois, faute d'avoir réussi à obtenir les cristaux de l'*holo*-protéine, cette supputation n'aura pu être expérimentalement vérifiée.

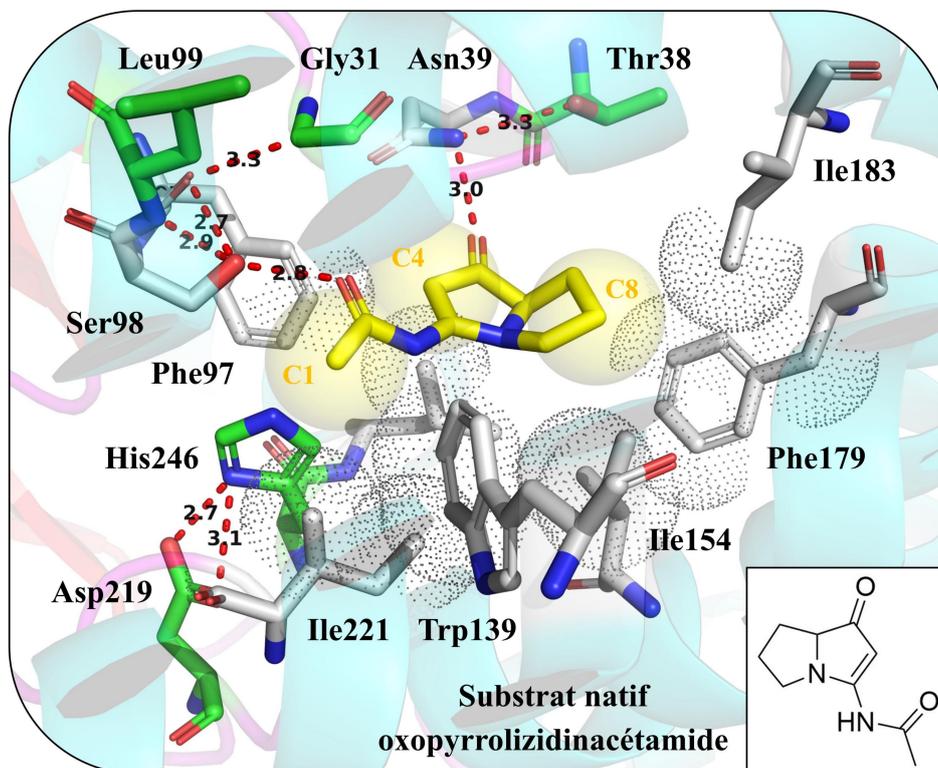


Figure 97. Liaison du substrat oxopyrrolizidinacétamide dans le site actif de l'hydrolase XhpG par modélisation moléculaire. Des interactions hydrophobes (sphères) et des liaisons hydrogènes (tirets pointillés) formées au niveau de l'hétérocycle oxopyrrolizidine et de la chaîne acétamide positionnent et orientent le substrat natif par rapport à la triade catalytique (Ser98, Asp219 et His246). Les résidus directement impliqués dans la reconnaissance du substrat *docké* sont indiqués en gris (*i.e.* Asn39, Phe97, Ser98, Trp139, Ile154, Phe179, Ile183, Ile221 et Leu247), et ceux engagés dans des interactions avec d'autres acides aminés sont représentés en vert (*i.e.* Gly31, Thr38, Met88, Leu99, Asp219 et His246).

I.2. Comparaison avec la structure cristallographique d'une enzyme homologue

I.2.A. Organisation structurale du sous-domaine α -hélicoïdal et de la région charnière

Les enzymes ayant un repliement α/β hydrolase partagent une structure tertiaire et des propriétés catalytiques communes. Afin de vérifier cette observation, la structure de l'hydrolase XhpG a alors été comparée à celle de la protéine homologue la plus proche, l'énol-lactonase PcaD (PDB : 2XUA) identifiée avec le serveur Dali (Bains *et al.*, 2011) (**Figure 98**). En effet, ces enzymes adoptent un repliement identique, comme en témoigne le RMSD calculé de 1,4 Å (212 atomes C α). Néanmoins, l'enzyme XhpG diffère notablement de l'homologue PcaD en termes de position, d'orientation et de longueur relative des hélices α du sous-domaine hélicoïdal (**Figure 98A**). La région charnière inter-domaine diverge aussi considérablement au niveau de la composition en résidus (NTAARIG au sein de l'énol-lactonase PcaD) (**Figure 98B**). Contrairement à certains membres de la super-famille des hydrolases pour lesquels une boucle flexible assure des ajustements conformationnels nécessaires à la fixation du substrat et à la libération du produit, ce *linker* semble rigide avec un facteur d'agitation thermique moyen inférieur à 20–25 Å² dans les monomères XhpG et PcaD (Otyepka et Damborský, 2002). Cette constatation suggère que la coiffe de l'enzyme XhpG pourrait ne pas alterner entre des états ouverts et fermés, une hypothèse cohérente avec la diffusion directe du substrat dans le tunnel du site actif même en l'absence d'un tel mouvement charnière.

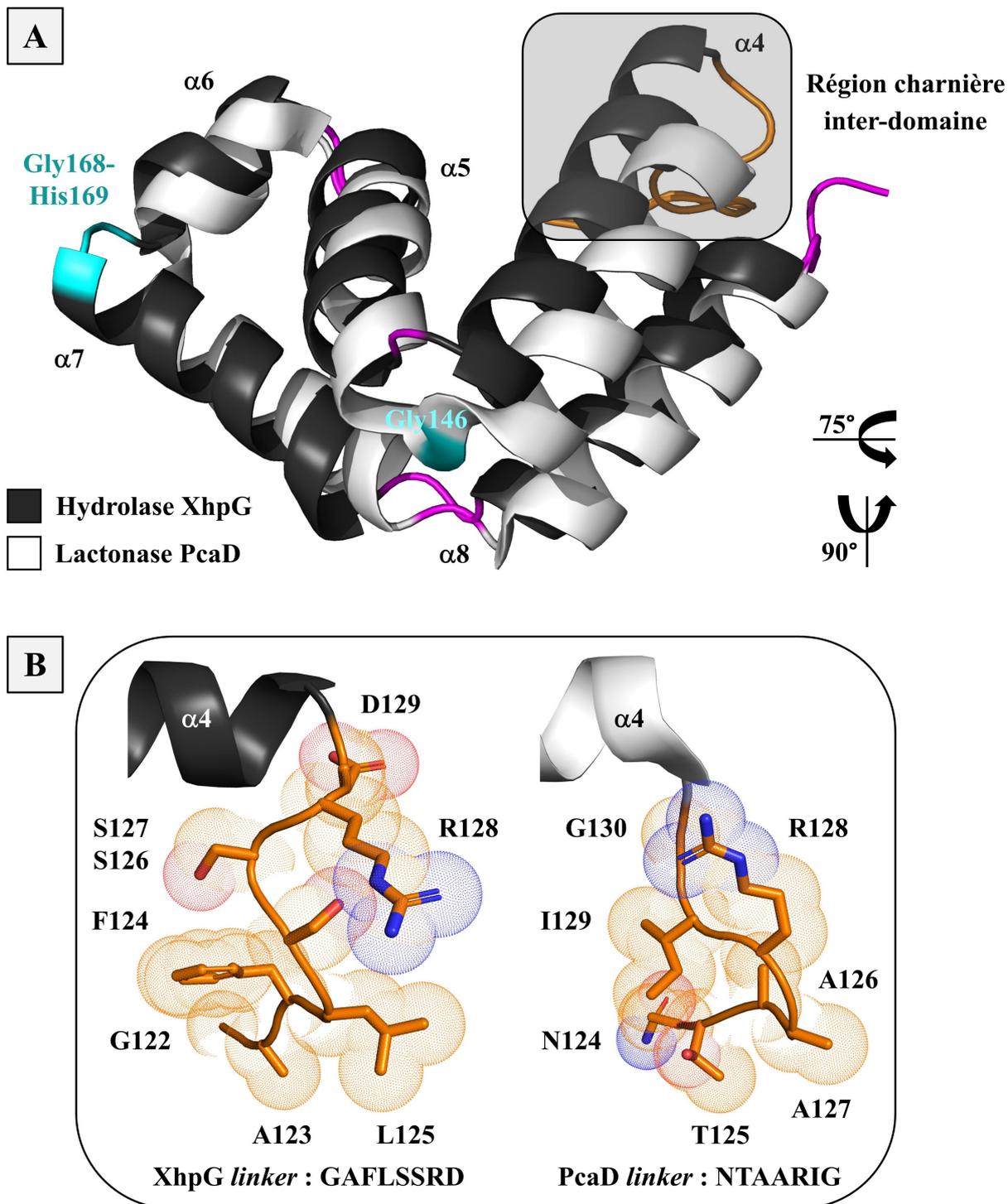


Figure 98. Superposition structurale du sous-domaine α -hélicoïdal des enzymes XhpG et PcaD. (A) Les régions de la coiffe des protéines homologues, représentées en noir ou blanc, contiennent chacune cinq hélices α , annotées $\alpha 4$ à $\alpha 8$. Les sous-domaines hélicoïdaux diffèrent au niveau de la position et de l'orientation de l'hélice $\alpha 4$, de la jonction entre la boucle $\alpha 4$ - $\alpha 5$ et l'hélice adjacente ($\alpha 6$) (un résidu correspondant à Gly146 est absent dans XhpG), ainsi que dans la longueur relative de l'hélice $\alpha 7$ avec un tour supplémentaire (Gly168-His169), comme cela est mis en évidence en cyan. Les boucles situées entre les hélices α sont représentées en magenta dans les structures, sauf la région charnière inter-domaine indiquée en orange. (B) Le *linker* $\beta 6$ - $\alpha 4$ détermine la position et orientation de l'hélice $\alpha 4$ adjacente. Au sein des deux structures, la boucle présente une rigidité importante, comme en attestent les facteurs *B* d'agitation thermique compris entre 15 Å² et 25 Å².

I.2.B. Dimérisation de l'enzyme en solution et modélisation de conformères

Afin d'étayer cette hypothèse, l'hydrolase XhpG a été étudiée par SAXS et les données ont été analysées avec les programmes ATSAS (Manalastas-Cantos *et al.*, 2021) (**Figure 99**). Le traitement initial avec PRIMUS démontre que l'enzyme est dimérique en solution, comme en atteste une masse moléculaire observée de 50–58 kDa (calculée : 58,7 kDa). La concordance entre la courbe théorique calculée pour la structure dimérique (CRY SOL) et les données expérimentales ($\chi^2 = 1,68$) confirme ce résultat (Franke *et al.*, 2017) (**Figure 99A**). Dans le modèle correspondant, les hélices $\alpha 4$ – $\alpha 6$ du sous-domaine hélicoïdal médient l'association entre chaque monomère, ce qui est conforme au rôle connu de cette région dans le processus d'oligomérisation au sein du clade des α/β hydrolases (Bains *et al.*, 2011). En revanche, divers homologues interagissent par le biais d'un feuillet β antiparallèle, ce qui illustre différents auto-assemblages sous-jacents aux fonctions catalytiques (Fushinobu *et al.*, 2002 ; Habe *et al.*, 2003). Au niveau du repliement tertiaire, l'affinement de la structure dimérique, dérivée de 8QZU, avec le programme SREFLEX améliore l'ajustement à la courbe expérimentale de 47 % ($\chi^2 = 1,14$, RMSD = 1,91 Å) (Panjkovich et Svergun, 2016) (**Figure 99B**). Cette étape a permis d'introduire des mouvements conformationnels de structures secondaires régulières et boucles, sans pour autant générer une forme ouverte dans les modèles obtenus. Ainsi, ces résultats soutiennent que l'état fermé de l'*apo*-protéine est aussi pertinent en solution, et s'avère probablement l'unique forme de XhpG, du moins en l'absence du substrat oxopyrrolizidinacétamide.

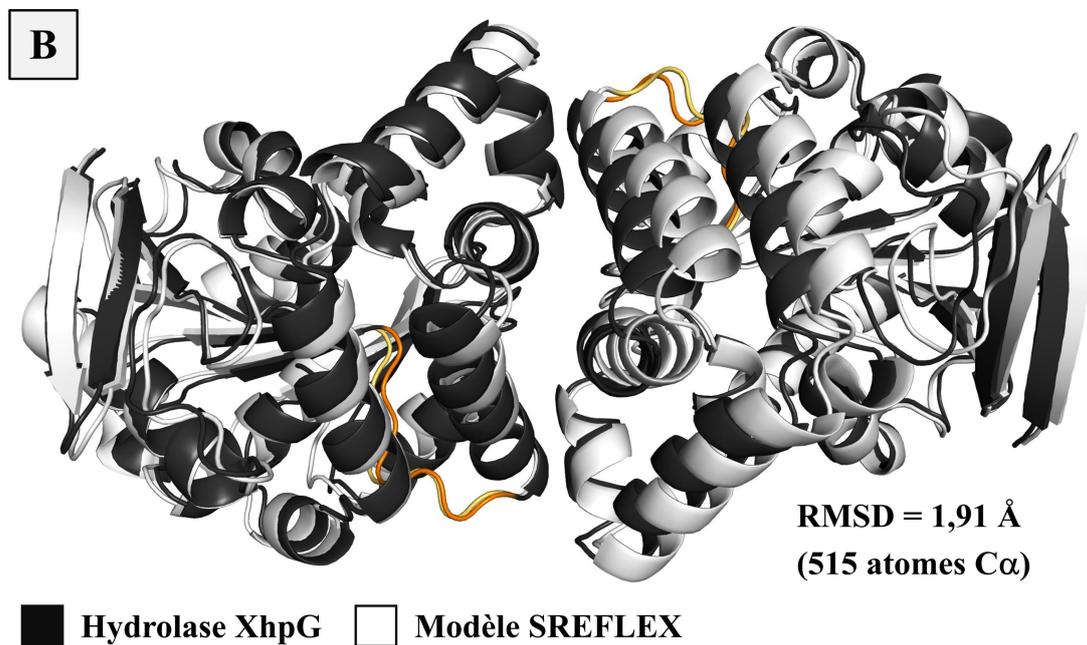
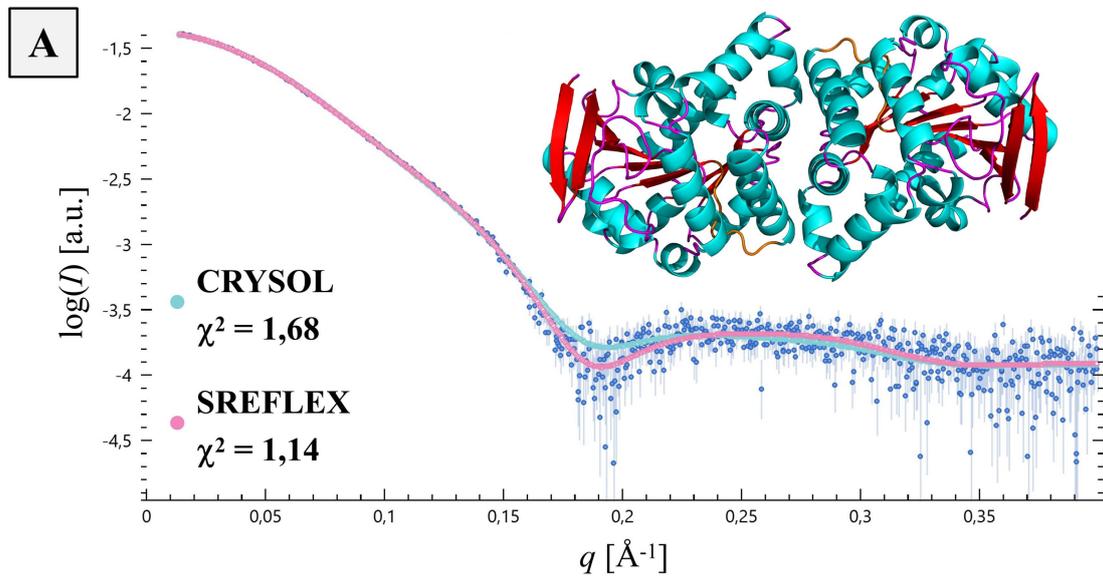


Figure 99. Dimérisation de l'hydrolase XhpG en solution et modélisation d'un conformère. (A) La courbe obtenue avec le programme CRYSOLOG (Franke *et al.*, 2017) montre un accord important ($\chi^2 = 1,68$) entre les données théoriques de diffusion calculées pour la structure dimérique et la courbe expérimentale, ce qui confirme que l'hydrolase XhpG est dimérique en solution. (B) Les modèles générés avec le programme SREFLEX (Panjkovitch et Svergun, 2016) suggèrent des ajustements conformationnels mineurs du sous-domaine hélicoïdal et de la région charnière inter-domaine ($< 2,5$ Å), ce qui corrobore l'hypothèse d'un état fermé de l'enzyme, du moins en l'absence du substrat.

I.2.C. Configuration spatiale et composition du site actif

L'analyse comparative a également montré que les sous-domaines contribuent de manière inégale à la formation du site actif à leur interface, avec une prédominance du cœur catalytique dans le cas de l'enzyme XhpG ou de la région α -hélicoïdale pour l'homologue PcaD (**Figure 100**). Par ailleurs, le tunnel d'accès à la cavité hydrophobe de XhpG est distinct de la plupart des hydrolases α/β , dont l'énol-lactonase, une observation plutôt inhabituelle pour des protéines aux repliements identiques (**Figures 100 et 101**). En dépit de ces différences notables, l'organisation globale du site actif semble assez similaire, en particulier avec la sérine nucléophile enfouie profondément sous la surface de la protéine (17 Å (XhpG) et 12 Å (PcaD)) (**Figures 100 et 101**).

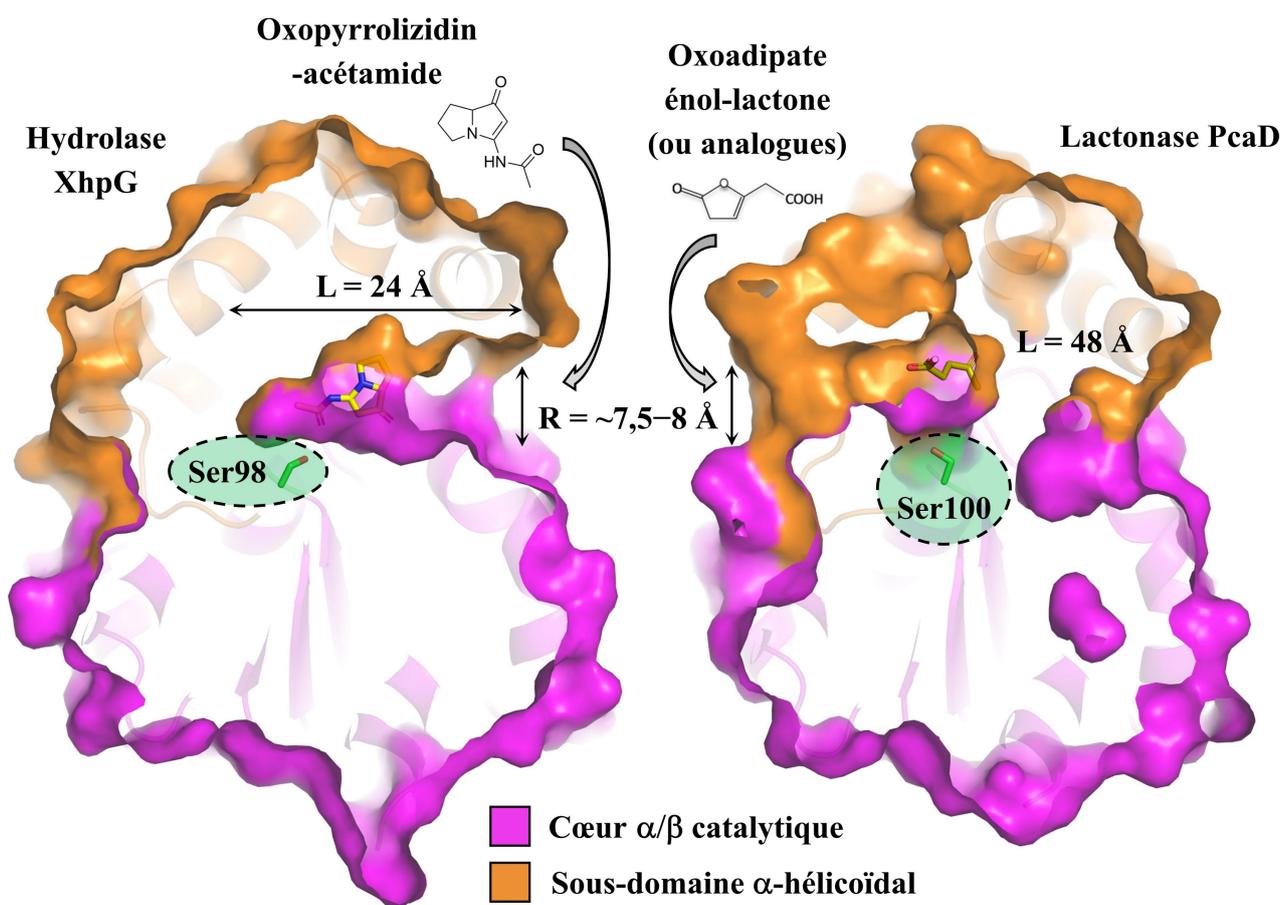


Figure 100. Comparaison de la configuration spatiale du site actif entre les enzymes homologues. Dans chaque cas, le noyau catalytique et le sous-domaine α -hélicoïdal, représentés en magenta et orange respectivement, forment un site actif étendu à leurs interfaces (185 Å² (XhpG) et 135 Å² (PcaD)). Les enzymes diffèrent nettement quant à la localisation du tunnel d'accès du substrat, et leurs sites actifs présentent différentes dimensions et formes. Cependant, l'agencement de la poche catalytique reste similaire, en l'occurrence avec le résidu sérine nucléophile du cœur α/β (Ser98 (XhpG) et Ser100 (PcaD)) situé à une distance d'environ 15 Å de la surface de la protéine.

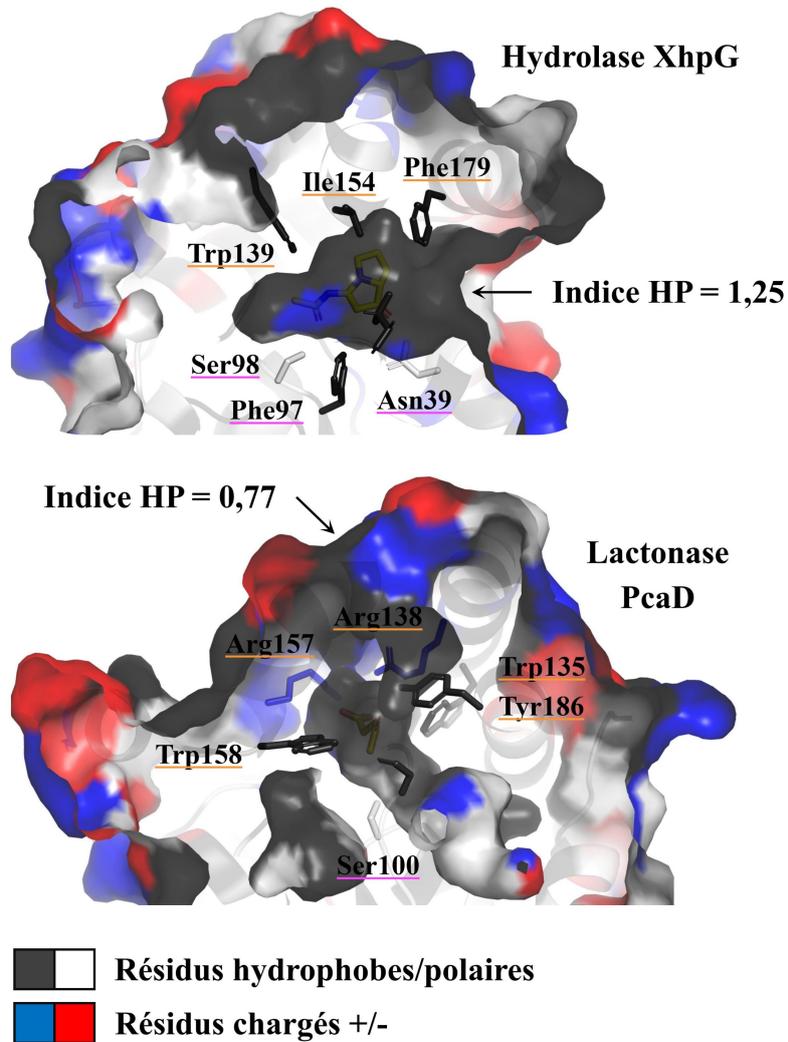


Figure 101. Comparaison de la composition en résidus du site actif entre les enzymes homologues. La cavité occluse présente une nature hydrophobe avec un indice d'hydropathie positif (1,25 (XhpG) et 0,77 (PcaD)), calculé à partir du programme PDBsum (Laskowski *et al.*, 2018). De multiples résidus aliphatiques ou aromatiques du sous-domaine α -hélicoïdal, ainsi que polaires (Asn39 (XhpG)) ou chargés (Arg138, Arg157 (PcaD)), contribuent à la formation du site actif et assurent la liaison du ligand. Les résidus hydrophobes, polaires, ou chargés +/- sont montrés en noir, blanc, bleu et rouge, respectivement.

I.2.D. Conservation de la triade catalytique et liaison des ligands au sein du site actif

Les résidus catalytiques des enzymes XhpG et PcaD se positionnent presque identiquement avec des valeurs de $C\alpha$ -RMSD comprises entre 0,5 Å et 0,6 Å (**Figure 102**). Cependant, la chaîne latérale du résidu Ser98 présente un rotamère distinct dans le modèle de *docking* : le groupement hydroxyle est aligné le long de l'axe du « trou de l'oxyanion », une conformation énergétiquement favorable à la liaison du substrat en accord avec le score Vina (outil CB-Dock2) (Liu *et al.*, 2022) (**Figure 102**). Par conséquent, le site actif occlus et conservé de ces enzymes est peu exposé aux solvants, ce qui suscite certaines questions au sujet de la spécificité du substrat et de sa flexibilité conformationnelle.

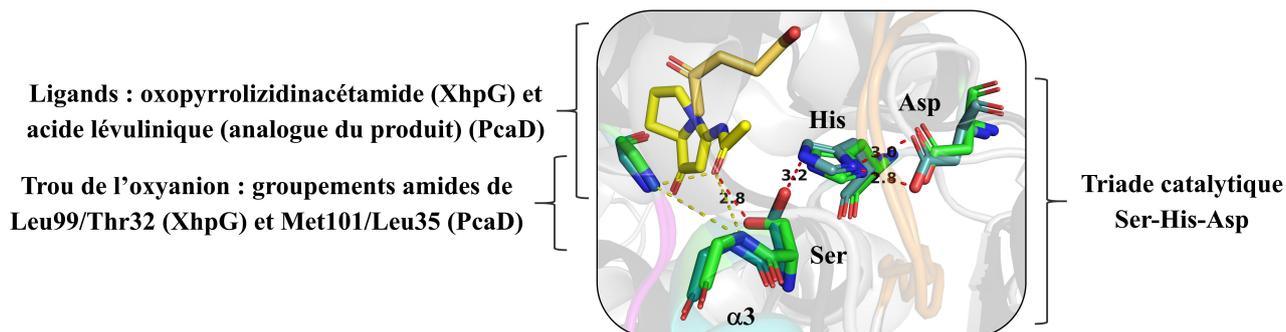


Figure 102. Conservation de la triade catalytique au sein des enzymes homologues. Le site actif comporte la triade conservée Ser-His-Asp, dont les atomes $C\alpha$ se superposent avec des RMSDs de 0,6 Å, 0,5 Å et 0,6 Å, respectivement entre chaque structure. Les rotamères observés des résidus nucléophiles sont différents, avec le groupement hydroxyle de Ser98 (XhpG) aligné le long de l'axe du trou de l'oxyanion, et celui de Ser100 (PcaD) orienté vers l'atome $N\epsilon$ de la base générale His. Un autre atome de celle-ci ($N\delta$) participe à la liaison hydrogène avec la fonction carboxylate du résidu Asp, comme la majorité des membres de la super-famille des α/β hydrolases (Bains *et al.*, 2011). Le trou de l'oxyanion comprend les groupements amides de la chaîne principale des résidus Leu99/Thr32 (XhpG) ou Met101/Leu35 (PcaD).

Afin d'explorer les bases moléculaires de la liaison des ligands, le mécanisme de reconnaissance du substrat oxopyrrolizidinacétamide *docké* au niveau du site actif de l'hydrolase XhpG a été mis en parallèle de celui de l'analogue du produit, l'acide lévulinique lié au sein de l'énol-lactonase PcaD (Figure 103). Cette analyse a démontré des caractéristiques communes, notamment l'encombrement stérique des poches du site actif, ainsi que les déterminants de liaison fournis par chacun des sous-domaines central et hélicoïdal. En particulier, la chaîne latérale indole du résidu Trp139 restreint le diamètre de la cavité hydrophobe étendue, à l'instar du résidu Trp135 dans PcaD, ce qui impose des contraintes géométriques aux ligands complexés (Figures 101 et 103). Le résidu Phe159 est conservé dans l'homologue (Trp158), mais ces acides aminés assurent des rôles différents : Phe159 maintient l'orientation du groupement phényle du résidu Phe179 adjacent, tandis que Trp158 entre directement en contact avec la chaîne linéaire de l'acide lévulinique (Figure 103). Les autres résidus constitutifs des sites de fixation ne sont globalement pas conservés entre les enzymes, ce qui est cohérent avec les structures chimiques divergentes des substrats. Un parallèle pourrait néanmoins être établi entre les résidus Asn39 (XhpG) et Ser34 (PcaD), étant donné que la chaîne latérale polaire de ce résidu pourrait aussi former une liaison hydrogène avec la fonction carbonyle du substrat oxoadipate énol-lactone (Figure 103). Dans l'ensemble, bien que la région de la coiffe et le site actif de l'hydrolase XhpG présentent une certaine variabilité structurale par rapport à l'homologue, dans les deux cas, le sous-domaine α -hélicoïdal contribue significativement à la sélectivité de substrat et pourrait faciliter la libération du produit depuis le cœur catalytique conservé.

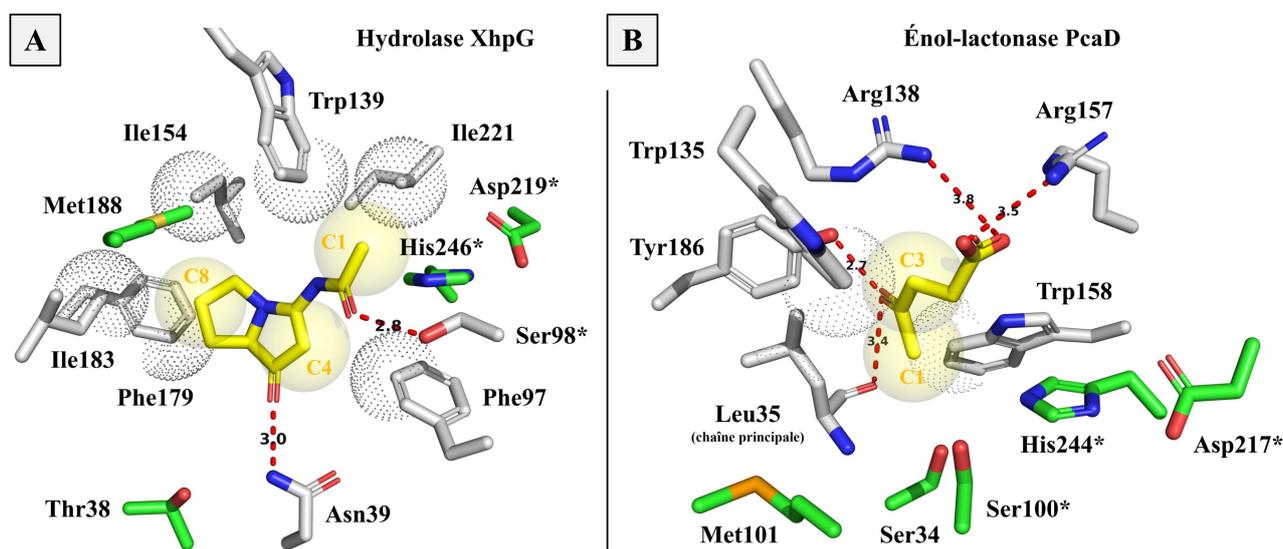


Figure 103. Comparaison du mode de liaison des ligands au sein du site actif entre les enzymes homologues. Des interactions hydrophobes (sphères) et liaisons hydrogènes (tirets) assurent spécifiquement la reconnaissance du substrat *docké* oxopyrrolizidinacétamide (A) ou de l'analogue du produit acide lévulinique (B) dans le site actif des homologues structuraux XhpG et PcaD, respectivement. Les résidus directement impliqués dans la fixation du substrat *docké* sont indiqués en gris et ceux engagés dans des interactions avec d'autres résidus le sont en vert. Les résidus catalytiques sont annotés avec un astérisque (XhpG/PcaD : Ser98*/Ser100*, Asp219*/Asp217*, His246*/His244*).

SECTION II. Élucidation des relations structure-fonction de la déshydratase AzeD

Dans l'objectif de comprendre le mécanisme catalytique et la spécificité de substrat d'une autre enzyme post-NRPS, la déshydratase AzeD, des approches analogues de biochimie et de biophysique ont été mises en œuvre et ont abouti à l'élucidation de la structure cristallographique. Initialement, le résidu His130 du motif xHxxxDG consensus a été proposé comme étant essentiel à la réaction après que la séquence a été alignée avec des domaines C homologues (**Figures 104 et 105**). Suite à cette recherche préliminaire, le gène a été cloné dans le plasmide pET28a et exprimé en système bactérien hétérologue (*E. coli* BL21 (DE3)). La protéine recombinante a alors été purifiée par chromatographie d'affinité sur colonne Ni²⁺-sépharose, suivie d'une filtration sur gel (**Figure 106**). À cause de l'indice d'instabilité important de l'enzyme, calculé avec l'outil ProtParam (ExpPASy), l'étiquette 6×His n'a résolument pas été clivée (Gasteiger *et al.*, 2005 ; Gamage *et al.*, 2019). Les échantillons obtenus ont été analysés par DLS, avant d'être soumis à des tests de cristallisation (**Figure 107**).

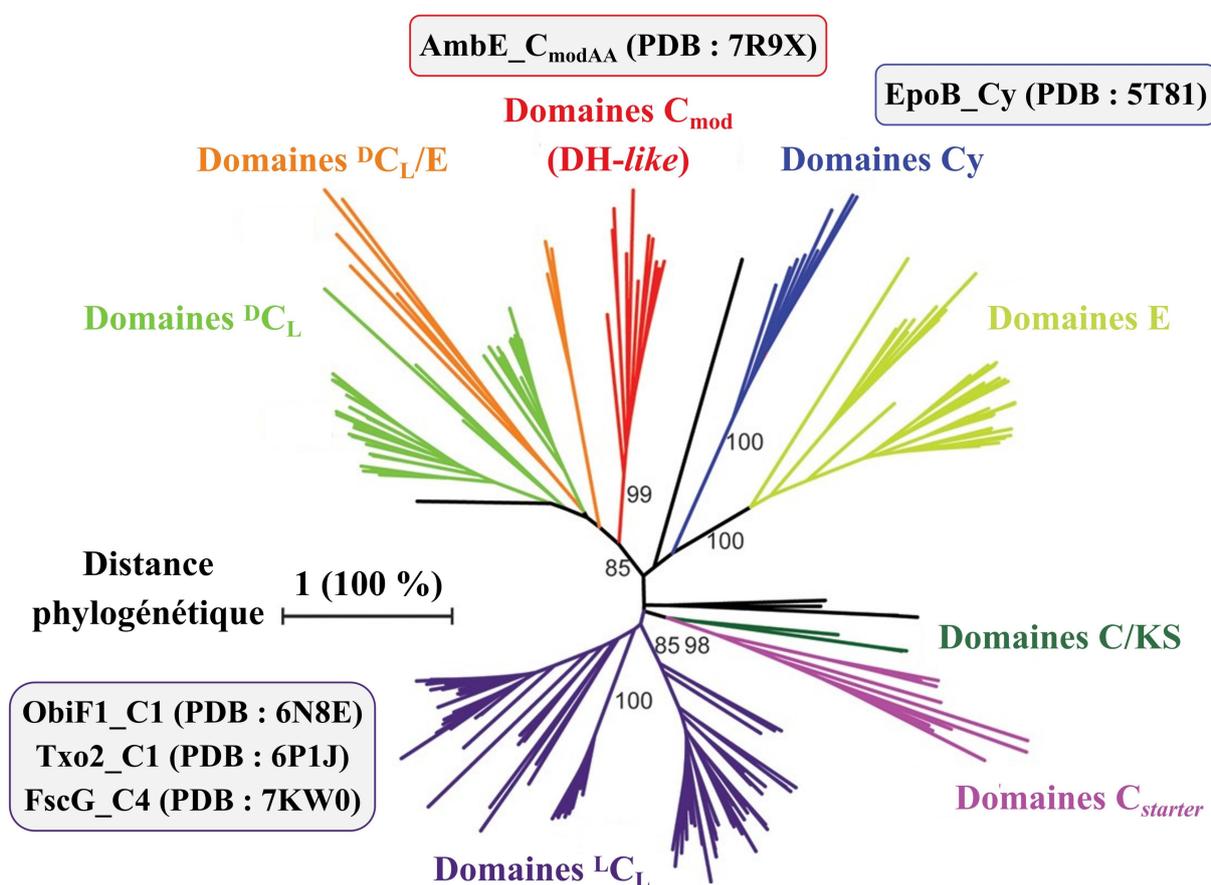


Figure 104. Classification phylogénétique de domaines homologues à la déshydratase AzeD (adaptée de Patteson *et al.*, 2022). Les séquences de 200 domaines C de diverses fonctions ont été regroupées dans un arbre phylogénétique selon la méthode du maximum de vraisemblance (Patteson *et al.*, 2022). L'échelle indique le pourcentage moyen attendu de substitutions de résidus par site. Les valeurs de confiance (85–100) ont été calculées pour chacun des sous-groupes évolutifs par une méthode d'inférence statistique (*i.e. bootstrap*) (Patteson *et al.*, 2022). La séquence de la déshydratase AzeD a été alignée avec certains domaines de condensation L_{C_L}, de modification de type DH-like (déshydratation) ou d'hétérocyclisation (voir **Figure 105**).

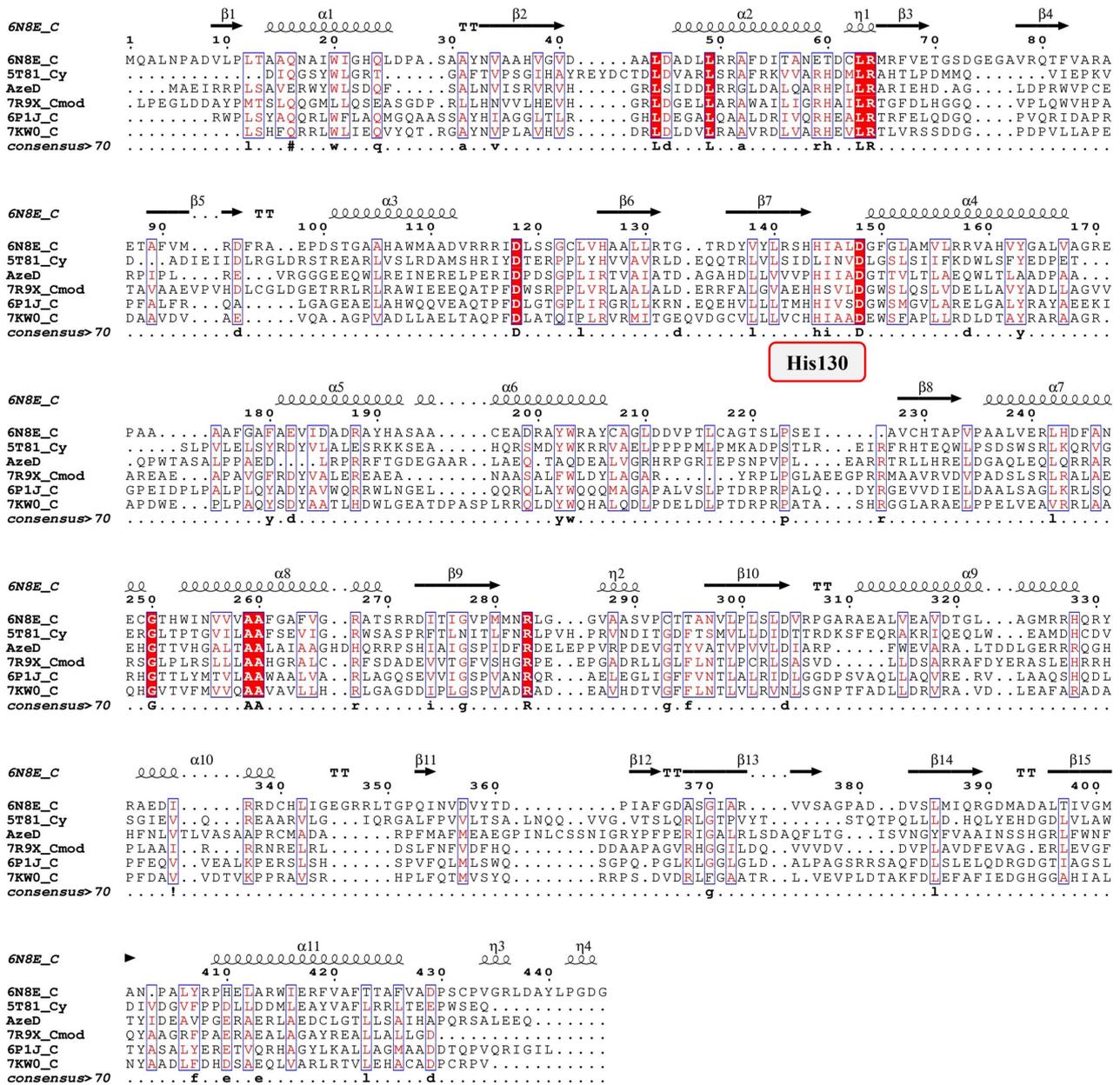


Figure 105. Alignement séquentiel des homologues de la déshydratase AzeD. La séquence primaire de l'enzyme a été analysée par alignement par rapport aux séquences de divers homologues, incluant des domaines C intégratifs ayant une fonction de condensation (PDB : 6P1J, 7KW0, 6N8E), d'hétérocyclisation (PDB : 5T81) ou de déshydratation (PDB : 7R9X). Les domaines C de NRPS ont été identifiés à l'aide de la banque PDB et du serveur de comparaison structurale Dali (Berman *et al.*, 2000 ; Holm *et al.*, 2023). Les homologues présentent une assez faible identité de séquence (environ 17–22 %) et de nombreux éléments d'insertion, mais partagent quand même des déterminants impliqués dans la liaison du substrat et/ou la réaction catalytique. En l'occurrence, la déshydratase AzeD contient les résidus His130 et Asp134 conservés du motif xxHxxxDx, bien que l'histidine ne le soit pas strictement. Par ailleurs, ils possèdent des structures secondaires conservées et, de ce fait, montrent un Z-score de similarité de repliement important, de l'ordre de 30 %.

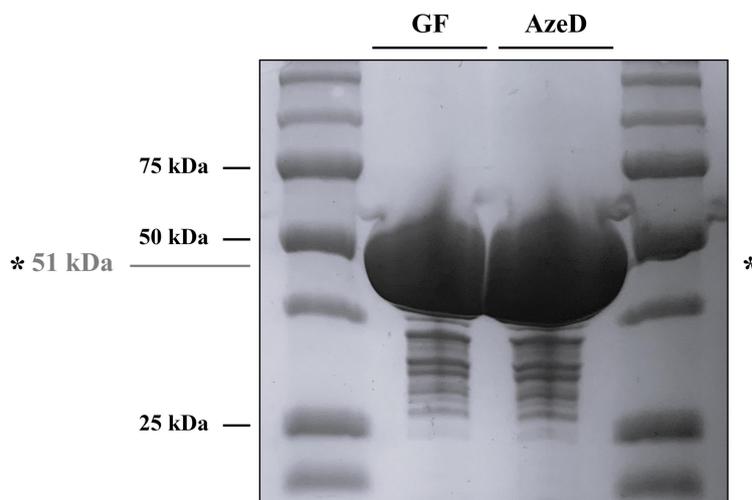


Figure 106. Purification de la déshydratase AzeD par chromatographie liquide. L'enzyme AzeD a été purifiée par chromatographie d'affinité sur ions nickel immobilisés (non-représentée) et chromatographie d'exclusion stérique (notée GF). Le rendement final de la purification a été estimé à 34 mg.L⁻¹. Un total de 50 µg de protéines a été chargé sur un gel de polyacrylamide et, après avoir révélé celui-ci, l'espèce attendue est majoritaire, mais apparaît néanmoins sujette à une protéolyse en des bandes de faibles masses moléculaires (10–50 ng), à l'instar de certains domaines C.

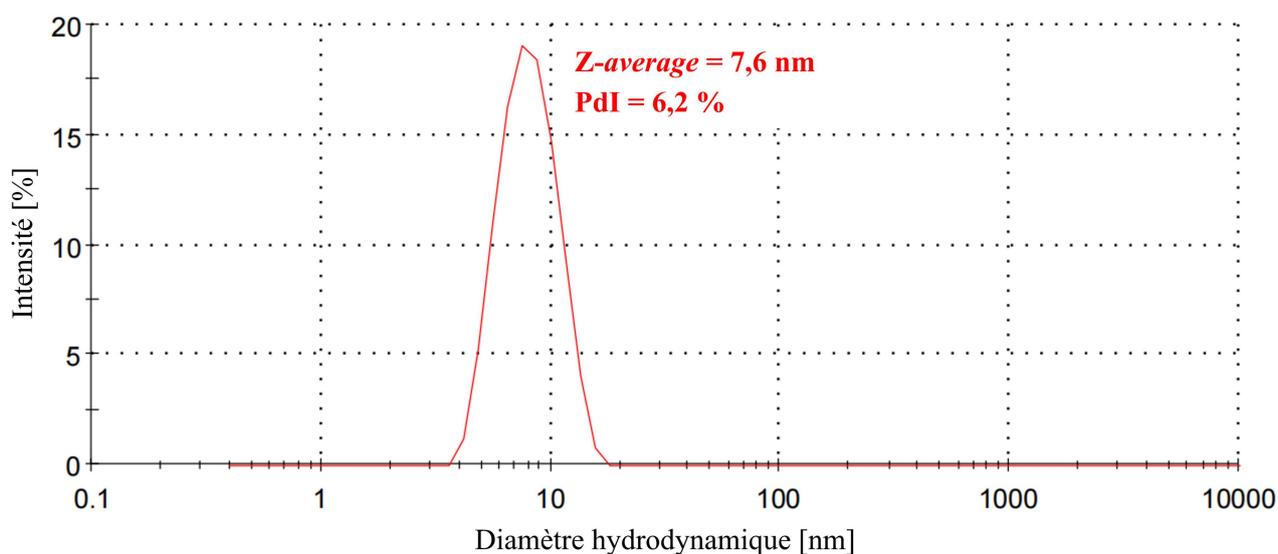
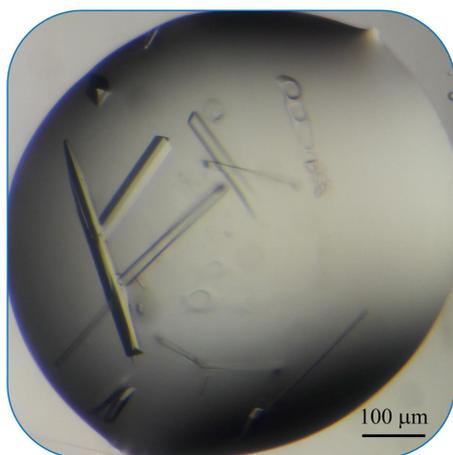


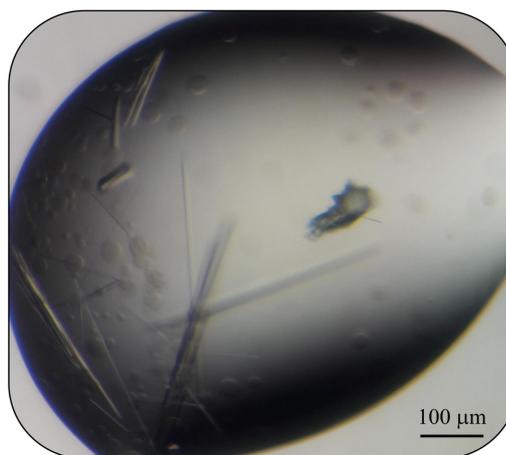
Figure 107. Analyse de la déshydratase AzeD par DLS. Les résultats de DLS indiquent que l'enzyme a un diamètre hydrodynamique moyen de 8 ± 2 nm et que l'échantillon est homogène avec un faible indice de polydispersité de 6,2 %, malgré la protéolyse apparente en conditions dénaturantes (voir **Figure 106**).

II.1. Résolution de la structure cristallographique

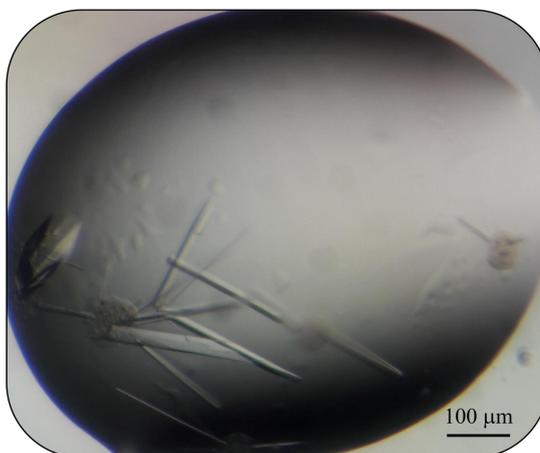
Au cours d'essais préliminaires, l'enzyme AzeD a été cristallisée sous forme d'une *apo*-protéine par diffusion de vapeur en gouttes assises (**Figure 108**). Parmi les conditions propices, quelques-unes ont été affinées à l'échelle microscopique et les cristaux obtenus ont été analysés par diffraction des rayons X afin d'enregistrer un ensemble de données de haute qualité. Suite à leur traitement initial, la structure tridimensionnelle de la déshydratase AzeD autonome de *P. aeruginosa* a été déterminée par remplacement moléculaire (AlphaFold2) à une résolution de 1,55 Å (PDB : 8S6L) (**Annexe PDB**). L'unité asymétrique du cristal contient un monomère de 48 kDa, dont les résidus flexibles au niveau des régions N- et C-terminales (Met1, Arg435–Gln441) sont absents de la structure résolue.



100 mM bis-Tris méthane, pH 6,5
25 % (m/v) PEG-3350



100 mM HEPES, pH 7,5
25 % (m/v) PEG-3350



100 mM bis-Tris méthane, pH 6,5
20 % (m/v) PEG-5000 monométhyléther

Figure 108. Cristallisation de la déshydratase AzeD par diffusion de vapeur. Les cristaux de la macromolécule ont été obtenus à 20 mg.mL⁻¹ dans 3 conditions du kit Index (Hampton Research), l'une étant composée de 100 mM de bis-Tris méthane à pH 6,5 et de 25 % (m/v) de PEG-3350, par diffusion de vapeur en nano-gouttes assises dans des plaques MRC. Les cristaux présentent le plus souvent une forme de prisme à base rectangulaire et une longueur de 300–500 μm. HEPES : acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique, PEG : polyéthylène glycol.

II.1.A. Description du repliement global

À l'instar des domaines C, l'enzyme AzeD dérive d'une chloramphénicol acétyltransférase (CAT) ancestrale par duplication, fusion et maturation d'un protomère en une structure pseudo-dimérique, où les lobes N- et C-terminaux sont séparés par une fente centrale dans un arrangement en *V-shape* (Leslie *et al.*, 1988 ; Bloudoff et Schmeing, 2017) (**Figure 109**). Le sous-domaine NT comporte un feuillet β principal à six brins ($\beta 4$ - $\beta 5$ - $\beta 6$ - $\beta 1$ - $\beta 11$ - $\beta 12$), dont deux contigus ($\beta 11$ -DA- $\beta 12$) font partie d'un β -*crossover*, ainsi que quatre hélices α et deux brins β antiparallèles ($\beta 2$ - $\beta 3$) périphériques. Le sous-domaine CT possède quant à lui un unique feuillet β à six brins torsadés ($\beta 7$ - $\beta 14$ - $\beta 13$ - $\beta 10$ - $\beta 8$ - $\beta 9$) et flanqué de six hélices α . Les lobes s'assemblent au moyen d'éléments structuraux conservés : l'hélice $\alpha 5$ localisée à l'extrémité basale et une hélice 3_{10} enchâssée dans la région NT (α -*crossover*), ainsi que le motif boucle-brins-boucle (β -*crossover*) (*i.e.* 29 résidus, Ile359–Tyr387) situé entre les brins $\beta 10$ et $\beta 13$ (Tan *et al.*, 2020) (**Figure 109**).

Les domaines C caractérisés à ce jour présentent un état plus ou moins ouvert à l'interface entre les lobes, une variation structurale qui pourrait avoir une incidence sur les interactions protéiques au sein des systèmes NRPS (Bloudoff et Schmeing, 2017). Dans la structure de AzeD, la conformation étant assez fermée, le β -*crossover* étendu pourrait agir comme un « élément de verrouillage » (*latch*) et ainsi restreindre l'accès au tunnel du site actif (Samel *et al.*, 2007 ; Bloudoff et Schmeing, 2017). Cependant, aucun changement d'orientation significatif d'une quelconque structure secondaire, qui pourrait le cas échéant révéler un phénomène de liaison au substrat, n'a été jusqu'à présent rapporté pour les domaines C connus, exception faite de Txo1_C2 de la NRPS impliquée dans la biosynthèse de l'antibiotique teixobactine (Tan *et al.*, 2020). Plus précisément, l'interaction avec le domaine PCP accepteur entraîne un mouvement de l'hélice $\alpha 1$ vers l'extérieur et le repliement induit de la boucle adjacente, ce qui expose l'interface hydrophobe du domaine C au solvant et permet ainsi l'accès du substrat au site actif (Tan *et al.*, 2020).

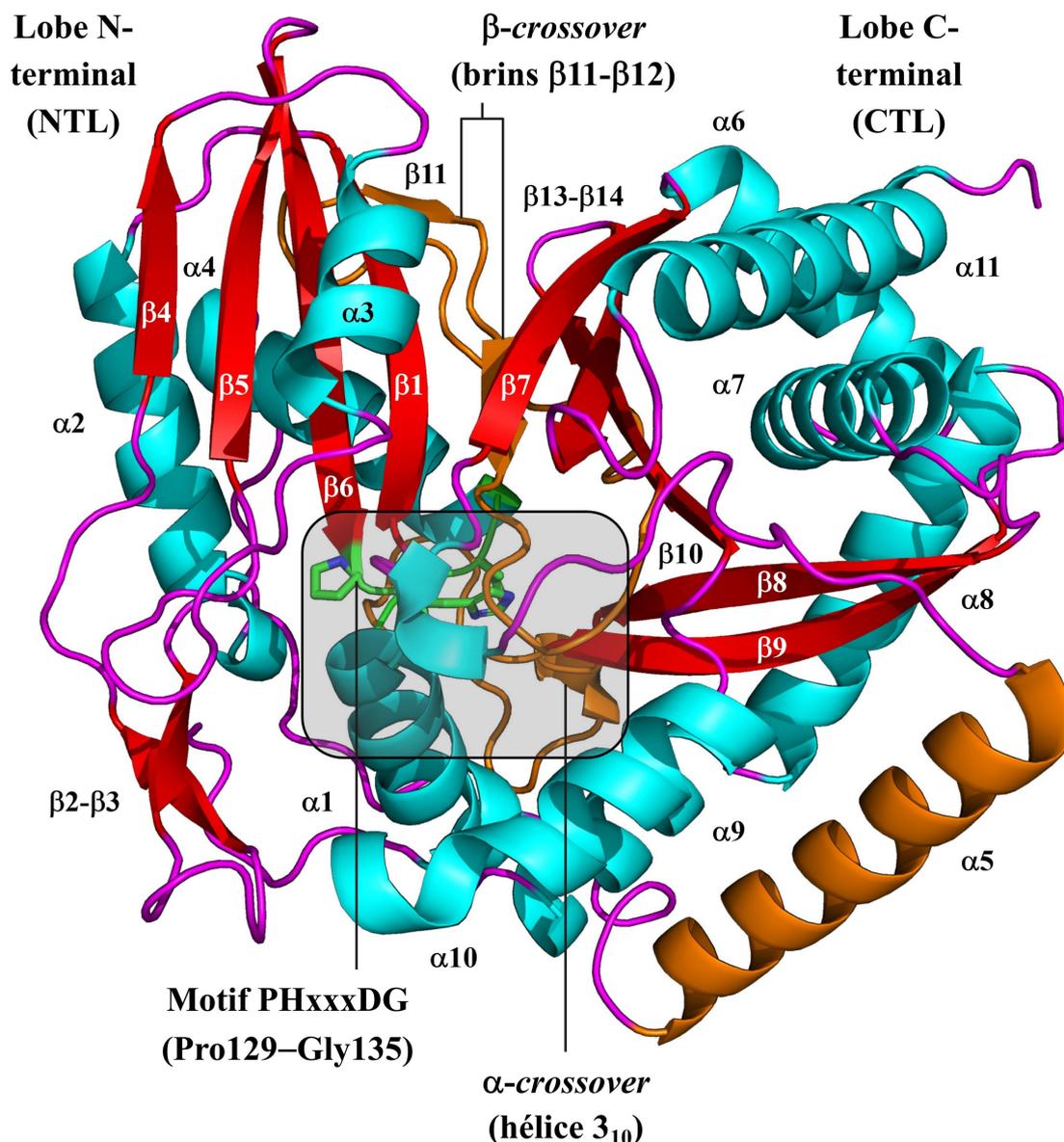


Figure 109. Représentation de la structure cristallographique de la déshydratase AzeD. Le monomère de l'enzyme AzeD se compose de lobes N- et C-terminaux (NTL/CTL) séparés par une ouverture centrale d'une faible amplitude. Le sous-domaine NT (~ 200 résidus) contient un feuillet β majeur, constitué de six brins ($\beta 4$ - $\beta 5$ - $\beta 6$ - $\beta 1$ - $\beta 11$ - $\beta 12$) et flanqué de quatre hélices α , ainsi qu'un feuillet β mineur formé des brins $\beta 2$ et $\beta 3$ antiparallèles. Le sous-domaine CT comprend un unique feuillet β à six brins torsadés ($\beta 7$ - $\beta 14$ - $\beta 13$ - $\beta 10$ - $\beta 8$ - $\beta 9$) et entouré de six hélices α . Les lobes s'assemblent *via* des éléments structuraux conservés : l'hélice $\alpha 5$ située à l'extrémité basale et une hélice 3_{10} enchâssée dans la région NT (α -crossover), ainsi que des brins contigus $\beta 11$ - $\beta 12$ localisés dans une section étendue (Ile359-Tyr387) (β -crossover) à l'interface. Le motif PHxxxDG (boucle $\beta 6$ - $\alpha 4$, lobe NT) inclut des résidus ayant un rôle structural (Pro129, Asp134) ou catalytique (His130). Les hélices α et les brins β sont représentés en cyan et en rouge, respectivement. Les éléments de structure interlobulaires sont indiqués en orange et le motif catalytique est illustré en vert. PDB : 8S6L.

II.1.B. Organisation structurale du site actif et liaison du substrat

Afin d'explorer les bases moléculaires de la liaison du substrat, des analyses de *docking in silico* ont été réalisées avec le programme CB-Dock2 (Liu *et al.*, 2022). Le site actif de la déshydratase AzeD se situe à l'interface des lobes et, dans l'état *apo* de l'enzyme, le canal hydrophobe est plutôt accessible (7–8 Å de diamètre) au substrat (**Figure 110**). Le motif catalytique PHxxxDG (boucle β 6- α 4) contient les résidus conservés à rôles structuraux (Asp134-Gly135) ou fonctionnels (His130) des domaines de condensation (Patteson *et al.*, 2022). Dans la structure de AzeD, l'acide aminé Pro129 remplace un autre résidu histidine conservé et induit une courbure au niveau de la séquence, ce qui permet à la base générale His130 de s'approcher étroitement du substrat (3,0 Å) (**Figure 111**). Les chaînes latérales de certains résidus, dont Trp14, Leu324, Val328, Thr380 et Val389, contactent les atomes de carbone C1, C2 ou C9 du L-azétidomonamide *via* des interactions hydrophobes, tandis que les groupements des résidus Tyr15, Thr290 et Asn385 établissent chacun une liaison hydrogène forte avec une fonction amide ou carbonyle du ligand (**Figure 111**). Toutefois, ces résultats pourraient ne pas refléter correctement la position et/ou l'orientation du substrat, mais plutôt servir de base à une proposition de mécanisme réactionnel.

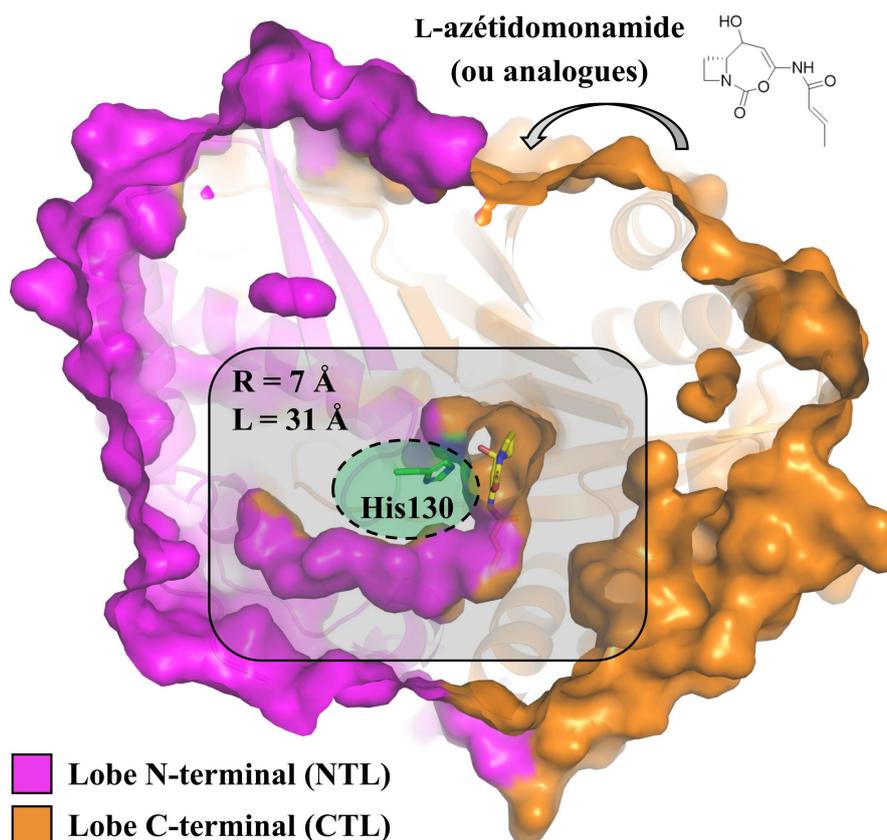


Figure 110. Configuration spatiale du site actif de la déshydratase AzeD. Les lobes NT et CT, présentés en magenta et orange respectivement, créent un site actif étendu à l'interface. Le tunnel d'entrée unique a des dimensions $R = 7 \text{ \AA}$ et $L = 31 \text{ \AA}$, et le substrat occupe sa partie proximale. La base His130 du lobe NT se trouve à une distance d'environ 11 \AA de la surface de la protéine.

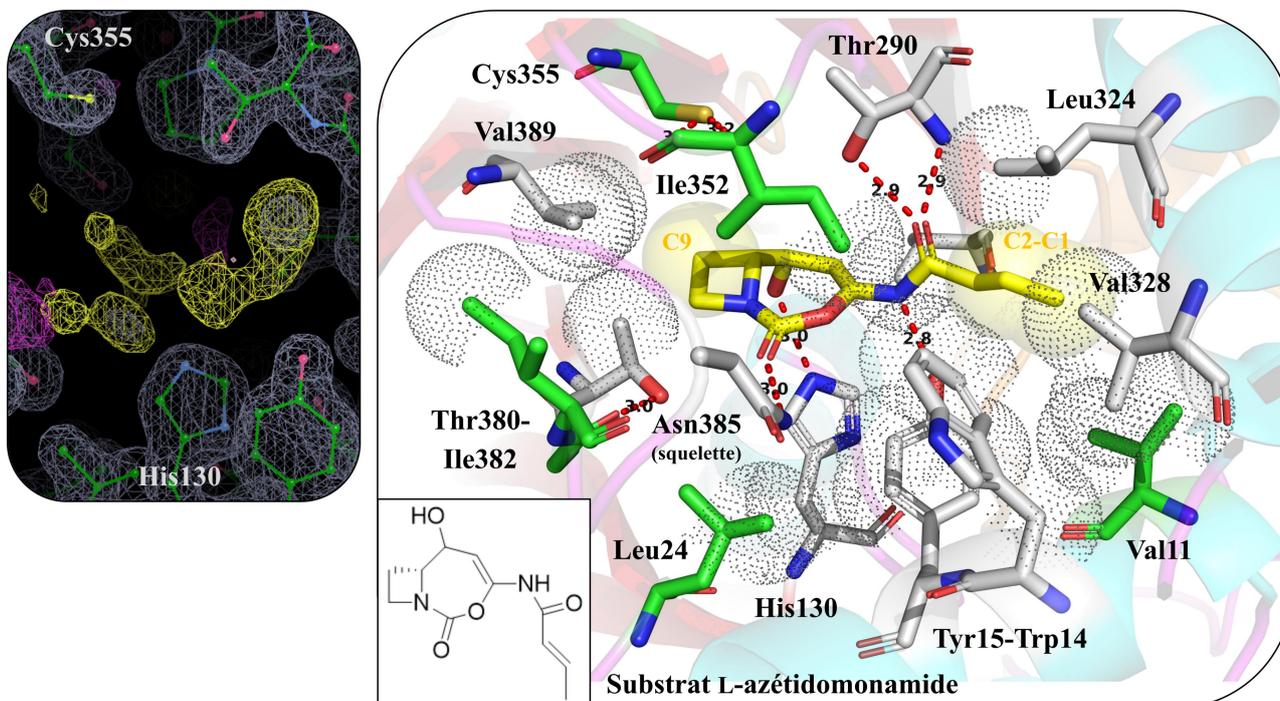


Figure 111. Liaison du substrat L-azetidomonamide dans le site actif de la déshydratase AzeD par modélisation moléculaire. Des interactions hydrophobes (sphères) et des liaisons hydrogènes (tirets pointillés) formées au niveau de l'hétérocycle azetidopyrroïdine et/ou de la chaîne buténamide positionnent et orientent le substrat par rapport au résidu His130 catalytique. Les résidus directement impliqués dans la reconnaissance du substrat *docké* sont indiqués en gris (*i.e.* Trp14, Tyr15, His130, Val288, Thr290, Leu324, Val328, Thr380, Asn385 et Val389). Les résidus engagés dans des interactions avec d'autres acides aminés sont représentés en vert (*i.e.* Val11, Leu24, Ile352, Cys355 et Ile382).

Le mécanisme de déshydratation du substrat L-azetidomonamide correspond vraisemblablement à une élimination bimoléculaire (**Annexe 2**). La réaction nécessite une conformation antipériplanaire du proton acide (C8-H) et de l'hydroxyle (C7-OH), la stéréochimie naturelle lorsque ce groupement nucléofuge a une configuration *R*. Dans ce cas, un couple acide/base assure de manière synchrone la déprotonation du centre tertiaire (carbone), la formation de la double liaison et l'élimination de l'eau. Des expériences de mutagenèse dirigée suivies des tests d'activité correspondants ont démontré que le résidu His130 agit en tant que base générale, ce qui est cohérent avec le rôle identique proposé au sein de domaines déshydratants homologues (Gaudelli *et al.*, 2015 ; Patteson *et al.*, 2022). Par contre, le catalyseur acide est là aussi moins évident, mais, du fait de sa proximité, Cys355 pourrait assurer cette fonction. Des investigations complémentaires sur d'autres résidus du site actif potentiellement impliqués dans le mécanisme réactionnel sont en cours.

II.2. Comparaison avec la structure cristallographique de domaines homologues

II.2.A. Organisation structurale des éléments interlobulaires et conformation en solution

Bien que les domaines C partagent un repliement caractéristique, certaines modifications quant aux éléments structuraux ou au motif catalytique élargissent le répertoire des réactions alternatives, avec l'estérification, l'épimérisation, la cyclisation, la déshydratation, et le recrutement d'enzymes partenaires (Balibar *et al.*, 2005 ; Rausch *et al.*, 2007 ; Koketsu *et al.*, 2010 ; Gaudelli *et al.*, 2015 ; Peschke *et al.*, 2016). Afin de vérifier cette observation, la structure de la déshydratase AzeD a alors été comparée à celle de proches homologues, en l'occurrence des domaines intégratifs ayant un rôle classique de condensation (Txo2_C1, PDB : 6P1J) ou atypique de déshydratation (AmbE_C_{modAA}, PDB : 7R9X), identifiés à partir du serveur Dali (Tan *et al.*, 2020 ; Patteson *et al.*, 2022). En effet, ces protéines adoptent un repliement canonique de type CAT, comme l'attestent les RMSD calculés de 4,1 Å (Txo2_C1) et 6,3 Å (AmbE_C_{modAA}). Par contre, l'enzyme AzeD diffère sensiblement des homologues au niveau de l'amplitude entre les lobes, ainsi qu'en termes de position, d'orientation et de longueur relative des éléments structuraux à l'interface (**Figure 112**). Au sein de la déshydratase AzeD, le β -*crossover* étendu ou des structures secondaires flexibles pourraient réguler l'accessibilité du site actif au substrat discret.

Afin de soutenir davantage cette hypothèse, la déshydratase AzeD a été étudiée par SAXS et les données ont été analysées avec les outils ATSAS (Manalastas-Cantos *et al.*, 2021) (**Figure 113**). Le traitement avec PRIMUS démontre que l'enzyme est monomérique en solution, comme en témoigne une masse moléculaire observée d'environ 48 kDa (calculée : 51,2 kDa). La correspondance entre la courbe théorique calculée pour la structure PDB (CRY SOL) et les données SAXS brutes ($\chi^2 = 3,17$) confirme ce constat (Franke *et al.*, 2017) (**Figure 113A**). L'état oligomérique contraste toutefois avec la structure homodimérique décrite pour le domaine de déshydratation AmbE_C_{modAA}, mais demeure conforme au caractère des multienzymes NRPS. Concernant le repliement global, l'affinement de la structure avec le programme SREFLEX améliore l'ajustement à la courbe expérimentale de 119 % ($\chi^2 = 1,45$, RMSD = 1,58 Å) (Panjkovich et Svergun, 2016) (**Figure 113B**). Néanmoins, aucun des modèles ne présente le moindre mouvement majeur au niveau de la région interlobulaire. Ainsi, ces résultats prouvent que la conformation fermée de la protéine dans le cristal est également cohérente en solution et apparaît comme l'état plausible de AzeD, en l'absence du substrat L-azétidomonamide ou d'enzymes biosynthétiques partenaires.

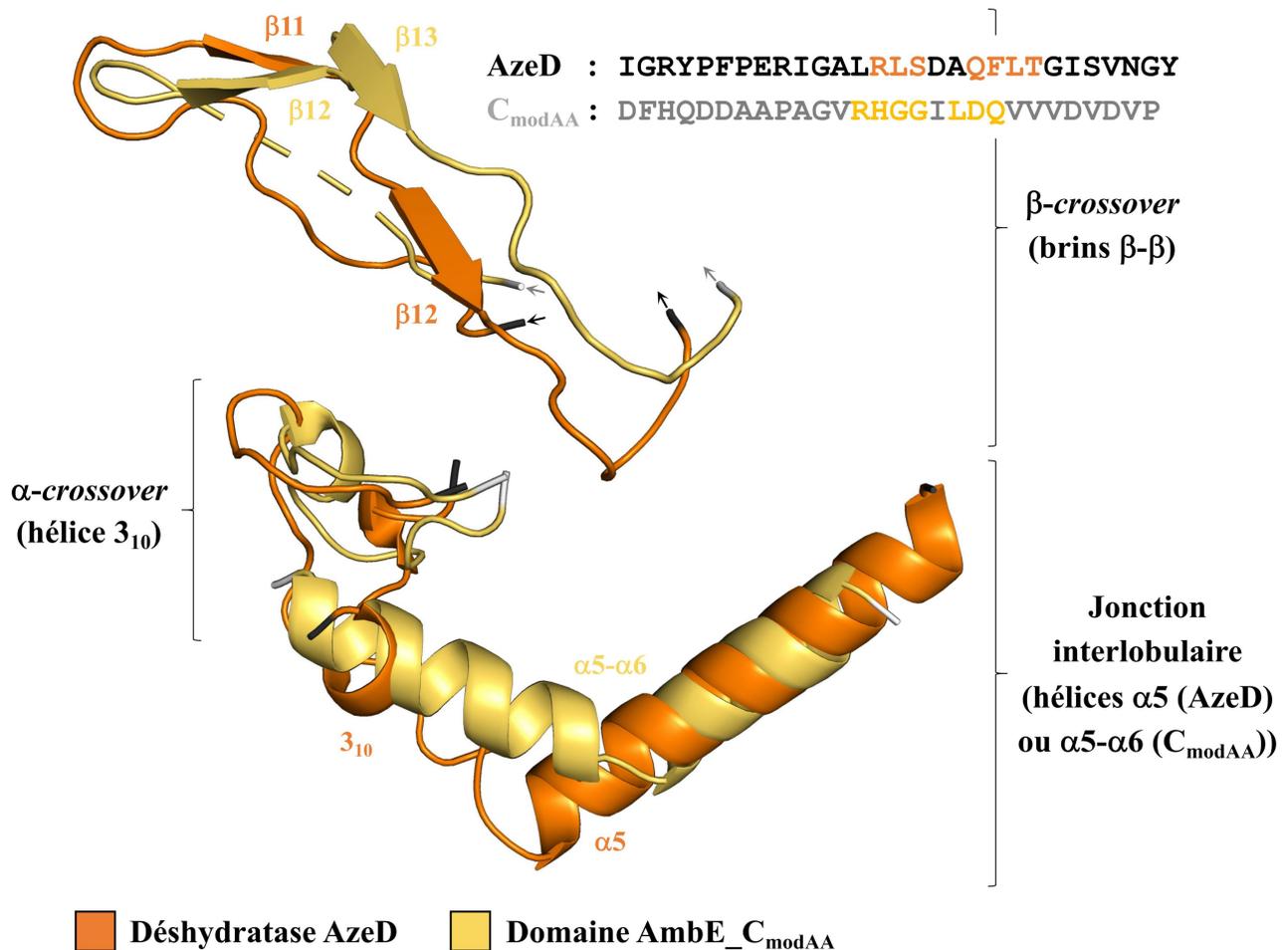


Figure 112. Superposition structurale des éléments interlobulaires de la déshydratase AzeD et du domaine AmbE C_{modAA}. Des hélices α et les croisements α -/ β - de la déshydratase AzeD et du domaine AmbE_C_{modAA}, représentés en orange et jaune respectivement, permettent le maintien entre les lobes à l'interface de chaque structure. Les éléments de structure en question diffèrent au niveau de la région charnière α -hélicoïdale, vu qu'une hélice 3₁₀ résiduelle est présente dans AzeD, au lieu d'une hélice α classique, et que l'hélice principale contient quatre tours supplémentaires. De plus, ils se distinguent selon l'orientation du α -crossover, une structure vestigiale nécessaire à l'interaction avec le domaine PCP des NRPS, et dans la composition du β -crossover, dont certains acides aminés (QDDAA) sont absents chez l'homologue.

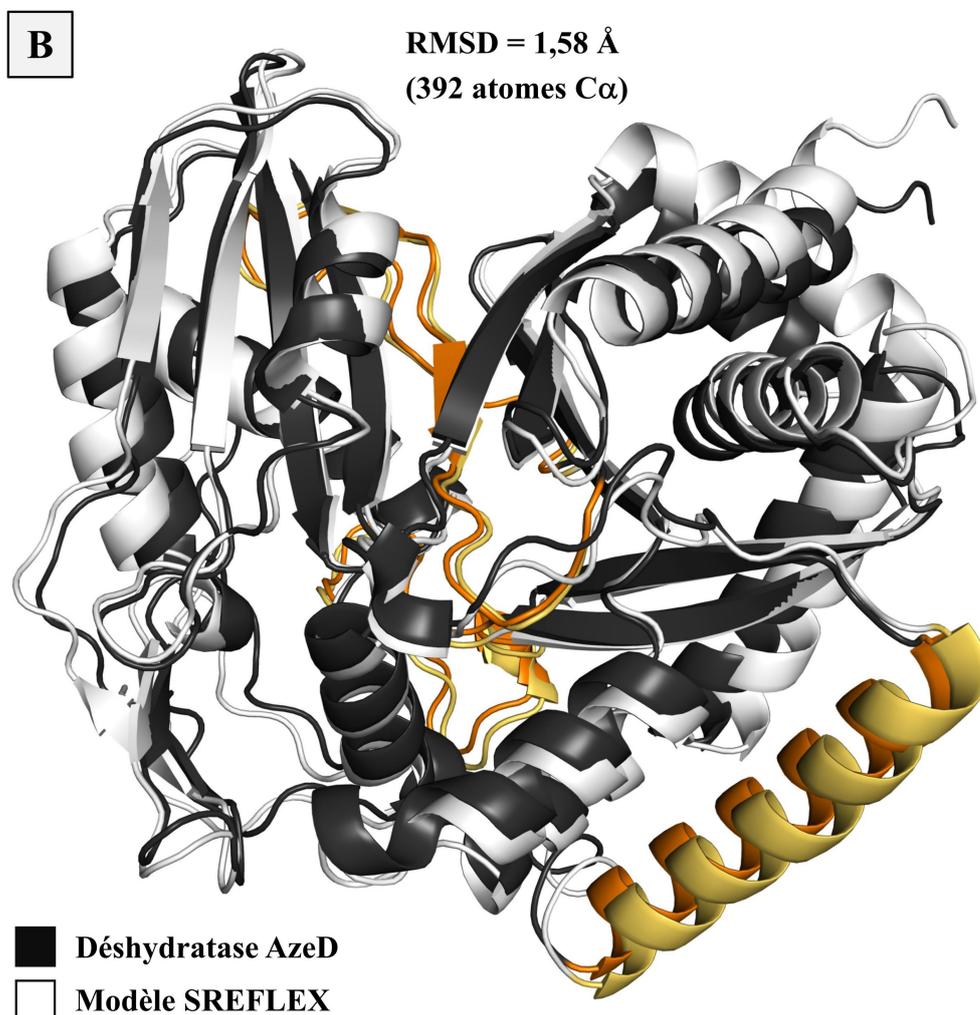
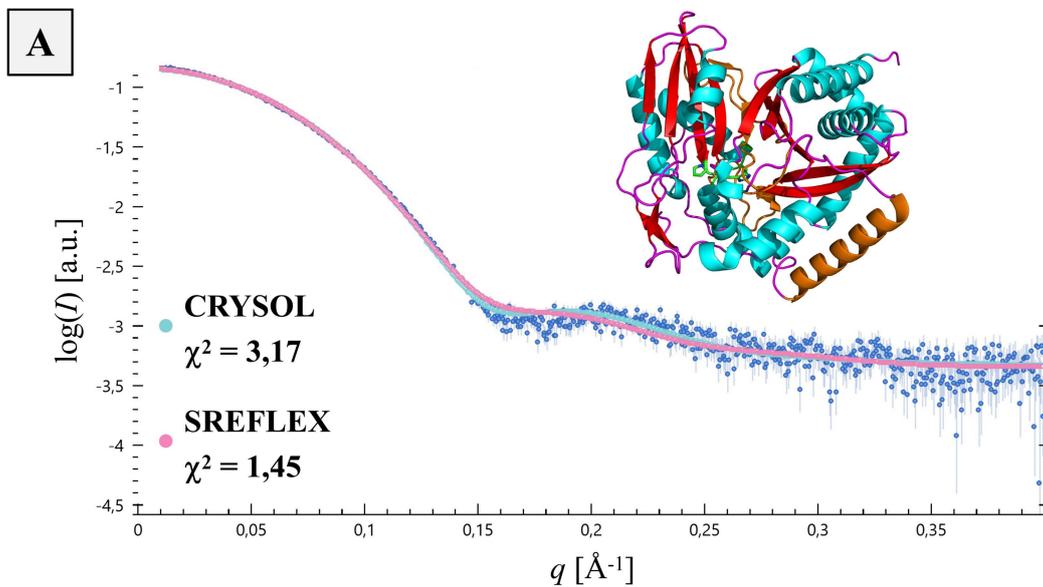


Figure 113. Conformation de la déshydratase AzeD en solution et modélisation d'un conformère. (A) La courbe obtenue avec le programme CRY SOL (Franke *et al.*, 2017) montre un accord important ($\chi^2 = 3,17$) entre les données théoriques de diffusion calculées pour la structure et la courbe expérimentale, ce qui confirme que la déshydratase AzeD est monomérique en solution. (B) Les modèles générés avec le programme SREFLEX (Panjkovitch et Svergun, 2016) suggèrent des ajustements conformationnels mineurs des éléments structuraux interlobulaires et des extrémités N- et C-terminales ($< 2,5$ Å), ce qui appuie l'hypothèse d'un état fermé de l'enzyme, du moins en l'absence du substrat.

II.2.B. Investigation du mécanisme catalytique

Des expériences de mutagenèse dirigée du domaine C_{modAA} et les tests d'activité correspondants ont permis d'identifier plusieurs résidus nécessaires à la fonction de déshydratation (Patteson *et al.*, 2022). En effet, la seule mutation des résidus His1496 ou Asp1501 (motif C3 HHxxxD) provoque l'agrégation du domaine isolé *in vitro*, et une mutation combinée avec la base His contiguë (H1497A) annihile la biosynthèse du produit, ce qui corrobore leur rôle dans le repliement local ou la catalyse (Patteson *et al.*, 2022). Néanmoins, la mutation du second résidu His des domaines C intégratifs de multiples NRPS impacte différemment leur activité respective, allant d'une réduction modeste à une suppression complète, ce qui suggère que sa fonction est susceptible d'être compensée par d'autres résidus du site actif (Bergendahl *et al.*, 2002 ; Roche et Walsh, 2003). En tous cas, étant donné que seuls quelques autres résidus semblent essentiels, les enzymes ayant une activité de déshydratation, dont AzeD, exploitent vraisemblablement une paire distincte de résidus acide/base afin de catalyser la réaction d'élimination (Patteson *et al.*, 2022).

SECTION III. Caractérisation biophysique des complexes PCP-C de la NRPS AzeB

Dans le but d'obtenir, de caractériser et de cristalliser des complexes non-natifs PCP1-C2-PCP2 de la NRPS AzeB, une stratégie originale a été développée à partir de la construction de plasmides d'expression de protéines chimériques jusqu'à l'analyse des macromolécules non-covalentes et aux tentatives de cristallisation. Tout d'abord, les vecteurs ont été conçus *in silico* avec le programme SnapGene et leurs séquences ont été générées par PCR ou chauffage-refroidissement (**Figure 114**). Les fragments ont alors été assemblés par une réaction de Golden Gate avec l'enzyme de restriction de type IIS *BsaI* et l'ADN ligase T4, et les produits attendus ont été détectés par PCR sur colonies des souches *E. coli* DH5 α transformées avec les mélanges réactionnels (**Figure 115**).

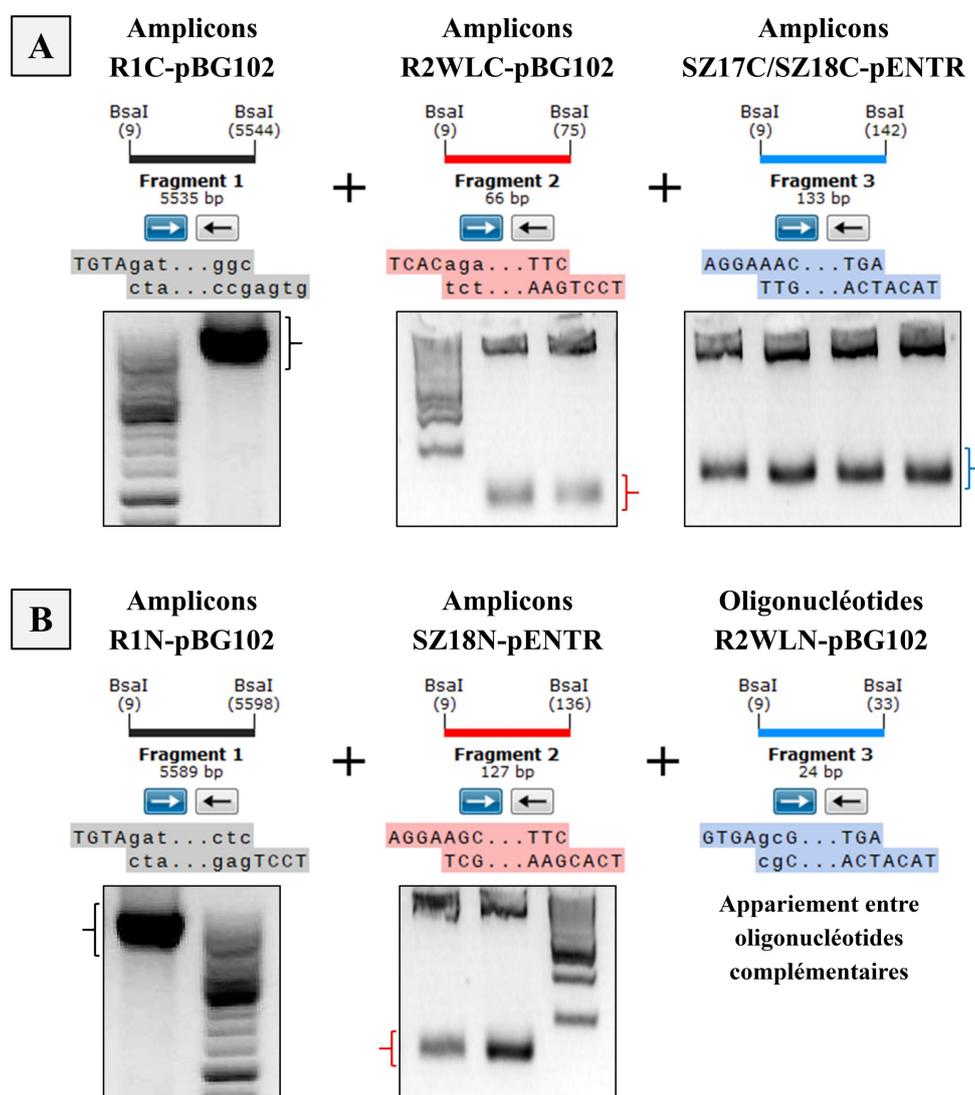


Figure 114. Amplification des séquences d'assemblage des plasmides pSZ par PCR. Dans le cadre de l'assemblage des plasmides pSZ17WLC et pSZ18WLC (A), ou pSZ18WLN (B), leurs fragments ont été préalablement amplifiés par PCR (*i.e.* R1C, R1N, R2WLC, SZ17C, SZ18C et SZ18N) et/ou générés par chauffage-refroidissement (R2WLN) avec les amorces ou oligonucléotides adéquats. pb : paires de bases.

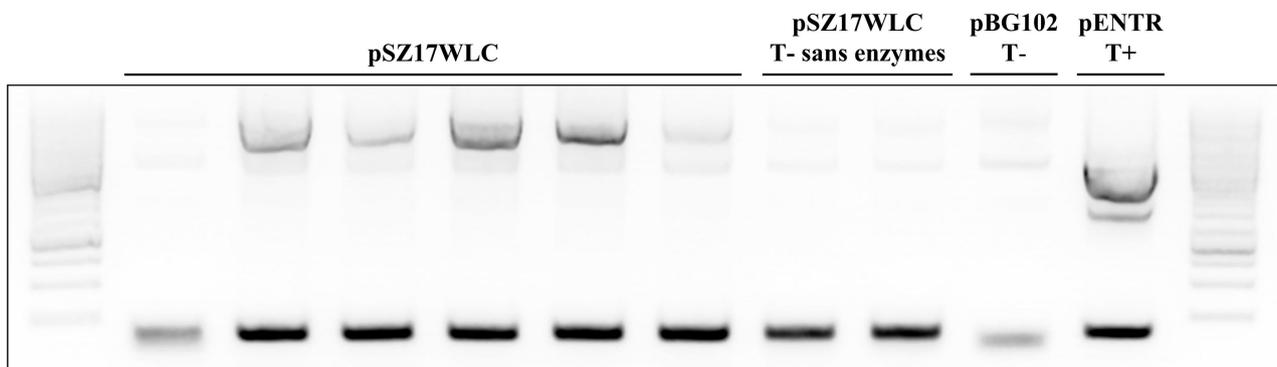


Figure 115. Détection du plasmide pSZ17WLC par PCR sur colonies de souches *E. coli* DH5 α recombinantes. Les amplicons de hauts poids moléculaires (+ 5500 pb) observés pistes 2–5 correspondent au produit attendu, ce qui indique l’assemblage réussi du plasmide pSZ17WLC. Le produit PCR n’est au contraire pas visualisé au niveau des différentes pistes des témoins négatifs (*i.e.* mélange réactionnel de Golden Gate sans enzymes, plasmide matriciel pBG102, chacun introduit chez *E.coli*), ce qui confirme le résultat probant du criblage. Les amplicons de 133 pb observés dans tous les échantillons correspondent au SYNZIP17, mais n’auraient pas dû être présents à la piste du plasmide pBG102 (témoin négatif), ce qui traduit un certain manque de spécificité de la paire d’amorces utilisées dans cette expérience.

Les séquences géniques PCP1-C2 et PCP2 ont été amplifiées par PCR, clonées dans les plasmides pSZ préparés à cet effet, et les produits obtenus ont été analysés par digestion enzymatique (**Figure 116**). Cette étape a servi à vérifier la fonctionnalité des nouveaux vecteurs construits et l’intégrité des inserts. Les constructions correspondantes ont ensuite été exprimées chez l’hôte hétérologue *E. coli* BL21 (DE3) et les protéines recombinantes purifiées par chromatographie en phase liquide selon le protocole standardisé (**Figure 117**). Cette étape a ainsi permis de produire les protéines chimériques fusionnées avec les SZ17 (PCP1-C2) ou SZ18 (PCP2) à leurs extrémités N- ou C-terminales, et d’en confirmer l’homogénéité en solution.

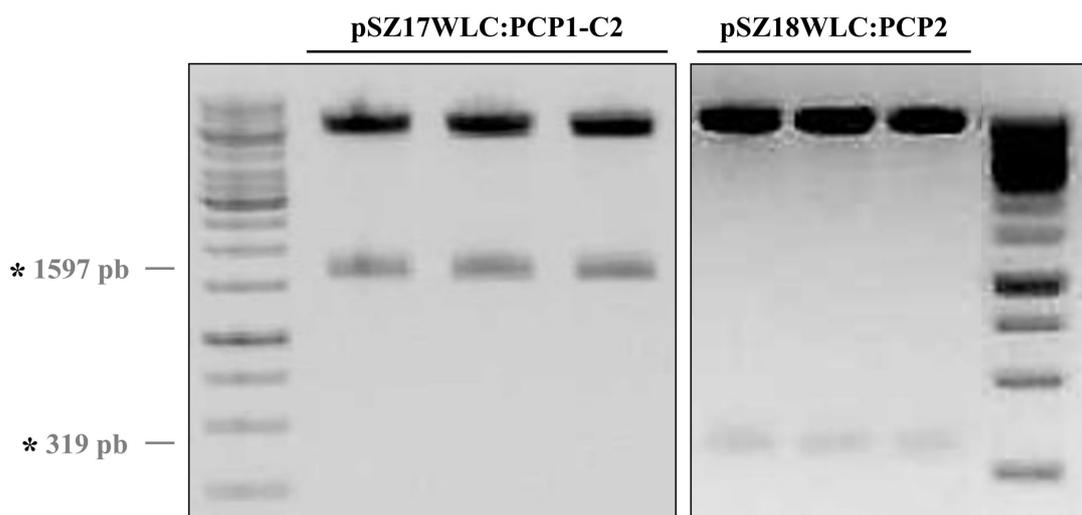


Figure 116. Digestion des plasmides pSZ recombinants par les enzymes *Bam*HI/*Eco*RI. Les vecteurs recombinants ont été analysés par digestion avec les enzymes de restriction *Bam*HI et *Eco*RI ayant été utilisées dans l’étape de clonage. Les bandes observées à environ 5000 pb correspondent aux plasmides recombinants linéarisés dans chaque échantillon, tandis que les bandes visualisées aux tailles intermédiaires correspondent aux inserts PCP1-C2 (1597 pb) ou PCP2 (319 pb) clonés chez l’hôte *E. coli* DH5 α .

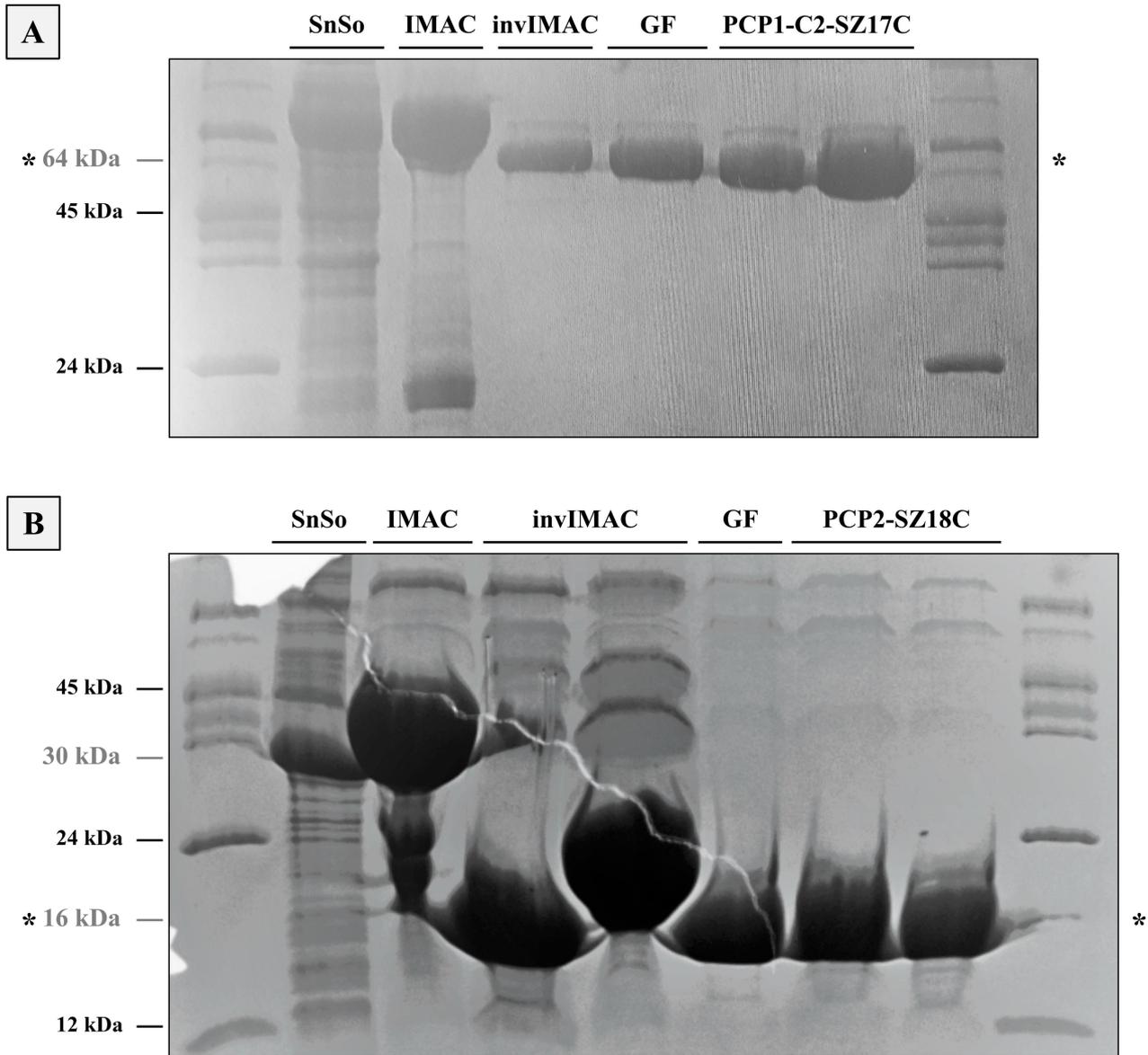


Figure 117. Purification des domaines chimériques par chromatographie liquide. Les domaines PCP1-C2-SZ17C (A) et PCP2-SZ18C (B) d'intérêt ont chacun été purifiés par chromatographie IMAC, suivie d'une dialyse en présence de la protéase HRV-3C pour cliver les étiquettes SUMO et 6×His. Suite à une seconde étape d'IMAC (notée invIMAC), où les protéines ont été éluées par lavage, les domaines ont été purifiés par chromatographie d'exclusion stérique (notée GF). Le rendement final de la purification a été estimé à 8 mg.L⁻¹ pour la protéine PCP1-C2-SZ17C et à 52 mg.L⁻¹ pour la protéine PCP2-SZ18C. De manière analogue, le domaine SZ18N-PCP2, non-représenté, a été purifié et le rendement final a été évalué à 44 mg.L⁻¹.

Les échantillons de protéines obtenus ont été analysés par DLS, ce qui a permis de confirmer la monodispersité des espèces PCP1-C2-SZ17C, PCP2-SZ18C et SZ18N-PCP2, et également par ITC, ce qui a permis de déterminer l’affinité de liaison des complexes macromoléculaires (**Figure 118**). En effet, dans chaque cas, la courbe de titrage isotherme montre que la variation d’enthalpie libre (ΔG) est négative, ce qui signifie que l’interaction, exergonique, ne nécessite aucune énergie d’activation. Grâce à la paire de SZ17:SZ18 ($K_d < 10$ nM), l’association des partenaires PCP1-C2-PCP2 apparaît assez forte au sein des complexes correspondants ($K_{d1} = 245$ nM, $K_{d2} = 463$ nM). Les résultats acquis par cette méthode pourraient être affinés *via* la modification des concentrations finales, ou l’inversion des protéines entre la cellule d’échantillonnage et la seringue, et validés par une autre technique, par exemple la résonance plasmonique de surface (SPR).

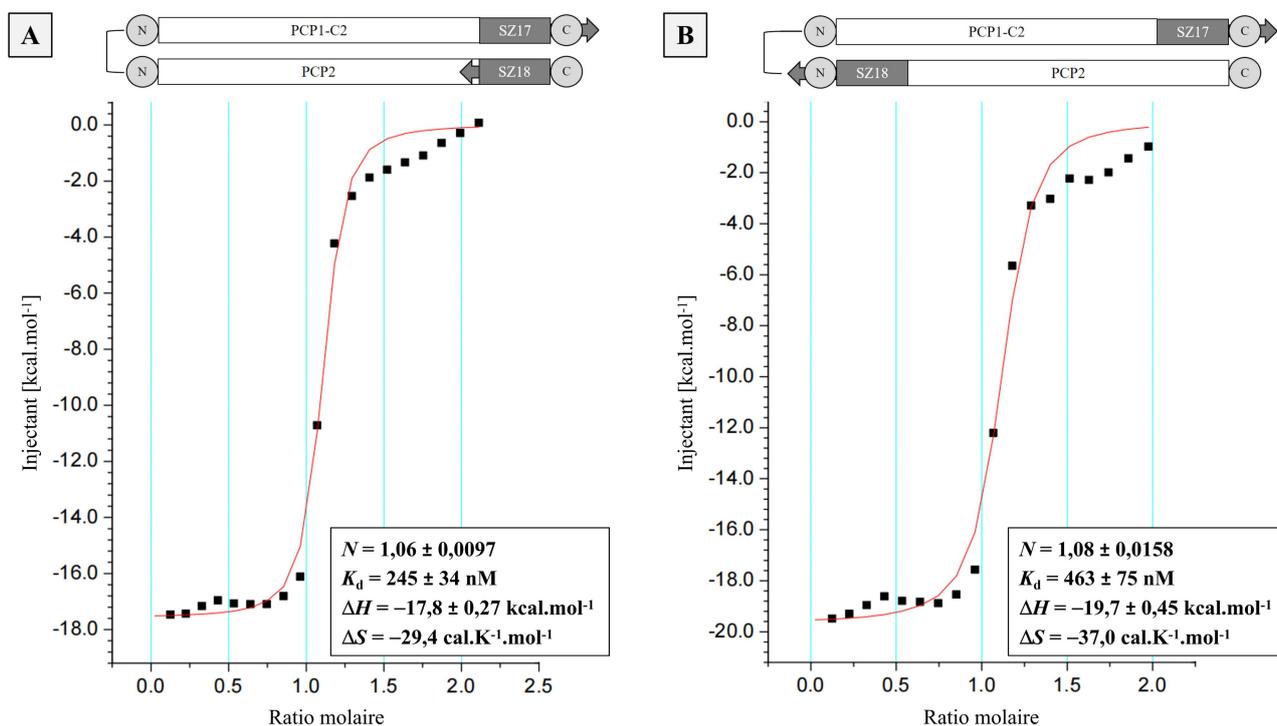


Figure 118. Analyse des complexes PCP1-C2-PCP2 par ITC. Les résultats obtenus par ITC mettent en évidence que les protéines chimériques PCP1-C2-SZ17C et PCP2-SZ18C (A) ou PCP1-C2-SZ17C et SZ18N-PCP2 (B) interagissent spontanément ($\Delta G < 0$) avec une stœchiométrie proche de $N = 1$ et une force de liaison assez importante ($K_d < 500$ nM), et forment les complexes non-covalents correspondants. Néanmoins, à cause de leur masse moléculaire de 80 kDa, cette association demeure d’une affinité 20–50 fois moindre par rapport à celle calculée à moins de 10 nM pour la seule paire de SZ17:SZ18 (9,9 kDa) par polarisation de fluorescence (Thompson *et al.*, 2012).

Les complexes PCP1-C2-PCP2 non-natifs ont été étudiés par SAXS, et les données de diffusion traitées puis analysées avec les programmes de la suite ATSAS 3.0 (Manalastas-Cantos *et al.*, 2021) (**Tableau XIV et Figure 119**). Des enveloppes structurales ont notamment été modélisées *ab initio* (DAMMIN), moyennées (DAMAVER) et filtrées (DAMFILT) en une forme unique minimale avec une variance acceptable (NSD = 0,67) (Volkov et Svergun, 2003). Un modèle computationnel obtenu avec le logiciel AlphaFold2 y a alors été superposé avec une concordance convenable (NSD = 4,65) (SUPCOMB) (Kozin et Svergun, 2001) (**Figure 119**). À défaut d’avoir réussi à obtenir des cristaux des macromolécules, ce résultat donne un aperçu général d’une structure tridimensionnelle et d’une conformation plausible, bien que l’orientation du complexe ternaire demeure sujette à interprétation, comme en témoigne un score d’ambiguïté important (1,69) (AMBIMETER) (Petoukhov et Svergun, 2015). De surcroît, la correspondance entre la courbe théorique calculée (SREFLEX) et les données expérimentales s’avère assez médiocre ($\chi^2 = 10,34$), ce qui suggère qu’une multitude de conformères coexistent en solution (Panjkovich et Svergun, 2016).

Tableau XIV. Enregistrement et traitement des données SAXS des complexes PCP1-C2-PCP2.

Traitement des données	PCP1-C2-SZ17C:PCP2-SZ18C	PCP1-C2-SZ17C:SZ18N-PCP2
Traitement de la courbe de Guinier		
$I(0)$ [cm^{-1}]	$0,110 \pm 3,4 \cdot 10^{-4}$	$0,089 \pm 4,0 \cdot 10^{-4}$
R_g [Å]	$39,93 \pm 0,14$	$36,99 \pm 0,25$
q_{\min} [Å ⁻¹]	0,025	0,024
$q \cdot R_g$ max	1,29	1,11
Coefficient de corrélation R^2	0,97	0,92
Traitement de la fonction $P(r)$		
q_{\min} – q_{\max} [Å ⁻¹]	0,02–0,33	0,02–0,33
$I(0)$ [cm^{-1}]	0,110	0,090
R_g [Å]	43,64	38,82
D_{\max} [Å]	153,39	132,55
Coefficient de corrélation R^2	0,75	0,88
Report de la masse moléculaire [kDa]		
Invariant de Porod (Q_p)	86,34	81,72
MoW	80,93	78,10
Volume de corrélation (V_c)	76,88	78,70
Inférence bayésienne	78,50	76,35
Échelle absolue	n.d.	n.d.

La courbe de Guinier dans une gamme linéaire de faibles q (0,025-0,05 Å⁻¹) ($q \cdot R_g < 1,3$) et la fonction $P(r)$ permettent d’évaluer les paramètres $I(0)$, R_g et/ou D_{\max} des complexes PCP1-C2-PCP2. En particulier, le rayon de giration atteint environ 40 Å et la distance interatomique maximale plus de 130 Å dans chaque cas. Par ailleurs, différentes méthodes permettent d’estimer les masses moléculaires, lesquelles sont cohérentes avec celles attendues (80,0 kDa et 80,2 kDa).

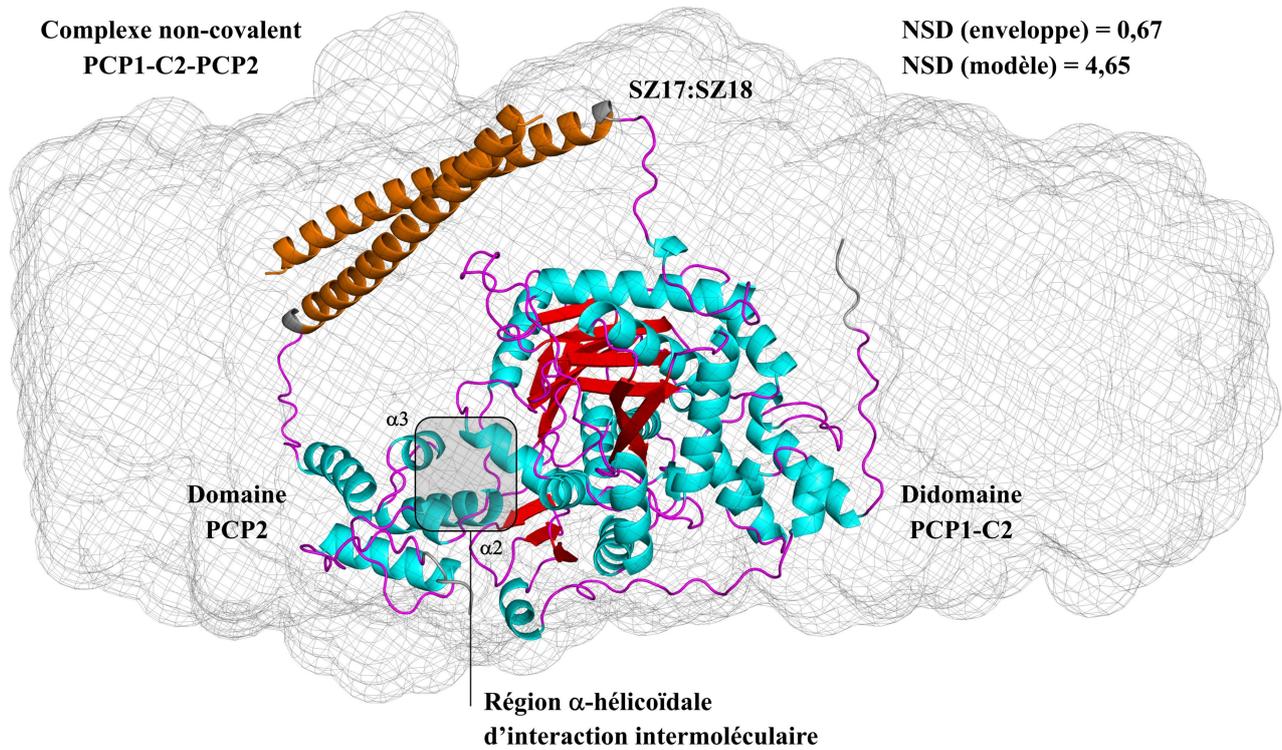


Figure 119. Interaction des multidomains PCP1-C2-SZ17C et PCP2-SZ18C et conformation d'un complexe par modélisation moléculaire. Dans le cadre de la modélisation de complexes ternaires, 10 enveloppes structurales ont été calculées et moyennées en une forme finale à faible résolution (39,1 Å) avec les programmes DAMs (Volkov et Svergun, 2003). Le paramètre de disparité spatiale (NSD) de 0,67 (idéalement inférieur à 1) indique une faible variabilité entre les modèles initiaux. Une structure atomique du complexe non-natif PCP1-C2-SZ17C:PCP2-SZ18C a été calculée avec le logiciel AlphaFold2, affinée par rapport à la courbe expérimentale (SREFLEX), et superposée à l'enveloppe moléculaire avec un accord correct (NSD = 4,65) (SUPCOMB) (Kozin et Svergun, 2001 ; Panjkovich et Svergun, 2016 ; Franke *et al.*, 2017). Dans le modèle correspondant, la paire de SYNZIP assure une interaction non-covalente entre les partenaires, comme attendu, et les hélices $\alpha 2$ et $\alpha 3$ du domaine PCP2 stabilisent le lobe N-terminal du domaine C2. Les domaines PCP sont néanmoins sujets à des mouvements conformationnels majeurs au sein du complexe, vu que 49 catégories de forme s'avèrent compatibles avec la distribution des électrons (AMBIMETER) (Petoukhov et Svergun, 2015), dont les plus étendues le semblent davantage.

Étant donné que la stratégie mise au point n'a jusqu'alors pas permis de cristalliser le domaine C2 flexible, une approche différente a été pensée, avec la co-expression de protéines chaperonnes, telles qu'une β -glucosidase (PDB : 1UG6) ou une β -lactamase (PDB : 1M40), ayant chacune été résolue à moins de 1 Å (Minasov *et al.*, 2002 ; Lokanath *et al.*, 2003). Au même titre que pour les complexes PCP1-C2-PCP2, des essais de cristallisation du domaine C2 pourront être initiés en présence de ces partenaires protéiques.

CONCLUSION ET PERSPECTIVES

Dans cette thèse, nous présentons la caractérisation structurale de l'hydrolase XhpG, une enzyme post-NRPS ayant un rôle de maturation d'un produit oxopyrrolizidinamine chez l'entomopathogène *Xenorhabdus hominickii*. La structure cristallographique résolue révèle en premier lieu le repliement canonique de la super-famille des α/β hydrolases et l'organisation du sous-domaine hélicoïdal, dont de multiples résidus hydrophobes définissent l'architecture profonde du site actif. Parallèlement, les résultats des expériences de SAXS mettent en lumière l'état dimérique de l'*apo*-protéine en solution, ainsi que le caractère rigide de la région charnière inter-domaine et la conformation fermée résultante. Par ailleurs, l'analyse computationnelle du mécanisme de liaison du substrat à l'interface, positionné et orienté dans un arrangement contraint à proximité de la triade catalytique Ser-His-Asp, démontre une dépendance envers des interactions hydrophobes et liaisons hydrogènes, à l'instar de la lactonase PcaD homologue. En dépit d'une certaine variabilité structurale du sous-domaine α -hélicoïdal et du site actif entre ces enzymes, la région de la coiffe détermine en grande partie la spécificité de substrat et pourrait faciliter la libération du produit depuis le cœur α/β . Afin de confirmer ces interprétations, des essais avancés de co-cristallisation de la macromolécule en présence d'analogues fonctionnels et des expériences de mutagenèse dirigée de résidus importants, suivies de tests d'activité enzymatique, pourraient être effectués à moyen terme. De cette façon, des variantes possédant une gamme élargie de substrats, une énantiosélectivité augmentée ou inversée, ou une thermostabilité accrue, ouvriraient la voie à de nouvelles applications biotechnologiques, avant tout dans le système modèle.

De manière analogue, nous rapportons la caractérisation d'une autre enzyme de modification, la déshydratase AzeD, responsable de l'élimination bimoléculaire (E2) d'un groupement hydroxyle du produit azétidomonamide au sein du pathogène humain *Pseudomonas aeruginosa*. Dans un premier temps, la structure cristallographique déterminée dévoile un repliement typique des chloramphenicol acétyltransférases (CAT) en deux lobes avec une fente centrale, ainsi que l'organisation des éléments de maintien au niveau de l'interface. En accord avec la conformation fermée de l'*apo*-protéine dans l'unité asymétrique et comme pour la majorité des domaines C intégratifs homologues, l'analyse des courbes SAXS ne met en évidence aucun mouvement majeur du β -*crossover* étendu ou de structures secondaires flexibles sous-jacent à un contrôle de l'accessibilité du site actif. Par rapport au mode de reconnaissance du substrat au sein de cette cavité hydrophobe, les résultats de *docking* moléculaire le montrent en étroite liaison avec la base His130, incluse dans le motif catalytique C3 xPHxxxDG, et d'autres résidus à rôle structural (Pro129) ou stabilisateur. Néanmoins, le résidu acide impliqué dans la reprotonation reste inconnu et, afin de lever cette ambiguïté, des investigations complémentaires concernant le mécanisme réactionnel *in vitro* sont actuellement menées avec les collaborateurs.

Au sein de leur système biosynthétique respectif, les enzymes XhpG et AzeD sont les premières à avoir été caractérisées sur le plan structural. Dans ce contexte, des approches bio-informatiques et expérimentales combinées, telles que l'analyse de la co-expression génique des clusters, suivie de la précipitation de protéines partenaires et de l'identification des complexes supramoléculaires obtenus, permettraient d'établir les réseaux d'interactions correspondants. Sous réserve de la purification de ces entités, des techniques biophysiques aux visées multiples, comme l'ITC, l'interférométrie (BLI), la fluorimétrie différentielle à balayage (DSF), le dichroïsme circulaire (CD) ou la thermophorèse à micro-échelle (MST) permettraient de les étudier de manière approfondie. À ce propos, une question pertinente en suspens concerne la spécificité, l'affinité et la stabilité des interactions putatives entre les NRPS et BVMO au cours de la libération du dipeptide intermédiaire commun aux voies des AP. Malgré la purification de la méga-synthase AzeB de *P. aeruginosa* et la cristallisation fortuite de la mono-oxygénase AzeC, le rendement limité ($1-3 \text{ mg.L}^{-1}$) et/ou le problème de non-reproductibilité des cristaux entravent la poursuite avec succès de recherches avancées (**Figure 120**). Par conséquent, les perspectives de travail porteraient sur l'amélioration des conditions d'obtention de ces enzymes biosynthétiques et sur la compréhension des relations structure-fonction-interaction. De tels résultats contribueraient à clarifier la régulation de l'activité enzymatique et à faciliter la modélisation des flux des composés intermédiaires en aval du point de jonction métabolique.

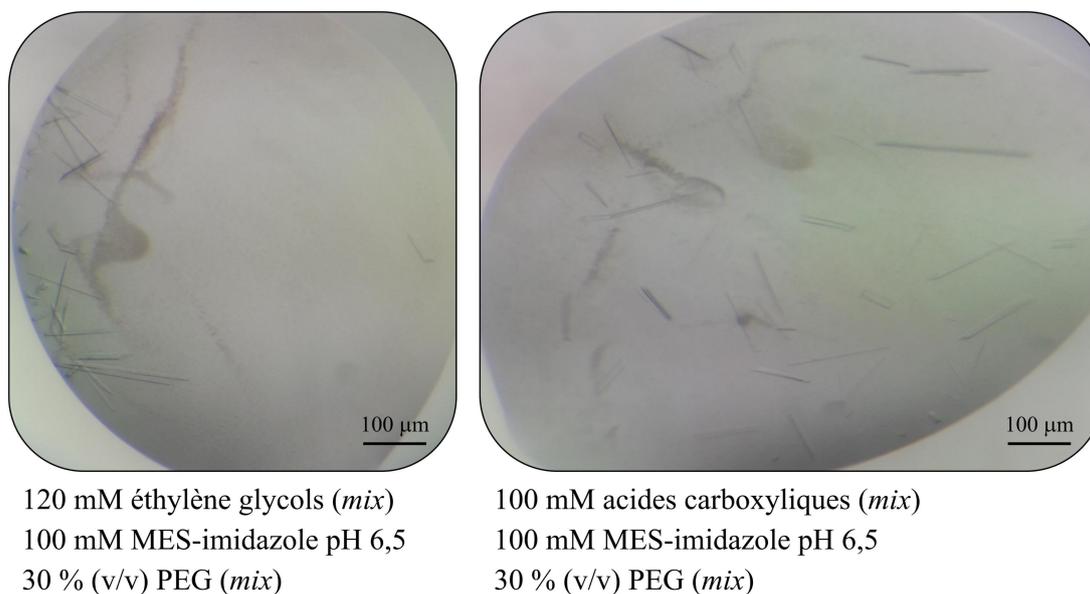
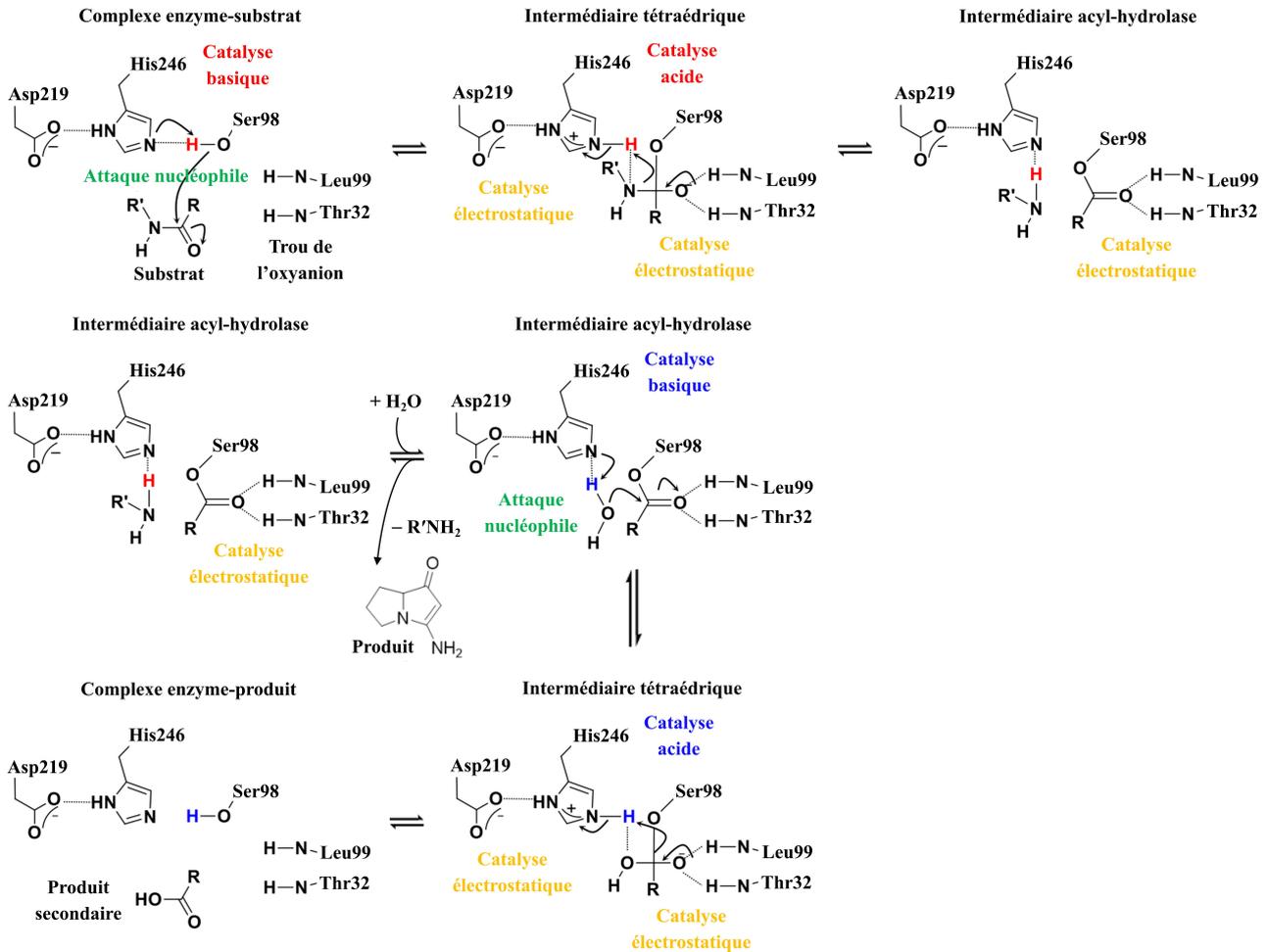


Figure 120. Cristallisation de la BVMO AzeC par diffusion de vapeur. Les cristaux de l'enzyme ont été obtenus à $1,5 \text{ mg.mL}^{-1}$ en présence du cofacteur NADPH (2 mM) dans des conditions du kit Morpheus (Molecular Dimensions) par diffusion de vapeur en nano-gouttes. Ces cristaux en forme de longues baguettes ne peuvent néanmoins pas être affinés en l'état à cause de la faible quantité de protéines dans chaque lot et de la non-reproductibilité des essais préliminaires. BVMO : mono-oxygénase de Baeyer-Villiger, NADPH : nicotinamide adénine dinucléotide phosphate, MES : acide 2-(*N*-morpholino) éthane sulfonique, *mix* : mélange équimolaire.

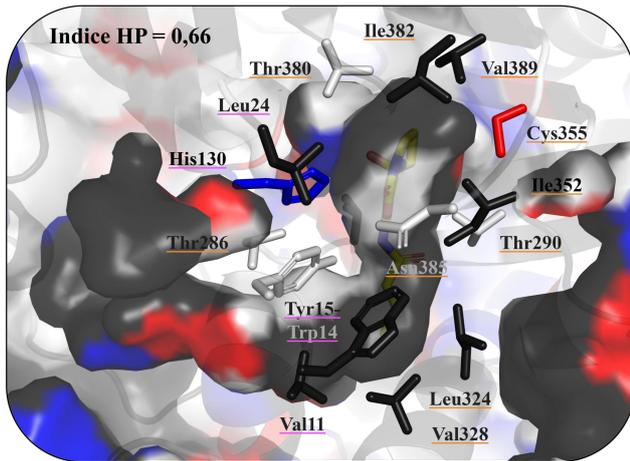
Une autre perspective intéressante serait d'explorer la compartimentation spatiale des enzymes des systèmes modèles chez les hôtes *X. hominickii* et *P. aeruginosa*. Étant donné que les interactions au sein des complexes protéiques sont susceptibles de jouer un rôle primordial dans la séquestration subcellulaire des partenaires, des analyses en microscopie à fluorescence permettraient de fournir de précieuses indications quant à l'activation de ces voies métaboliques en réponse à des changements environnementaux (*quorum-sensing*). Par exemple, au cours d'une infection du pathogène humain *P. aeruginosa*, des acyl-homosérine lactones (AHL) régulent positivement l'expression du BGC *aze* et un produit final (azétidomonamide B) favorise le développement des biofilms bactériens. Ainsi, ce travail de fond concernant l'organisation macromoléculaire des NRPS et une co-localisation avec des enzymes biosynthétiques permettrait d'établir divers effets sur le fonctionnement de ces systèmes, en termes d'efficacité catalytique, ainsi que sur le devenir des métabolites et leurs activités biologiques dans des processus critiques (*e.g.* virulence, production de toxines, résistance aux antibiotiques).

En conclusion, les connaissances acquises au sujet des relations structure-fonction-interaction des enzymes biosynthétiques et de leur possible organisation en méga-complexes, avec la NRPS en tant qu'entité centrale, contribueront indubitablement à l'ingénierie des systèmes artificiels. Compte tenu du large potentiel pharmaceutique des AP bactériens, ces travaux permettront à terme de générer des dérivés bioactifs originaux, notamment comme effecteurs contre le pathogène virulent *P. aeruginosa*.

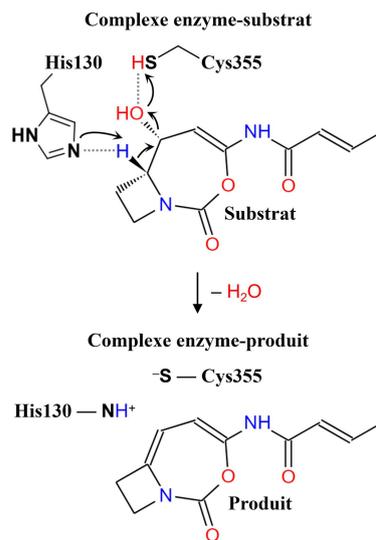
ANNEXES



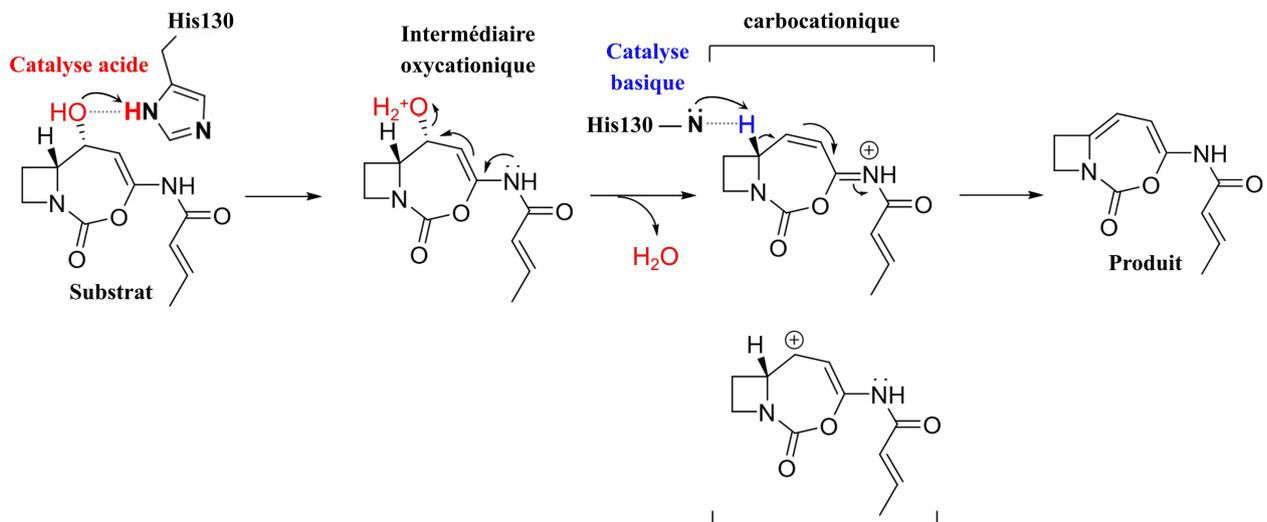
Annexe 1. Proposition du mécanisme réactionnel de l'hydrolase XhpG (adaptée de K. J. Weissman). Le mécanisme réactionnel correspond incontestablement à celui établi à l'origine pour les protéases à sérine. Dans un premier temps, la triade catalytique Ser-His-Asp forme un relai de charge, ce qui active le nucléophile Ser98. Au sein de la structure, celui-ci est inclus dans le motif GFSLG, dont les résidus glycine y imposent une forte courbure, ce qui permet au nucléophile de s'approcher à proximité du substrat. La fonction hydroxyle du résidu Ser98 attaque alors le groupement carbonyle de la liaison amide du ligand oxopyrrolizidinacétamide, ce qui crée un intermédiaire tétraédrique (sp^3) instable. Le trou de l'oxyanion comprend les groupements amides des résidus Leu99/Thr32 adjacents, et stabilise l'intermédiaire transitoire *via* des interactions électrostatiques. Cela génère un intermédiaire acyl-hydrolyse, entraîne le clivage de liaison amide et assure la libération du produit ($-R'NH_2$). Ensuite, une molécule d'eau activée catalyse une nouvelle attaque nucléophile sur le groupement carbonyle de l'intermédiaire acyl-hydrolyse, ce qui crée un autre intermédiaire tétraédrique. Dans un dernier temps, les atomes de carbone et d'oxygène de celui-ci reforment une double liaison, ce qui provoque la libération d'un second produit de type acide carboxylique et la régénération du site actif.

A

■ Résidus hydrophobes/polaires
 ■ Résidus chargés +/-

B**C**

Complexe enzyme-substrat



Annexe 2. Proposition de mécanismes réactionnels de la déshydratase AzeD. (A) La poche catalytique présente une nature hydrophobe avec un indice HP positif, calculé à partir de l'outil PDBsum (Laskowski *et al.*, 2018). De multiples résidus aliphatiques ou aromatiques de chacun des lobes contribuent à la formation du site actif et assurent la liaison du ligand. Les résidus hydrophobes, polaires, et/ou chargés +/- sont montrés en noir, blanc, bleu et rouge, respectivement. (B) Le mécanisme de déshydratation du substrat L-azétidomonamide correspond vraisemblablement à une élimination bimoléculaire (E2). La réaction nécessite une conformation antipériplanaire du proton acide et de l'hydroxyle (C7-OH), la stéréochimie naturelle lorsque ce groupement nucléofuge a une configuration *R*. Dans ce cas uniquement, un couple acide/base, positionné de part et d'autre de la cavité hydrophobe (*e.g.* His130/Cys355), assure de manière synchrone la déprotonation du centre tertiaire (carbone), la formation de la double liaison et l'élimination de l'eau. (C) Le mécanisme réactionnel pourrait alternativement correspondre à une élimination monomoléculaire (E1). En ce sens, le résidu His130 agit à la fois en tant qu'acide et base générale, ce qui crée un intermédiaire de type carbocation, stabilisé par mésomérie dans un système conjugué étendu, et permet la libération du produit.

1 **Pyrrolizwilline, a unique bacterial alkaloid assembled by a nonribosomal peptide**
2 **synthetase and non-enzymatic dimerization**

3 Juliana Effert^{1,&}, Margaretha Westphalen^{2,&}, Andrea Calderari^{3,&}, Yi-Ming Shi^{1,4}, Isam
4 Elamri^{1,2}, Soumaya Najah⁵, Peter Grün¹, Yanyan Li⁵, Arnaud Gruez^{3,*}, Kira J.
5 Weissman^{3,*}, Helge B. Bode^{1,2,6,7,8,*}

6

7 1 Max Planck Institute for Terrestrial Microbiology, Department of Natural Products
8 in Organismic Interactions, 35043, Marburg, Germany.

9 2 Molecular Biotechnology, Department of Biosciences, Goethe University
10 Frankfurt, 60438, Frankfurt am Main, Germany.

11 3 Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France.

12 4 Key Laboratory of Quantitative Synthetic Biology, Center for Synthetic
13 Biochemistry, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of
14 Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China.

15 5 Unit MCAM, UMR 7245 CNRS-Muséum National d'Histoire Naturelle (MNHN),
16 75005 Paris, France.

17 6 Center for Synthetic Microbiology (SYNMIKRO), Phillips University Marburg,
18 35043 Marburg, Germany.

19 7 Department of Chemistry, Phillips University Marburg, 35043 Marburg, Germany.

20 8 LOEWE Centre for Translational Biodiversity Genomics (LOEW-TBG) &
21 Senckenberg Gesellschaft für Naturforschung, 60325, Frankfurt am Main,
22 Germany.

23

24 * Corresponding authors: arnaud.gruez@univ-lorraine.fr; kira.weissman@univ-lorraine.fr;
25 helge.bode@mpi-marburg.mpg.de

26 & These authors contributed equally to this work

27

28

29 **Abstract**

30 Pyrrolizidine alkaloids (PAs) are a structurally diverse group of heterocyclic specialized
31 metabolites characterized by a core structure comprising a hexahydro-1*H*-pyrrolizine.
32 PAs are synthesized through two main pathways. In plants, assembly occurs via a
33 homospermidine synthase, and in bacteria, through combined action of a nonribosomal
34 peptide synthetase and a Baeyer-Villiger monooxygenase. While the toxic properties of
35 plant-derived PAs and their prevalence in animal and human foods have been
36 extensively studied, the biological roles and biosynthesis of more complex bacterial PAs
37 are not well understood. Here, we report the identification and characterization of a
38 bacterial biosynthetic gene cluster from *Xenorhabdus hominickii*, *xhpA–G*, which is
39 responsible for producing the PA pseudo-dimer pyrrolizwilline. Analysis of *X. hominickii*
40 promoter exchange mutants together with heterologous expression of *xhpA–G* in *E. coli*,
41 revealed a set of pathway intermediates, two of which were chemically synthesized, as
42 well as multiple derivatives. This information was leveraged to propose a detailed
43 biosynthetic pathway to pyrrolizwilline. Furthermore, we have characterized the
44 hydrolase XhpG, the key enzyme in the conversion of the pathway intermediate
45 pyrrolizixenamide to pyrrolizwilline, using X-ray crystallography and small-angle X-ray
46 scattering (SAXS).

47

48 Introduction

49 Pyrrolizidine alkaloids (PA) are widespread natural products, which are defined by their
50 pyrrolizidine core structure (a bicyclic aliphatic hydrocarbon which consists of two ortho-
51 fused five-membered rings with a bridgehead nitrogen).^[1] Most of the prominent PAs
52 are produced in plants, where the key biosynthetic enzyme is the homospermidine
53 synthase (HSS). HSS concatenates two putrescine units to yield homospermidine,
54 which is then transformed into the canonical PA skeleton by a series of oxidations and
55 Mannich reactions.^[1,2] Plant PAs are toxic to both humans and animals, acting as both
56 hepatotoxins (e.g. retrorsine^[3]) and genotoxins (e.g. europine^[4]).

57 The biosynthetic origin of the approximately 40 known bacterial PAs^[5] is distinct, and
58 their ecological roles and activities are largely unknown. Notably, however, the PA-
59 derived natural product brabantamide, carrying a tetrahydro-1*H*,3*H*-pyrrolo[1,2-*c*]oxazol-
60 3-one core instead of the hexahydro-1*H*-pyrrolizine, inspired the development of a first-
61 in-class inhibitor of phospholipase A₂, a promising approach for the treatment of
62 atherosclerosis.^[6]

63 The skeleton of microbial PA derivatives is constructed by a common pathway
64 involving a nonribosomal peptide synthetase (NRPS) and a Baeyer-Villiger
65 monooxygenase (BVMO).^[7] NRPSs are multifunctional mega-enzymes that assemble
66 peptide chains via a series of coordinated enzymatic steps.^[8] In these systems, the
67 catalytic and peptidyl carrier protein (PCP) domains are organized into functional
68 modules. For example, in the biosynthesis of pyrrolizixenamide (**1**) in the
69 entomopathogenic bacterium *Xenorhabdus stockiae*,^[7] the bimodular NRPS PxaA
70 extends *N*-acylated serine with proline, followed by dehydration of Ser and cyclization to
71 yield a 5,6-bicyclic indolizidine (**3**) (Figure 1). A BVMO, PxaB, then catalyzes oxygen
72 insertion leading to a 5,7-bicyclic intermediate (**4**). Finally, spontaneous ring contraction
73 and decarboxylation afford pyrrolizixenamide-C₆ (**1**).^[7] Interestingly, the outcome of the
74 BVMO reaction seems to be dependent on the acyl residue, as the PxaB homolog BraC
75 requires 3-hydroxy acyl residues for brabantamide core formation.^[9] Further examples
76 of bacterial PAs built around a central pyrrolizidine include the bohemamines,^[10]
77 clazamycins,^[11] legonmycins,^[5, 12] jenamidines,^[13] and the structurally related

78 azetidomonamide B,^[14] which differ mainly in the acyl side chain or the amino acids
79 used as building blocks. The pyrrolizidine core can be further diversified as exemplified
80 by the various bohemamine derivatives including spithioneines,^[15] quinoheanine,^[16]
81 dibohemamines^[17] and the 1,3-oxazin-6-one^[18] derivatives. While the biosynthesis of the
82 pyrrolizidine core has been studied in several systems,^[5, 7, 12, 14] the subsequent
83 reactions leading to PA structural diversification are not well understood.

84 Here we describe the discovery and biosynthesis of pyrrolizwilline (**5**) from
85 *Xenorhabdus hominickii*, which is an unusual pseudo-dimer derived from
86 pyrrolizixenamide-C₂ (**2**). The identity of its biosynthetic cluster (*xhp*) was confirmed by
87 heterologous expression in *E. coli*, which also allowed for the identification of further
88 derivatives as well as biosynthesis intermediates, two of which were chemically
89 synthesized. Based on analysis of *X. hominickii* promoter exchange mutants, we
90 propose a detailed pathway to pyrrolizwilline. Finally, we report the atomic resolution
91 crystal structure of a key enzyme, the hydrolase XhpG, which initiates the multi-step
92 reaction series by which **2** is elaborated to **5**.

93

94 Results

95 In the context of our long-term interest in discovering bacterial PA compounds with
96 novel structures, we examined our in-house database of >100 *Xenorhabdus* and
97 *Photorhabdus* genomes and identified *Xenorhabdus hominickii* DSM 17903 as the only
98 strain encoding two biosynthetic gene clusters (BGCs) with an architecture resembling
99 that of the previously described PA pyrrolizixenamide-C₆ (**1**) BGC^[7] (Figure 1A). The
100 first BGC consists of two genes, *pxzA* and *pxzB*, encoding a bimodular NRPS and a
101 BVMO.^[7] We focused our attention, however, on the unique *xhp* BGC which is located
102 at a distance of 2,905 kbp from the *pxz* BGC, as it contains five additional genes
103 (*xhpC–xhpG*) next to those encoding the NRPS and BVMO (*xhpA*, *xhpB*). Analysis
104 using the CompArative GENE Cluster Analysis Toolbox (CAGECAT)^[19] revealed that the
105 *xhpA–xhpG* BGC is unique in its gene composition, suggesting that the encoded
106 pathway might reasonably lead to the new pyrrolizidine structure (Figure 1B).

107 Identification of the products of the *pxzAB* and *xhpA–G* BGCs

108 Initial analysis of extracts of wild type (WT) *X. hominickii* revealed the production of **2**
109 (Figure S1, Table S1) by PxzAB, but no other related compound potentially deriving
110 from the *xhp* cluster. Therefore, to identify the corresponding natural product, we
111 performed promoter exchange experiments^[20] in order to activate its expression. For
112 this, we swapped the native promoter of *xhpA* with an arabinose-inducible P_{BAD}
113 promoter to generate the strain *X. hominickii* P_{BAD_xhpA} , carrying out the same
114 exchange into the *pxz* cluster as a control. Cultivation and induction of *X. hominickii*
115 P_{BAD_xhpA} cultures allowed for identification of **5** (Figure 2A), henceforth named
116 pyrrolizwilline, via liquid chromatography coupled to high resolution mass spectrometry
117 (LC-HRMS) (m/z calc'd for $C_{21}H_{24}N_4O_3$ $[M+H]^+$ = 381.1921; obs'd m/z = 381.1918;
118 Δ ppm = -0.8). Importantly, this mass was not observed in the control cultures (Figure
119 2A). The overall structure of **5** was characterized by NMR following large scale
120 cultivation and purification (Figure S2). This analysis revealed that **5** harbors two
121 pyrrolizidines fused with a central pyridine-4-carboxylic group that is linked to a valine
122 (Figure 2A). The presence of L-valine in **5** was confirmed by intact incorporation of
123 deuterated D_8 -L-valine (Figure S3). Furthermore, heterologous expression of *xhpAB* and
124 *pxzAB* in *E. coli* (Figure 2B) yielded **2** as the main product (Table S1 and Figure S4),
125 suggesting that **5** is derived from **2**. In addition, a variant of **2** incorporating a propionate
126 derived ethyl group (**6**) was detected as a minor product (Table S1 and Figure S4).

127

128 The pyrrolizwilline biosynthetic pathway

129 The first two steps of the Xhp and Pxz pathways leading to **2** are identical. Specifically,
130 the NRPSs PxzA/XhpA assemble a 5,6-bicyclic indolizidine intermediate (**3**, Figure S5),
131 followed by oxygen insertion by the BVMOs PxzB/XhpB, subsequent ring contraction
132 and decarboxylation (as described previously in Figure 1).

133 We hypothesized that the *N*-acyl hydrolase XhpG then plays an important role in
134 initiating the diversification from **2** to **5**. As initial confirmation for this idea, an *xhpG*
135 inactivation mutant ($\Delta xhpG$) of *X. hominickii* P_{BAD_xhpA} was found unable to produce **5**,

136 but accumulated **2** as expected, as assessed by LC-HRMS (Figure S6). Furthermore,
137 addition of *xhpG* to the *E. coli* strains expressing *xhpAB* and *pxzAB* under the control of
138 a P_{BAD} promoter, led to a new compound **7**, with a mass consistent with the deacylated
139 form of **2** (Table S1, Figure 3 and Figure S7). For structure confirmation and *in vitro*
140 biochemical assays, we additionally synthesized **2** and **7** starting from commercial N-
141 benzyloxycarbonyl-L-proline following established protocols^[21] (Figure S8) and
142 confirmed their structure by NMR (**2**: Table S6 and Figure S9; **7**: Table S7 and Figure
143 S10) and HR-MS analysis. Interestingly, XhpG also accepts the product of PxaAB^[7] –
144 pyrrolizixenamide with a C₆ acyl side chain (**1**) – as a substrate (Figure S7). Next,
145 heterologous expression of different combinations of *xhpAB* and *xhpG* with the genes
146 encoding the putative short-chain dehydrogenase/reductases (SDRs) *xhpC* and *xhpD*
147 revealed that only when the reductase XhpD is present, the intermediate **8** from the
148 reduction of **7** is detected (Figure S11). As a proof of the reaction order, heterologous
149 expression of *xhpABD* with *xhpG* resulted in production of **8** (Figure S12), whereas
150 expression of *xhpABD* only did not yield the reduced form of **2**. Hence, we concluded
151 that XhpG first deacetylates **2** to **7** followed by the reduction of **7** to **8** by XhpD (Figure
152 3).

153 To further investigate pyrrolizwilline biosynthesis, we expressed the entire *xhp* cluster
154 (*xhpA–G*) heterologously in *E. coli*. LC-HRMS analysis of culture extracts allowed us to
155 detect the intermediates **8**, **9**, **10**, **11**, and **12a** (Table S1 and Figure S13). Together,
156 these results allowed us to propose the following complete biosynthetic pathway to **5**
157 (Figure 3): After elimination of water from **8**, the resulting **9** is extended with glyoxylic
158 acid via nucleophilic attack of the electron-rich carbon in **9**, forming **13**. Glyoxylic acid is
159 a common metabolite in the glyoxylate cycle in bacteria.^[22] This step is likely non-
160 enzymatic, given the reactivity of **9**. Subsequent dehydration of **13** followed by
161 nucleophilic attack by a second pyrrolizidine moiety (**9**) generates dimeric **15**. After
162 intramolecular nucleophilic attack by an amine to form the central ring and release of
163 ammonia, proton loss of intermediate **17** subsequently results in stable intermediate **11**.
164 We hypothesize that **11** gets either reduced to **10** as a side product or is oxidized to **12a**,
165 which is then activated by an unknown CoA ligase to the corresponding CoA ester **18**.
166 Finally, the freestanding C domain XhpF catalyzes condensation of **18** with L-valine to

167 yield **5**. We hypothesize that the putative CoA ligase responsible for activation of **12a** is
168 missing in *E. coli*, and therefore the final product **5** is not observed.

169 Remarkably, we were able to detect two additional derivatives of **12** – **12b** and **12c** –
170 when *xhpA–G* were expressed in *E. coli* (Figure 3), which correspond to methyl and
171 hydroxyethyl-substitutions on the pyridine ring, respectively (Table S1 and Figure S13).
172 This result may reflect the fact that different aldehydes are present in the cell and can
173 therefore react with **9**. Specifically, acetaldehyde and lactaldehyde, common
174 metabolites in *E. coli* metabolism, could lead to the formation of **12b** and **12c** (Figure
175 S13), respectively.^[23] Similarly, feeding the culture with propionaldehyde led to the
176 production of a fourth derivative, **12d**, confirming that free aldehydes do indeed react
177 with **9** (Figure S13). When the synthetic model compound 5-amino-2,3-dihydro-1*H*-
178 pyrrolizin-1-one (**19**), an oxo-derivative of **9**, was combined with glyoxylic acid or
179 acetaldehyde *in vitro*, **20** and **21** were produced spontaneously (Figure 4 and Figure
180 S14), consistent with the proposed intrinsic reactivity of **9**, further supporting our
181 hypothesis. Notably, **9** is a close homologue of the clazamycins isolated from
182 *Streptomyces* species. Clazamycins carry a chlorine and a bridgehead-OH on the
183 pyrrolizidine core, in contrast to **9** which is formed by spontaneous oxidation in air.^[11, 24]

184

185 **Biochemical characterization of XhpG**

186 Having demonstrated a key role for XhpG as an *N*-acyl hydrolase in the pathway to **5**,
187 we next aimed to examine its function in detail. For this, we purified the WT protein and
188 an active site S98A mutant (Figure S16 and S17 *vide infra*), and incubated the enzymes
189 with synthetic pyrrolizixenamide **2** *in vitro*. LC-HRMS analyses of the reaction mixtures
190 showed that the WT XhpG produced significant amounts of a new compound with a
191 mass corresponding to the hydrolytic product **7** (Table S1 and Figure S15). Comparison
192 of the retention time and tandem MS spectra with those of synthetic **7** standard
193 confirmed the identity of the product (Figure S15). While **7** was observed in both the
194 negative controls and in the presence of XhpG_S98A which we ascribe to the intrinsic

195 instability of **2** under the assay conditions, turnover to **2** was substantially greater in the
196 presence of WT XhpG, confirming a role for Ser98 in the catalysis.

197

198 **Purification and crystallization of the mutant XhpG S98A and structure** 199 **determination**

200 We next aimed to solve the structure of a binary complex of XhpG_S98A in the
201 presence of its native substrate. After confirming the monodispersity of the sample by
202 dynamic light scattering (DLS) (Figure S18), we crystallized the *apo* form of the enzyme
203 by vapor diffusion and solved its structure by molecular replacement at a resolution of
204 1.18 Å (PDB ID: 8QZU) (Table S8). The asymmetric unit contains a monomer of 29 kDa
205 molecular weight. Disordered residues are quite rare in the structure: the Gly1 and Pro2
206 of the residual tag, as well as Asp268, Val269 and Pro270 residues of the C-terminal
207 loop are missing in the electron density maps. Overall, 98.5% (259) of the residues are
208 located in the favored region of the Ramachandran plot, and 99.6% (221) display
209 favored rotamers. Several side chains exhibit multiple conformations and one of the
210 nine proline residues (Pro91) adopts a *cis* conformation in the structure.

211 We next attempted to co-crystallize a binary XhpG/substrate complex by soaking.
212 This experiment was unsuccessful, a result which we hypothesized was due to the
213 limited amount of substrate available and its high instability, the low solvent content of
214 the crystals, and the inaccessibility of the active site in the crystal packing.

215 **Description of the overall XhpG structure and organization of the active site**

216 The XhpG enzyme belongs to the superfamily of α/β hydrolases which catalyze a wide
217 range of reactions and share a characteristic fold comprising a catalytic core and an α -
218 helical sub-domain.^[25] The core consists of an α/β -sheet, formed by eight β -strands
219 shielded from the solvent by four α -helices on one side and by two α -helices on the
220 other (Figure 5).^[26] The β -sheet displays the canonical topology 12435678 with strand
221 β_2 antiparallel to the others, and an overall highly twisted structure, to the extent that
222 the first (β_1) and last strands (β_8) are orthogonal.^[27] The five α -helical sub-domain
223 (α_4 – α_8), which is connected to the hydrolase core by the 8-residue β_6 - α_4 loop

224 (GAFSSRD), protrudes from the structure and acts as a 'cap' on the active site (Figure
225 5).

226 To understand how XhpG is adapted to its specific function, we compared its
227 structure to that of its closest homologue identified using the Dali server, the PcaD enol-
228 lactonase (PDB ID: 2XUA)^[28] (Figure S19), which catalyzes the conversion of the
229 monocyclic 3-oxoadipate-enol-lactone into the linear β -ketoadipate. The calculated
230 RMSD value of 1.4 Å (212 C α atoms) confirms that the two enzymes adopt an overall
231 α/β hydrolase fold. However, the XhpG hydrolase differs from PcaD in terms of the
232 position, orientation and relative length of the α -helices constituting the cap sub-domain
233 (Figure S19). The inter-domain hinge region also diverges substantially in residue
234 composition (NTAARIG in PcaD) (Figure S19). In contrast to certain members of the α/β
235 hydrolase superfamily for which a flexible loop permits the conformational adjustments
236 of the α -helical sub-domain necessary for substrate accommodation and product
237 release,^[29] this linker in both XhpG and PcaD monomers appears quite rigid with an
238 average temperature factor below 25 Å². This observation suggests that the cap sub-
239 domain of XhpG may not alternate between open and closed configurations,^[28] a
240 hypothesis consistent with the relative accessibility of the substrate channel even in the
241 closed form.

242 We next analyzed XhpG S98A in the solution phase by small-angle X-ray scattering
243 (SAXS) (Figure S20, S21 and Table S9). Initial data treatment using PRIMUS showed
244 that the enzyme is dimeric in solution (observed molecular weight: 50–58 kDa vs.
245 calculated MW: 58.7 kDa)^[30] (Table S9), a result supported by the strong agreement
246 between the theoretical curve computed for a dimeric structure (CRY SOL) and the
247 experimental data ($\chi^2 = 1.68$)^[31] (Figure S21). In the dimeric model, helices $\alpha 4$ – $\alpha 6$ of the
248 helical sub-domain mediate the association between the monomers, which is consistent
249 with the known role of this region in promoting oligomerisation within the α/β hydrolase
250 clade.^[28] At the level of tertiary fold, refining the initial model using the SREFLEX
251 program^[32] to incorporate conformational adjustment of structural elements improved
252 the fit to the experimental curve by 47% ($\chi^2 = 1.14$) (Figure S21). However, in none of
253 the improved models did we observe an open form of the enzyme. Thus, globally, these
254 results confirmed that the closed state observed for the *apo* protein in the crystal is

255 also relevant in solution, and is likely to be the only form adopted by the enzyme, at
256 least in the absence of substrate **2**.

257 In the crystal structure, the XhpG active site is located at the interface between the
258 core and cap sub-domains, and in the *apo* form of the enzyme, the 8 Å diameter
259 hydrophobic channel appears relatively accessible to the substrate. For comparison, in
260 the case of lipases, the α -helical sub-domain obstructs the entry tunnel, so movement of
261 the cap region is required to expose the active site.^[27, 33] In the XhpG structure, the
262 conserved catalytic triad consists of residues Ser98, Asp219 and His246, each of which
263 is located in a distinct non-structured region: a γ -turn (β 5- α 3 loop) known as the
264 'nucleophile elbow' with the Ser at its center, and the loops β 7- α 9 (Asp) and β 8- α 10
265 (His).^[27] The reaction mechanism undoubtedly corresponds to that established for
266 serine proteases, in which the catalytic triad is complemented by an oxyanion hole.^[34] In
267 the XhpG hydrolase, the Ser98 residue is situated within the consensus sequence
268 GFSLG. The Gly residues are located at the ends of helix α 3 and strand β 5, and impose
269 a sharp turn on this motif which allows the nucleophile to closely approach the substrate
270 (2.8 Å). Structural alignment with PcaD shows that the main chains in the region of helix
271 α 3 (residues 98 – 114) as well as the loop β 3- α 1 (residues 29 – 36) superimpose,
272 identifying the NH moieties of residues Thr32 and Leu99 as forming the oxyanion hole
273 in XhpG (Figure S22).

274 Comparison to PcaD shows furthermore that the catalytic α/β core and cap domains
275 contribute unequally to the formation of the active site, with the core predominating in
276 the case of XhpG and the helical cap region in that of PcaD (Figure S23). Moreover, the
277 substrate entry point into the hydrophobic cavity of XhpG is distinct from most
278 hydrolases including PcaD, which is an unusual observation for proteins which exhibit
279 identical folds (Figure S23). Despite these notable differences, the overall organization
280 of the active site appears quite similar, in particular with the nucleophilic Ser buried
281 unusually deeply below the protein surface (17 Å (XhpG) and 12 Å (PcaD) (Figure S23))
282 and thus poorly solvent exposed, and all catalytic residues positioned almost identically
283 ($C\alpha$ -RMSD values are in the range of 0.5 Å–0.6 Å (Figure S24)).

284

285 Although the catalytic mechanism of $\alpha\beta$ hydrolases is conserved, their substrates are
286 highly variable,^[35] and thus it was also of interest to identify the motifs contributing to
287 specificity. To address the basis for recognition of the bicyclic pyrrolizixenamide **2** by
288 XhpG, we performed *in silico* docking experiments using the CB-Dock2 program^[36]
289 (Figure 5B). This modelling revealed that the side chains of aliphatic and aromatic
290 residues Phe97, Trp139, Ile154, Phe179, Ile183, Ile221 and Leu247 bind the
291 heterocycle via hydrophobic interactions, while the functional groups of residues Asn39
292 and Ser98 establish hydrogen bonds to the carbonyl groups. A mutational study
293 revealed that Asn39 is involved in catalysis, as changing it to Asp or Ala diminished the
294 turnover, compared to the negative control. However, not to the same extent as a
295 change of the catalytic Ser-98 to Ala (Figure S25). To identify particularities of **2**
296 recognition, these results were compared to the identified binding mode of the levulinic
297 acid product analogue by PcaD^[28] (Figure S24). This comparison highlighted several
298 common recognition features, notably the overall hindered character of the active site
299 pockets, as well as the multiple binding determinants furnished by both the core and α -
300 helical cap sub-domains. In particular, the indole side chain of Trp139, like that of
301 Trp135 in PcaD, forms a steric bottleneck in the extended hydrophobic cavity, which
302 imposes geometric constraints on the complexed ligands (Figure S24). Phe159 is also
303 conserved in PcaD (Trp158), but the corresponding amino acids ensure different roles:
304 Phe159 maintains the orientation of the Phe179 side chain (Figure S24, left panel),
305 while Trp158 directly contacts the linear chain of levulinic acid (Figure S24, right panel).
306 The other residues comprising the recognition sites are overall not strongly conserved
307 between the two enzymes, which is consistent with the divergent chemical structures of
308 their substrates. A parallel could nonetheless be established between residues Asn39
309 (XhpG) and Ser34 (PcaD), as the Ser34 side chain might also make a hydrogen bond to
310 the carbonyl group of the oxoadipate enol-lactone substrate (Figure S24). Overall,
311 although the cap region and the active site of the XhpG hydrolase exhibit some
312 structural variability relative to PcaD, in both cases, the α -helical sub-domain contributes
313 to substrate selection, and could additionally mediate the release of the product from
314 the conserved catalytic α/β core.

315

316 **Conclusion**

317 In this work, we have successfully isolated a highly unusual bacterial PA pseudo-dimer,
318 pyrrolizwilline (**5**), as well as identified the BGC *xhpA–G* encoding its biosynthesis. Via
319 heterologous expression in *E. coli*, we have shown in addition that the NRPS XhpA and
320 the BVMO XhpB collaborate to form the PA core structure **2**, which is subsequently
321 elaborated to **5** by a series of reactions initiated by the *N*-acyl hydrolase XhpG. This
322 process involves a key non-enzymatic step that makes use of the intrinsic reactivity of
323 the pyrrolizidine intermediate to react with a biologically available aldehyde, which is
324 reminiscent of dibohemamine^[37] biosynthesis. However, in the biosynthesis of **5**, a
325 structurally more complex aldehyde than formaldehyde is used in the dimerization of
326 natural products,^[38] which offers further possibilities for modification. Moreover, multiple
327 reactions involving both pathway specific and general-purpose enzymes are required to
328 accomplish pyrrolizwilline biosynthesis, highlighting the complexity and diversity of
329 bacterial PA pathways.

330 The *N*-acyl hydrolase XhpG plays an essential role in initiating the reactions leading
331 to pyrrolizwilline. Elucidation of the XhpG crystal structure at high-resolution allowed us
332 to identify architectural adaptations relative to its closest structural homologue PcaD, that
333 likely underpin its substrate specificity. These notably include organizational
334 divergences in the α -helical sub-domain which caps the active site and the hinge region
335 linking it to the catalytic core, as well as differences in the location of the tunnel leading
336 into the active site and the active site shapes and sizes. The combined effect of these
337 modifications is to allow XhpG to deploy a topologically distinct set of residues for
338 substrate binding. Nonetheless, the active sites of both enzymes are largely
339 hydrophobic in character, and incorporate steric constraints to help in orienting the
340 substrates relative to a shared catalytic apparatus. This information can now be
341 leveraged to rationally alter XhpG substrate choice in order to redirect other PA
342 pathways towards alternative products.

343 To the best of our knowledge, we describe here for the first time a biosynthetic route
344 involving both non-enzymatic and enzymatic reactions leading to the cross-linking of
345 two bacterial pyrrolizidine alkaloid moieties in a natural context. PA compounds are

346 valuable starting points for drug development, and PA dimers are particularly interesting,
347 as a previous report showed that bohemamine dimers display cytotoxicity against
348 cancer cell lines, a property lacking in the monomers.^[37] Hence, identification of the
349 pyrrolizwilline biosynthetic pathway reported herein serves as a useful starting point for
350 further PA diversification approaches. Indeed, heterologous expression of this pathway
351 coupled with precursor feeding has already resulted in four new-to-nature pyrrolizwilline
352 derivatives.

353

354 **Acknowledgements**

355 Funding for this work was provided by the Max-Planck Society (to H.B.B.), a DFG grant
356 (BO 1834/10-1) to H.B.B., the Agence Nationale de la Recherche (ANR) to Y.L. and
357 K.J.W. (ANR-20-CE92-0038-02 NRPSBacAza, funded jointly with DFG grant BO
358 1834/10-1), and the Centre National de la Recherche Scientifique (CNRS) (to Y.L. and
359 K.J.W.). We thank the B2S facility (UMS2008/US40 IBSLor, Université de Lorraine) for
360 crystallization and x-ray equipment, Dr. M. Savko and Dr. W. Shepard for help with
361 diffraction data acquisition on the beamline Proxima-2A (SOLEIL, synchrotron), Dr. E.
362 Lakis for helpful discussion, and Dr. Carsten Kegler for plasmids.

363

364 **Keywords:** bacterial natural products • biosynthesis • pyrrolizidine alkaloids •
365 $\alpha\beta$ hydrolase • X-ray crystallography

366

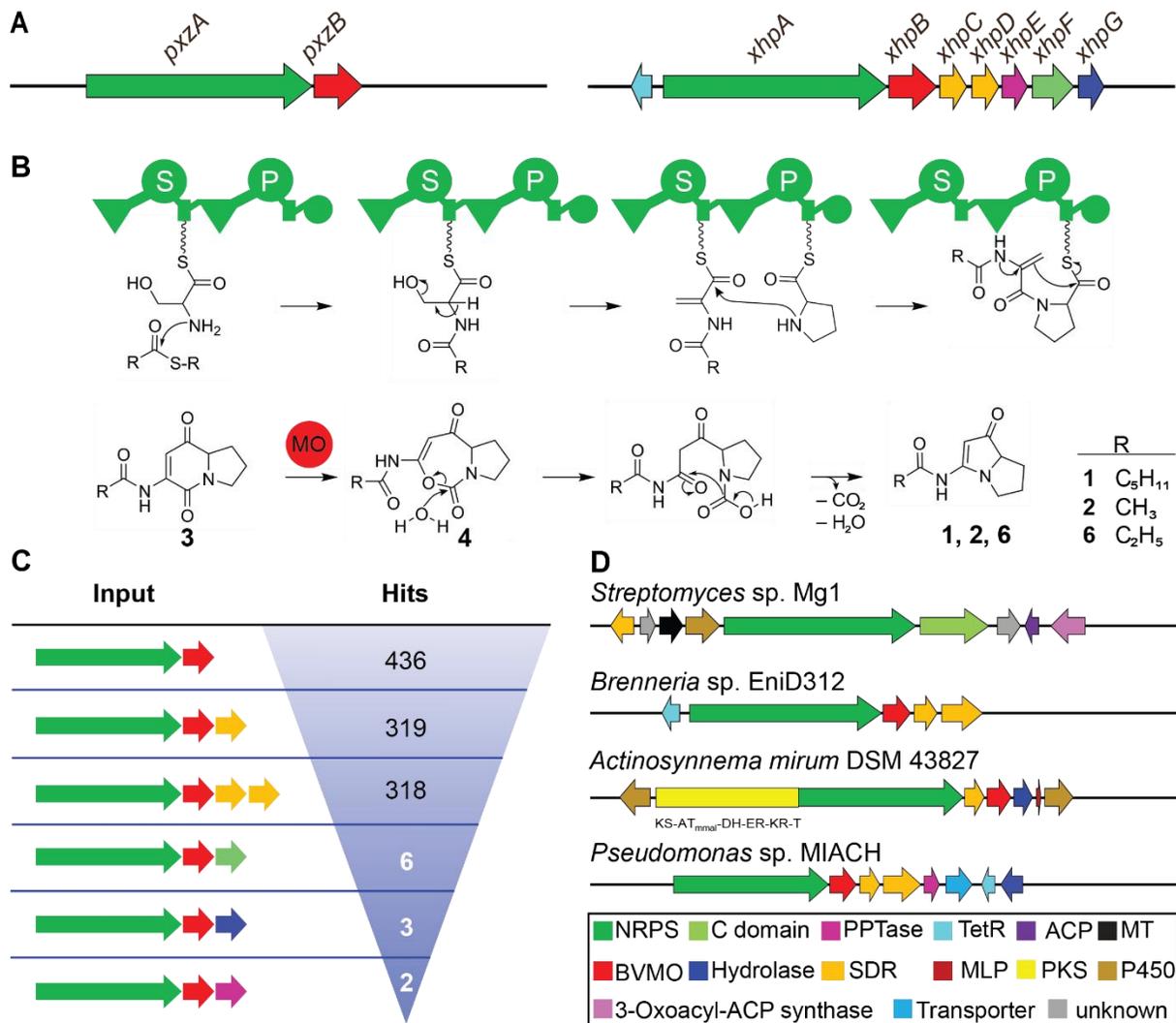
- 367 [1] R. Moreira, D. M. Pereira, P. Valentão, P. B. Andrade, *Int. J. Mol. Sci.* **2018**, *19*, 1668.
368 [2] a) E. Roeder, *Pharmazie*. **1995**, *50*, 83-98; b) M. M. Zakaria, T. Stegemann, C. Sievert, L. H. Kruse,
369 E. Kaltenecker, U. Girreser, S. S. Çiçek, M. Nimtz, D. Ober, *Plant Cell*. **2022**, *34*, 2364-2382.
370 [3] W. Li, T. Cheng, T. Jiang, M. Zhou, B. Gong, G. Zhao, J. Li, R. Tan, X. Yang, K. Joshi, Y. Peng, M.
371 Cheng, T. Liu, D. O. Wang, J. Zheng, *Chem. Biol. Interact.* **2022**, *365*, 110047.
372 [4] N. S. A. Hadi, E. E. Bankoglu, H. Stopper, *Arch. Toxicol.* **2023**, *97*, 295-306.
373 [5] W. J. M. Lewis, D. M. Shaw, J. Robertson, *Beilstein J. Org. Chem.* **2021**, *17*, 334-342.
374 [6] a) I. L. Pinto, H. F. Boyd, D. M. B. Hickey, *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2015-2017; b) Y.
375 Schmidt, M. van der Voort, M. Crüseemann, J. Piel, M. Josten, H.-G. Sahl, H. Miess, J. M.
376 Raaijmakers, H. Gross, *ChemBioChem*. **2014**, *15*, 259-266; c) C. W. Johnston, R. Zvanych, N.
377 Khyzha, N. A. Magarvey, *ChemBioChem*. **2013**, *14*, 431-435.

- 378 [7] O. Schimming, V. L. Challinor, N. J. Tobias, H. Adihou, P. Grün, L. Poschel, C. Richter, H. Schwalbe,
379 H. B. Bode, *Angew. Chem. Int. Ed.* **2015**, *54*, 12702-12705.
- 380 [8] M. A. Fischbach, C. T. Walsh, *Chem. Rev.* **2006**, *106*, 3468-3496.
- 381 [9] M. Einsiedler, K. Lamm, J. F. Ohlrogge, S. Schuler, I. J. Richter, T. Lübken, T. A. M. Gulder, *J. Am.*
382 *Chem. Soc.* **2024**.
- 383 [10] a) Q. Zhang, K. K. Schrader, H. N. ElSohly, S. Takamatsu, *J. Antibiot.* **2003**, *56*, 673-681; b) D. E.
384 Nettleton, Jr., D. M. Balitz, T. W. Doyle, W. T. Bradner, D. L. Johnson, F. A. O'Herron, R. H.
385 Schreiber, A. B. Coon, J. E. Moseley, R. W. Myllymaki, *J. Nat. Prod.* **1980**, *43*, 242-258; c) T. S.
386 Bugni, M. Woolery, C. A. Kauffman, P. R. Jensen, W. Fenical, *J. Nat. Prod.* **2006**, *69*, 1626-1628.
- 387 [11] a) L. A. Dolak, C. DeBoer, *J. Antibiot. (Tokyo)*. **1980**, *33*, 83-84; b) Y. Horiuchi, S. Kondo, T. Ikeda,
388 D. Ikeda, K. Miura, M. Hamada, T. Takeuchi, H. Umezawa, *J. Antibiot.* **1979**, *32*, 762-764.
- 389 [12] S. Huang, J. Tabudravu, S. S. Elsayed, J. Travert, D. Peace, M. H. Tong, K. Kyeremeh, S. M. Kelly, L.
390 Trembleau, R. Ebel, M. Jaspars, Y. Yu, H. Deng, *Angew. Chem. Int. Ed.* **2015**, *54*, 12697-12701.
- 391 [13] J.-F. Hu, D. Wunderlich, R. Thiericke, H.-M. Dahse, S. Grabley, X.-Z. Feng, I. Sattler, *J. Antibiot.*
392 **2003**, *56*, 747-754.
- 393 [14] Z. Hong, A. Bolard, C. Giraud, S. Prévost, G. Genta-Jouve, C. Deregnacourt, S. Häussler, K.
394 Jeannot, Y. Li, *Angew. Chem. Int. Ed.* **2019**, *58*, 3178-3182.
- 395 [15] P. Fu, J. B. MacMillan, *Org. Lett.* **2015**, *17*, 3046-3049.
- 396 [16] B. Jiang, W. Zhao, S. Li, H. Liu, L. Yu, W. Niu, H. He, L. Wu, *J. Antibiot.* **2018**, *71*, 965-967.
- 397 [17] a) B. Jiang, W. Zhao, S. Li, H. Liu, L. Yu, Y. Zhang, H. He, L. Wu, *J. Nat. Prod.* **2017**, *80*, 2825-2829;
398 b) P. Fu, A. Legako, S. La, J. B. MacMillan, *Chem. Eur. J.* **2016**, *22*, 3491-3495.
- 399 [18] P. Fu, S. La, J. B. MacMillan, *J. Nat. Prod.* **2016**, *79*, 455-462.
- 400 [19] C. L. M. Gilchrist, T. J. Booth, B. van Wersch, L. van Grieken, M. H. Medema, Y.-H. Chooi,
401 *Bioinform. Adv.* **2021**, *1*.
- 402 [20] E. Bode, A. O. Brachmann, C. Kegler, R. Simsek, C. Dauth, Q. Zhou, M. Kaiser, P. Klemmt, H. B.
403 Bode, *ChemBioChem.* **2015**, *16*, 1115-1119.
- 404 [21] B. B. Snider, J. R. Duvall, *Org. Lett.* **2004**, *6*, 1265-1268.
- 405 [22] H. L. Kornberg, *Biochem. J.* **1966**, *99*, 1-11.
- 406 [23] a) D. P. Clark, J. E. Cronan, *J. Bacteriol.* **1980**, *144*, 179-184; b) L. N. Ornston, M. K. Ornston, *J.*
407 *Bacteriol.* **1969**, *98*, 1098-1108; c) L. Baldomà, J. Aguilar, *J. Biol. Chem.* **1987**, *262*, 13991-13996.
- 408 [24] F. Jamshaid, V. V. R. Kondakal, C. D. Newman, R. Dobson, H. João, C. R. Rice, J. M. Mwansa, B.
409 Thapa, K. Hemming, *Tetrahedron.* **2020**, *76*, 131570.
- 410 [25] D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I.
411 Silman, J. Schrag, *Protein Eng. Des. Sel.* **1992**, *5*, 197-211.
- 412 [26] T. D. H. Bugg, *Bioorg. Chem.* **2004**, *32*, 367-375.
- 413 [27] P. D. Carr, D. L. Ollis, *Protein Pept. Lett.* **2009**, *16*, 1137-1148.
- 414 [28] J. Bains, L. Kaufman, B. Farnell, M. J. Boulanger, *J. Mol. Biol.* **2011**, *406*, 649-658.
- 415 [29] M. Otyepka, J. Damborský, *Protein Sci.* **2002**, *11*, 1206-1217.
- 416 [30] K. Manalastas-Cantos, P. V. Konarev, N. R. Hajizadeh, A. G. Kikhney, M. V. Petoukhov, D. S.
417 Molodenskiy, A. Panjkovich, H. D. Mertens, A. Gruzinov, C. Borges, *J. Appl. Cryst.* **2021**, *54*, 343-
418 355.
- 419 [31] D. Franke, M. Petoukhov, P. Konarev, A. Panjkovich, A. Tuukkanen, H. Mertens, A. Kikhney, N.
420 Hajizadeh, J. Franklin, C. Jeffries, *J. Appl. Cryst.* **2017**, *50*, 1212-1225.
- 421 [32] A. Panjkovich, D. I. Svergun, *Phys. Chem. Chem. Phys.* **2016**, *18*, 5707-5719.
- 422 [33] D. J. Ericsson, A. Kasrayan, P. Johansson, T. Bergfors, A. G. Sandström, J.-E. Bäckvall, S. L.
423 Mowbray, *J. Mol. Biol.* **2008**, *376*, 109-119.
- 424 [34] L. Hedstrom, *Chem. Rev.* **2002**, *102*, 4501-4524.
- 425 [35] S. Yang, Z. Qin, X. Duan, Q. Yan, Z. Jiang, *J. Lipid Res.* **2015**, *56*, 1616-1624.

- 426 [36] Y. Liu, X. Yang, J. Gan, S. Chen, Z.-X. Xiao, Y. Cao, *Nucleic Acids Res.* **2022**, *50*, W159-W164.
427 [37] P. Fu, A. Legako, S. La, J. B. MacMillan, *Chemistry*. **2016**, *22*, 3491-3495.
428 [38] Y. Fan, J. Shen, Z. Liu, K. Xia, W. Zhu, P. Fu, *Nat. Prod. Rep.* **2022**, *39*, 1305-1324.

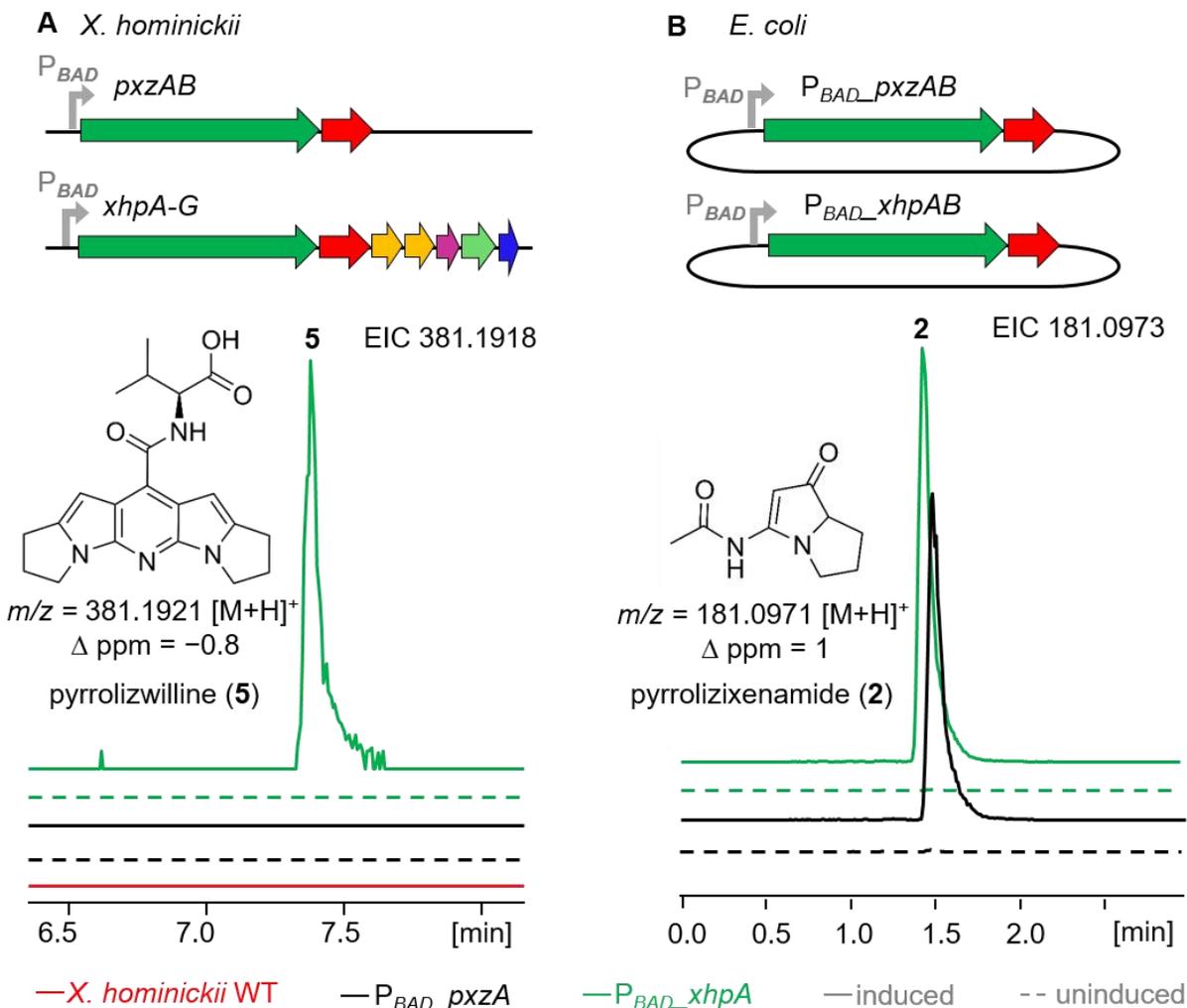
429

430



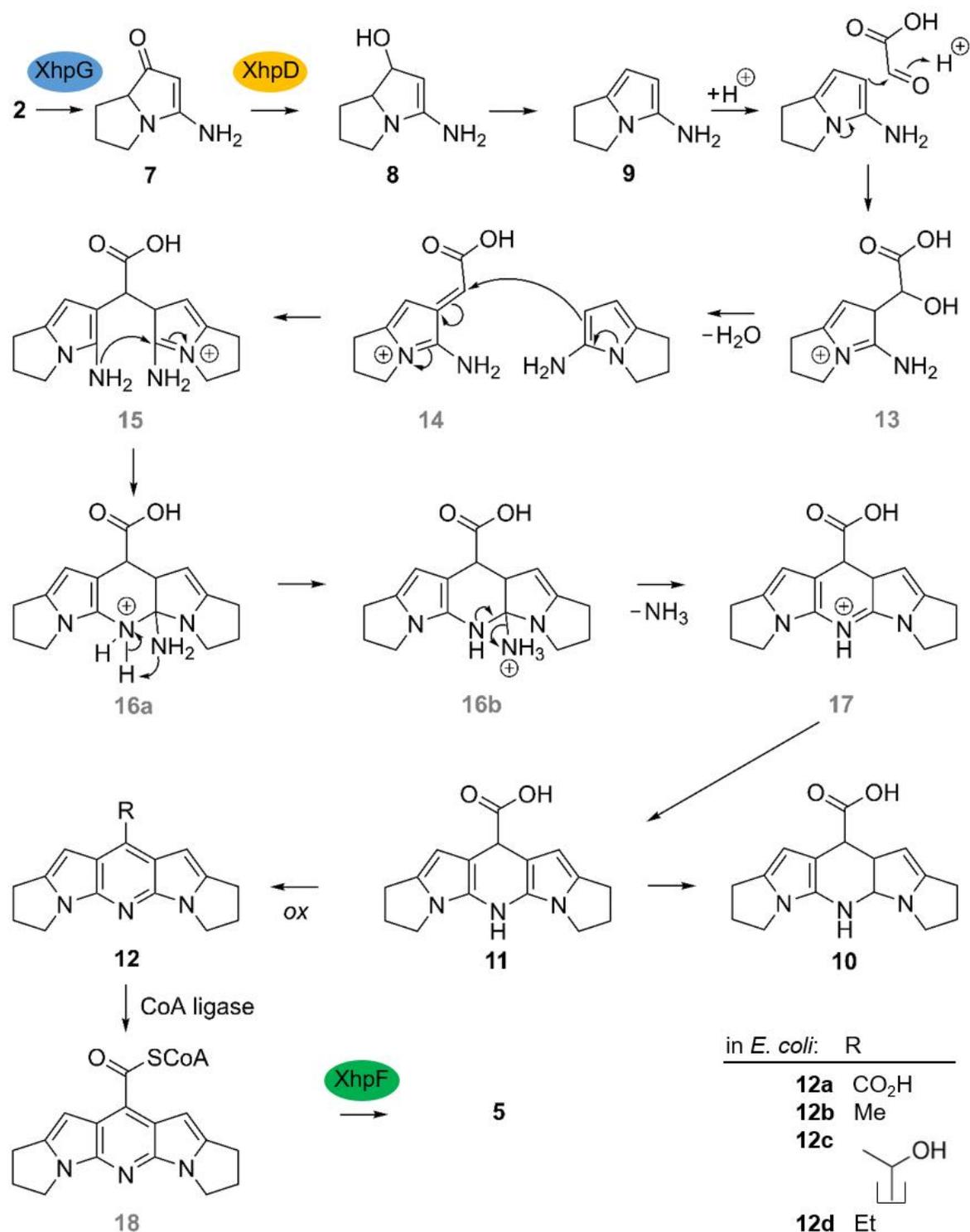
431
 432 **Figure 1.** Identification of the biosynthetic gene clusters *pxz* and *xhp* in the
 433 *Xenorhabdus hominickii* genome. **A)** Schematic overview of the *pxz* and *xhp* BGCs. **B)**
 434 Biosynthesis of pyrrolizixenamides- C_2 (**2**) by a bimodular NRPS followed by catalysis by
 435 a Baeyer-Villiger monooxygenase.^[7] **C)** Results of CAGECAT analysis^[19] using different
 436 *xhp* gene combinations as input, showing the uniqueness of the gene combinations in
 437 the *xhp* cluster. **D)** Manually curated examples of additional BGCs containing a
 438 homologue of *xhpA*. Legend: NRPS, nonribosomal peptide synthetase; BVMO, Baeyer-
 439 Villiger monooxygenase; TetR, TetR family regulator; SDR, short-chain
 440 dehydrogenase/reductase; PPTase, phosphopantetheinyl transferase; MLP, MbtH-like
 441 protein; ACP, acyl carrier protein; MT, methyltransferase; PKS, polyketide synthase.

442



443
 444 **Figure 2.** Characterization of the BGCs *xhpA-G* and *pxzAB* from *X. hominickii*. **A)**
 445 Introduction of the P_{BAD} promoter upstream of *xhpA* (green) results in production of
 446 pyrrolizwilline **5**. Structure and extracted ion chromatogram (EIC) of **5** (m/z calc'd for
 447 $C_{21}H_{24}N_4O_3 [M+H]^+ = 381.1921$; obs'd $m/z = 381.1918$; $\Delta \text{ppm} = -0.8$) are shown.
 448 **B)** Heterologous expression of *pxzAB* (black) and *xhpAB* (green) in *E. coli* results in the
 449 production of pyrrolizixenamide- C_2 **2**. Structure and EIC of **2** (m/z calc'd for $C_9H_{12}N_2O_2$
 450 $[M+H]^+ = 181.0971$; obs'd. $m/z = 181.0973$; $\Delta \text{ppm} = 1$) are shown. Dashed lines:
 451 uninduced cultures (—), solid lines: cultures induced with 0.2% arabinose (+). For color
 452 codes of genes see Figure 1.

453

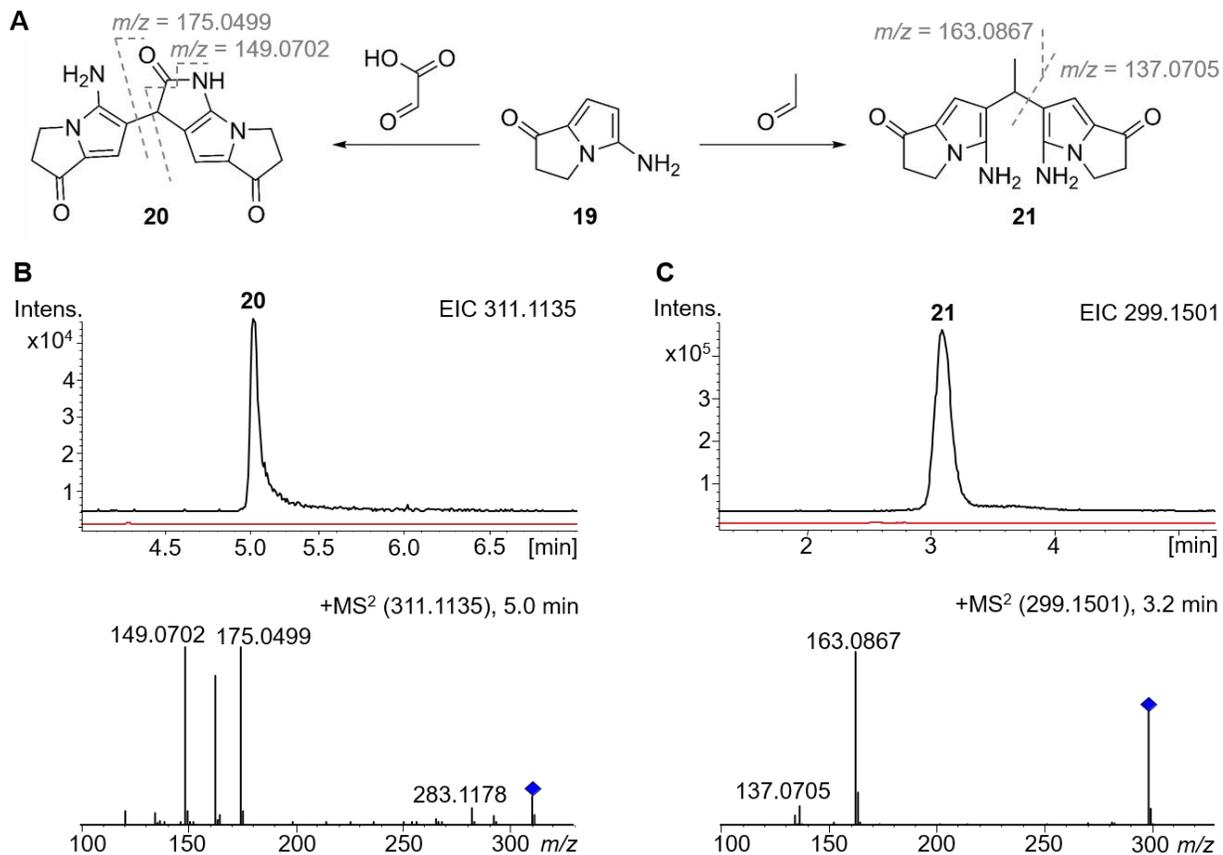


454

455 **Figure 3.** Postulated biosynthetic pathway to pyrrolizwilline. XhpA and XhpB provide **2**,
 456 which is *N*-deacylated by the hydrolase XhpG and then reduced by the reductase XhpD.
 457 Next, the pyrrolizidine moiety **9** is extended by glyoxylic acid. After dehydration followed

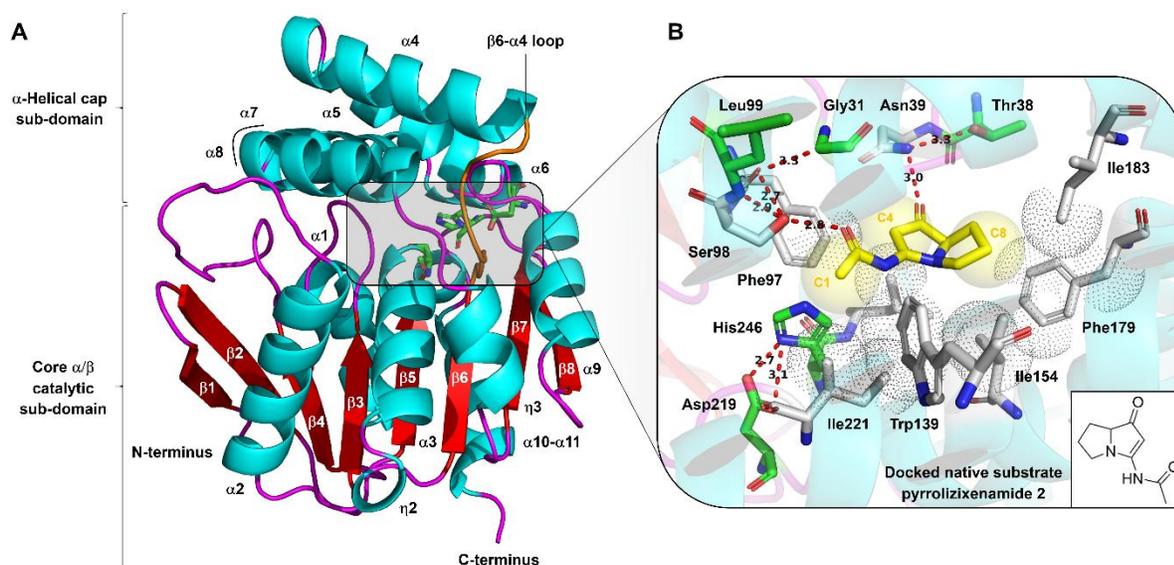
458 by a second addition of **9**, molecular rearrangement, and the loss of NH₃, **17** is formed.
459 Deprotonation of intermediate **17** subsequently leads to the formation of **11**, which can
460 be oxidized to **12** or reduced to **10**. Next, **12** is activated by a CoA ligase to give **18**.
461 XhpF catalyzes condensation of **18** with L-valine to yield **5**. Heterologous expression of
462 XhpA–G resulted in the accumulation of a series of derivatives of **12**, whereas **12d** was
463 only produced when propionaldehyde was fed to the production culture. The
464 compounds indicated with grey numbers are hypothetical, while the identities of those in
465 black are supported by HRMS analysis (see supplementary information). For color
466 codes of proteins see Figure 1.

467



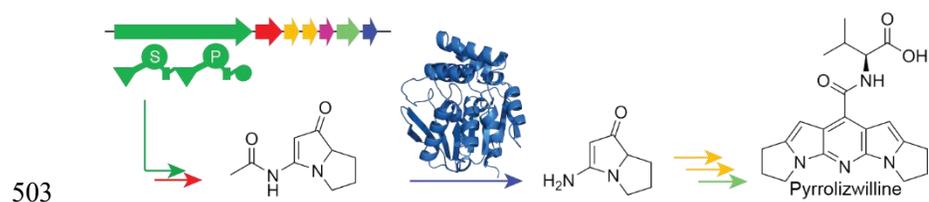
468
 469 **Figure 4.** *In vitro* cross-linking experiment with the model compound 5-amino-2,3-
 470 dihydro-1*H*-pyrrolizin-1-one **19**, glyoxylic acid and acetaldehyde in THF. **A**) Schematic
 471 overview of **19** reacting with glyoxylic acid leading to the formation of **20**, and of **19**
 472 reacting with acetaldehyde leading to the formation of **21** (Figure S14) with proposed
 473 structure for the key MS² fragmentation. **B**) A peak (black chromatogram) consistent
 474 with cross-linking of two molecules of **19** with glyoxylic acid was obtained, as judged by
 475 comparison to the reaction without glyoxylic acid (red). MS/MS analysis of **20** (m/z
 476 calc'd for C₁₆H₁₅N₄O₃ [M+H]⁺ = 311.1139; obs'd m/z = 311.1135, Δ ppm = -1.2). Three
 477 MS² fragments can be observed: m/z calc'd for C₁₅H₁₄N₄O₂ [M+H]⁺ = 283.1189; m/z
 478 calc'd for C₉H₆N₂O₂ [M+H]⁺ = 175.0502, and m/z calc'd for C₈H₈N₂O [M+H]⁺ = 149.0709.
 479 **C**) A peak (black chromatogram) consistent with cross-linking of two molecules of **19**
 480 with acetaldehyde was obtained, as judged by comparison to the reaction without
 481 acetaldehyde (red). MS/MS analysis of **21** (m/z calc'd for C₁₆H₁₈N₄O₂ [M+H]⁺ =
 482 299.1502; obs'd m/z = 299.1501, Δ ppm = -0.4). The MS² fragment m/z calc'd for

483 $C_9H_{10}N_2O$ $[M+H]^+ = 163.0866$ and the MS^2 fragment m/z calc'd for $C_7H_8N_2O$ $[M+H]^+ =$
484 137.0709 , can be observed.



485
 486 **Figure 5.** Crystal structure of the XhpG hydrolase solved at 1.18 Å resolution and
 487 computational model for substrate binding. **A)** The catalytic core of the XhpG monomer
 488 comprises eight β-strands flanked by six α-helices, and the adjacent α-helical cap sub-
 489 domain, which is connected to the core via the β6-α4 hinge loop, contains five α-helices
 490 (α4–α8) and covers the active site. The α-helices and the 3_{10} helices (indicated with the
 491 symbol η) are shown in cyan, and the β-strands in red. PDB ID: 8QZU. **B)** Binding mode
 492 of the substrate pyrrolizixenamide **2** (yellow) as predicted by molecular docking using
 493 the CB-Dock2 program.^[36] This analysis revealed a series of hydrophobic interactions
 494 (spheres) and hydrogen bonds (dotted lines) to the pyrrolizidine heterocycle and
 495 acetamide moiety, which position the substrate appropriately relative to the conserved
 496 catalytic triad (Ser98, Asp219 and His246). Residues directly involved in recognition of
 497 the docked substrate are indicated in grey (i.e. Asn39, Phe97, Ser98, Trp139, Ile154,
 498 Phe179, Ile183, Ile221 and Leu247), while those engaged in interactions with other
 499 residues are displayed in green (i.e. Gly31, Thr38, Met88, Leu99, Asp219 and His246).
 500
 501

502 **Entry of the Table of Contents**



504 The biosynthesis of pyrrolizwilline produced by XhpA–G was elucidated in *Xenorhabdus*
505 *hominickii* using a promoter exchange approach together with heterologous expression
506 of *xhpA–G* in *E. coli*. The hydrolase XhpG was identified as the key enzyme in the
507 conversion of the intermediate pyrrolizixenamides to pyrrolizwilline and therefore
508 structurally characterized.

509

.Supplementary Information

Pyrrrolizwilline, a unique bacterial alkaloid assembled by a nonribosomal peptide synthetase and non-enzymatic dimerization

Juliana Effert^{1,&}, Margaretha Westphalen^{2,&}, Andrea Calderari^{3,&}, Yi-Ming Shi^{1,4}, Isam Elamri^{1,2}, Soumaya Najah⁵, Peter Grün¹, Yanyan Li⁵, Arnaud Gruez^{3,*}, Kira J. Weissman^{3,*}, Helge B. Bode^{1,2,6,7,8,*}

1 Max Planck Institute for Terrestrial Microbiology, Department of Natural Products in Organismic Interactions, 35043, Marburg, Germany.

2 Molecular Biotechnology, Department of Biosciences, Goethe University Frankfurt, 60438, Frankfurt am Main, Germany.

3 Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France.

4 Key Laboratory of Quantitative Synthetic Biology, Center for Synthetic Biochemistry, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China.

5 Unit MCAM, UMR 7245 CNRS-Muséum National d'Histoire Naturelle (MNHN), 75005 Paris, France.

6 Center for Synthetic Microbiology (SYNMIKRO), Phillips University Marburg, 35043 Marburg, Germany.

7 Department of Chemistry, Phillips University Marburg, 35043 Marburg, Germany.

8 LOEWE Centre for Translational Biodiversity Genomics (LOEW-TBG) & Senckenberg Gesellschaft für Naturforschung, 60325, Frankfurt am Main, Germany.

* Corresponding authors: arnaud.gruez@univ-lorraine.fr; kira.weissman@univ-lorraine.fr; helge.bode@mpi-marburg.mpg.de

& These authors contributed equally to this work

33 **Supplementary Information**

34 **A. Materials and Methods**

35 **Bioinformatic and phylogenetic analysis**

36 Initial bioinformatic characterization of the *xhp* BGC was carried out using
37 antiSMASH (v3.0.5),^[1] followed later by reanalysis using updated versions (most
38 recent v5.1.0).^[2] ClusterBlast analysis from the antiSMASH 5.1.0 platform was used
39 for identification of related BGCs, in addition to manual curation of further
40 examples.^[3] As input for the CAGECAT analysis,^[4] nucleotide sequences of the
41 different genes were used in various combinations, as shown in Figure 1. Results
42 were filtered manually to include only those BGCs containing all of the input gene
43 sequences.^[4]

44 **Cultivation of strains**

45 Strains were cultivated in LB (10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl)
46 and XPP3 (see recipe below) media at 28 °C (*Xenorhabdus*) or 37 °C (*E. coli*)
47 overnight, with shaking at 200 rpm. For cultivation on LB agar plates, 1.5% (w/v)
48 agar was added to the medium. When appropriate, antibiotics kanamycin (50
49 µg/mL), chloramphenicol (34 µg/mL), spectinomycin (100 µg/mL) and/or ampicillin
50 (100 µg/mL) were added. For induction of promoter exchange strains or for
51 induction of heterologous expression in *E. coli*, 0.2% (w/v) L-arabinose was added
52 to the culture at the start of cultivation. For long-term storage of strains, 900 µL of
53 overnight culture were mixed with 600 µL of 50% (v/v) glycerol and stored at
54 –80 °C.^[5]

XPP3 (1L):

10 g	Glycerol
20 ml	M9 salt A
20 ml	M9 salt B
20 g	Proteose peptone No.3
1 g	Na-pyruvate

after autoclaving

2 ml	Vitamin solution
1 ml	Trace element solution

M9 salt A (1 L):	350 g	K ₂ HPO ₄
	100 g	KH ₂ PO ₄

M9 salt B (1 L):	29.4 g	Sodium citrate
	50 g	Ammonium sulfate (NH ₄) ₂ SO ₄
	5 g	Magnesium sulfate MgSO ₄
Vitamin solution (1 L):	10 mg	Folic acid
	6 mg	Biotin
	200 mg	<i>p</i> -Aminobenzoic acid
	1 g	Thiamine-HCl
	1.2 g	Pantothenic acid
	2.3 g	Nicotinic acid
	12 g	Pyridoxine-HCl
	20 mg	Vitamin B12
Trace element solution (1 L):	100 mg	Riboflavin
	40 mg	ZnCl ₂
	200 mg	FeCl ₃ × 6 H ₂ O
	10 mg	CuCl ₂ × 2 H ₂ O
	10 mg	MnCl ₂ × 4 H ₂ O
	10 mg	Na ₂ B ₄ O ₇ × 10 H ₂ O
	10 mg	(NH ₄) ₆ Mo ₇ O ₂₄ × 4 H ₂ O

55

56 **Genomic DNA**

57 Genomic DNA was purified using either the Gentra Puregene Yeast/Bact. Kit
 58 (Qiagen), or the Monarch® Genomic DNA Purification Kit (NEB), according to the
 59 manufacturers' protocols.

60 **Plasmid DNA**

61 For plasmid purification, either the alkaline lysis protocol modified from Birnboim
 62 and Doly 1979,^[6] or the Invisorb® Plasmid Spin Mini Two kit (Stratec Molecular
 63 GmbH) was used. For the alkaline lysis protocol, 2 mL overnight culture were
 64 harvested at 21,000 x *g* for 30 s. The cell pellet was resuspended in 250 µL buffer I
 65 (50 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), with addition of 40 µg/mL
 66 RNase A prior to use) by vortexing. Then, 250 µL buffer II (200 mM NaOH, 1% SDS)
 67 were added, and the samples were inverted ten times. After adding 250 µL buffer III
 68 (3 M KOAc (pH 5.5)), the samples were inverted again ten times and centrifuged at
 69 full speed for 5 min. The supernatant was transferred to fresh tubes containing
 70 650 µL isopropanol and the samples were inverted twenty times. The samples were
 71 centrifuged at 21,000 x *g* for 5 min, the supernatant was discarded and the pellet
 72 was washed with 500 µL 70% ethanol at 21,000 x *g* for 2 min. The supernatant was

73 discarded and the DNA pellets were air-dried prior to resuspension in 40 μL H_2O .
74 Plasmid digest with restriction endonucleases was conducted according to the
75 manufacturer's instructions.^[5]

76 **Polymerase chain reaction**

77 For amplification of DNA, polymerase chain reaction (PCR) was used. Polymerases
78 S7 Fusion polymeraseTM (Mobidiag), Q5[®] High Fidelity DNA polymerase (New
79 England Biolabs), PhusionTM High Fidelity DNA polymerase (Thermo Fisher
80 Scientific) were used according to the manufacturers' instructions. PCR was
81 conducted with Lab Cycler Gradient (Sensoquest GmbH) or peqSTAR 96X
82 Universal (VWR Peqlab) thermocyclers. Taq Colony PCR was used for verification
83 of assembled plasmids in *E. coli* or plasmid insertions in *Xenorhabdus*. PCR
84 reactions contained Taq DNA polymerase and 10 \times ThermoPolTM buffer (20 mM
85 Tris-HCl (pH 8.8), 10 mM $(\text{NH}_4)_2\text{SO}_4$, 10 mM KCl, 2 mM MgSO_4 , 0.1 % TritonR X-
86 100). As a first step, the colony to be queried was resuspended in 50 μL
87 1xThermoPolTM buffer and the suspension was incubated at 95 $^\circ\text{C}$ for 15 min in a
88 thermocycler. 1 μL of the suspension was then used for PCR (25 μL) according to
89 the guidelines for Taq DNA polymerase (Thermo Fisher Scientific). In case of PCR
90 amplification of the full length plasmid backbone, a subsequent digest was
91 performed with *DpnI* (Thermo Fisher Scientific), and the PCR product was purified
92 from 1.5% (w/v) agarose gel.^[5]

93 **Hot fusion cloning**

94 For plasmid assembly, NEBuilder[®] HiFi DNA Assembly mix (New England Biolabs)
95 was used according to the manufacturer's instructions.

96 ***E. coli* transformation**

97 *E. coli* cells were transformed via electroporation.^[7] For preparation of
98 electrocompetent cells, a LB culture was inoculated 1:100 from an overnight culture
99 and grown to an $\text{OD}_{595} = 0.6$. The following steps were conducted on ice and with
100 4 $^\circ\text{C}$ cold 10% glycerol. The cells were harvested at 1,968 $\times g$ at 4 $^\circ\text{C}$ for 15 min,
101 and then the cells were washed with 4/5th culture volume of 10% glycerol at 1,968 $\times g$
102 at 4 $^\circ\text{C}$ for 15 min. This washing step was repeated with 1/25th culture volume and
103 1/50th culture volume of 10% glycerol. Finally, the cells were resuspended in 1/500th
104 culture volume and 50 μL aliquots were stored at -80 $^\circ\text{C}$. Prior to electroporation,

105 cell aliquots were thawed on ice and approx. 50–100 ng pDNA were added. For
106 electroporation, 1 mm cuvettes and either a GenePulser Xcell™ (Bio-Rad
107 Laboratories GmbH) or a GenePulser® II (Bio-Rad Laboratories GmbH) set at
108 parameters 1250 V, 25 µF, 200 Ω, was used. Following the electroporation step, 1
109 mL LB medium was added and the cells were incubated at 37 °C with shaking for 1
110 h. Finally, the cells were harvested at 21,000 x g for 5 min, resuspended in 200 µL
111 LB, plated on LB agar plates with appropriate antibiotics, and incubated overnight.^[5]

112 **Promoter exchange**

113 Promoter exchange mutants of *Xenorhabdus* which carry the P_{BAD} in front of the
114 target gene, were generated used a previously published protocol.^[5, 8]

115 **Heat shock transformation of *Xenorhabdus***

116 A fresh LB culture was inoculated 1:100 from a *Xenorhabdus* overnight culture, and
117 grown with shaking at 30 °C to an OD₅₉₅ = 0.6–0.8. 1 mL of this
118 culture/transformation was removed, and the cells pelleted at 12,298 x g at 4 °C for
119 1 min. The cells were washed with 1 mL XTB1 buffer (100 µL 1 M Tris-HCl (pH 6.5),
120 600 µL CaCl₂, 9.3 mL H₂O, filter sterilized (stored at 4 °C)), and then centrifuged at
121 12,298 x g at 4 °C for 1 min. XTB1 was prepared fresh 30 min prior to cell harvest,
122 and the buffer and cells were kept on ice for the following steps. Next, the cells were
123 resuspended in 200 µL XTB1, and 7 µL DMSO were added. After 5 min, another 7
124 µL DMSO were added and after further 5 min incubation on ice, approx. 150 ng
125 plasmid DNA were added. The cells were then kept on ice for 30 min. For heat
126 shock, cells were incubated at 37 °C for exactly 2 min. The cells were then
127 transferred to ice for 2 min, and then 800 µL SOC medium (2% tryptone, 0.5%
128 yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM
129 glucose) were added, and the mixture incubated with shaking at 30 °C for 1 h. Next,
130 the entire culture volume was poured on an LB agar plate (50 µg/mL kanamycin),
131 and incubated for 2 days at 30 °C.^[5]

132 **Cloning and directed mutagenesis of the XhpG hydrolase gene sequence**

133 The XhpG hydrolase gene from *Xenorhabdus hominickii* DSM17903 was amplified
134 by PCR with suitable primers (Merck) and Phusion Hot Start II Polymerase (Thermo
135 Fisher Scientific) on a Mastercycler Pro thermocycler (Eppendorf). Primers were
136 designed to match the ends of the targeted regions, with a minimal melting

137 temperature (T_m) difference and a terminal G/C base, as well as to introduce *Bam*HI
138 (5'-G↓GATCC-3') and *Hind*III (5'-A↓AGCTT-3') restriction enzyme sites (Table S4).
139 Amplicons were purified by agarose gel electrophoresis and extracted from the
140 reaction mixture using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-
141 Nagel). The extraction products were digested with 5 U of restriction enzymes
142 (Thermo Fisher Scientific) in Fast Digest buffer at 37 °C for 1 h. Inserts were cloned
143 into pBG102 plasmid (Center for Structural Biology, Vanderbilt University,
144 Tennessee) which codes for an N-terminal SUMO-His₆ tag to yield pBG102:*xhpG*,
145 or pET28b (Novagen) to generate the expression plasmid pET28-XhpG.
146 Constructed plasmids were prepared using the NucleoSpin® Plasmid DNA kit
147 (Macherey-Nagel). The sequences of the inserts were confirmed by sequencing
148 (Eurofins Genomics).

149 To determine residues for site-directed mutagenesis, XhpG homologs were
150 identified using the PDB database^[9] and the Dali structural comparison server^[10]
151 and their sequences were analyzed by multiple alignment using the ClustalW
152 program.^[11] The corresponding Figure S16 was created using the ESPript
153 program.^[12] A site-specific mutation at the codon of the conserved catalytic residue
154 Ser98 was introduced by mutagenic PCR with appropriate oligonucleotides (Table
155 S4) and Phusion High-Fidelity Polymerase. The parental plasmid pBG102:*xhpG*
156 DNA was digested with 10 U of *Dpn*I restriction enzyme (Thermo Fischer Scientific)
157 at 37 °C for 1 h. Following bacterial transformation and plasmid preparation steps,
158 the mutation was confirmed by sequencing.

159 **Expression and purification of His₆-XhpG**

160 A preculture of *E. coli* BL21(DE3) harboring pET28-XhpG was grown in 10 mL LB
161 medium supplemented with kanamycin (50 µg/mL) at 37 °C with shaking overnight,
162 and subsequently used to inoculate 1 L LB with kanamycin (50 µg/mL). This culture
163 was grown at 37 °C with shaking until the OD₆₀₀ reached 0.7–1. After a cold shock
164 on ice for 30 min, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to the
165 culture at a final concentration of 0.1 mM to induce protein production. Cells were
166 allowed to grow at 16 °C overnight before being harvested. The pellets were
167 resuspended in 25 mL buffer A (50 mM Tris-HCl (pH 7.4), 500 mM NaCl, 10 mM
168 imidazole) supplemented with 0.1 mg/mL DNAase I, and then lysed by French
169 Press. After centrifugation at 21,160 x *g* for 45 min, the cleared supernatant was

170 applied to a self-packed column containing 5 mL Ni-NTA resin (Qiagen) and purified
171 by nickel-affinity chromatography by gravity. The column was pre-equilibrated with
172 buffer A. After sample loading, the column was washed with 30 mL buffer containing
173 30 mM imidazole, and His₆-tagged XhpG was eluted step-wise with 10 mL of each
174 buffer containing 100, 250 and 500 mM imidazole, respectively. After verification by
175 SDS-PAGE, correct fractions (250 mM imidazole) were combined and concentrated
176 using a Vivaspin 15R concentrator (Sartorius). The protein was then desalted using
177 a PD-10 column (Cytiva), exchanged into the storage buffer (20 mM Tris-HCl (pH
178 7.4), 300 mM NaCl and 10% glycerol) and further concentrated. Aliquots of the
179 purified protein were flash frozen in liquid N₂ and stored at -80 °C until use. The
180 overall yield of pure XhpG was 4.3 mg per liter of culture.

181 **Production, expression and purification of XhpG S98A**

182 Recombinant *E. coli* BL21(DE3) bacteria transformed with the recombinant plasmid
183 pBG102:*xhpG*/S98A were inoculated into 0.1 L of LB medium supplemented with
184 kanamycin (50 µg/mL) and the cultures were then incubated at 37 °C with agitation
185 for 18 h. Several 2 L cultures of selective LB medium were seeded with 2 mL of a
186 pre-culture and grown to an OD_{600nm} of 0.7, and production of the recombinant
187 proteins was induced by addition of 0.5 mM IPTG. Cultures were incubated at 20 °C
188 with agitation for 24 h, and then the cells were harvested by centrifugation (4000 g,
189 4 °C, 30 min). Cellular pellets were resuspended in 30 mL of lysis buffer (50 mM
190 Tris-HCl (pH 7.8), 300 mM NaCl, 20 mM imidazole, 10% glycerol) and then the
191 bacteria were lysed by 4 cycles of sonication on ice (15 kPsi, 4 °C, 2 min) (Sonifier
192 S-250D, Branson Ultrasonics). Lysates were concentrated by centrifugation (48,000
193 g, 4 °C, 30 min) and the supernatant was filtered through a sterile membrane (0.22
194 µm) using a syringe (Merck).

195 Next, the supernatant was loaded onto a HisTrap affinity chromatography column
196 (Ni²⁺-sepharose, 10 mL) (GE Healthcare) equilibrated in buffer A (50 mM Tris-HCl
197 (pH 7.8), 300 mM NaCl, 20 mM imidazole, 10% glycerol) on an ÄKTA Avant system
198 (Cytiva). The column was washed with 10 column volumes of lysis buffer at 5
199 mL/min, and then the protein eluted with buffer B containing 0.3 M imidazole at an
200 identical flow rate. The XhpG hydrolase, which was expressed with a SUMO-His₆
201 tag, was then dialyzed in the presence of human rhinovirus 3C protease (HRV-3C)
202 (10 µg/mL, 4 °C, 16 h). The sample was reinjected onto the HisTrap column, and

203 the target protein obtained by washing with buffer A at 5 mL/min. Fractions
204 containing the protein were concentrated using an Amicon Ultracel-10 concentrator
205 (Merck). The sample was then injected through a loop (10 mL) onto a Superdex-75
206 size exclusion chromatography column (130 mL) (GE Healthcare), fitted onto an
207 ÄKTA Pure machine (Cytiva), and equilibrated in purification buffer (20 mM Tris-HCl
208 (pH 7.8), 300 mM NaCl, 5% glycerol). The recombinant protein was eluted in the
209 buffer at a flow rate of 1 mL/min, concentrated using an Amicon Ultracel-10
210 concentrator to 11 mg/mL, and then flash-frozen in liquid nitrogen and stored at
211 $-80\text{ }^{\circ}\text{C}$. Final sample concentration was measured at 280 nm using a NanoDrop
212 3000C spectrophotometer (Thermo Fisher Scientific), with extinction coefficients
213 calculated using the EXPASy ProtParam tool^[13]. Protein purity was determined by
214 10% SDS-PAGE (Figure S17).

215 **Determination of XhpG activity *in vitro***

216 A total of 100 μL reaction mixture containing 50 mM Tris-HCl (pH 7.5), 2 mM
217 substrate (**2**) and 85 μM His₆-XhpG was incubated overnight at 30 $^{\circ}\text{C}$ alongside a
218 negative control under the same conditions containing boiled protein. To stop the
219 reaction, 50 μL of methanol was added to 50 μL of the mixture. Production of **7** was
220 analyzed by ultra-high-performance liquid chromatography coupled to high
221 resolution mass spectrometry (UHPLC-HRMS) on an Ultimate 3000-RSLC system
222 (Thermo Scientific) connected to an electrospray ionization-quadrupole-time-of-flight
223 (ESI-Q-TOF) instrument (Maxis II ETD, Bruker Daltonics). Separation was achieved
224 using an RSLC Polar Advantage II Acclaim column (2.2 μm , 120 \AA , 2.1 \times 100 mm;
225 Thermo Scientific) with the following gradient of mobile phase A (milliQ water +
226 0.1% formic acid) and B (acetonitrile + 0.08% formic acid), at 300 $\mu\text{L}/\text{min}$: linear
227 increase from 2–95% B for 16 min, 95% B for 1 min, decrease to 2% B for 1 min,
228 equilibration for 3 min. MS detection was carried out in positive ion mode in the
229 range m/z 50–1500. The source parameters were as follows: nebulizer gas 35 psi,
230 dry gas 8 L/min, capillary voltage 3500 V, end plate offset 500 V, temperature
231 200 $^{\circ}\text{C}$. Analyses were performed in MS only mode or in data-dependent auto-
232 MS/MS mode using collision-induced dissociation. The MS instrument was
233 calibrated for each run using a sodium formate solution consisting of 10 mM sodium
234 hydroxide in isopropanol/0.2% formic acid (1:1 v/v). The data were analyzed using
235 Data Analysis 4.4 software (Bruker Daltonics).

236 **Crystallization of recombinant XhpG S98A**

237 Prior to crystallization trials, the monodispersity of XhpG S98A was confirmed by
238 dynamic light scattering using a Zetasizer Nano-ZS (Malvern Panalytical) (Figure
239 S18). The crystallization conditions were screened using the sitting-drop vapor
240 diffusion method in MRC plates (96 wells) (Greiner Bio-One) with a Mosquito robot
241 (SPT Labtech). Initial hits were obtained with the Structure Screen 1 kit (Molecular
242 Dimensions). The optimal condition consisted of 200 mM magnesium acetate, 100
243 mM sodium cacodylate (pH 6.5), and 20% PEG-8000 (18 °C, 1–2 days of
244 incubation). Crystals were then manually reproduced using the hanging-drop vapor
245 diffusion technique in Linbro® plates (24 wells) (Crystalgen), after mixing 1 µL of
246 protein solution and 2 µL of crystallization buffer (and also at a 2:4 ratio).

247 **X-ray diffraction, construction and validation of the XhpG S98A crystal** 248 **structure**

249 High quality crystals were harvested with nylon cryoloops of suitable sizes
250 (CrystalCap SPINE HT, Hampton Research), then soaked in a cryoprotectant
251 solution (mother liquor with substrate (10 mM) and 30% ethylene glycol) for 1–10
252 min, and directly cryo-cooled in nitrogen gas flow at –180 °C. X-ray diffraction data
253 were collected at the SOLEIL synchrotron (Saint-Aubin, France) on the PROXIMA-
254 2A beamline. For this, crystals were successively mounted on a high-performance
255 κ -goniometer, then irradiated with a high-energy X-ray beam (15 keV (0.83 Å))
256 using the oscillation method over a standard angular range with an amplitude of
257 0.1°. Diffraction images were captured using a single-photon counting EIGER X 9M
258 detector (750 image.s⁻¹) (Dectris). The crystals belong to the C2 space group
259 (monoclinic system) with the unit cell parameters $a = 48.83$ Å, $b = 68.54$ Å, $c =$
260 75.48 Å, $\beta = 98.11^\circ$. Datasets were indexed and integrated using the XDS software,
261 and scaled using pointless and aimless programs (CCP4-WinCoot).^[14] Their quality
262 was assessed by relying on relevant parameters such as maximum resolution,
263 signal-to-noise ratio, completeness, multiplicity and correlation coefficient $CC_{1/2}$
264 (Table S6).

265 Initial phases were determined by molecular replacement using the MOLREP
266 program,^[15] based on a structural model generated using the AlphaFold2
267 software.^[16] The structure was constructed and refined by iterative cycles of
268 manual reconstruction in the electron density maps using the WinCoot program,^[17]

269 followed by restrained refinement at 1.18 Å using REFMAC5^[18] (Table S6). The final
270 model geometry was validated using MolProbity^[19] and the wwPDB OneDep
271 server.^[20] The structure of the XhpG hydrolase S98A mutant was deposited in the
272 PDB database under the accession code 8QZU.

273 Binary complex structures with the substrate were subsequently modelled by
274 molecular docking using the CB-Dock2 program.^[21] The corresponding figures were
275 prepared using the PyMOL program.^[22]

276 **Small-angle X-ray scattering of recombinant proteins**

277 Small-angle X-ray scattering (SAXS) data on the mutant enzyme in solution were
278 recorded at the SOLEIL synchrotron on the SWING beamline (Saint-Aubin, France).
279 The raw scattering frames were reduced through normalisation of the intensity $I(q)$
280 to an absolute scale and buffer subtraction using the FOXTROT program.^[23] The
281 intensity $I(0)$ values and radius of gyration R_g values were plotted as a SAXS profile
282 against the elution time of the enzyme. All frames which exhibited similar R_g values
283 as a function of $I(0)$ were then averaged, and then processing and analysis of the
284 averaged data carried out using PRIMUS,^[24] CRY SOL,^[25] and SREFLEX,^[26] from
285 the ATSAS 3.0 programs suite (Table S7).

286 ***In vitro* cross-linking experiment**

287 The compound 5-amino-2,3-dihydro-1*H*-pyrrolizin-1-one (purchased from WuXi
288 AppTec (Shanghai, China)) was used for *in vitro* cross-linking. For this, 0.4 mg 5-
289 amino-2,3-dihydro-1*H*-pyrrolizin-1-one were dissolved in 600 µL THF, and 100 µL of
290 1 mM glyoxylic acid or 100 µL of 1 mM acetaldehyde was added. The mixture was
291 incubated with shaking at room temperature for 72 h. After incubation, the mixture
292 was dried under vacuum and the product was dissolved in acetonitrile. The samples
293 were analyzed by LC-HRMS at a final dilution of 1:10 relative to the original reaction
294 mixture, using a protocol based on Fu *et al.*^[27]

295 **Sample preparation for LC-MS analysis**

296 Unless stated otherwise, production cultures were grown in LB medium or XPP3
297 medium and incubated with shaking at 30 °C (*Xenorhabdus*) or 22 °C (*E. coli*) for
298 72 h. Metabolite production was induced at the beginning of cultivation using 0.2%
299 L-arabinose. In certain cases, propionaldehyde (1 mM or 0.1 mM) was added after 6
300 h of incubation to the *E. coli* expression cultures. After 72 h, 100 µL of culture were

301 mixed with 900 μ L of MeOH. After incubation for 30 min with shaking at 1200 rpm,
302 the mixture was centrifuged for 20 min at 21,000 \times g. 100 μ L of supernatant were
303 then transferred into an HPLC vial and subjected to LC-MS analysis.^[5]

304 **LC-MS analysis**

305 Routine HPLC-MS analysis was conducted on a Dionex Ultimate 3000 (Thermo
306 Fisher Scientific) RP-HPLC coupled to an AmaZonX (Bruker) ESI-IT mass
307 spectrometer. 5 μ L samples were injected onto a C18 Acquity UPLC BEH column
308 (Waters) and subjected to a 16 min gradient of 5–95% ACN/0.1% formic acid in
309 H₂O/0.1 % formic acid at a flow rate of 0.4 mL/min. The mass spectrometer was set
310 to positive ionization mode with a mass range of m/z 100–1200 and cone voltage of
311 4500 V. HR data were acquired using a Bruker Impact IITM ESI-Q-OTOF
312 instrument in positive ionization mode with a mass range m/z 100–1200 and cone
313 voltage 1500 V, coupled to a Dionex Ultimate 3000 (Thermo Fisher Scientific) RP-
314 HPLC. 5 μ L sample were injected onto a C18 ACQUITY UPLC BEH column and
315 subjected to the same analytical conditions as above. The data were analyzed
316 using the Data Analysis 4.3 (Bruker) software. If not stated otherwise, HPLC-UV-MS
317 chromatograms sharing the same X-axis have the same signal intensity. Further
318 HPLC analysis was carried out using an Elute LC series 1300 (Bruker) equipped
319 with a C18 column (ACQUITY UPLC BEH C-18, 130 Å, 2.1 mm \times 50 mm, Waters).
320 Separation was conducted at a flow rate of 0.4 mL/min using acetonitrile (ACN) and
321 water containing 0.1% formic acid (v/v) using a 5–95% ACN gradient over 18 min
322 with at a flow rate of 0.4 mL/min. In this case, high-resolution mass spectrometric
323 analyses were performed using an ESI ion-trap mass spectrometer (timsTOF fleX
324 MALDI 2, Bruker). ESI-MS spectra were recorded in positive-ion-mode with the
325 mass range from m/z 100–2000, and ultraviolet (UV) signal at 190–800 nm.

326 **Purification of pyrrolizwilline (5)**

327 To purify pyrrolizwilline (5), *X. hominickii* P_{BAD_xhpA} was cultivated at 30 °C with
328 shaking for 72 h in LB (50 μ g/mL kanamycin, 0.2 % L-arabinose) containing 2%
329 Amberlite® XAD-16 resin. The resin was harvested from 4 \times 1 L culture volume by
330 filtration and extracted in three rounds with methanol under constant magnetic
331 stirring. The total of approx. 600 mL MeOH extracts was dried under reduced
332 pressure (dried crude extract yield = 3.3 g). For purification, the crude extract was
333 resuspended in 2 mL solvent (70% DMSO, 20% MeOH, 10% isopropanol).

334 Prefractionation was conducted using a 1260 Infinity II preparative HPLC (Agilent)
335 instrument equipped with a ZORBAX Eclipse XDB C18 column (80 Å, 5 µm, 9.4 ×
336 250 mm; Agilent), via a 20 min gradient 35–55 % ACN/0.1% formic acid in H₂O/
337 0.1% formic acid (0–15 min 35–55% ACN, 15–20 min 35% ACN). The
338 prefractionation yield was 13.8 mg. For further purification, we used a 1260 Infinity II
339 semi-preparative LC-MS instrument (Agilent) equipped with a ZORBAX Eclipse Plus
340 C18 column (95 Å, 5 µm, 4.6 × 250 mm; Agilent) and a 20 min gradient of 50–55%
341 ACN/0.1% formic acid in H₂O/ 0.1% formic acid (0–15 min 50–55% ACN, 15–20 min
342 50% ACN). The final yield of pure **5** was 2.3 mg, and the sample was solved in
343 DMSO-d₆ for ¹H (500 MHz, Bruker) and ¹³C (125 MHz, Bruker) NMR analysis. The
344 software MestReNova (Mestrelab Research) was used for data analysis.^[5]

345 **Chemical synthesis of compounds 2 and 7**

346 For structure confirmation and biochemical assays compounds pyrrolizixenamide **2**
347 and its deacylated analogue **7** were synthesized in a multistep synthesis. The
348 synthesis started from commercial N-benzyloxycarbonyl-L-proline via Cbz-proline
349 N-hydroxysuccinimide ester **A** and vinylogous urea **B** as described previously
350 (Figure S8).^[28]

351 The synthesis of Cbz-proline N-hydroxysuccinimide ester **A** was initiated by a
352 Steglich esterification of the commercial available N-benzyloxycarbonyl-L-proline (1
353 g, 4.4 mmol, 1 eq.) that was stirred 5 minutes under N₂ in 25 mL dry
354 dimethylformamide (DMF) at 0 °C with N-hydroxysuccinimide (NHS; 1.02 g, 8.9
355 mmol, 2.2 eq.) and catalytic quantities of 4-dimethylaminopyridine (DMAP).
356 Coupling reagent 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDCI; 0.82 g, 5.3
357 mmol, 1.2 eq.) was added to this suspension and stirred at room temperature under
358 same conditions for a further 24 h. The yellow solution was then diluted with
359 dichloromethane (DCM) and washed with concentrated NaHCO₃ solution, brine and
360 H₂O. After drying with MgSO₄, the solvent was removed *in vacuo*. To purify the
361 target molecule **A**, flash chromatography on silica gel and a stepwise DCM/MeOH
362 gradient (99:1, 49:1, 19:1) was performed. Subsequently lyophilisation afforded the
363 target precursor as white needles: Reaction yield: 0.9 g (2.51 mmol, 58.1%); TLC
364 (DCM/MeOH, 49/1): R_f: 0.79; Flash chromatography, t_R: 2.7 min.

365 t-Butyl cyanoacetate (4.8 mL, 33.9 mmol, 2 eq.) was added dropwise (very carefully)
366 to a solution of sodium hydride (NaH; 960 mg, 8.0 mmol, 60% dispersion in mineral

367 oil) in 100 mL of dry benzene under inert atmosphere at RT. The obtained slurry
368 was stirred for 2 h and treated with **A** (3.3 g, 9.6 mmol, 1 eq.) in 30 mL of dry
369 benzene. The resulting mixture was stirred for an additional 18 h at RT, quenched
370 with distilled water and extracted finally with diethylether. Subsequently, the
371 aqueous layer was acidified with 2 M HCl resulting in the formation of a white
372 precipitate, that dissolved upon addition of dichlormethane. The layers were
373 separated and the aqueous layer was extracted with DCM. The organic layers were
374 combined together and dried over MgSO₄ and purified by flash chromatography to
375 give desired Cbz-carboxylate intermediate **B**: Reaction yield (2 steps): 0.8 g (3.3
376 mmol, 35%); TLC (DCM/MeOH, 19/1; NH₂-staining with ninhydrine) R_f: 0.32; Flash
377 chromatography, t_R: 17 min.

378 The acylation of **B** was achieved by deprotonation with sodium hydride (NaH; 5.5
379 mg, 0.23 mmol, 2.2 eq. 60% dispersion in mineral oil) under N₂. Deprotonation
380 reaction proceeded until no hydrogen evolution was observed (about 15 min). Then
381 acetyl chloride (22.4 μl, 0.32 mmol, 3 eq.) was added and the resulting mixture was
382 stirred for an additional 3 h at RT under inert gas to avoid undesired hydrolysis of
383 AcCl. To remove the tert-butoxycarbonyl protecting group, a solution of acylated
384 analogue was prepared (20 mg, 0.084 mmol in 1.7 ml DCM) and cooled down to 0
385 degree. To this solution was added 0.3 ml TFA and stirred for 10 h at RT.
386 Subsequently the solvent was removed and diethylether was then added dropwise
387 until white needles were precipitated. Flash chromatography on silica gel
388 (DCM/MeOH) gave the desired compound **2**: Reaction yield (2 steps): 80 mg (0.4
389 mmol, 11%); TLC (DCM/MeOH. 9/1): R_f: 0.66; Flash chromatography (DCM/MeOH),
390 t_R: 9.8 min.

391 For the synthesis of **7**, the crude carboxylate **B** was dissolved in MeOH and stirred
392 with 10% Pd/C (680mg, 0.64 mmol) under 1 atm of H₂ at ambient temperature for
393 10 h. The mixture was filtered through Celite and concentrated to give unpurified
394 intermediate, which cyclized on standing for 24 h at RT to give crude **7**. Flash
395 chromatography on silica gel (DCM/MeOH) gave the desired compound **7**: Reaction
396 yield (3 steps): 120 mg (0.8 mmol, 15%); TLC (DCM/MeOH, 9/1; NH₂-staining with
397 ninhydrine) R_f: 0.42.; Flash chromatography, t_R: 4.5 min.

398

399

400 **B. Supplementary tables**401 **Table S1.** Calculated and measured $[M+H]^+$ ions of listed compounds with Δ ppm.

Compound	Sum formula	$[M+H]^+$ ion obs'd., <i>m/z</i>	$[M+H]^+$ ion calc'd, <i>m/z</i>	Δ ppm
1	C ₁₃ H ₂₀ N ₂ O ₂	237.1603	237.1597	2.4
2	C ₉ H ₁₂ N ₂ O ₂	181.0973	181.0971	1
3	C ₁₀ H ₁₂ N ₂ O ₃	209.0922	209.0920	0.7
5	C ₂₁ H ₂₄ N ₄ O ₃	381.1918	381.1921	-0.8
6	C ₁₀ H ₁₄ N ₂ O ₂	195.1128	195.1128	0
7	C ₇ H ₁₀ N ₂ O	139.0864	139.0865	-2
8	C ₇ H ₁₂ N ₂ O	141.1020	141.1022	-1.6
9	C ₇ H ₁₀ N ₂	123.0906	123.0916	-8.5
10	C ₁₆ H ₁₉ N ₃ O ₂	286.1548	286.1550	-0.8
11	C ₁₆ H ₁₇ N ₃ O ₂	284.1394	284.1394	0
12a	C ₁₆ H ₁₅ N ₃ O ₂	282.1231	282.1237	-2.1
12b	C ₁₆ H ₁₇ N ₃	252.1496	252.1495	0.5
12c	C ₁₇ H ₁₉ N ₃ O	282.1601	282.1601	0
12d	C ₁₇ H ₁₉ N ₃	266.1654	266.1652	0.9
20	C ₁₆ H ₁₅ N ₄ O ₃	311.1135	311.1139	-1.2
21	C ₁₆ H ₁₈ N ₄ O ₂	299.1502	299.1503	-0.3

402

403 **Table S2.** Strains used in this project.

Strain	Genotype/Description	Reference
	<i>E. coli</i> DH10B F- <i>mcrA</i> , $\Delta(mrr-hsdRMS-mcrBC)\Phi 80/lacZ\Delta M15,\Delta lacX74$, <i>recA1</i> , <i>endA1</i> , <i>araD139</i> , $\Delta(araleu)7697$ <i>galU</i> , <i>galK</i> , <i>rpsL</i> , <i>nupG</i> , λ -	Invitrogen
	<i>E. coli</i> DH10B:: <i>mtaA</i> <i>E. coli</i> DH10B with $\Delta entD::mtaA$	[29]
10010	<i>E. coli</i> DH10b <i>mtaA</i> + <i>pACYC_P_{BAD}_pxaAB</i> <i>E. coli</i> DH10B with $\Delta entD::mtaA$ + <i>pACYC_P_{BAD}_pxaAB</i>	This work
10002	<i>E. coli</i> DH10b <i>mtaA</i> + <i>pACYC_P_{BAD}_pxzAB</i> <i>E. coli</i> DH10B with $\Delta entD::mtaA$ + <i>pACYC_P_{BAD}_pxzAB</i>	This work
10001	<i>E. coli</i> DH10b <i>mtaA</i> + <i>pACYC_P_{BAD}_pxzA</i> <i>E. coli</i> DH10B with $\Delta entD::mtaA$ + <i>pACYC_P_{BAD}_pxzA</i>	This work
10005	<i>E. coli</i> DH10b <i>mtaA</i> + <i>pACYC_P_{BAD}_xhpAB</i> <i>E. coli</i> DH10B with $\Delta entD::mtaA$ + <i>pACYC_P_{BAD}_xhpAB</i>	This work
10004	<i>E. coli</i> DH10b <i>mtaA</i> + <i>pACYC_P_{BAD}_xhpA</i> <i>E. coli</i> DH10B with $\Delta entD::mtaA$ + <i>pACYC_P_{BAD}_xhpA</i>	This work
10007	<i>E. coli</i> DH10b <i>mtaA</i> + <i>pCOLA_P_{BAD}_xhpC-G</i> <i>E. coli</i> DH10B with $\Delta entD::mtaA$ + <i>pCOLA_P_{BAD}_xhpC-G</i>	This work
10022	<i>E. coli</i> DH10b <i>mtaA</i> + <i>pCOLA_P_{BAD}_xhpG</i> <i>E. coli</i> DH10B with $\Delta entD::mtaA$ + <i>pCOLA_P_{BAD}_xhpG</i>	This work
10261	<i>E. coli</i> DH10b <i>mtaA</i> + <i>pACYC_P_{BAD}_xhpABC</i> <i>E. coli</i> DH10B with $\Delta entD::mtaA$ + <i>pACYC_P_{BAD}_xhpABC</i>	This work
10262	<i>E. coli</i> DH10b <i>mtaA</i> + <i>pACYC_P_{BAD}_xhpABD</i> <i>E. coli</i> DH10B with $\Delta entD::mtaA$ + <i>pACYC_P_{BAD}_xhpABD</i>	This work
10263	<i>E. coli</i> DH10b <i>mtaA</i> + <i>pACYC_P_{BAD}_xhpABCD</i> <i>E. coli</i> DH10B with $\Delta entD::mtaA$ w + <i>pACYC_P_{BAD}_xhpABCD</i>	This work
MW272	<i>E. coli</i> DH10b <i>mtaA</i> + <i>pCOLA_P_{BAD}_xhpA</i> <i>E. coli</i> DH10B:: <i>mtaA</i> + <i>pCOLA_P_{BAD}_xhpA</i>	This work
MW277	<i>E. coli</i> DH10b <i>mtaA</i> + <i>pCOLA_P_{BAD}_xhpAB</i> <i>E. coli</i> DH10B:: <i>mtaA</i> + <i>pCOLA_ara_xhpAB</i>	This work

	<i>E. coli</i> S17-1 λpir	Tpr Smr <i>recA</i> , <i>thi</i> , <i>pro</i> , <i>hsdR</i> -M+RP4: 2-Tc:Mu:Km Tn7 λpir	[30]
MW234	<i>E. coli</i> S17-1 λpir + pCEP_Kan_ <i>pxaA</i>	<i>E. coli</i> S17-1 λpir + pCEP_Kan_ <i>pxaA</i>	This work
	<i>E. coli</i> ST18	<i>E. coli</i> S17-1 λpir Δ <i>hemA</i>	[7]
MW199	<i>E. coli</i> ST18 + pCEP_kan_ <i>xhpA</i>	<i>E. coli</i> ST18 + pCEP_kan_ <i>xhpA</i>	This work
	<i>X. hominickii</i> WT	<i>X. hominickii</i> DSM 17903 WT	[31]
MW208	<i>X. hominickii</i> P _{BAD} <i>xhpA</i>	pCEP_kan_ <i>xhpA</i> is integrated, P _{BAD} <i>xhpA</i>	This work
MW368	<i>X. hominickii</i> + pCOLA_ara_ <i>xhom00317</i>	Contains plasmid pCOLA_ara_ <i>xhom00317</i>	This work

404

405

406 **Table S3.** Plasmids used in this project.

Plasmid	Description	Reference
pCOLA_ara_tacl	Cola <i>oriR</i> , Km ^R	[32]
pEB17	pDS132 based, R6K <i>ori</i> , <i>oriT</i> , Km ^R , <i>cipB</i> derivative with additional <i>Bgl</i> II site, <i>sacB</i>	[33]
pCEP_kan	pDS132 based, R6K <i>ori</i> , <i>oriT</i> , Km ^R , <i>araC</i> , P _{BAD}	[8]
pCEP_kan_xhpA	pCEP_kan with first 721 bp of <i>xhpA</i>	This work
pCOLA_P _{BAD} _xhpA	pCOLA_ara_tacl + <i>xhpA</i> inserted under P _{BAD} control	This work
pCOLA_P _{BAD} _xhpAB	pCOLA_ara_tacl + <i>xhpAB</i> inserted under P _{BAD} control	This work
pCOLA_P _{BAD} _xhom00317	pCOLA_ara_tacl + <i>xhom00317</i> inserted under P _{BAD} control	This work
pACYC_P _{BAD} _pxaAB	<i>ori</i> p15A, cm ^R , <i>araC</i> , P _{BAD} and <i>tacl</i> , I-SceI, I-CeuI, <i>pxaAB</i> inserted under P _{BAD} control	This work
pACYC_P _{BAD} _pxzAB	<i>ori</i> p15A, cm ^R , <i>araC</i> , P _{BAD} and <i>tacl</i> , I-SceI, I-CeuI, <i>pxzAB</i> inserted under P _{BAD} control	This work
pACYC_P _{BAD} _pxzA	<i>ori</i> p15A, cm ^R , <i>araC</i> , P _{BAD} and <i>tacl</i> , I-SceI, I-CeuI, <i>pxzA</i> inserted under P _{BAD} control	This work
pACYC_P _{BAD} _xhpAB	<i>ori</i> p15A, cm ^R , <i>araC</i> , P _{BAD} and <i>tacl</i> , I-SceI, I-CeuI, <i>xhpAB</i> inserted under P _{BAD} control	This work
pACYC_P _{BAD} _xhpA	<i>ori</i> p15A, cm ^R , <i>araC</i> , P _{BAD} and <i>tacl</i> , I-SceI, I-CeuI, <i>xhpA</i> inserted under P _{BAD} control	This work
pCK_0431 pACYC	<i>ori</i> p15A, cm ^R , <i>araC</i> , P _{BAD} and <i>tacl</i> , I-SceI, I-CeuI,	This work
pCK_0433 pCOLA	<i>ori</i> ColA, kan ^R , <i>araC</i> , P _{BAD} , <i>tacl</i> , I-SceI, I-CeuI, minus <i>Bsal</i>	This work
pCOLA_P _{BAD} _xhpC-G	<i>ori</i> ColA, kan ^R , <i>araC</i> , P _{BAD} , <i>tacl</i> , I-SceI, I-CeuI, minus <i>Bsal</i> , <i>xhpC-G</i> inserted under P _{BAD}	This work
pCOLA_P _{BAD} _xhpG	<i>ori</i> ColA, kan ^R , <i>araC</i> , P _{BAD} , <i>tacl</i> , I-SceI, I-CeuI, minus <i>Bsal</i> , <i>xhpG</i> inserted under P _{BAD}	This work
pACYC_P _{BAD} _xhpABC	<i>ori</i> ColA, kan ^R , <i>araC</i> , P _{BAD} , <i>tacl</i> , I-SceI, I-CeuI, minus <i>Bsal</i> , <i>xhpABC</i> inserted under P _{BAD}	This work
pACYC_P _{BAD} _xhpABD	<i>ori</i> ColA, kan ^R , <i>araC</i> , P _{BAD} , <i>tacl</i> , I-SceI, I-CeuI, minus <i>Bsal</i> , <i>xhpABD</i> inserted under P _{BAD}	This work
pACYC_P _{BAD} _xhpABCD	<i>ori</i> ColA, kan ^R , <i>araC</i> , P _{BAD} , <i>tacl</i> , I-SceI, I-CeuI, minus <i>Bsal</i> , <i>xhpABCD</i> inserted under P _{BAD}	This work

407

408 **Table S4.** Oligonucleotides used in this project. Restriction sites are underlined.

Plasmid/purpose	Oligonucleotide	Sequence 5' → 3'	PCR template
pCEP_Kan_xhpA	pCEP_bb_fw	ATGTGCATGCTCGAGCTC	pCEP_Kan
	pCEP_rv	ATGCTAGCCTCCTGTTAGC	
	MW12	TTTGGGCTAACAGGAGGCTAGCAT_ATGAATATTGCTACTGAA CTTGCATTTTC	<i>X. hominickii</i> gDNA
	MW8	TCTGCAGAGCTCGAGCATGCA_ACCCTCTATCTATTCCGCGC	
Verification pCEP_Kan_xhpA insertion	V_pCEP_fw	GCTATGCCATAGCATT TTTTATCCATAAG	
	MW9	GAGAAGTTGTTGCTCGCGAG	
pCEP_kan_pxaA (<i>X. hominickii</i>)	pCEP_bb_fw	ATGTGCATGCTCGAGCTC	pCEP_Kan
	pCEP_rv	ATGCTAGCCTCCTGTTAGC	
	MW11	TTTGGGCTAACAGGAGGCTAGCAT_ATGAAAAAATCACAAATT GTACCTCTTAC	<i>X. hominickii</i> gDNA
	MW5	TCTGCAGAGCTCGAGCATGCA_ATACGACGGTTCAGCAGAG G	
Verification pCEP_Kan_pxaA insertion	V_pCEP_fw	GCTATGCCATAGCATT TTTTATCCATAAG	
	MW6	GTGACATCGAACCATTGACG	
pCK_0431	ck0922	TAACTATAACGGTCCTAAGGTAGCGAACCTCAGGCATTTGAG AAGCACAC	pCK_0401
	ck0953	AATTACCCTGTTATCCCTACCCAATACGCAAACCGCCTC	pCK_0401 ^[34]
	ck0951	GTAGGGATAACAGGGTAATTATGACAACCTTGACGGCTACATC ATTCAC	
	ck0922-rev	CGCTACCTTAGGACCGTTATAGTTAGCAAAAAACCCCTCAAG ACCC	
pCK_0433	ck0952	CTAGGGATAACAGGGTAATTATGACAACCTTGACGGCTACATC ATTCAC	pCK_0406 ^[35]
	ck0922rev	AATTACCCTGTTATCCCTAGTCTAGAGTGATGGTGTCTGGGAA	

		TC	
	ck0960	CGGGTCTTGAGGGGTTTTTTGCTAACTATAACGGTCCTAAGG TAGCGAAG	pCK_0406
	ck0955	AATTACCCTGTTATCCCTAGTCTAGAGTGATGGTGTCTGGGAA TC	
pCOLA_P _{BAD} _xhpA	MW60	CATGGAATTCCTCCTGTTAGC	pCOLA_ara_tacl
	MW61	TGCAGGAGCTGTTGACAATTAATC	
	MW57	GGCTAACAGGAGGAATTCC_ATGAATATTGCTACTGAACTTG CATTTC	<i>X. hominickii</i> gDNA
	MW58	ATTGTCAACAGCTCCTGCAG_TTACTCATTTTCCAGTACCTCA C	
pCOLA_P _{BAD} _xhpAB	MW60	CATGGAATTCCTCCTGTTAGC	pCOLA_ara_tacl
	MW61	TGCAGGAGCTGTTGACAATTAATC	
	MW57	GGCTAACAGGAGGAATTCC_ATGAATATTGCTACTGAACTTG CATTTC	<i>X. hominickii</i> gDNA
	MW59	ATTGTCAACAGCTCCTGCAG_TTAATAAATATTTTTTCGTATTAT AATTATCCTGAAC	
pACYC_P _{BAD} _pxaAB	JE_G_for3.1	GTTTTTTTGGGCTAACAGGAGGAATTCATGAAAACCTTCACAA TTAGTACCTCTTACC	<i>X. stockiae</i> gDNA
	JE_G_rev6.1	GCAGCAGCCTAGGTTAATTAATTATAGGGTCTGGAGAGTAAG ATCTC	
pACYC_P _{BAD} _pxzAB	JE_G_for1.1	GTTTTTTTGGGCTAACAGGAGGAATTCATGAAAAAATCACAA ATTGTACCTCTTACTC	<i>X. hominickii</i> gDNA
	JE_G_rev2.1	GCAGCAGCCTAGGTTAATTAACCTCCTGAAAGAACTTAG AGATG	
pACYC_P _{BAD} _pxzA	JE_G_for1.1	GTTTTTTTGGGCTAACAGGAGGAATTCATGAAAAAATCACAA ATTGTACCTCTTACTC	<i>X. hominickii</i> gDNA
	JE_G_rev1.1	GCAGCAGCCTAGGTTAATTAATCATATTTTCTCCGAAGCCAA CAATAC	
pACYC_P _{BAD} _xhpAB	JE_G_for2.1	GTTTTTTTGGGCTAACAGGAGGAATTCATGAATATTGCTACT	<i>X. hominickii</i> gDNA

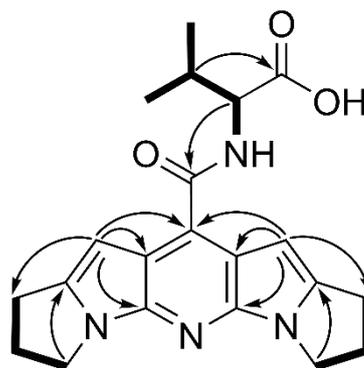
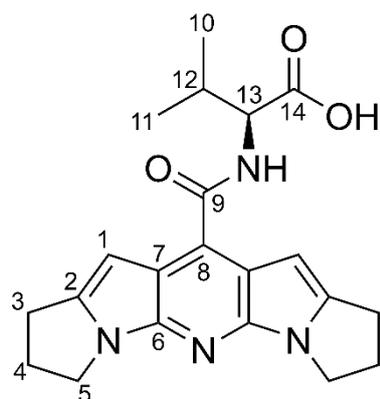
		GAACTTGCATTTTCG	
	JE_G_rev4.1	TTAATAAATATTTTTTCGTATTATAATTATCCTGAACAATAATATC	
pACYC_P _{BAD} _xhpA	JE_G_for2.1	GTTTTTTTGGGCTAACAGGAGGAATTCCATGAATATTGCTACT GAACTTGCATTTTCG	<i>X. hominickii</i> gDNA
	JE_G_rev3.1	GCAGCAGCCTAGGTTAATTAATTACTCATTTTCCAGTACCTCA CTTG	
pCOLA_P _{BAD} _xhpC-G	JE_G_for7	GTTTTTTTGGGCTAACAGGAGGAATTCCATGAATAAAAAAAGT ATTTTGATTACTGGGGC	<i>X. hominickii</i> gDNA
	JE_G_rev8	GCAGCAGCCTAGGTTAATTAATTAAGGCACATCCTCCGACAA AAG	
pCOLA_P _{BAD} _xhpG	JE_G_for_xhpG	GTTTTTTTGGGCTAACAGGAGGAATTCCATGAATAAAGTTAAA GTTGGCGATGCTC	<i>X. hominickii</i> gDNA
	JE_G_rev8	GCAGCAGCCTAGGTTAATTAATCATATTTCTCCGAAGCCAA CAATAC	
pACYC_P _{BAD} _xhpABC	JE_G_for2.1	GTTTTTTTGGGCTAACAGGAGGAATTCCATGAATATTGCTACT GAACTTGCATTTTCG	<i>X. hominickii</i> gDNA
	je0222	GTGGCAGCAGCCTAGGTTAATTAACTAACTTTCTCGAATTCT TCTGCC	
pACYC_P _{BAD} _xhpABC D	JE_G_for2.1	GTTTTTTTGGGCTAACAGGAGGAATTCCATGAATATTGCTACT GAACTTGCATTTTCG	<i>X. hominickii</i> gDNA
	je0225	GGTGGCAGCAGCCTAGGTTAATTAACTAAATTTTCTTGAATTC AGACTCC	
pACYC_P _{BAD} _xhpABD	JE_G_for2.1	GTTTTTTTGGGCTAACAGGAGGAATTCCATGAATATTGCTACT GAACTTGCATTTTCG	<i>X. hominickii</i> gDNA
	je0223	TTACCAATATACTTTTTCTTGTTCATGGTATCACCTTGTTATTTA ATAAATATTTTTTC	
	je0224	ATTTATTAATAACAAGGTGATACCATGACAAGAAAAAGTATA TTGGTAACTG	
	je0225	GGTGGCAGCAGCCTAGGTTAATTAACTAAATTTTCTTGAATTC AGACTCC	

pET28-XhpG	XhpG_NdeI	GGAATTCC <u>CATATGAATAAAGTTAAAGTTGGCGA</u>	<i>X. hominickii</i> gDNA
	XhpG_HindIII	CCCAAGCTTTTAAGGCACATCCTCCGACAA	
pBG102: <i>xhpG</i>	XhpG-F	TTTCC <u>GGATCC</u> ATGAATAAAGTTAAAGTTGGCG	<i>X. hominickii</i> gDNA
	XhpG-R	TTTCCAAGCTTTTAAGGCACATCCTCCGAC	
pBG102: <i>xhpG</i> /S98A	XhpG-S98A-F	TGTTGGTTTCGCACTTGGATCAGCGGTTGTGATAG	pBG102: <i>xhpG</i>
	XhpG-S98A-R	CTGATCCAAGTGCGAAACCAACAAGATCAAAAGG	

410 **Table S5.** ^1H and ^{13}C NMR data assignments for **5** in DMSO- d_6 and 2D NMR
 411 correlations within **5**.

412

No.	^{13}C	^1H
413 1	89.8, CH	6.34 (s)
414 2	143.7, C	-
415 3	24.1, CH_2	3.02 (t, 7.3)
416 4	27.2, CH_2	2.57 (m)
417 5	42.9, CH_2	4.13 (t, 6.9)
418 6	141.4, C	-
419 7	118.1, C	-
420 8	124.6, C	-
421 9	166.5, C	-
422 10	18.3, CH_3	0.95 (d, 6.7)
423 11	19.3, CH_3	0.99 (d, 6.7)
424 12	30.0, CH	2.22 (m)
425 13	57.9, CH	4.41 (br s)
13-NH	-	7.91 (br s)
14	173.0, C	-



— ^1H - ^1H COSY
 → HBMC

424

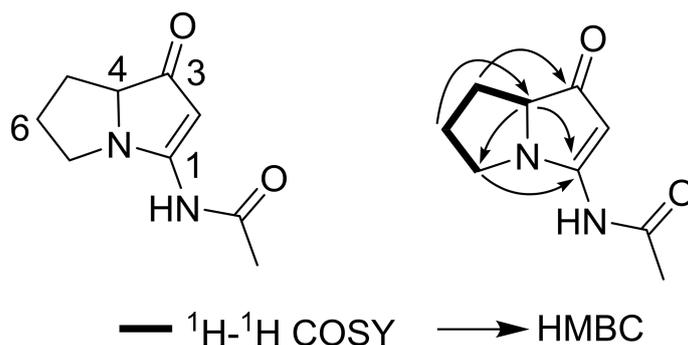
425

426 **Table S6.** ^1H (500 MHz) and ^{13}C (125 MHz) NMR spectroscopic data for compound
427 **2** in $\text{MeOH-}d_4$ (δ in ppm and J in Hz) and 2D NMR correlations of compound **2**.

428

no.	2	
	δ_{H} (mult., J)	δ_{C}
1		169.9
2	5.68 (s)	92.2
3		203.4
4	3.93 (dd, 9.0, 7.7)	69.0
5	1.53 (m)	26.1
	2.19 (m)	
6	2.15 (m)	27.5
7	3.22 (dt, 11.1, 7.7)	47.5
	3.43 (m)	
COCH_3	2.21 (s)	22.5
COCH_3		171.5

429



430

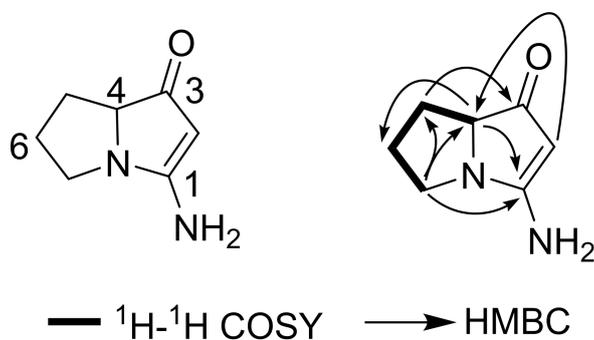
431

432 **Table S7.** ^1H (500 MHz) and ^{13}C (125 MHz) NMR spectroscopic data for compound
433 **7** in $\text{DMSO-}d_6$ (δ in ppm and J in Hz) and 2D correlations of compound **7**.

434

no.	7	
	δ_{H} (mult., J)	δ_{C}
1		177.2
2	4.15 (s)	83.1
3		195.7
4	3.61 (dd, 10.7, 5.1)	70.2
5	1.33 (m)	27.0
6	1.89 (m)	27.5
7	3.09 (m)	47.6

435



436

437

438 **Table S8.** Recording, processing and refinement of the crystallographic data for the
 439 XhpG S98A hydrolase.

440

Recording and processing data	XhpG S98A
Space group	C2
Cell dimensions	
a, b, c [Å]	48.83, 68.54, 75.48
α , β , γ [°]	90.00, 98.11, 90.00
Wavelength [Å]	0.826561
Resolution [Å]	39.50–1.18 (1.21–1.18)
<i>R</i> merge [%]	4.1 (148.0)
<i>R</i> pim [%]	1.7 (62.5)
<i>I</i> / σ <i>I</i>	17.9 (1.2)
Number of total observations	541379
Number of unique reflections	78423
Completeness [%]	98.0 (92.7)
Multiplicity	6.9 (6.4)
Wilson <i>B</i> -factors [Å ²]	21.248
Correlation coefficient <i>CC</i> _{1/2}	99.9 (67.4)
Refinement data	
Resolution [Å]	20.00–1.18
Number of reflections	74471
<i>R</i> , <i>R</i> _{free}	0.1663, 0.1963
Number of atoms	2286
Protein	2051
Other ligands/ions	4
Water	231
Average <i>B</i> -factors, all atoms [Å ²]	19.582
R.m.s deviations	
Bond lengths [Å]	0.0123
Bond angles [°]	1.7371
PDB ID	8QZU

441

442

443 **Table S9.** Recording and processing of the SAXS data acquired on the XhpG S98A
444 hydrolase.

445

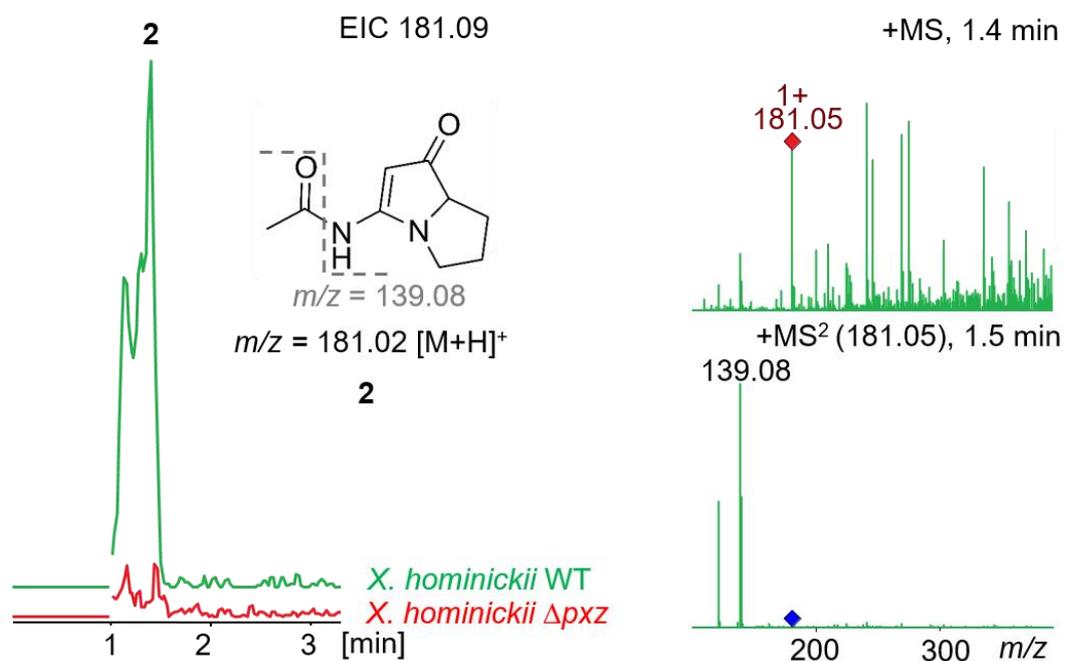
Recording and processing data	XhpG S98A
Guinier curve analysis	
$I(0)$ [cm^{-1}]	$0.042 \pm 6.7 \times 10^{-5}$
R_g [\AA]	26.94 ± 0.07
q_{min} [\AA^{-1}]	0.014
$q \cdot R_g$ max	1.18
Correlation coefficient R^2	0.92
<hr/>	
$P(r)$ function analysis	
q range [\AA^{-1}]	0.01–0.30
$I(0)$ [cm^{-1}]	0.042
R_g [\AA]	27.28
D_{max} [\AA]	87.92
Correlation coefficient R^2	0.91
<hr/>	
Molecular weight analysis [kDa]	
Porod invariant (Q_p)	50.96
MoW	55.91
Correlation volume (V_c)	48.64
Bayesian inference	53.15
Absolute scale	58.16

446

447

448

449 C. Supplementary Figures

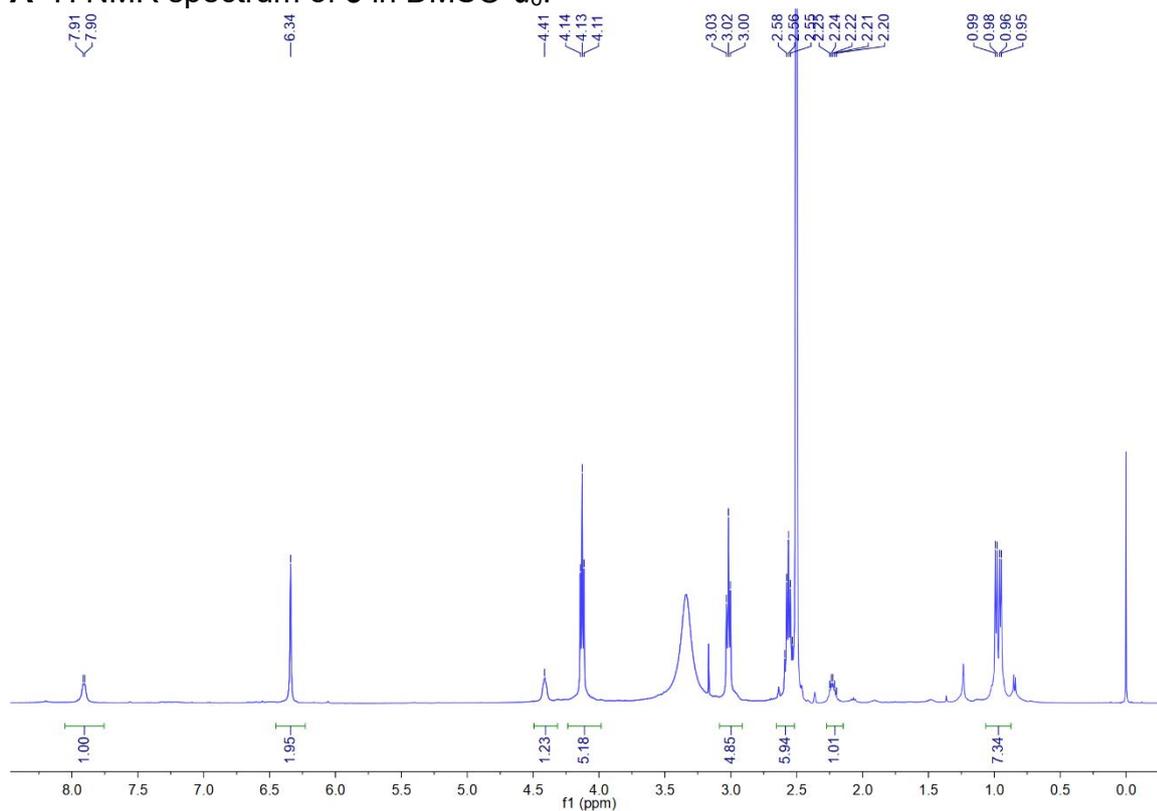


450

451 **Figure S1.** LC-MS/MS analysis for production of pyrrolizixenamide **2** (calc'd $m/z =$
452 181.09 [M+H⁺]) by *X. hominickii* wild type and Δpxz deletion mutant. Inset: proposed
453 structure for **2** indicating the major fragment ion ($m/z = 139.08$ [M+H⁺]).

454

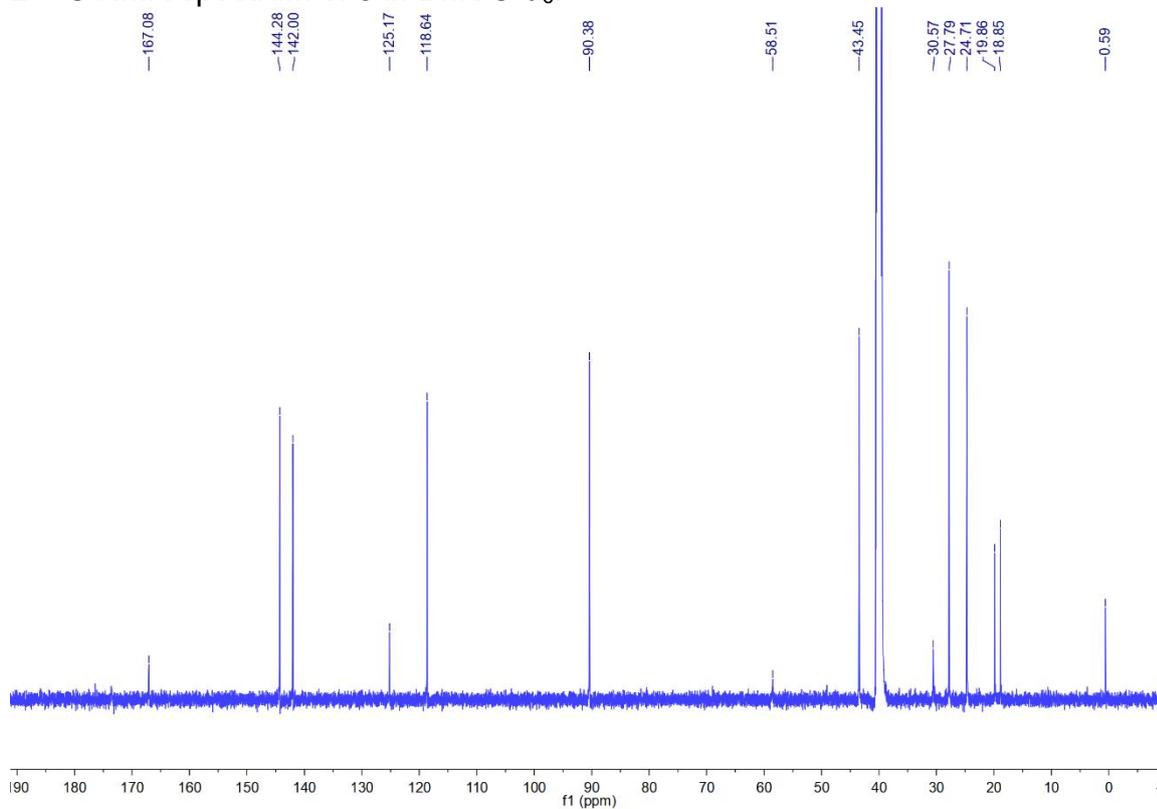
455 **A** ^1H NMR spectrum of **5** in $\text{DMSO-}d_6$.



456

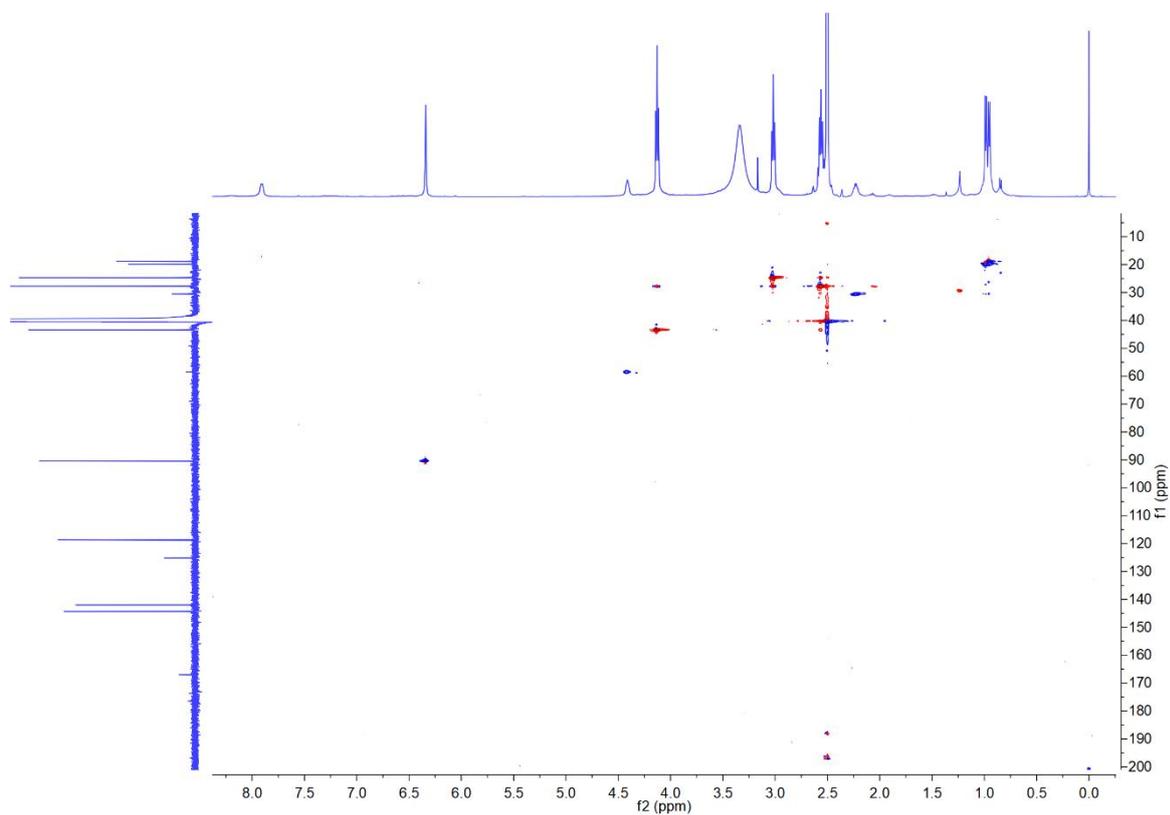
457

458 **B** ^{13}C NMR spectrum of **5** in $\text{DMSO-}d_6$.



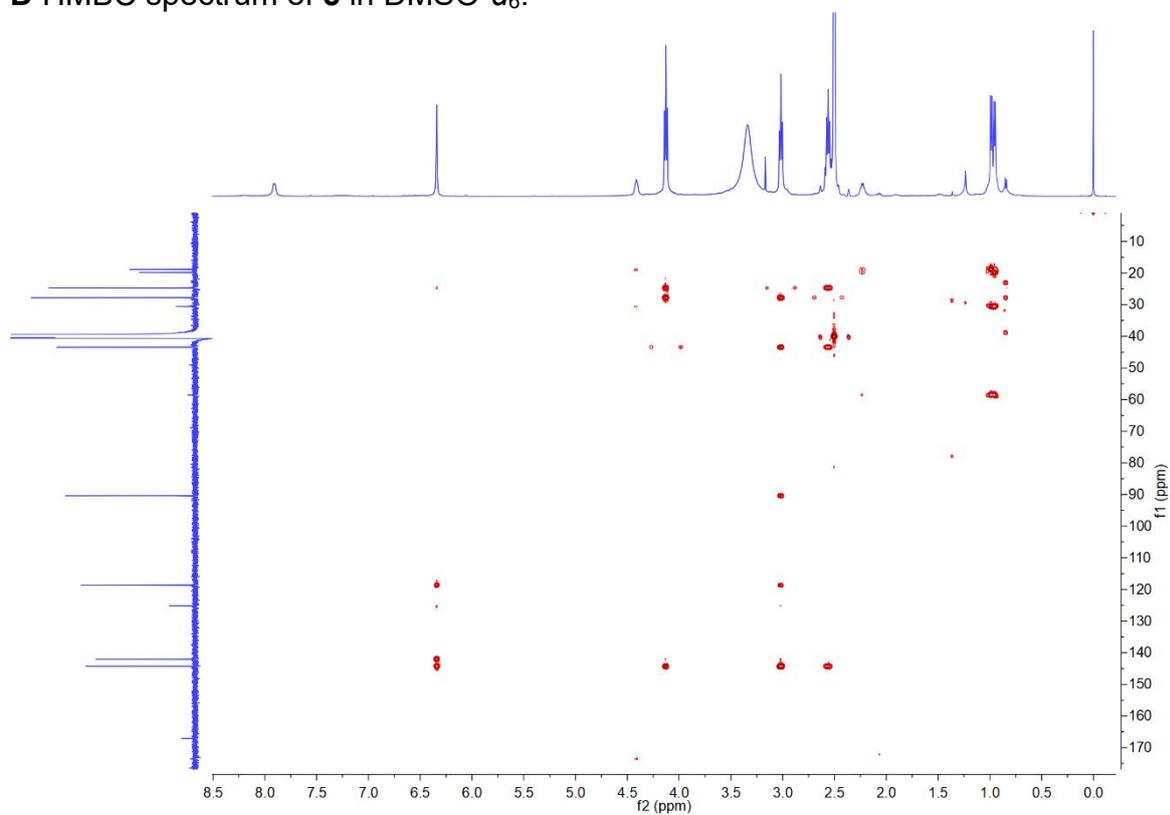
459

460 **C** HSQC spectrum of **5** in $\text{DMSO-}d_6$.



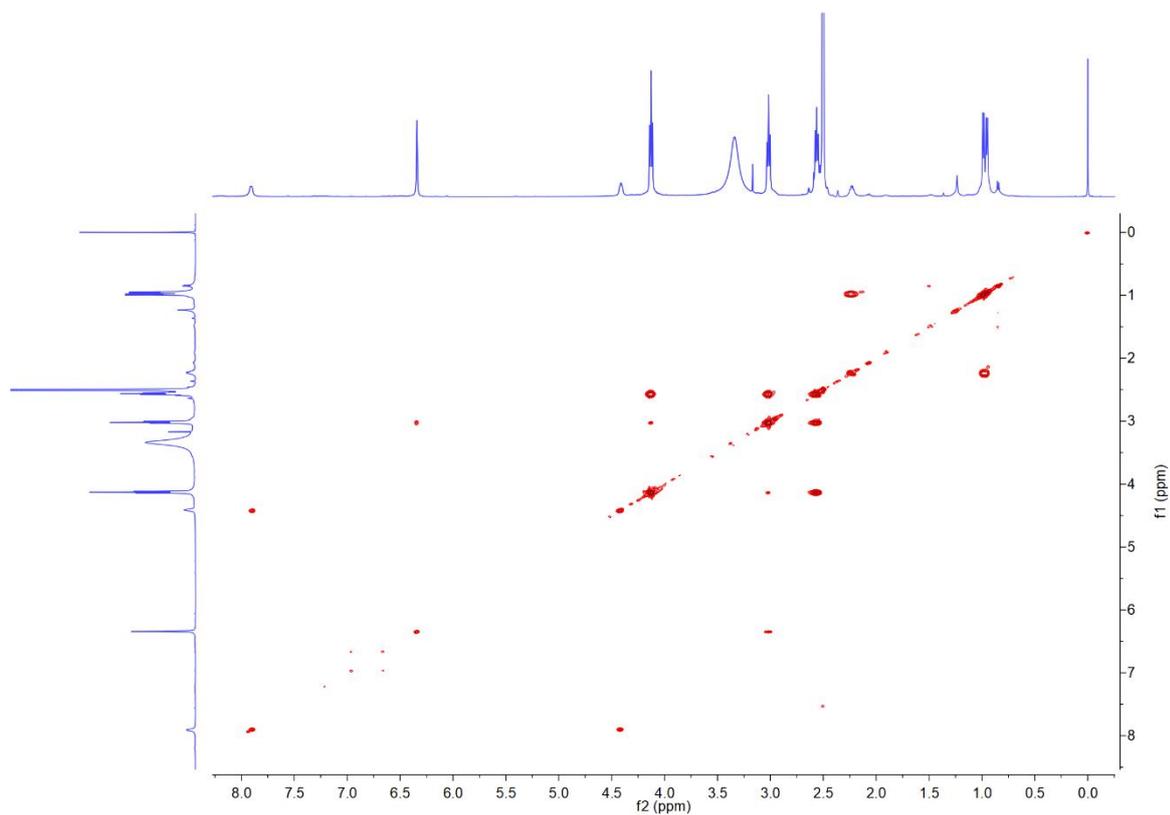
461
462
463

D HMBC spectrum of **5** in DMSO- d_6 .



464
465

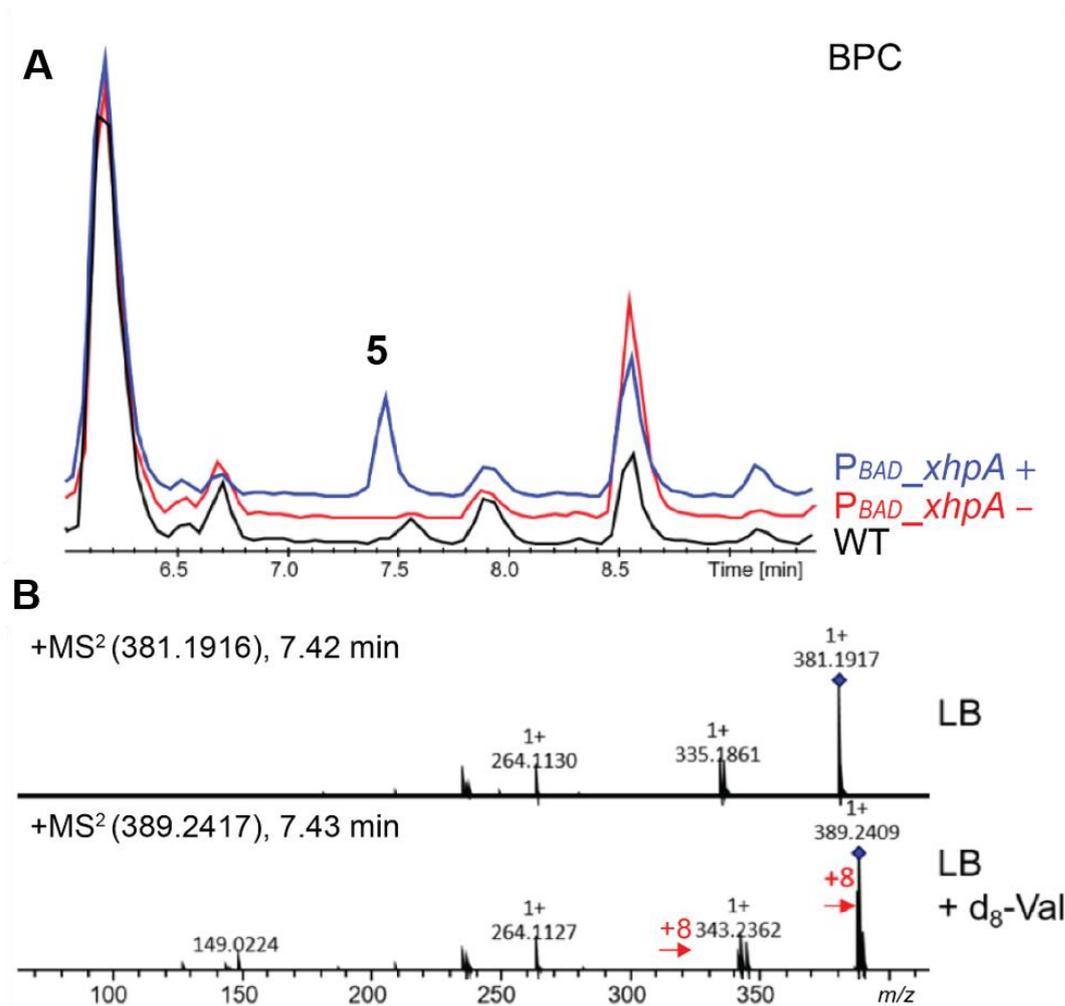
E ^1H - ^1H COSY spectrum of **5** in DMSO- d_6 .



466

467 **Figure S2.** NMR spectra of purified **5** in DMSO-d₆ (500 MHz).

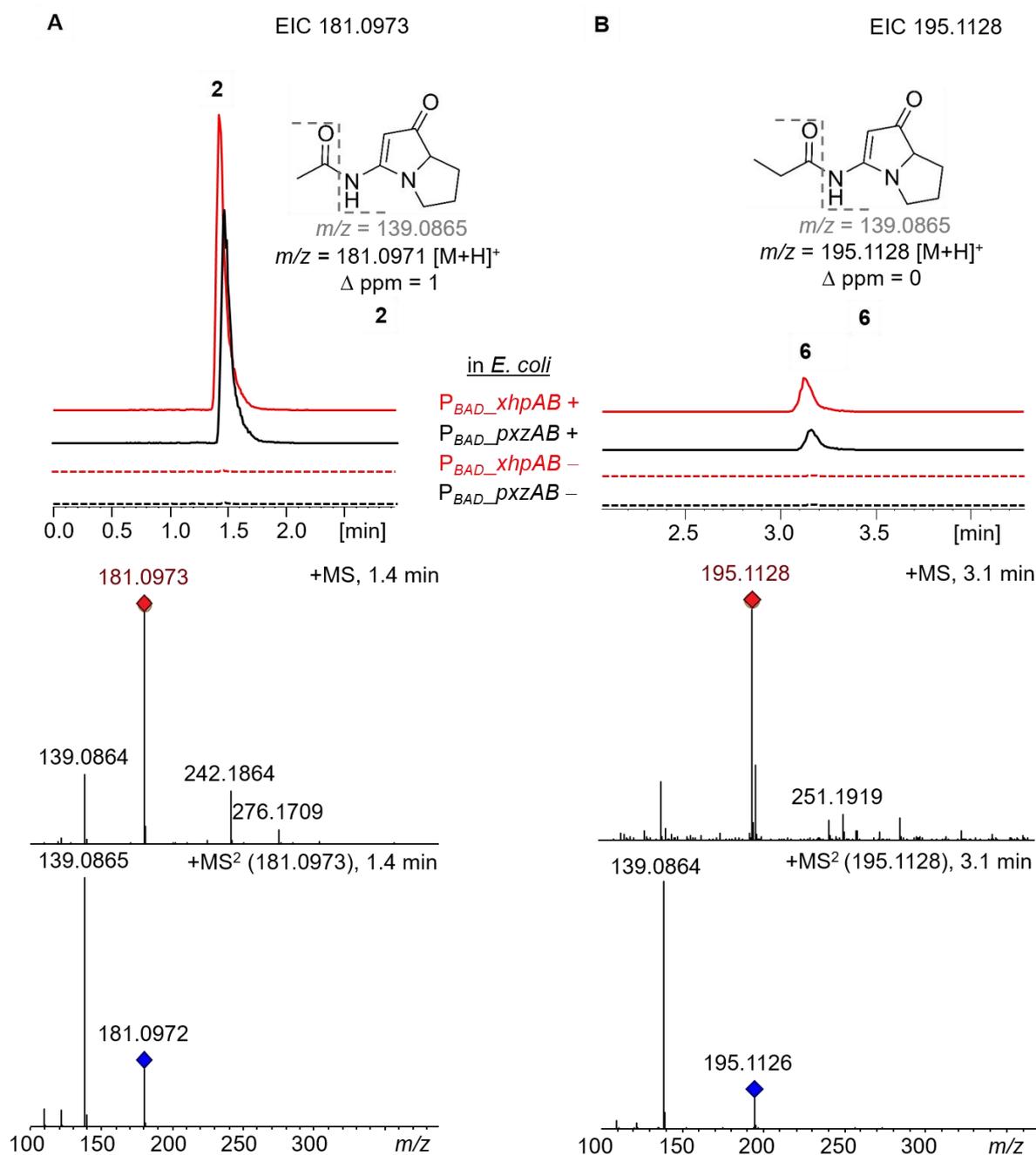
468



469

470 **Figure S3.** Insertion of a P_{BAD} promoter 5' to $xhpA$ in *X. hominickii* resulted in the
 471 production of **5**. **A)** BPC for *X. hominickii* WT, non-induced (-) and arabinose-
 472 induced (+) *X. hominickii* P_{BAD_xhpA} . **B)** MS² spectra of **5** extracted from *X.*
 473 *hominickii* P_{BAD_xhpA} grown in LB, and when grown in LB fed with d₈-L-valine. The
 474 mass of the parent ion and a major fragment increased by 8 Da, consistent with
 475 intact incorporation of the d₈-L-valine.

476



477

478 **Figure S4.** Heterologous expression of *xhpAB* (red) and *pxzAB* (black) in *E. coli*.

479 A) Expression of *xhpAB* and *pxzAB* results in the production of **2** with an acetyl side

480 chain as compared to uninduced control (dashed chromatograms), as judged by LC-

481 HRMS/MS analysis. Structure of **2** (m/z calc'd for $C_9H_{12}N_2O_2$ $[M+H]^+$ = 181.0971;

482 obs'd. m/z = 181.0973; Δ ppm = 1). Inset: proposed structure indicating the key

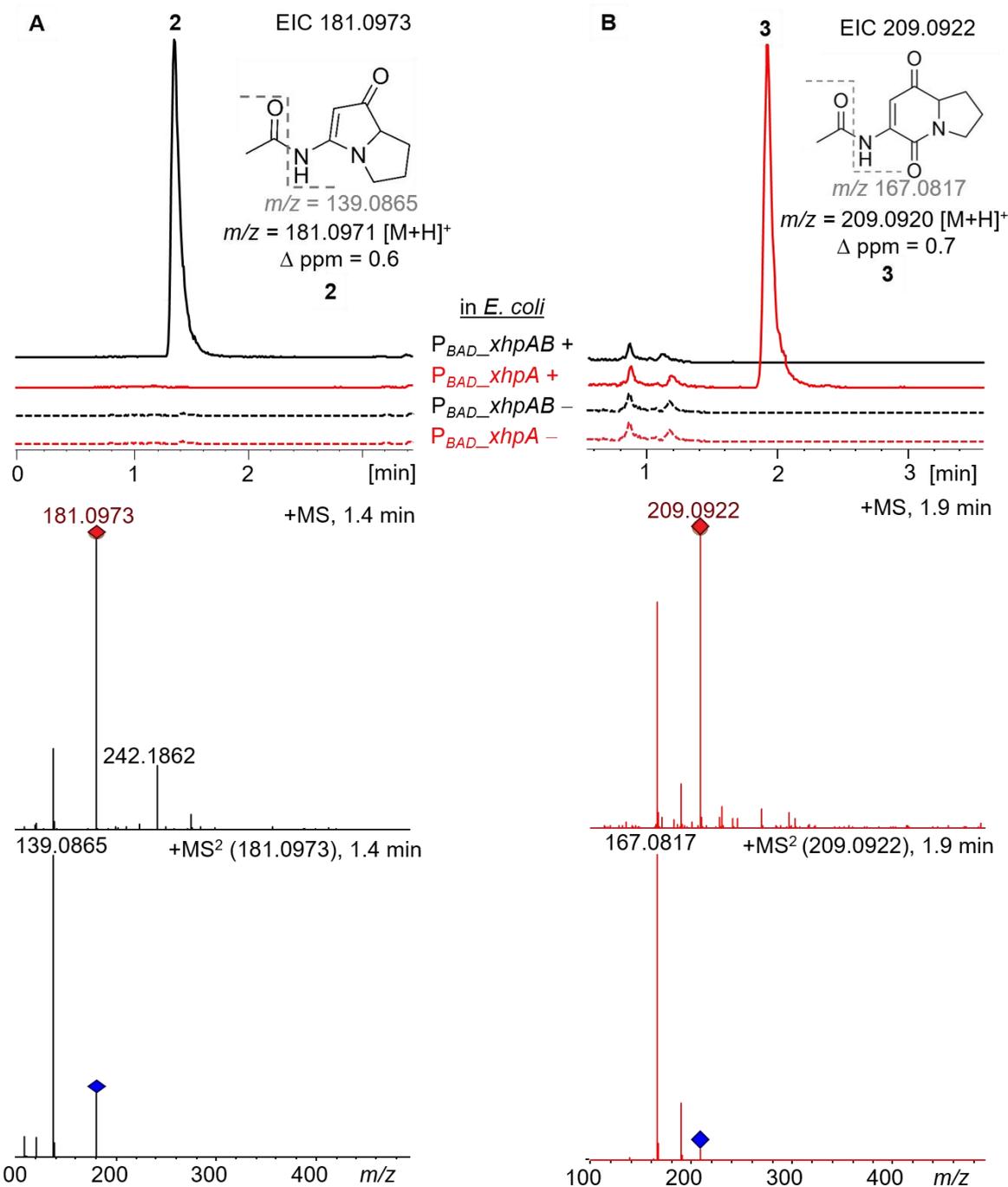
483 fragment ion (m/z calc'd for $C_7H_{10}N_2O$ $[M+H]^+$ = 139.0864). B) Expression of *xhpAB*

484 and *pxzAB* also leads to synthesis of minor amounts of **6** with a C_3 -acyl-side chain

485 as compared to uninduced control (dashed chromatograms), as judged by LC-

486 HRMS/MS analysis. Structure of **6** (m/z calc'd for $C_{10}H_{14}N_2O_2$ $[M+H]^+ = 195.1128$;
487 obs'd. $m/z = 195.1128$). This compound shares the same major fragment with **2** (m/z
488 calc'd for $C_7H_{10}N_2O$ $[M+H]^+ = 139.0864$), as shown (inset).

489



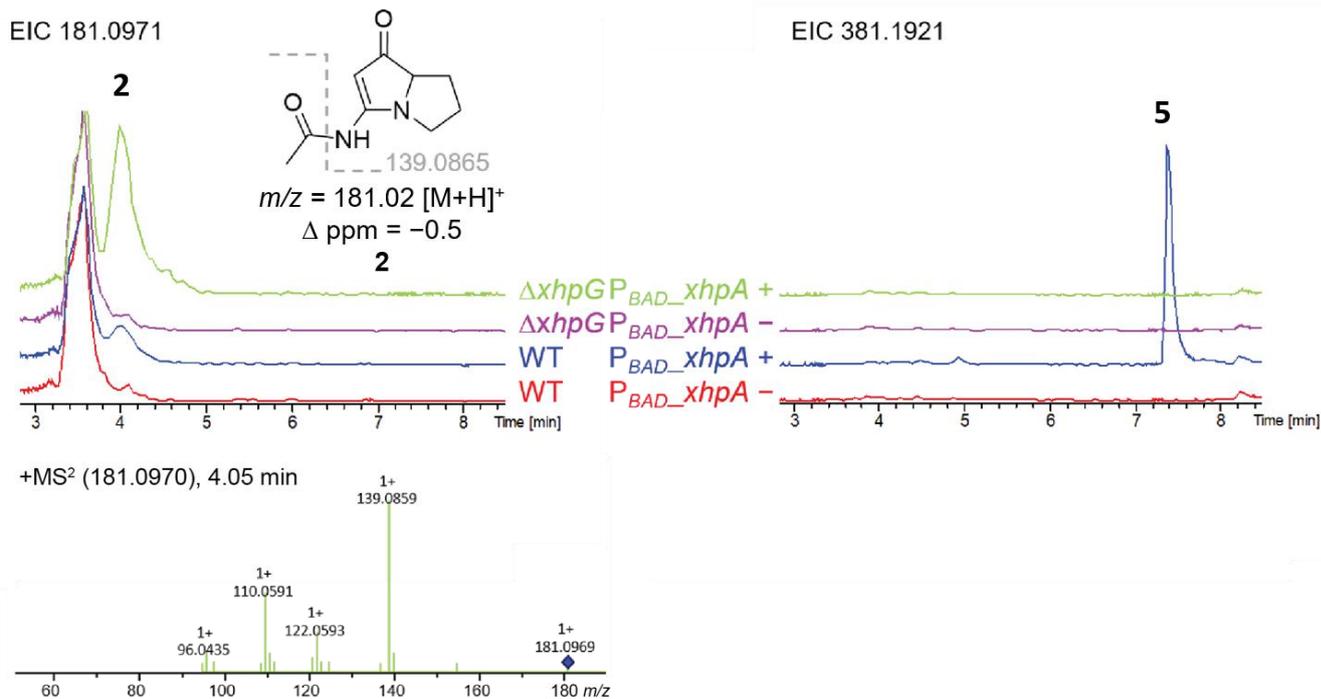
490

491 **Figure S5.** Heterologous expression of *xhpAB* (black) and *xhpA* (red) in *E. coli*.

492 **A)** Expression of *xhpAB* results in the production of **2** as compared to the uninduced
 493 control (dashed chromatograms), as judged by LC-HRMS/MS analysis. Structure of
 494 **2** (m/z calc'd for $C_9H_{12}N_2O_2 [M+H]^+ = 181.0971$; obs'd. $m/z = 181.0973$; $\Delta \text{ ppm} = 0.6$).
 495 A structure corresponding to the major fragment ion observed previously ($m/z =$
 496 $139.0865 [M+H]^+$) is shown. **B)** Expression of *xhpA* results in the production of **3** as
 497 compared to the uninduced control (dashed chromatograms), as judged by LC-

498 HRMS/MS analysis. Structure of **3** (m/z calc'd for $C_{10}H_{12}N_2O_3$ $[M+H]^+ = 209.0920$;
499 obs'd. $m/z = 209.0922$; Δ ppm = 0.7). Inset: proposed structure indicating the key
500 fragment ion (m/z calc'd for $C_8H_{10}N_2O_2$ $[M+H]^+ = 167.0815$).

501

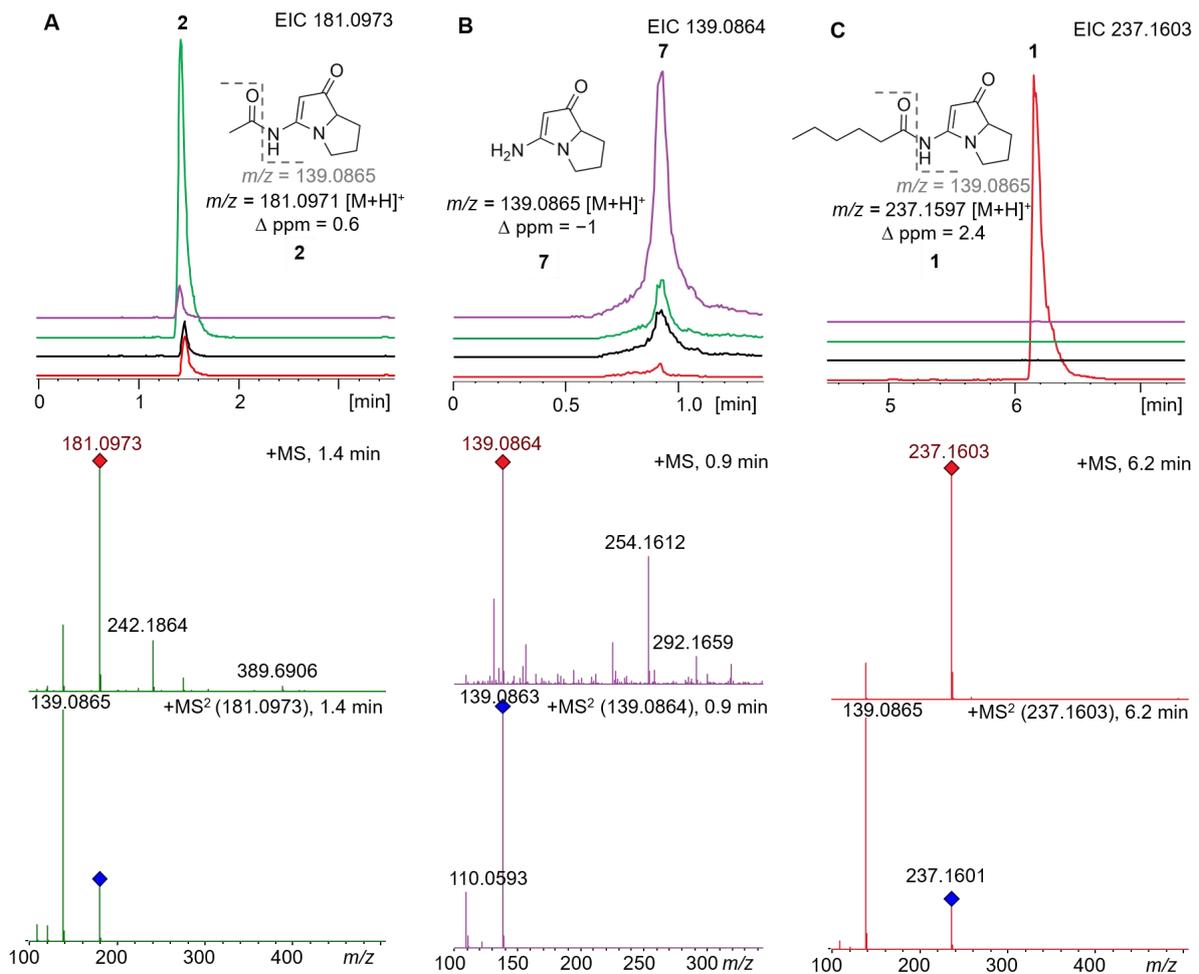


502

503 **Figure S6.** Pyrrolizixenamides-like compound production by the *xhpG* deletion mutant,
 504 as determined by LC-HRMS/MS analysis. *X. hominickii* $\Delta xhpG P_{BAD_xhpA}$ produces
 505 **2** and no pyrrolizwilline **5**. Structure of **2** (m/z calc'd for $C_9H_{12}N_2O_2 [M+H]^+ = 181.0971$;
 506 obs'd. $m/z = 181.0971$; $\Delta \text{ ppm} = 0.5$), showing the characteristic key fragment ($m/z =$
 507 $139.0865 [M+H]^+$) observed by MS². + arabinose induction, – no induction.

508

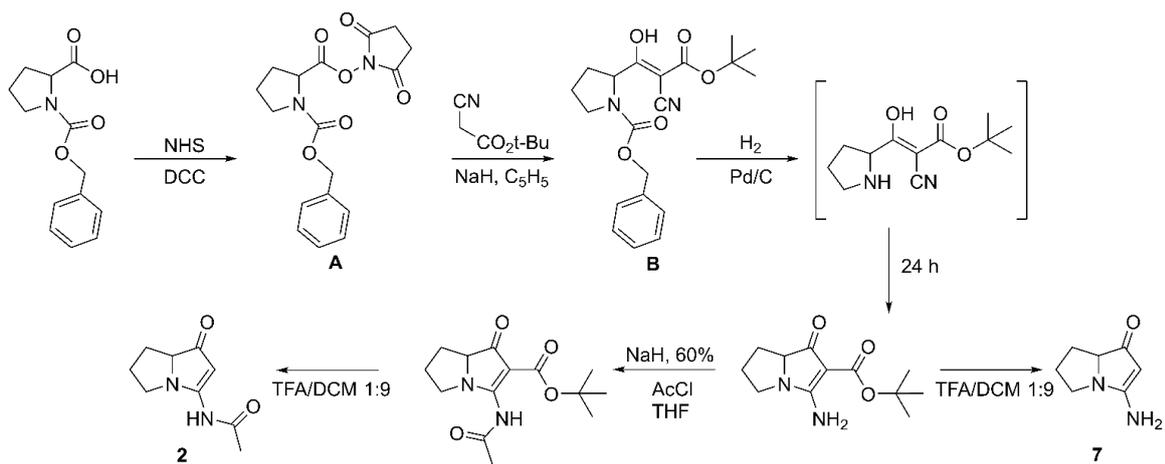
509



510

Heterologous expression in *E. coli*: P_{BAD_pxaAB} $P_{BAD_pxaAB} + P_{BAD_xhpG}$ P_{BAD_xhpAB} $P_{BAD_xhpAB} + P_{BAD_xhpG}$

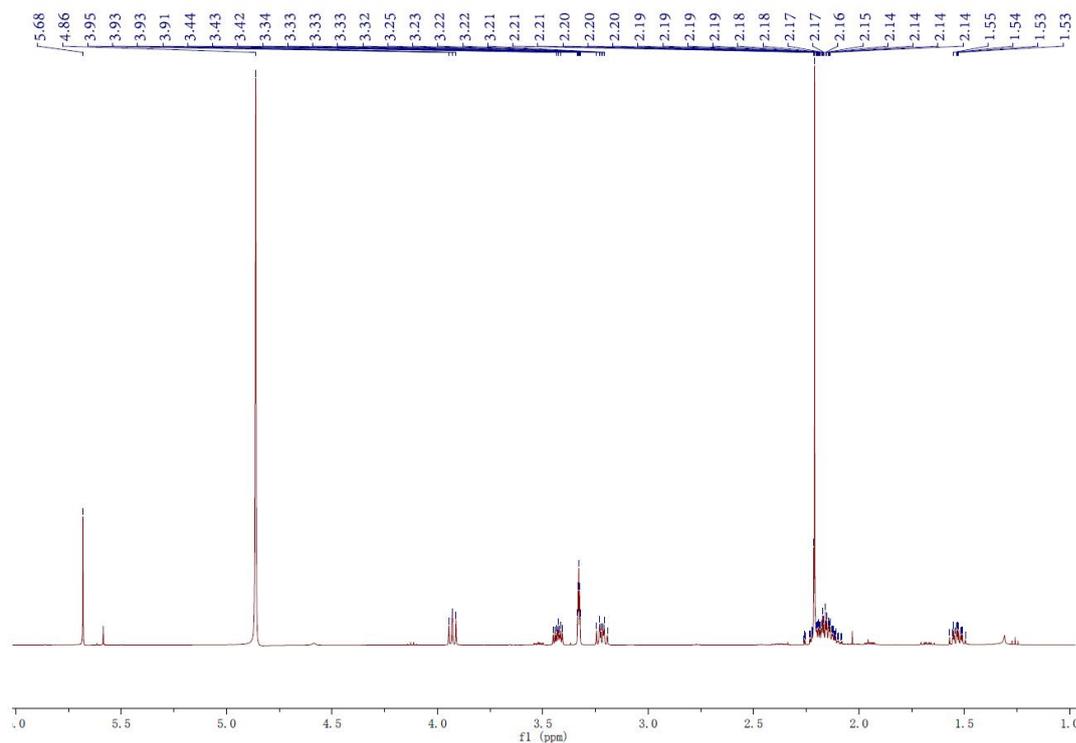
511 **Figure S7.** Investigation of the activity of XhpG. Expression of *xhpAB* + empty vector
 512 (green) in *E. coli* leads to (A) the production of **2**. Structure of **2** (m/z calc'd for
 513 $C_9H_{12}N_2O_2$ $[M+H]^+ = 181.0971$; obs'd $m/z = 181.0973$; $\Delta \text{ ppm} = 0.6$), indicating the
 514 key fragment (m/z calc'd for $C_7H_{10}N_2O$ $[M+H]^+ = 139.0864$) observed by MS². (B)
 515 Expression of *xhpAB* + *xhpG* (violet) leads to the production of **7**, as judged by LC-
 516 HRMS/MS analysis. Inset: Structure of **7** (m/z calc'd for $C_7H_{10}N_2O$ $[M+H]^+ =$
 517 139.0865 ; obs'd $m/z = 139.0864$); the characteristic fragment (m/z $[M+H]^+ = 110.0593$)
 518 was also observed for the synthetic standard (Figure S15). (C) Expression of *pxaAB*
 519 + empty vector (red) in *E. coli* leads to the production of **1**. Structure of **1** (m/z calc'd
 520 for $C_{13}H_{20}N_2O_2$ $[M+H]^+ = 237.1597$; obs'd $m/z = 237.1603$; $\Delta \text{ ppm} = 2.4$), which
 521 shares the same key fragment with **2** ($m/z = 139.0865$ $[M+H]^+$), as shown.
 522 Expression of *pxaAB* + *xhpG* (black) leads to the production of **7** (B) and no
 523 production of **1** (C).



524

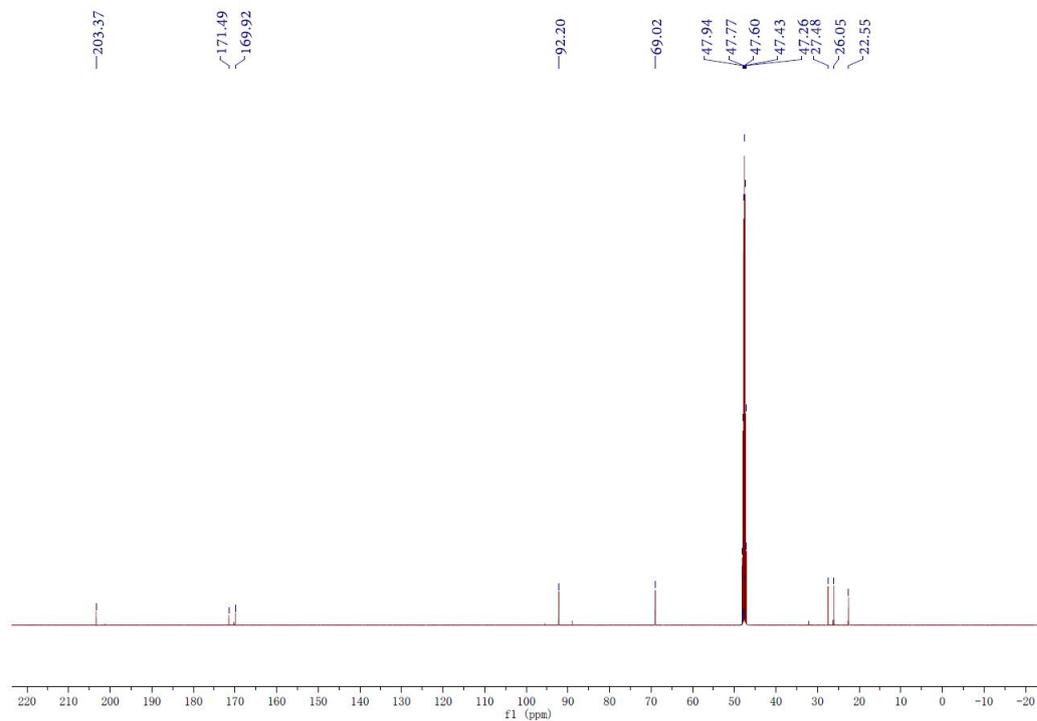
525 **Figure S8.** Synthesis scheme for compounds **2** and **7**.

526 **A** ^1H NMR (500 MHz) spectrum of compound **2** in $\text{MeOH-}d_4$.



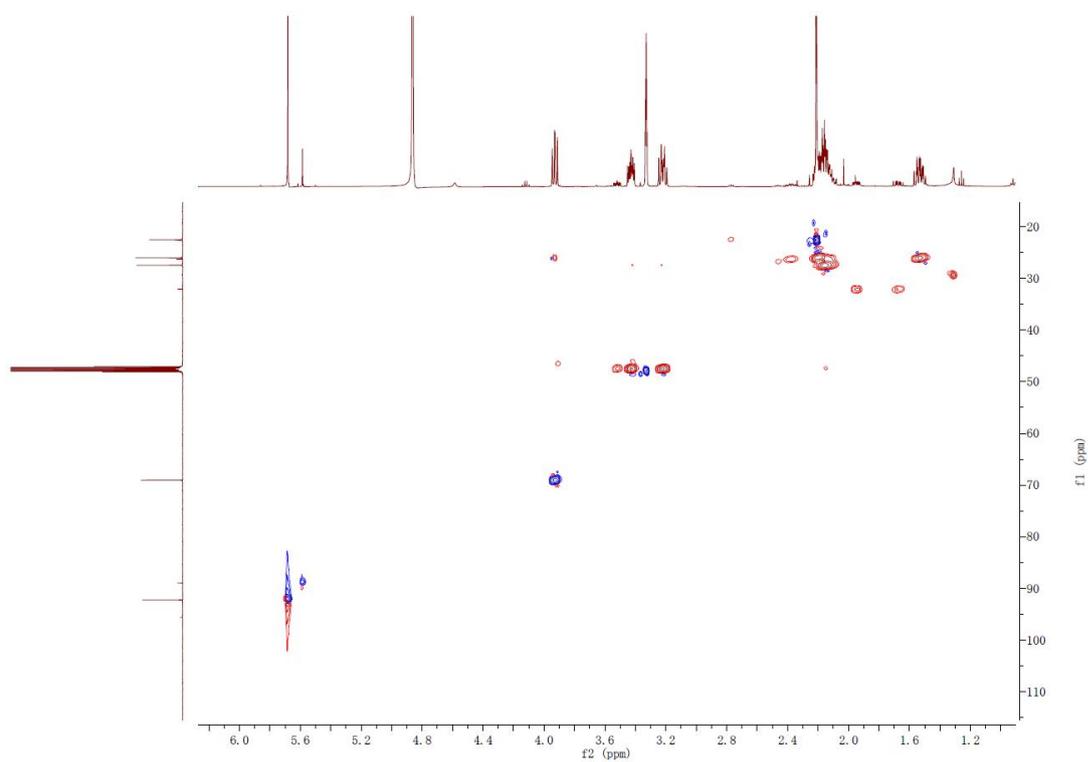
527

528 **B** ^{13}C NMR (125 MHz) spectrum of compound **2** in $\text{MeOH-}d_4$.



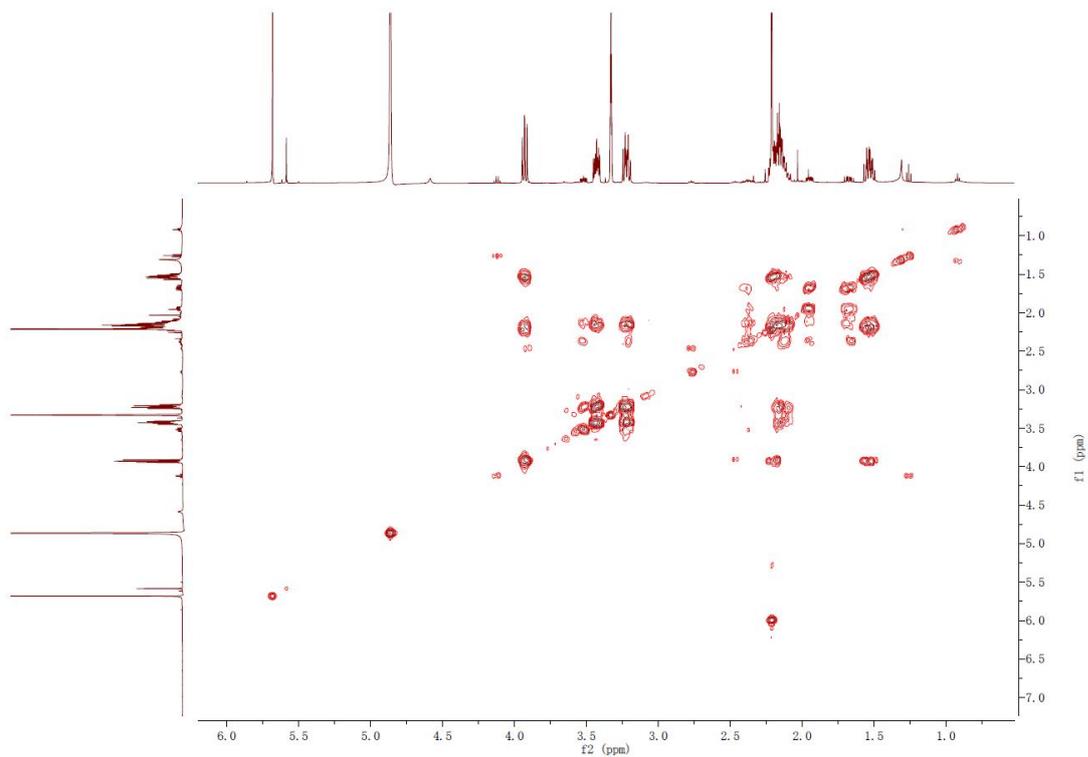
529

530 **C** HSQC spectrum of compound **2** in $\text{MeOH-}d_4$.



531

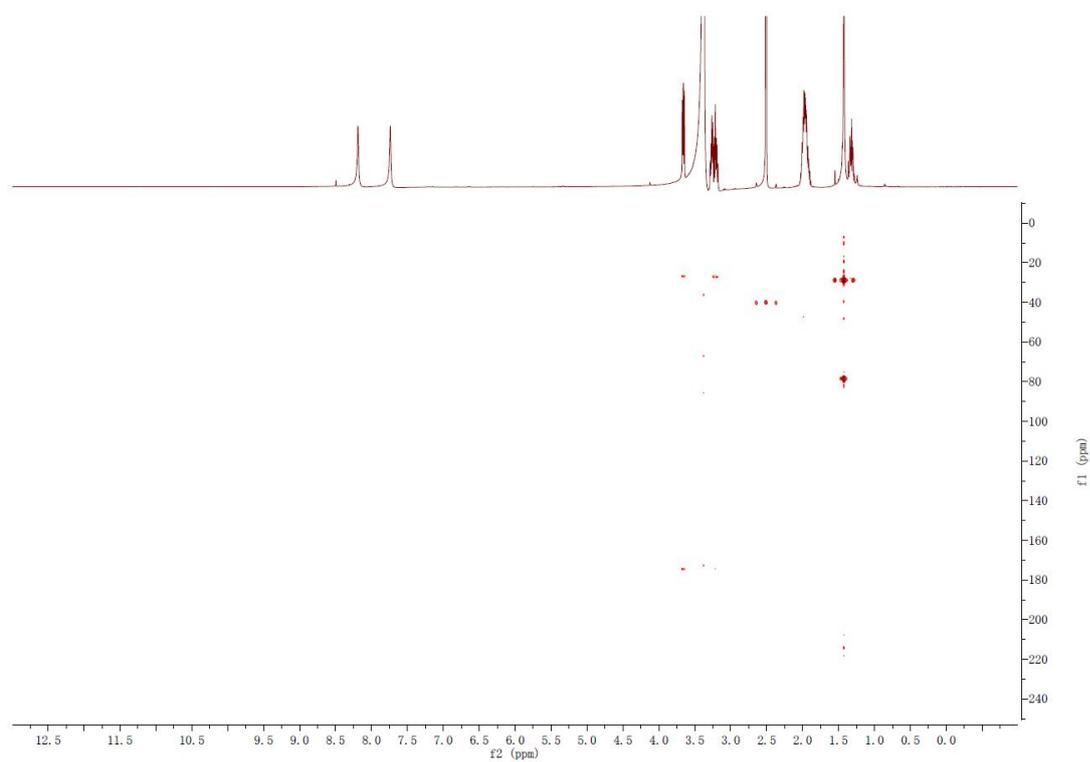
532 **D** ^1H - ^1H COSY spectrum of compound **2** in $\text{MeOH-}d_4$.



533

534 **E** HMBNMR spectrum of compound **2** in $\text{DMSO-}d_6$.

535

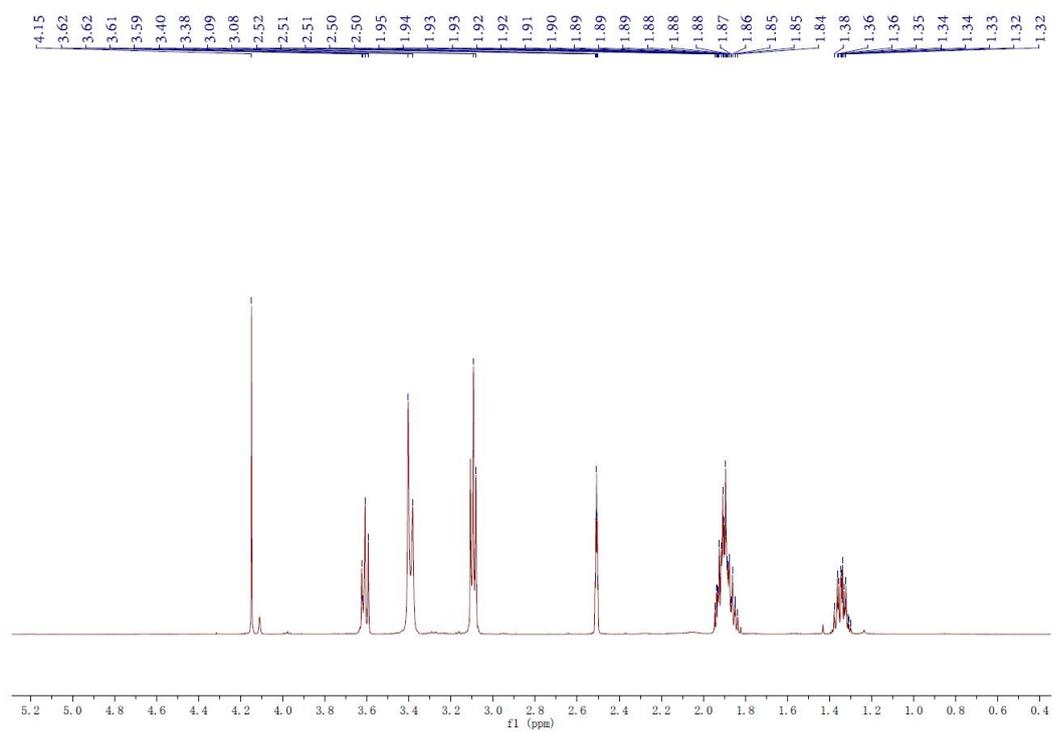


536

537 **Figure S9.** NMR spectra of purified **2**.

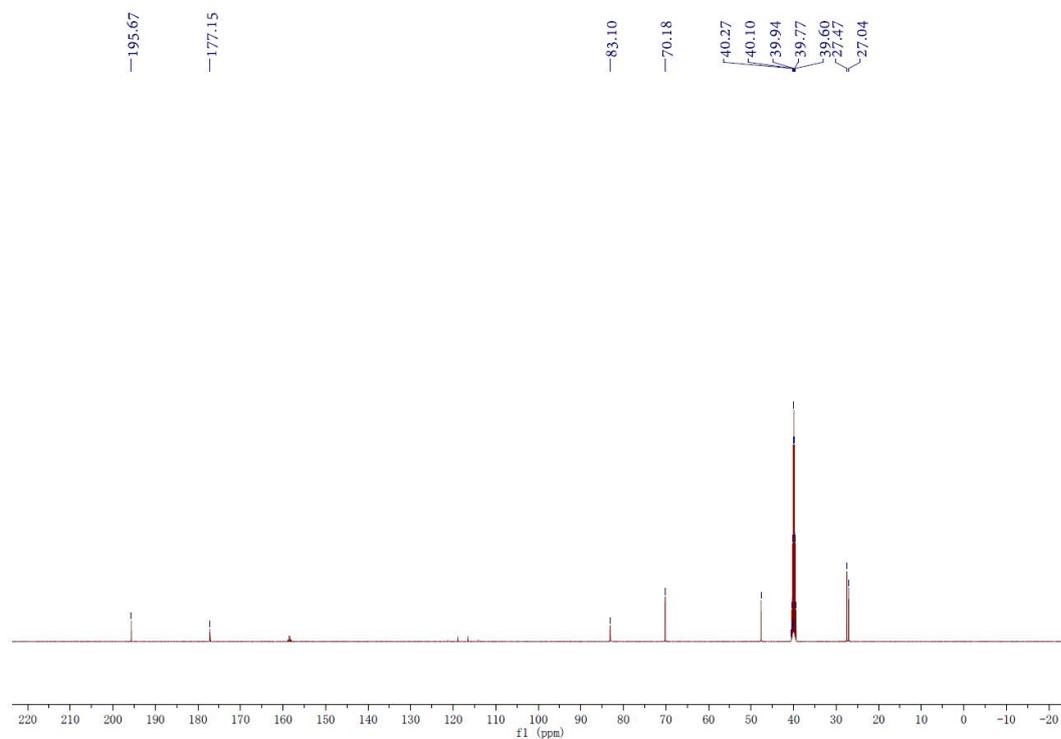
538

539 **A** ^1H NMR (500 MHz) spectrum of compound **7** in $\text{DMSO-}d_6$.



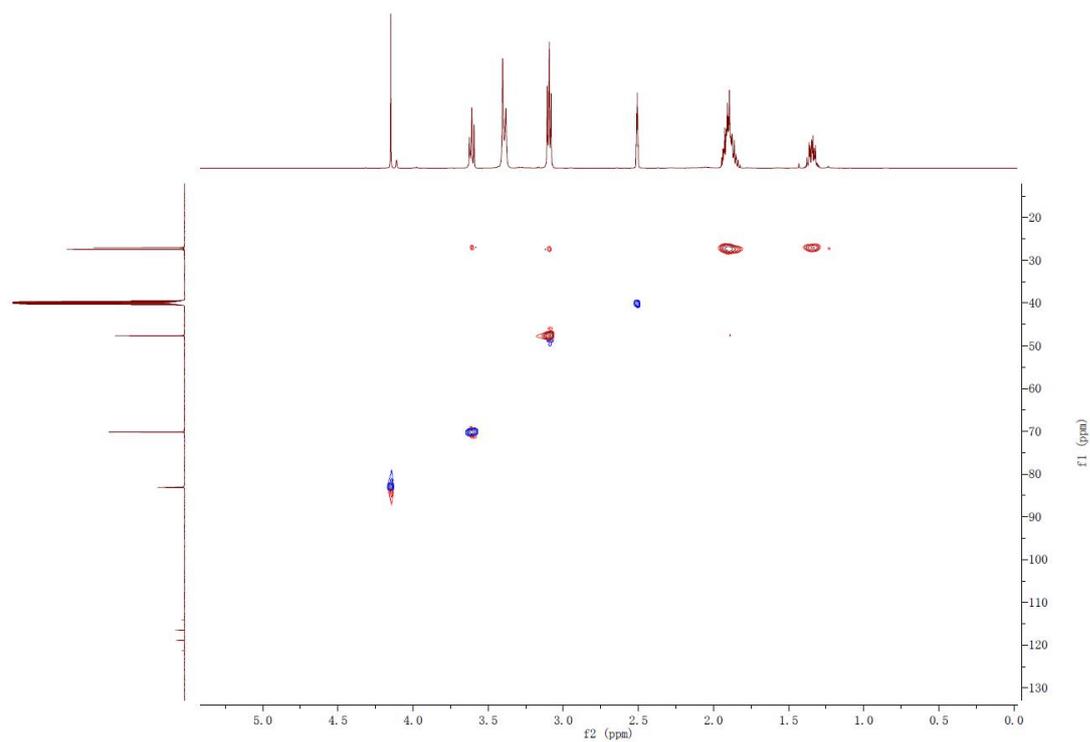
540

541 **B** ^{13}C NMR (125 MHz) spectrum of compound **7** in $\text{DMSO-}d_6$.



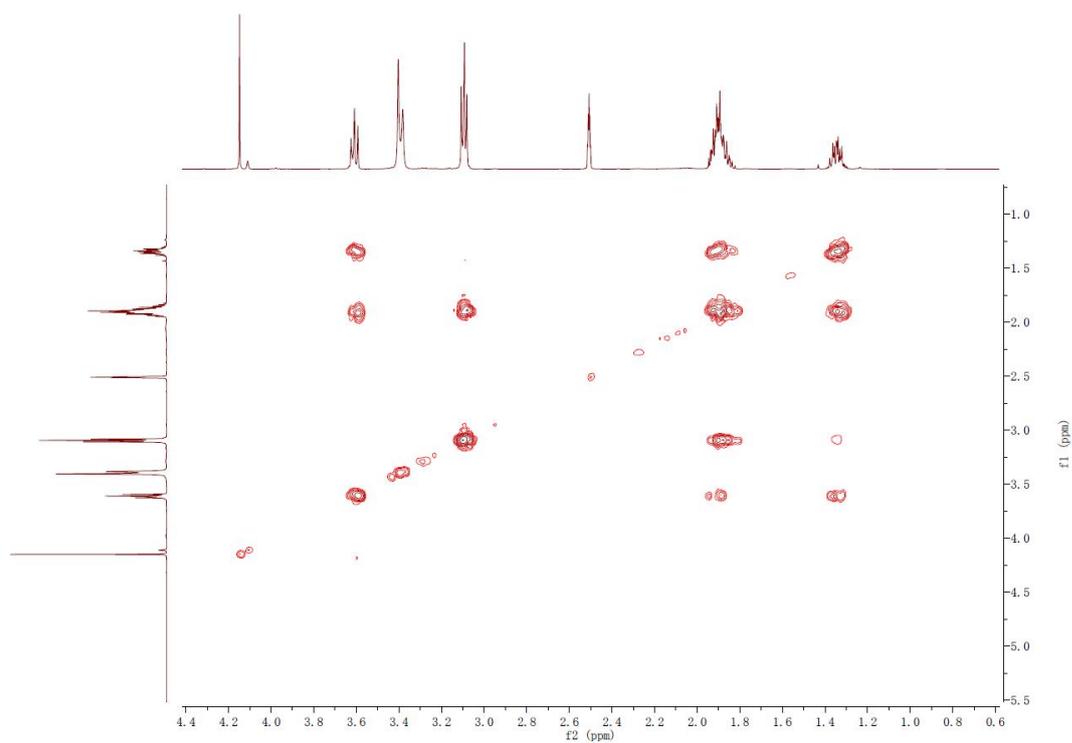
542

543 **C** HSQC spectrum of compound **7** in $\text{DMSO-}d_6$.



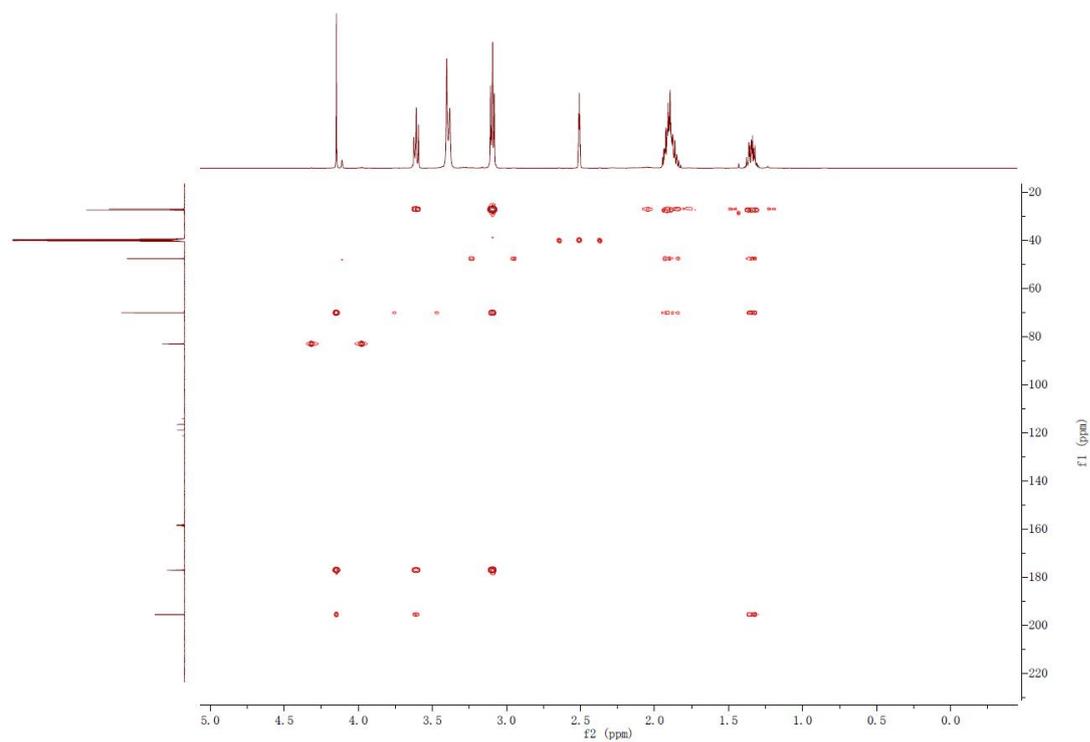
544

545 **D** ^1H - ^1H COSY spectrum of compound **7** in $\text{DMSO-}d_6$.



546

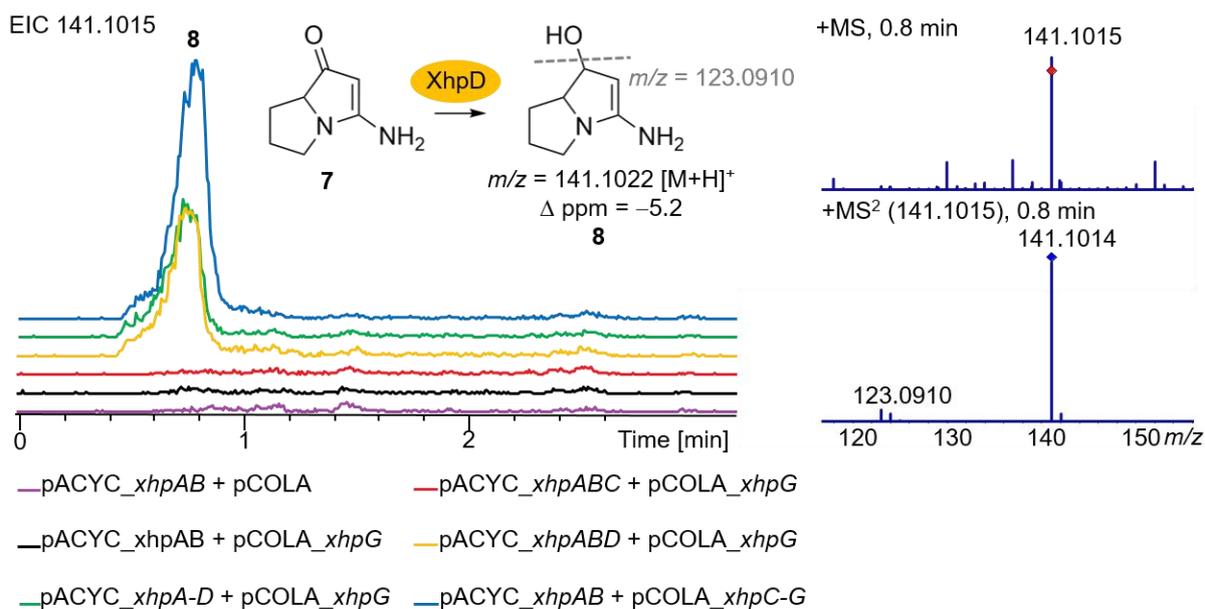
547 **E** HMBC spectrum of compound **7** in $\text{DMSO-}d_6$.



548

549 **Figure S10.** NMR spectra of purified **7**.

550

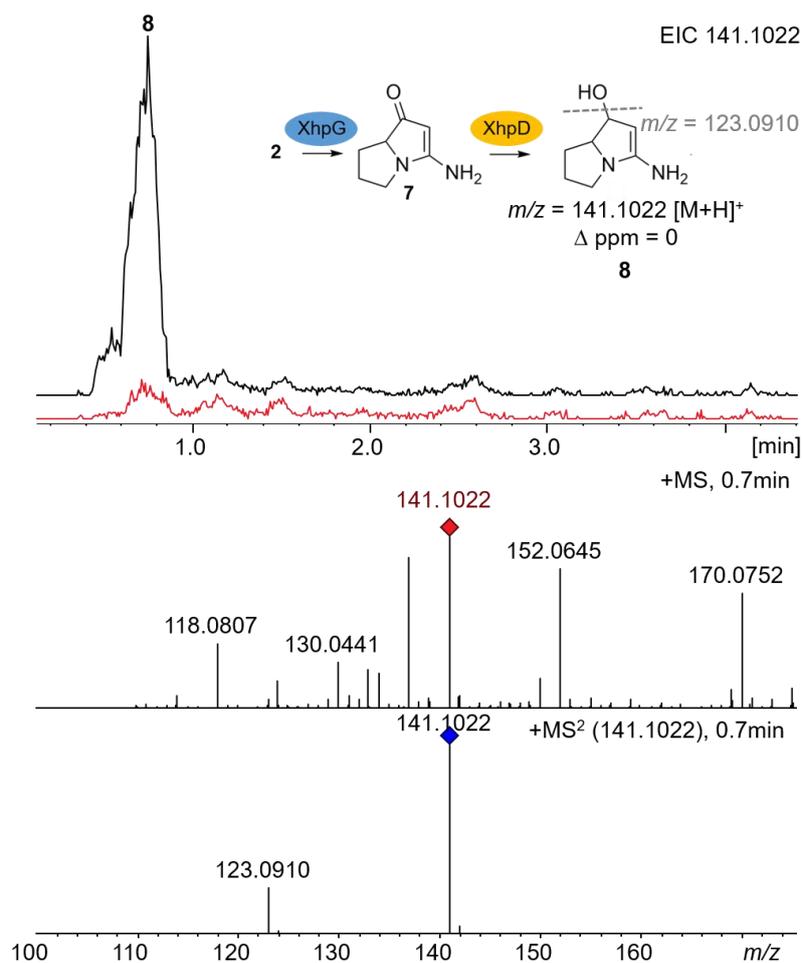


551

552 **Figure S11.** Heterologous expression of *xhpAB* and *xhpG* alongside different
 553 combinations of *xhpC* and *xhpD*, in *E. coli*. Shown are the EICs corresponding to
 554 product **8** (m/z calc'd for $C_7H_{12}N_2O [M+H]^+ = 141.1022$; obs'd $m/z = 141.1015$; $\Delta \text{ ppm}$
 555 $= -5.2$) obtained by reduction of **7**. Inset: proposed structure of **8** indicating the key
 556 MS² fragment (**8** -H₂O; m/z calc'd for $C_7H_{10}N_2 [M+H]^+ = 123.0916$).

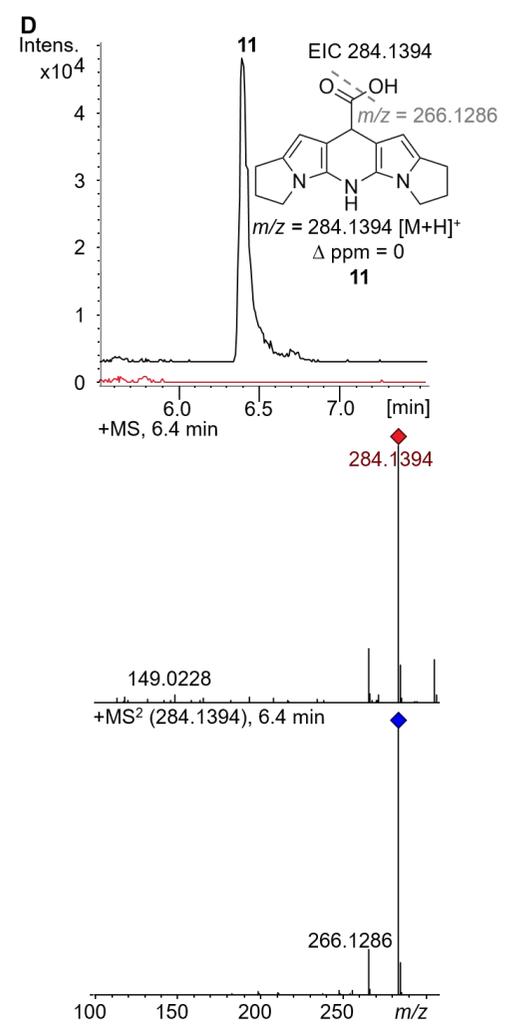
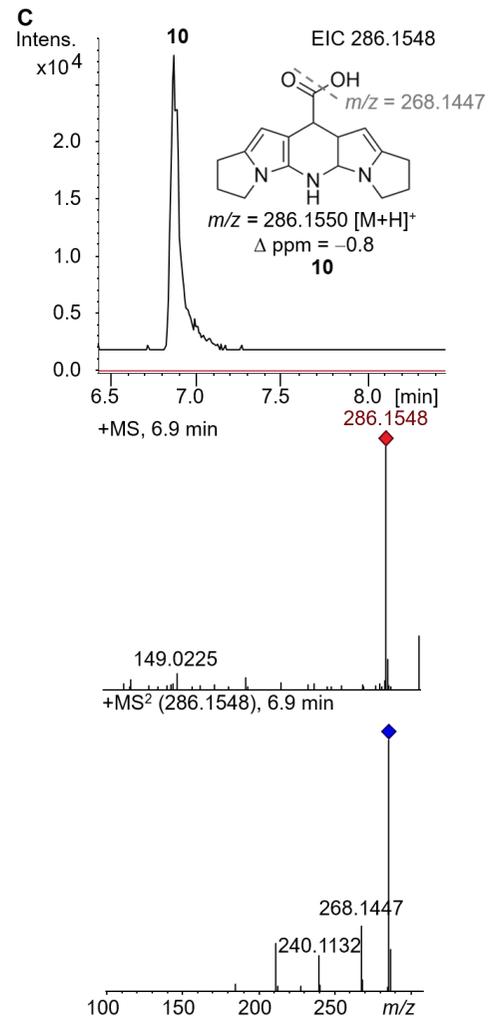
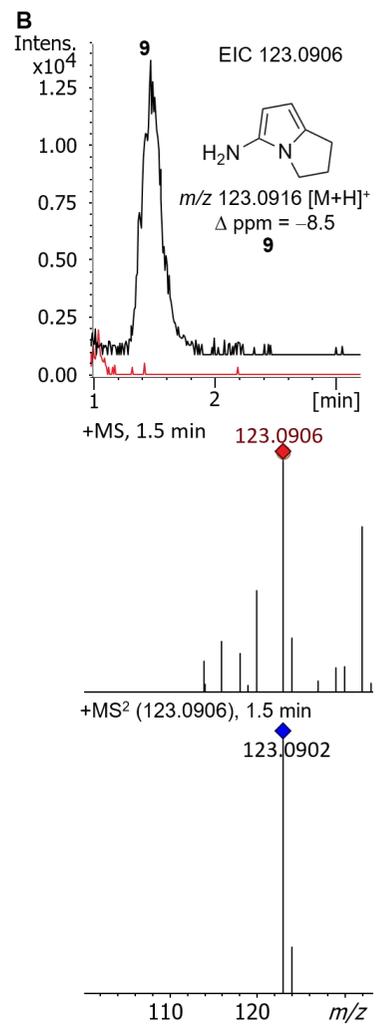
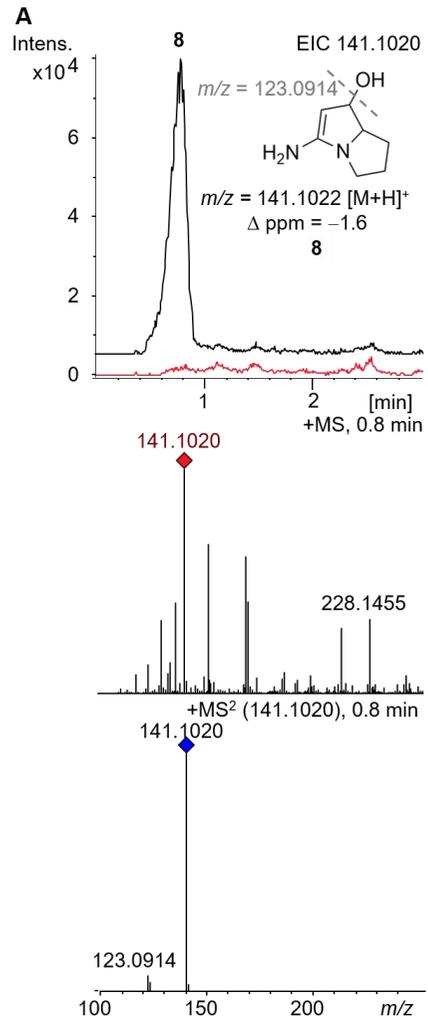
557

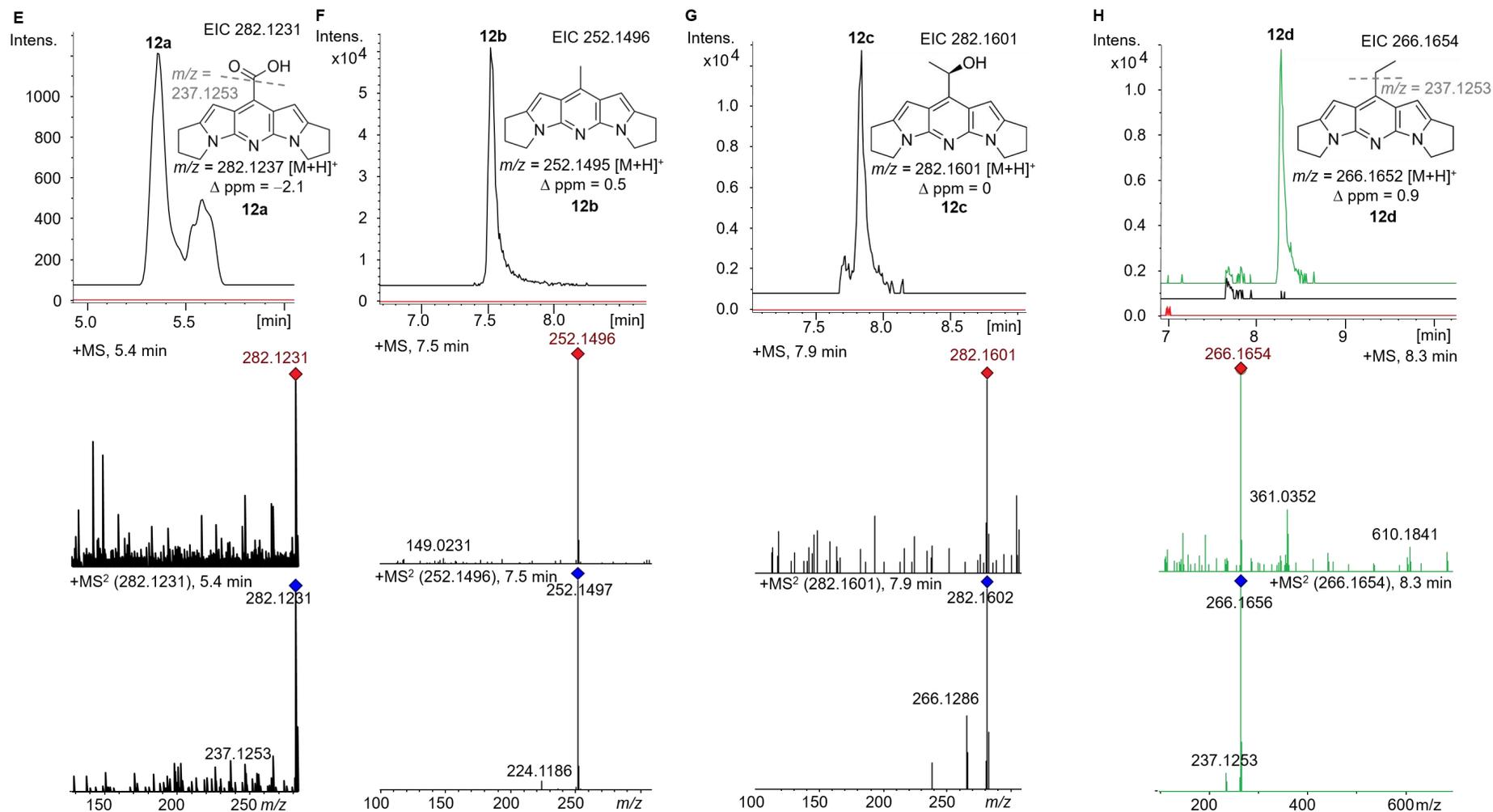
558



559

560 **Figure S12.** Heterologous expression of *xhpABD* with *xhpG* (black) relative to
 561 *xhpABD* with the empty vector control (red), in *E. coli*. Shown are the EICs
 562 corresponding to product **8** (m/z calc'd for $C_7H_{12}N_2O [M+H]^+ = 141.1022$; obs'd $m/z =$
 563 141.1022) obtained by reduction of **7**. Inset: proposed structure of **8** indicating the
 564 key MS^2 fragment (**8** $-H_2O$; m/z calc'd for $C_7H_{10}N_2 [M+H]^+ = 123.0916$).



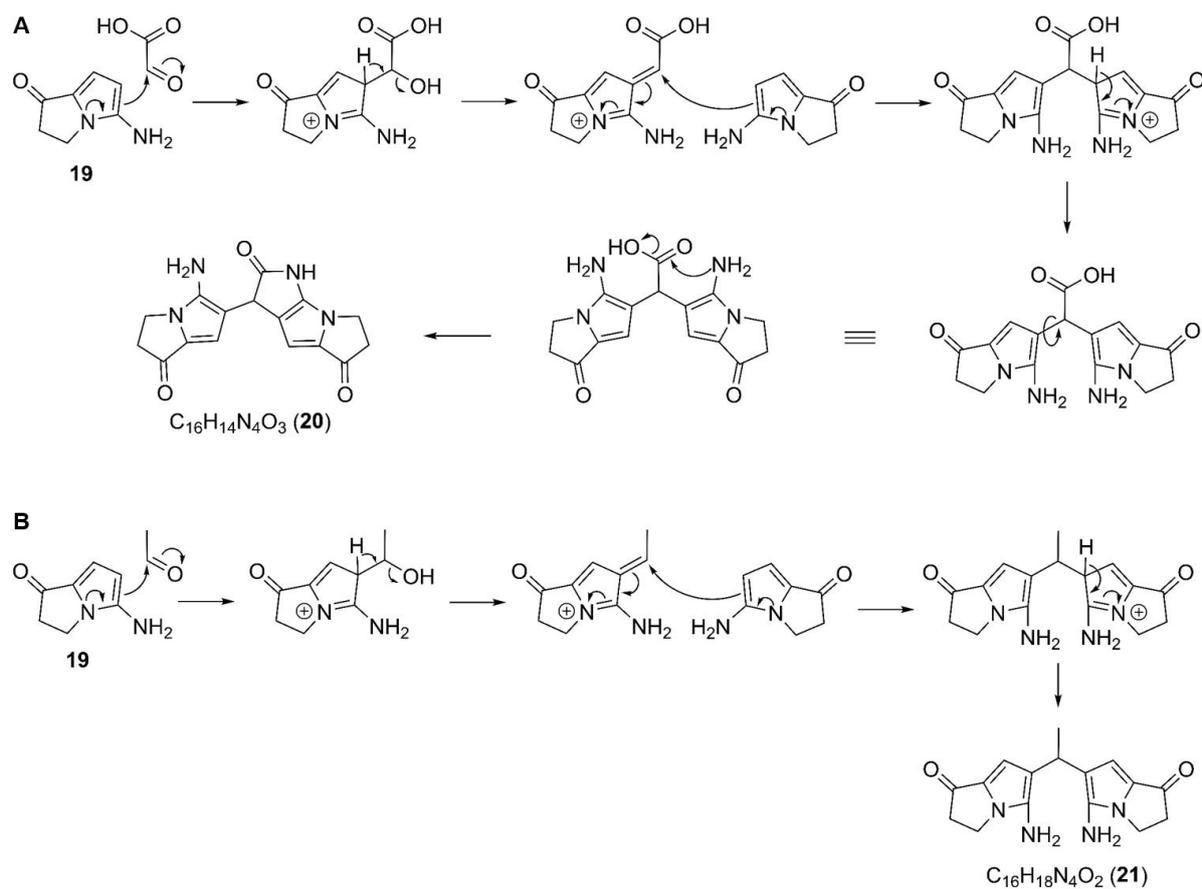


566

567 **Figure S13.** Extracted ion chromatograms used to calculate the predicted sum formulas of the pyrrolizwilline intermediates and
 568 derivatives. The EICs for the *E. coli* pACYC_xhpAB pCOLA empty vector extract (red) and the EICs for the *E. coli* pACYC_xhpAB
 569 pCOLA_xhpC-G extract (black) are shown. **A)** MS/MS analysis of **8**. The structure of **8** (m/z calc'd for C₇H₁₂N₂O [M+H]⁺ = 141.1022;

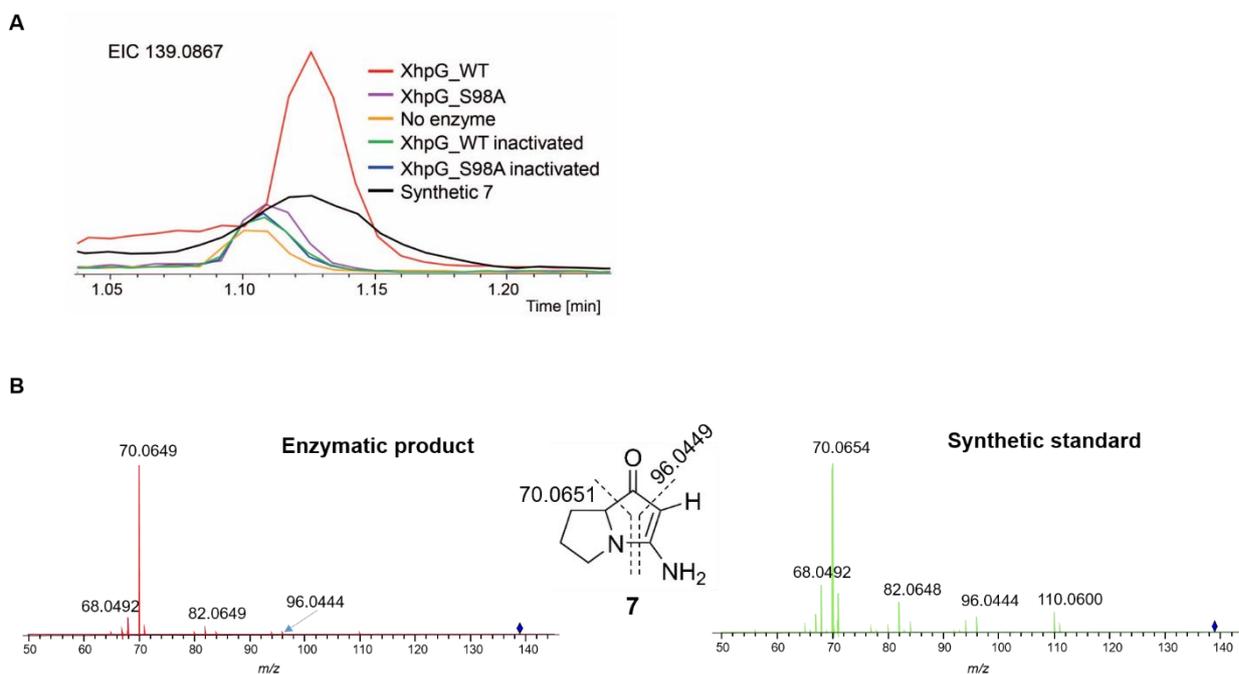
570 obs'd m/z = 141.1020; Δ ppm = -1.6) is shown, alongside a proposed structure and its suggested MS² fragment (m/z calc'd for
571 C₇H₁₀N₂ [M+H]⁺ = 123.0916). **B**) MS/MS analysis of **9**. The structure of **9** (m/z calc'd for C₇H₁₀N₂ [M+H]⁺ = 123.0916; obs'd m/z =
572 123.0906; Δ ppm = -8.5) is shown. **C**) MS/MS analysis of **10** (m/z calc'd for C₁₆H₁₉N₃O₂ [M+H]⁺ = 286.1550; obs'd m/z = 286.1548;
573 Δ ppm = -0.8). Inset: proposed structures for the key observed MS² fragments (m/z calc'd for C₁₆H₁₈N₃O⁺ [M]⁺ = 268.1444). **D**)
574 MS/MS analysis of **11**. The structure of **11** (m/z calc'd for C₁₆H₁₇N₃O₂ [M+H]⁺ = 284.1394; obs'd m/z = 284.1394) is shown. Inset:
575 proposed structure for the key MS² fragment (m/z calc'd for C₁₆H₁₆N₃O⁺ [M]⁺ = 266.1288). **E**) MS/MS analysis of **12a**. The structure
576 of **12a** (m/z calc'd for C₁₆H₁₅N₃O₂ [M+H]⁺ = 282.1237; obs'd m/z = 282.1231; Δ ppm = -2.1) is shown. Inset: proposed structure for
577 the MS² fragment (m/z calc'd for C₁₅H₁₅N₃⁺ [M]⁺ = 237.1260). **F**) MS/MS analysis of **12b** (m/z calc'd for C₁₆H₁₇N₃ [M+H]⁺ = 252.1495;
578 obs'd m/z = 252.1496; Δ ppm = 0.5). **G**) MS/MS analysis of **12c** (m/z calc'd for C₁₇H₁₉N₃O [M+H]⁺ = 282.1601; obs'd m/z =
579 282.1601). The characteristic fragment ion m/z [M+H]⁺ = 266.1286 was observed. **H**) Feeding of propionaldehyde to *E. coli*
580 expression culture (green chromatogram) results in the incorporation of propionaldehyde to yield **12d** (m/z calc'd for C₁₇H₁₉N₃
581 [M+H]⁺ = 266.1652; obs'd m/z = 266.1654; Δ ppm = 0.9). Inset: proposed structure for the MS² fragment (m/z calc'd for C₁₅H₁₅N₃⁺
582 [M]⁺ = 237.1260).

583



584

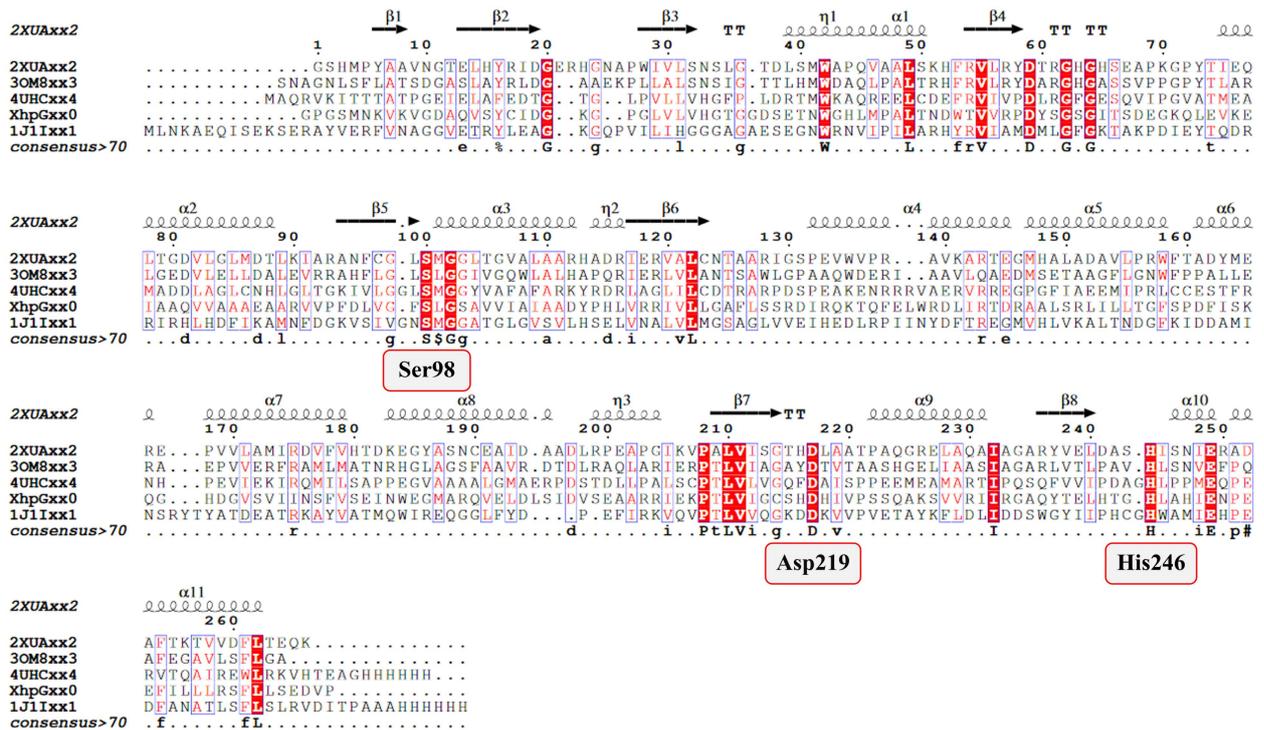
585 **Figure S14.** The model compound 5-amino-2,3-dihydro-1*H*-pyrrolizin-1-one (**19**)
 586 reacts with glyoxylic acid and acetaldehyde in THF. **A)** **19** reacts with glyoxylic acid
 587 to form **20**. **B)** **19** reacts with acetaldehyde to form **21**.



589

590

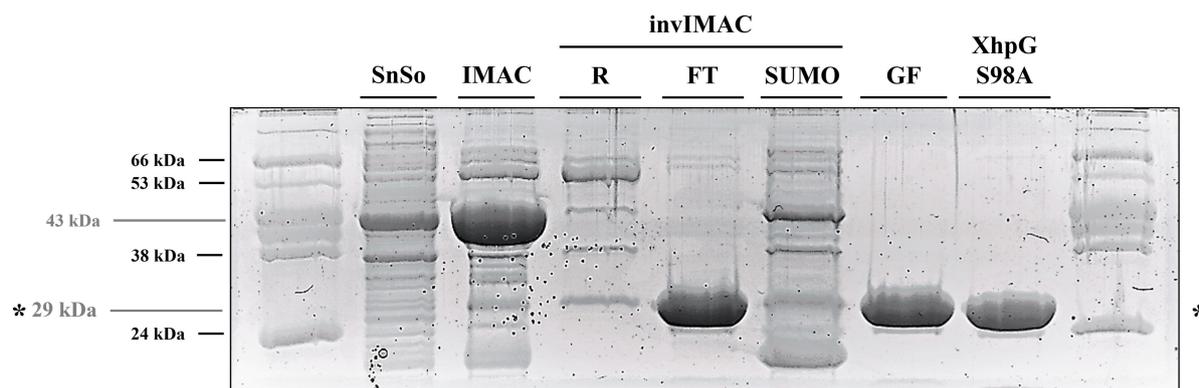
591 **Figure S15.** Characterization of XhpG activity *in vitro*. **A)** Extracted ion
 592 chromatograms (EIC) of $m/z = 139.0877$ $[M+H]^+$ (m/z calc'd for $C_7H_{10}N_2O$ $[M+H]^+ =$
 593 139.0864 ; obs'd $m/z = 139.0867$; Δ ppm = 2.1) corresponding to the hydrolytic
 594 product **7** under different assay conditions are shown. **B)** Comparison of MS² spectra
 595 of enzymatically produced and synthetic standard of **7**. Structures of the observed
 596 fragments are proposed (m/z calc'd for $C_4H_8N^+$ $[M]^+ = 70.0651$; m/z calc'd for
 597 $C_5H_6NO^+$ $[M]^+ = 96.0443$).



598

599 **Figure S16.** Sequence alignment of XhpG hydrolase homologues. The protein
 600 sequence of XhpG was aligned against sequences of homologous enzymes, including
 601 the enol-lactonase PcaD from *Burkholderia xenovorans* (PDB ID: 2XUA), a putative
 602 hydrolase from *Pseudomonas aeruginosa* PA01 (PDB ID: 3OM8), an esterase from
 603 *Thermogutta terrifontis* (PDB ID: 4UHC), and a serine hydrolase from
 604 *Janthinobacterium sp.* Strain J3 (PDB ID: 1J11) identified using the PDB database^[9]
 605 and the Dali structural comparison server.^[10] These enzymes all show mutual
 606 sequence and structural identity greater than 25%, with the exception of the serine
 607 hydrolase which has a structural similarity of 17%. In addition, these enzymes share
 608 sequential determinants involved in substrate binding and/or the catalytic reaction, in
 609 particular the residues Ser98, Asp219 and His246, which constitute a catalytic triad,
 610 and also conserved secondary structural elements. On the basis that the Ser98 residue
 611 was essential for catalysis, the corresponding codon was mutated in order to inactivate
 612 the enzyme.

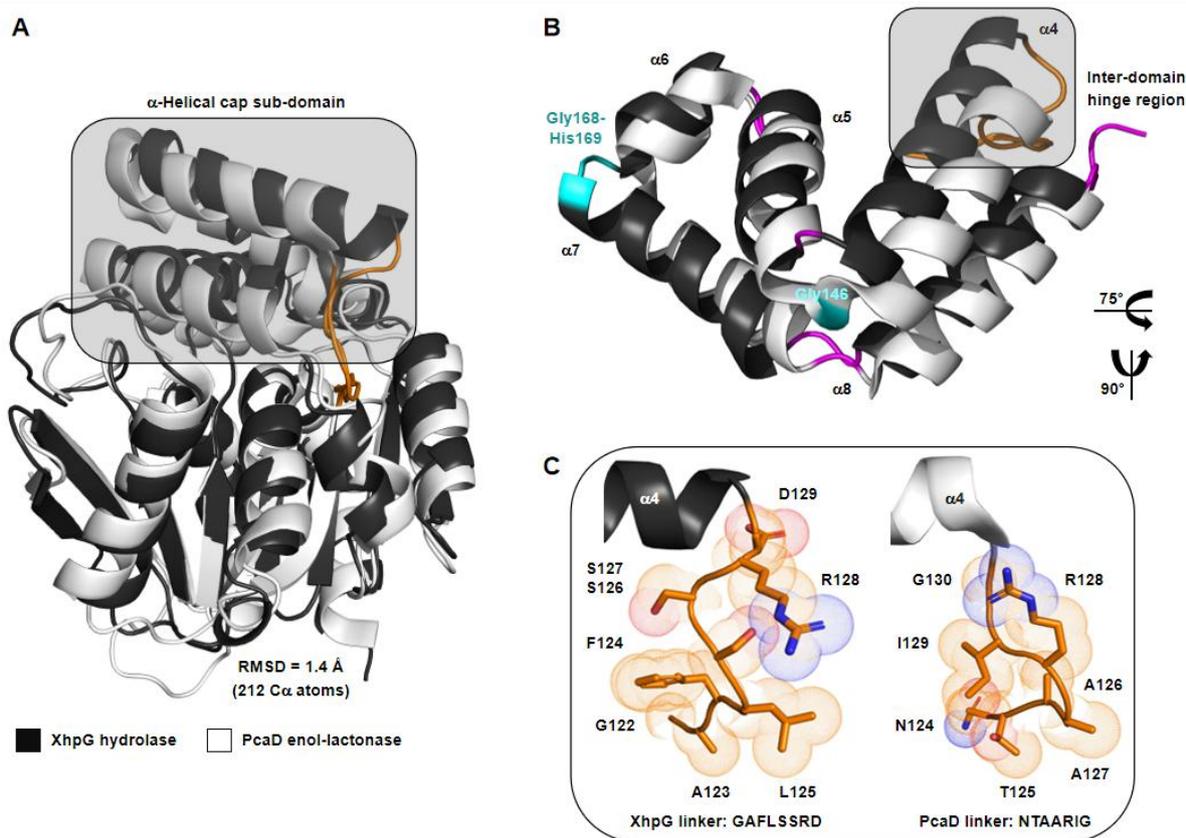
613



614

615 **Figure S17.** Purification of the mutant XhpG S98A hydrolase by liquid
 616 chromatography. The enzyme was purified by immobilised nickel ion affinity
 617 chromatography (noted IMAC), followed by dialysis in the presence of the HRV-3C
 618 protease, in order to cleave the SUMO and His₆ tags, and a second round of IMAC
 619 (denoted as invIMAC). Final polishing was achieved by size exclusion
 620 chromatography (labelled GF) on a Superdex-75 column, with the final yield
 621 estimated at 7 mg/L. SnSo: sonication supernatant; R: residual fraction of protein
 622 bound to resin before washing; FT: fraction of protein eluted by simple washing;
 623 SUMO: residual fraction of protein eluted by competition with 1 M imidazole;
 624 XhpGS98A: final XhpG S98A hydrolase sample.

625



633

634 **Figure S19.** Structural superposition of the XhpG and PcaD enzymes. **A)** The

635 homologues superimpose with a calculated RMSD value of 1.4 Å (212 C α atoms),

636 with minor deviations in the α -helical sub-domains. The XhpG hydrolase and the

637 PcaD enol-lactonase are depicted in black and white, respectively, with the inter-

638 domain hinge region indicated in orange. **B)** The cap sub-domains each contain five

639 α -helices, labelled α 4– α 8. Observable differences include the position and

640 orientation of the helix α 4, the junction between the α 4– α 5 loop and the adjacent

641 helix (a residue corresponding to Gly146 is absent in XhpG), and in the relative

642 length of the helix α 7 as an additional turn (Gly168-His169) is present in XhpG

643 (highlighted in cyan). The loops located between the α -helices are depicted in

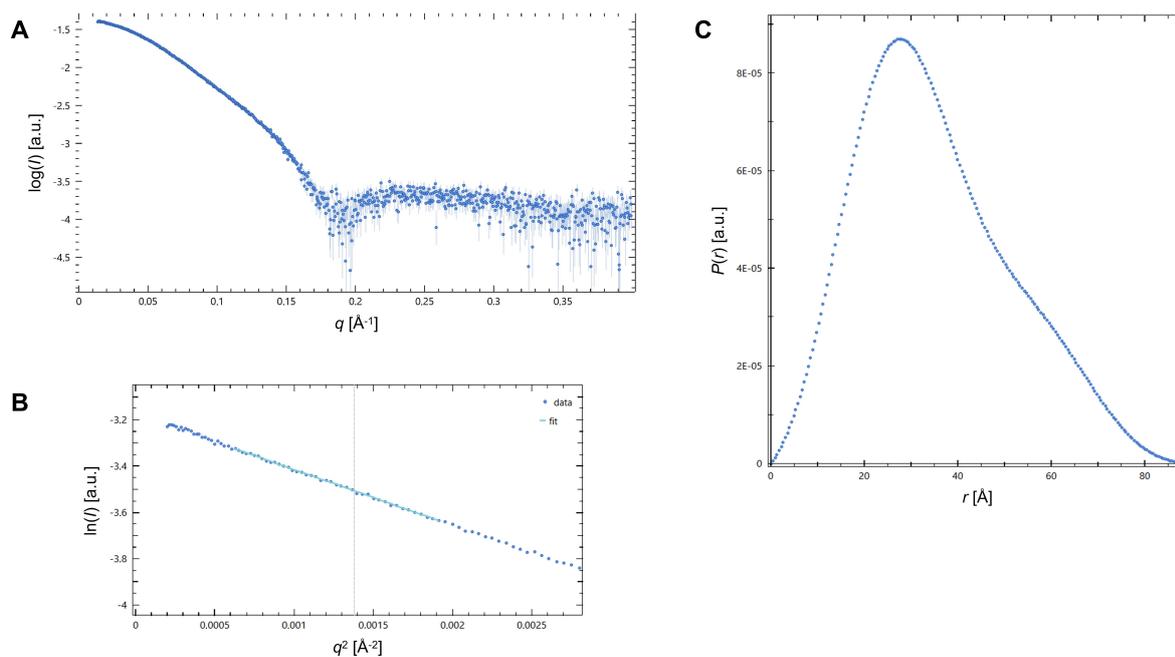
644 magenta in both structures, except for the inter-domain hinge loop, as mentioned. **C)**

645 The β 6– α 4 linker determines the position and orientation of the adjacent helix α 4. In

646 both crystal structures, the loop displays a high degree of rigidity, as evidenced by

647 average thermal agitation B -factors of 15–25 Å².

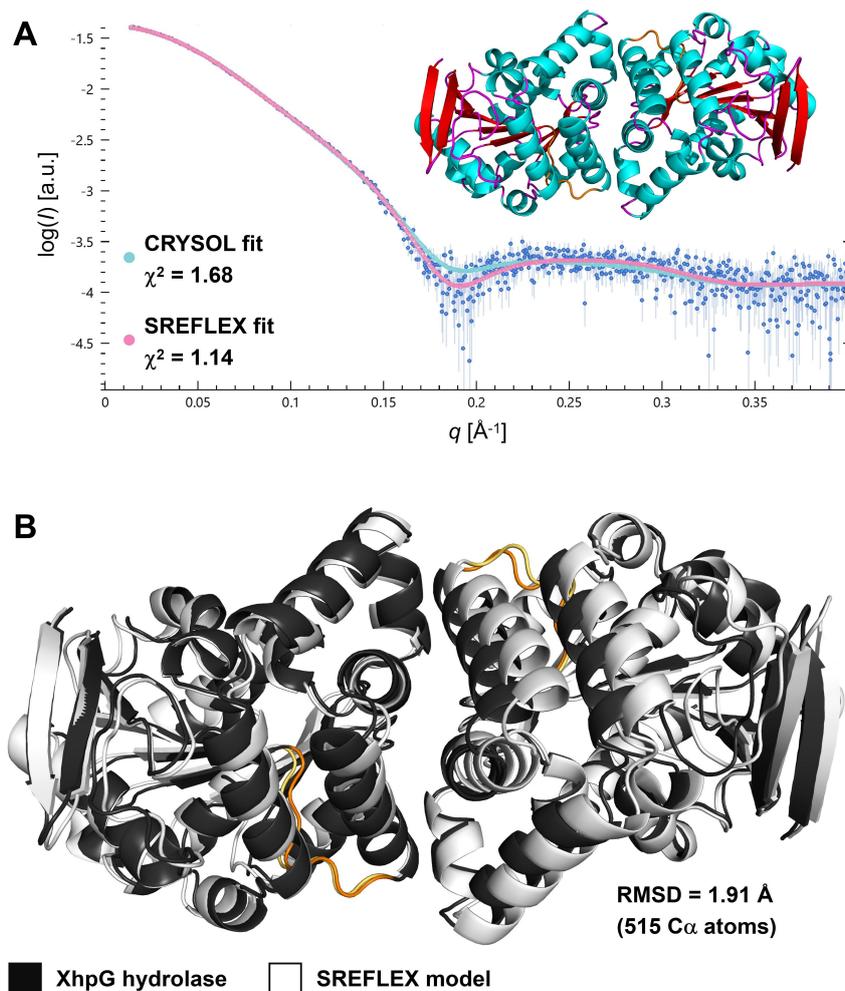
648



649

650 **Figure S20.** Analysis of the mutant XhpG S98A hydrolase by small-angle X-ray
 651 scattering (SAXS). **A**) The obtained curve expresses the logarithm of the diffusion
 652 intensity $\log(I)$ of the protein as a function of the scattering vector q in a characteristic
 653 range (3.6×10^{-3} – $5.5 \times 10^{-1} \text{\AA}^{-1}$) of the incident energy (0.77\AA , 16 keV) and the
 654 sample-to-detector distance (3 m). **B**) Analysis of linear data in the low q range of the
 655 Guinier curve afforded an initial estimate of the intensity at zero angle ($I(0)$) and the
 656 radius of gyration R_g . **C**) An inverse Fourier transform of the data up to 0.3\AA^{-1}
 657 provided access to the interatomic distance distribution within the molecule ($P(r)$
 658 function), as well as the parameters $I(0)$, R_g , and D_{\max} (maximum distance) (see
 659 Table S7).

660

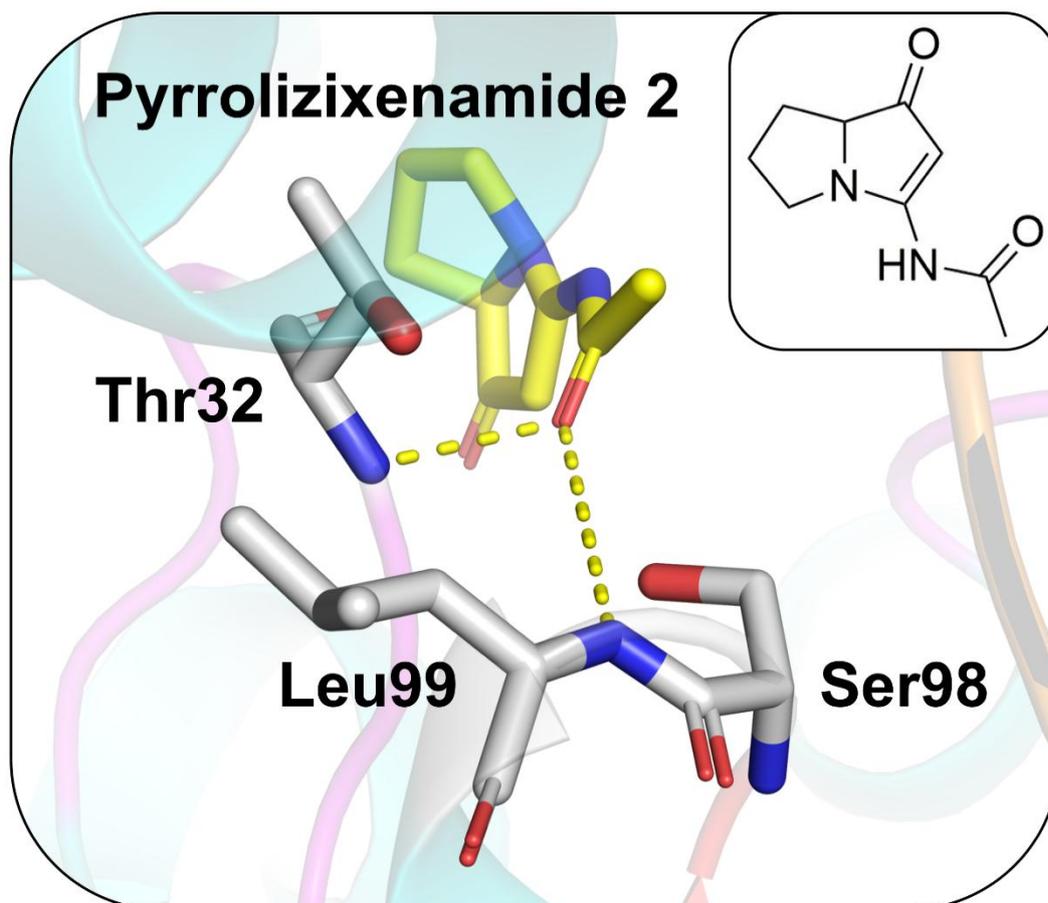


661

662 **Figure S21.** Dimerization of the XhpG hydrolase in solution and conformer modelling.

663 **A)** The curve obtained using CRYSQL^[25] shows good agreement ($\chi^2 = 1.68$) with the
 664 theoretical scattering data computed for a dimer and the experimental SAXS curve,
 665 which demonstrates that XhpG hydrolase is dimeric in solution. **B)** Molecular models
 666 generated using SREFLEX^[26] suggest that minor conformational adjustments occur
 667 in the α -helical sub-domain and the inter-domain hinge region (<2.5 Å) relative to the
 668 crystal structure. Globally, this analysis supports the proposal that XhpG adopts a
 669 closed structure, at least in the absence of substrate.

670

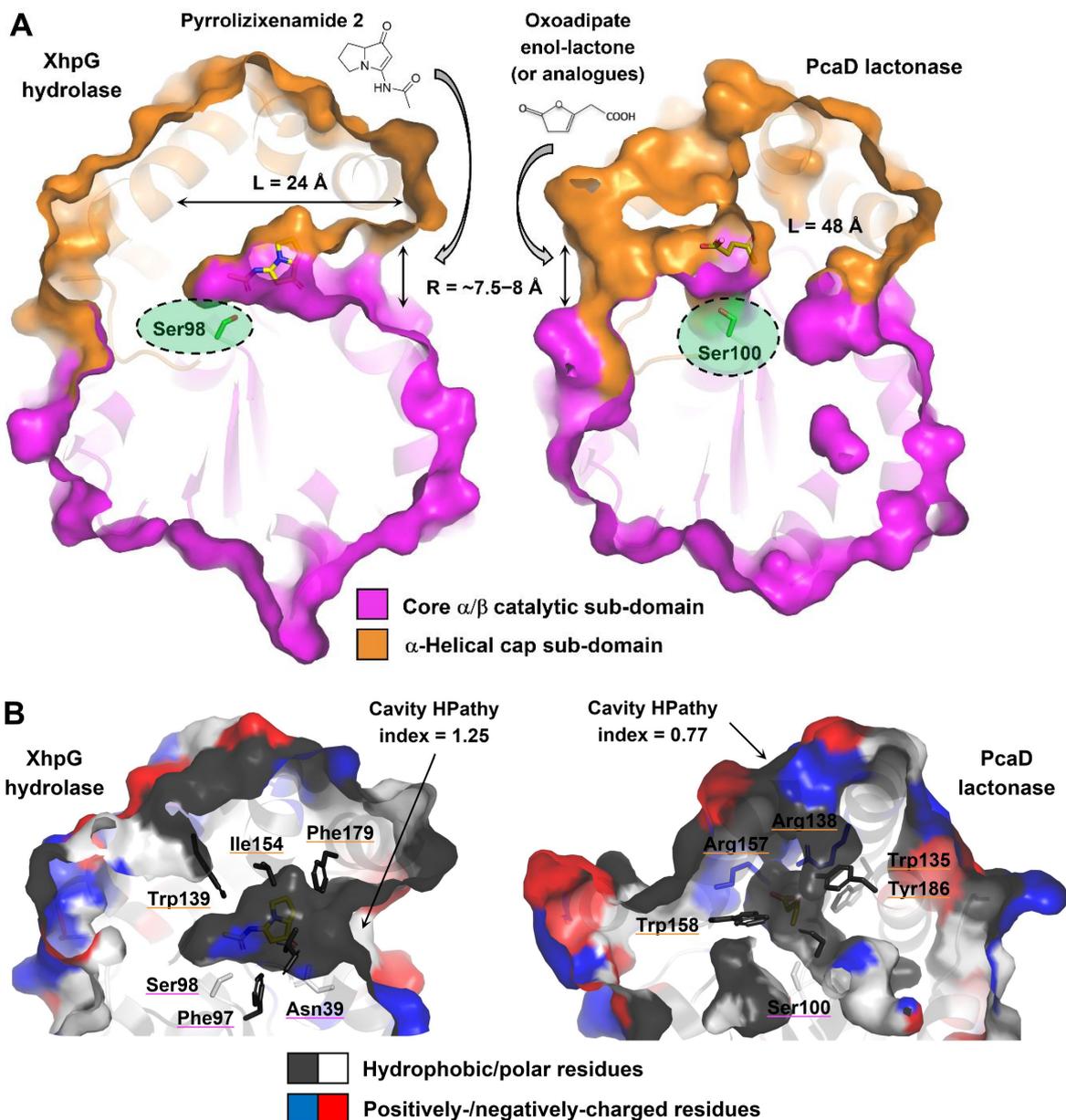


671

672 **Figure S22.** Residues comprising the oxanion hole in XhpG (Thr32 and Leu99),
673 shown relative to the catalytic serine (S98).

674

675



676

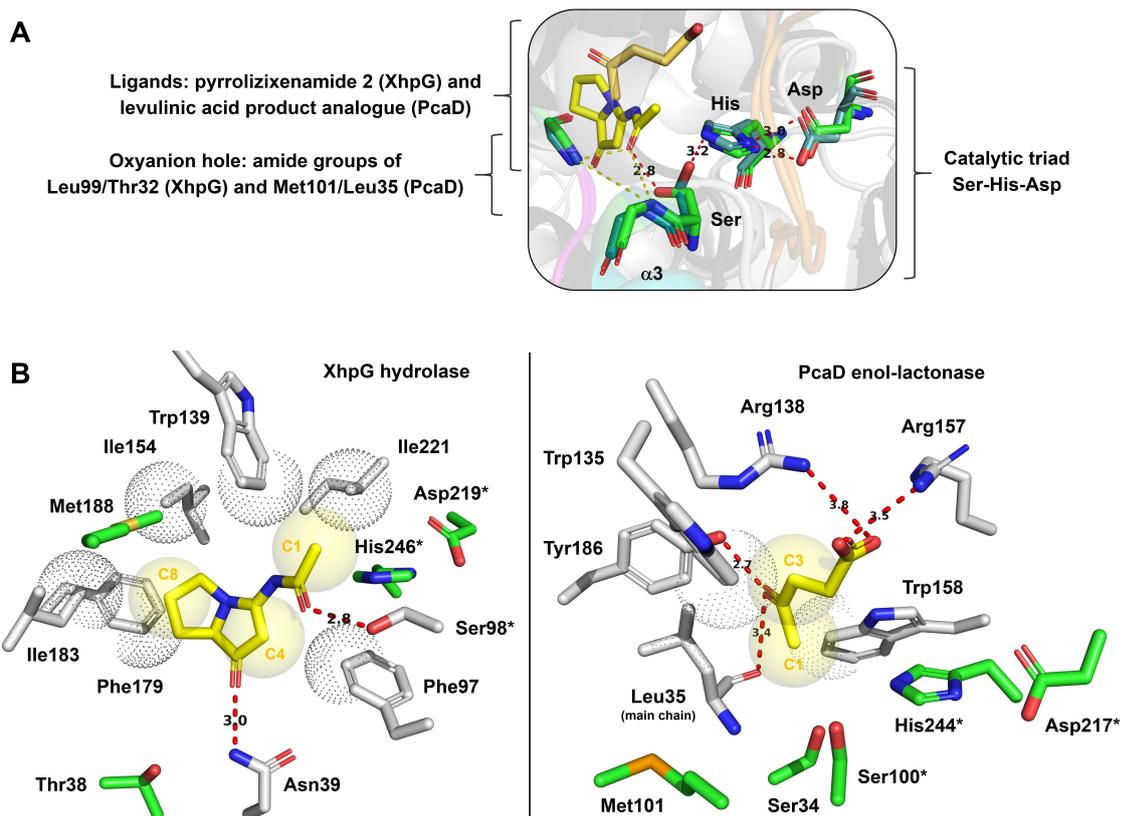
677

678 **Figure S23.** Comparison of the XhpG hydrolase and PcaD enol-lactonase structures.

679 **A)** The catalytic α/β core and α -helical sub-domains, depicted in magenta and
 680 orange respectively, form an extended active site at their interface (185 \AA^2 (XhpG)
 681 and 135 \AA^2 (PcaD)). The enzymes differ markedly in terms of the location of the
 682 substrate entry tunnel, and their active sites exhibit distinct dimensions and shapes.
 683 However, the structural organisation of the catalytic pocket remains similar, notably
 684 with the nucleophilic Ser residue of the α/β core (Ser98 (XhpG) and Ser100 (PcaD),
 685 green circles), situated at a distance of approximately 15 \AA from the protein surface.
 686 **B)** The occluded cavity in both enzymes is globally hydrophobic with a positive
 687 hydropathy index (1.25 (XhpG) and 0.77 (PcaD); calculated using the PDBsum

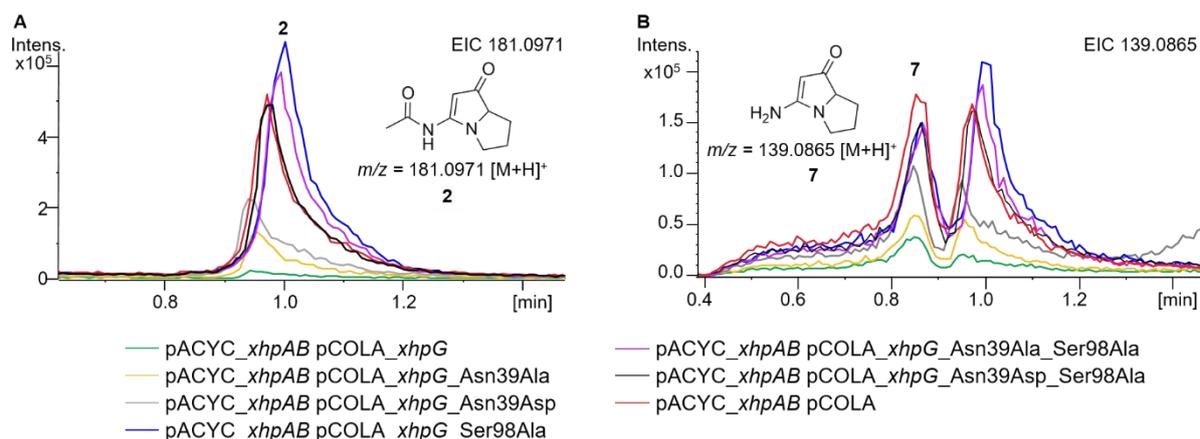
688 program^[36]). Multiple aliphatic and aromatic residues contributed by the α -helical
689 subdomain, as well as polar (XhpG) and charged amino acids (PcaD), mediate
690 ligand binding. Hydrophobic, polar, positively- and negatively-charged residues are
691 indicated in black, white, blue, and red, respectively.

692



693

694 **Figure S24.** Comparison of the ligand binding modes in the XhpG and PcaD active
 695 sites. **A**) The active site of each enzyme contains the conserved catalytic Ser-His-
 696 Asp triad, whose C α atoms superimpose with RMSDs of 0.6 Å, 0.5 Å and 0.6 Å,
 697 respectively, between the structural homologues. However, the observed rotamers of
 698 the nucleophilic residues are different, with the Ser98 hydroxyl group of XhpG
 699 aligned along the axis of the oxyanion hole, and that of PcaD (Ser100) oriented for a
 700 hydrogen bond with the N π atom of the His base. Another atom of this His residue
 701 (N π) participates in a hydrogen bond with the carboxylate group of the Asp residue,
 702 as in the majority of the α/β hydrolases clade members^[37]. The oxyanion hole
 703 comprises the amide groups of residues FSL (XhpG) and LSM (PcaD). The oxyanion
 704 holes are formed by the backbone NH of residues pairs Leu99/Thr32 (XhpG) and
 705 Met101/Leu35 (PcaD). **B**) Comparison between the binding of the pyrrolizixenamide
 706 substrate **2** in the XhpG active site, and the levulinic acid product analogue to PcaD.
 707 This analysis revealed that both hydrophobic (spheres) and polar interactions (dotted
 708 lines) (Asn39 and Ser98 in XhpG, and Arg138, Arg157 and Tyr186 in PcaD)
 709 contribute to recognition of the ligands. Residues directly involved in binding are
 710 displayed in grey and catalytic residues are indicated with an asterisk
 711 (Ser98*/Ser100*, Asp219*/Asp217*, His246*/His244*).



712

713 **Figure S25.** Heterologous expression of *xhpAB* alongside different mutants of
 714 *xhpG* in *E. coli*. Shown are the EICs corresponding to substrate **2** (m/z calc'd for
 715 $C_9H_{12}N_2O_2 [M+H]^+ = 181.0971$) (**A**) and product **7** (m/z calc'd for $C_7H_{12}N_2O [M+H]^+ =$
 716 139.0865) obtained by hydrolysis of **2** (**B**). The second peak in B at 1.0 min
 717 corresponds to fragment ion of **7** derived from fragmentation of **2** (see Figure S7).

718

719

720

721 References

- 722 [1] T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Brucoleri, S. Y. Lee, M. A. Fischbach, R.
 723 Müller, W. Wohlleben, R. Breitling, E. Takano, M. H. Medema, *Nucleic Acids Res.* **2015**, *43*,
 724 W237-W243.
- 725 [2] K. Blin, H. U. Kim, M. H. Medema, T. Weber, *Brief. Bioinform.* **2017**, *20*, 1103-1113.
- 726 [3] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber,
 727 *Nucleic Acids Res.* **2019**, *47*, W81-W87.
- 728 [4] C. L. M. Gilchrist, T. J. Booth, B. van Wersch, L. van Grieken, M. H. Medema, Y.-H. Chooi,
 729 *Bioinform. Adv.* **2021**, *1*.
- 730 [5] M. A. Westphalen, (Frankfurt am Main), **2021**.
- 731 [6] H. C. Birnboim, J. Doly, *Nucleic Acids Res.* **1979**, *7*, 1513-1523.
- 732 [7] W. J. Dower, J. F. Miller, C. W. Ragsdale, *Nucleic Acids Res.* **1988**, *16*, 6127-6145.
- 733 [8] E. Bode, A. O. Brachmann, C. Kegler, R. Simsek, C. Dauth, Q. Zhou, M. Kaiser, P. Klemmt, H. B.
 734 Bode, *ChemBioChem.* **2015**, *16*, 1115-1119.
- 735 [9] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E.
 736 Bourne, *Nucleic Acids Res.* **2000**, *28*, 235-242.
- 737 [10] L. Holm, A. Laiho, P. Törönen, M. Salgado, *Protein Sci.* **2023**, *32*, e4519.
- 738 [11] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F.
 739 Valentin, I. M. Wallace, A. Wilm, R. Lopez, *Bioinformatics.* **2007**, *23*, 2947-2948.
- 740 [12] X. Robert, P. Gouet, *Nucleic Acids Res.* **2014**, *42*, W320-W324.
- 741 [13] E. Gasteiger, C. Hoogland, A. Gattiker, S. e. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch, in
 742 *The Proteomics Protocols Handbook* (Ed.: J. M. Walker), Humana Press, Totowa, NJ, **2005**, pp.
 743 571-607.

- 744 [14] a) W. Kabsch, *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 125-132; b) M. D. Winn, C. C.
745 Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G.
746 Leslie, A. McCoy, *Acta Crystallogr. D. Biol. Crystallogr.* **2011**, *67*, 235-242.
- 747 [15] A. Vagin, A. Teplyakov, *J. Appl. Crystallogr.* **1997**, *30*, 1022-1025.
- 748 [16] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R.
749 Bates, A. Žídek, A. Potapenko, *Nature*. **2021**, *596*, 583-589.
- 750 [17] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*,
751 486-501.
- 752 [18] G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D.
753 Winn, F. Long, A. A. Vagin, *Acta Crystallogr. D. Biol. Crystallogr.* **2011**, *67*, 355-367.
- 754 [19] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W.
755 Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 12-
756 21.
- 757 [20] J. Y. Young, J. D. Westbrook, Z. Feng, R. Sala, E. Peisach, T. J. Oldfield, S. Sen, A. Gutmanas, D.
758 R. Armstrong, J. M. Berrisford, *Structure*. **2017**, *25*, 536-545.
- 759 [21] Y. Liu, X. Yang, J. Gan, S. Chen, Z.-X. Xiao, Y. Cao, *Nucleic Acids Res.* **2022**, *50*, W159-W164.
- 760 [22] L. Schrodinger, *Version* **2015**, *1*, 8.
- 761 [23] A. Thureau, P. Roblin, J. Pérez, *J. Appl. Crystallogr.* **2021**, *54*, 1698-1710.
- 762 [24] K. Manalastas-Cantos, P. V. Konarev, N. R. Hajizadeh, A. G. Kikhney, M. V. Petoukhov, D. S.
763 Molodenskiy, A. Panjkovich, H. D. Mertens, A. Gruzinov, C. Borges, *J. Appl. Crystallogr.* **2021**,
764 *54*, 343-355.
- 765 [25] D. Franke, M. Petoukhov, P. Konarev, A. Panjkovich, A. Tuukkanen, H. Mertens, A. Kikhney, N.
766 Hajizadeh, J. Franklin, C. Jeffries, *J. Appl. Crystallogr.* **2017**, *50*, 1212-1225.
- 767 [26] A. Panjkovich, D. I. Svergun, *Phys. Chem. Chem. Phys.* **2016**, *18*, 5707-5719.
- 768 [27] P. Fu, A. Legako, S. La, J. B. MacMillan, *Chemistry*. **2016**, *22*, 3491-3495.
- 769 [28] B. B. Snider, J. R. Duvall, *Org. Lett.* **2004**, *6*, 1265-1268.
- 770 [29] W. Gish, D. J. States, *Nat. Genet.* **1993**, *3*, 266-272.
- 771 [30] C. Fu, W. P. Donovan, O. Shikapwashya-Hasser, X. Ye, R. H. Cole, *PLoS One*. **2015**, *9*, e115318.
- 772 [31] P. Tailliez, S. Pagès, N. Ginibre, N. Boemare, *Int. J. Syst. Evol. Microbiol.* **2006**, *56*, 2805-2818.
- 773 [32] N. Neubacher, N. J. Tobias, M. Huber, X. Cai, T. Glatzer, S. J. Pidot, T. P. Stinear, A. L. Lütticke,
774 K. Papenfort, H. B. Bode, *Nat. Microbiol.* **2020**, *5*, 1481-1489.
- 775 [33] E. Bode, A. K. Heinrich, M. Hirschmann, D. Abebew, Y. N. Shi, T. D. Vo, F. Wesche, Y. M. Shi, P.
776 Grün, S. Simonyi, N. Keller, Y. Engel, S. Wenski, R. Bennet, S. Beyer, I. Bischoff, A. Buaya, S.
777 Brandt, I. Cakmak, H. Çimen, S. Eckstein, D. Frank, R. Fürst, M. Gand, G. Geisslinger, S. Hazir,
778 M. Henke, R. Heermann, V. Lecaudey, W. Schäfer, S. Schiffmann, A. Schöffler, R. Schwenk, M.
779 Skaljic, E. Thines, M. Thines, T. Ulshöfer, A. Vilcinskas, T. A. Wichelhaus, H. B. Bode, *Angew.*
780 *Chem. Int. Ed.* **2019**, *58*, 18957-18963.
- 781 [34] C. Kegler, H. B. Bode, *Angew. Chem. Int. Ed.* **2020**, *59*, 13463-13467.
- 782 [35] K. A. J. Bozhüyük, L. Präve, C. Kegler, L. Schenk, S. Kaiser, C. Schelhas, Y.-N. Shi, W.
783 Kuttelochner, M. Schreiber, J. Kandler, M. Alanjary, T. M. Mohiuddin, M. Groll, G. K. A.
784 Hochberg, H. B. Bode, *Science*. **2024**, *383*, eadg4320.
- 785 [36] R. A. Laskowski, J. Jabłońska, L. Pravda, R. S. Vařeková, J. M. Thornton, *Protein Sci.* **2018**, *27*,
786 129-134.
- 787 [37] J. Bains, L. Kaufman, B. Farnell, M. J. Boulanger, *J. Mol. Biol.* **2011**, *406*, 649-658.

788



Full wwPDB X-ray Structure Validation Report ⓘ

Oct 31, 2023 – 02:34 pm GMT

PDB ID : 8QZU
Title : XhpG hydrolase S98A
Deposited on : 2023-10-30
Resolution : 1.18 Å (reported)

This wwPDB validation report is for manuscript review

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<https://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtrriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)

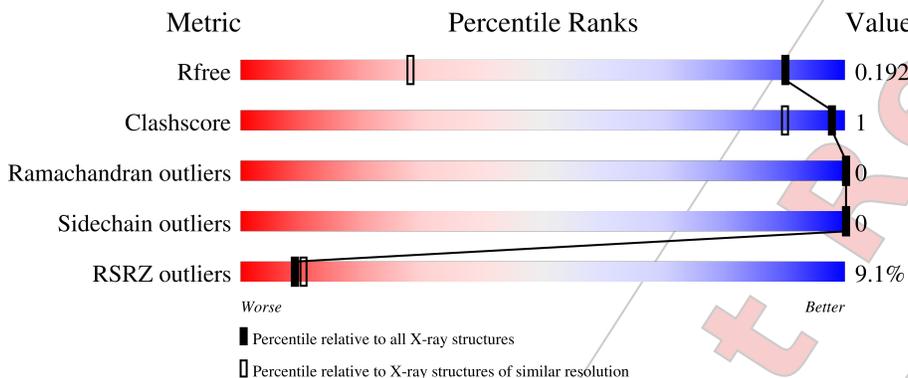
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.18 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1123 (1.20-1.16)
Clashscore	141614	1182 (1.20-1.16)
Ramachandran outliers	138981	1134 (1.20-1.16)
Sidechain outliers	138945	1134 (1.20-1.16)
RSRZ outliers	127900	1102 (1.20-1.16)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	270	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 2285 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

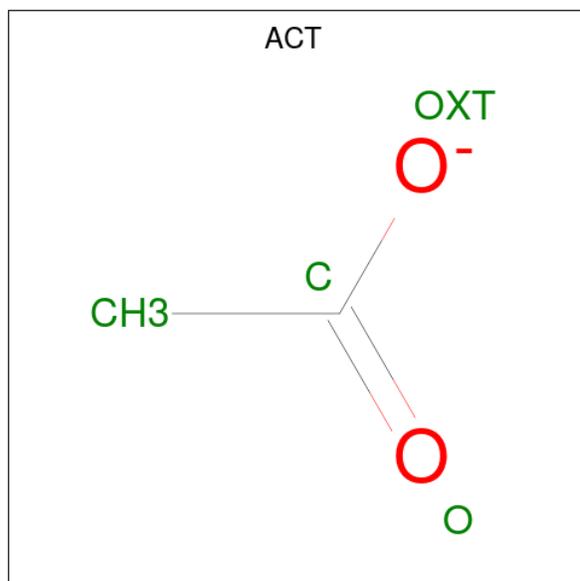
- Molecule 1 is a protein called Carboxylesterase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	265	2050	1297	363	385	5	37	2	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP A0A2G0QDP0
A	2	PRO	-	expression tag	UNP A0A2G0QDP0
A	3	GLY	-	expression tag	UNP A0A2G0QDP0
A	4	SER	-	expression tag	UNP A0A2G0QDP0
A	98	ALA	SER	engineered mutation	UNP A0A2G0QDP0

- Molecule 2 is ACETATE ION (three-letter code: ACT) (formula: C₂H₃O₂) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			4	2	2		

- Molecule 3 is water.

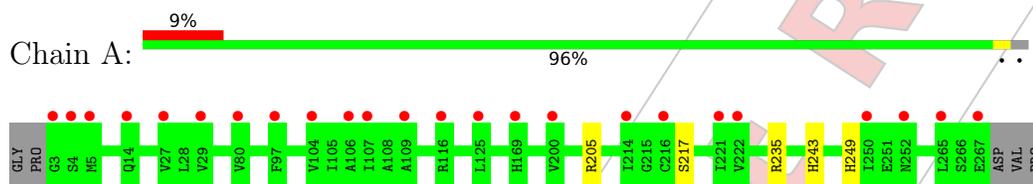
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	231	Total	O	0	0
			231	231		

For Manuscript Review

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Carboxylesterase



4 Data and refinement statistics i

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	48.83Å 68.54Å 75.48Å 90.00° 98.11° 90.00°	Depositor
Resolution (Å)	20.00 – 1.18 39.50 – 1.18	Depositor EDS
% Data completeness (in resolution range)	97.9 (20.00-1.18) 97.9 (39.50-1.18)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.19 (at 1.18Å)	Xtrriage
Refinement program	REFMAC 5.8.0405	Depositor
R, R_{free}	0.166 , 0.196 0.170 , 0.192	Depositor DCC
R_{free} test set	3921 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	17.6	Xtrriage
Anisotropy	0.382	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 49.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.98	EDS
Total number of atoms	2285	wwPDB-VP
Average B, all atoms (Å ²)	21.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.19% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: ACT

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.66	0/2089	0.84	0/2835

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2050	0	2057	3	0
2	A	4	0	3	0	0
3	A	231	0	0	0	0
All	All	2285	0	2060	3	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 1.

All (3) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:249:HIS:HD1	1:A:249:HIS:H	1.42	0.67

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:217:SER:OG	1:A:243:HIS:HA	2.09	0.52
1:A:205:ARG:O	1:A:235:ARG:NH1	2.46	0.45

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	265/270 (98%)	261 (98%)	4 (2%)	0	100 100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	222/224 (99%)	222 (100%)	0	100 100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
2	ACT	A	301	-	3,3,3	0.83	0	3,3,3	0.33	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

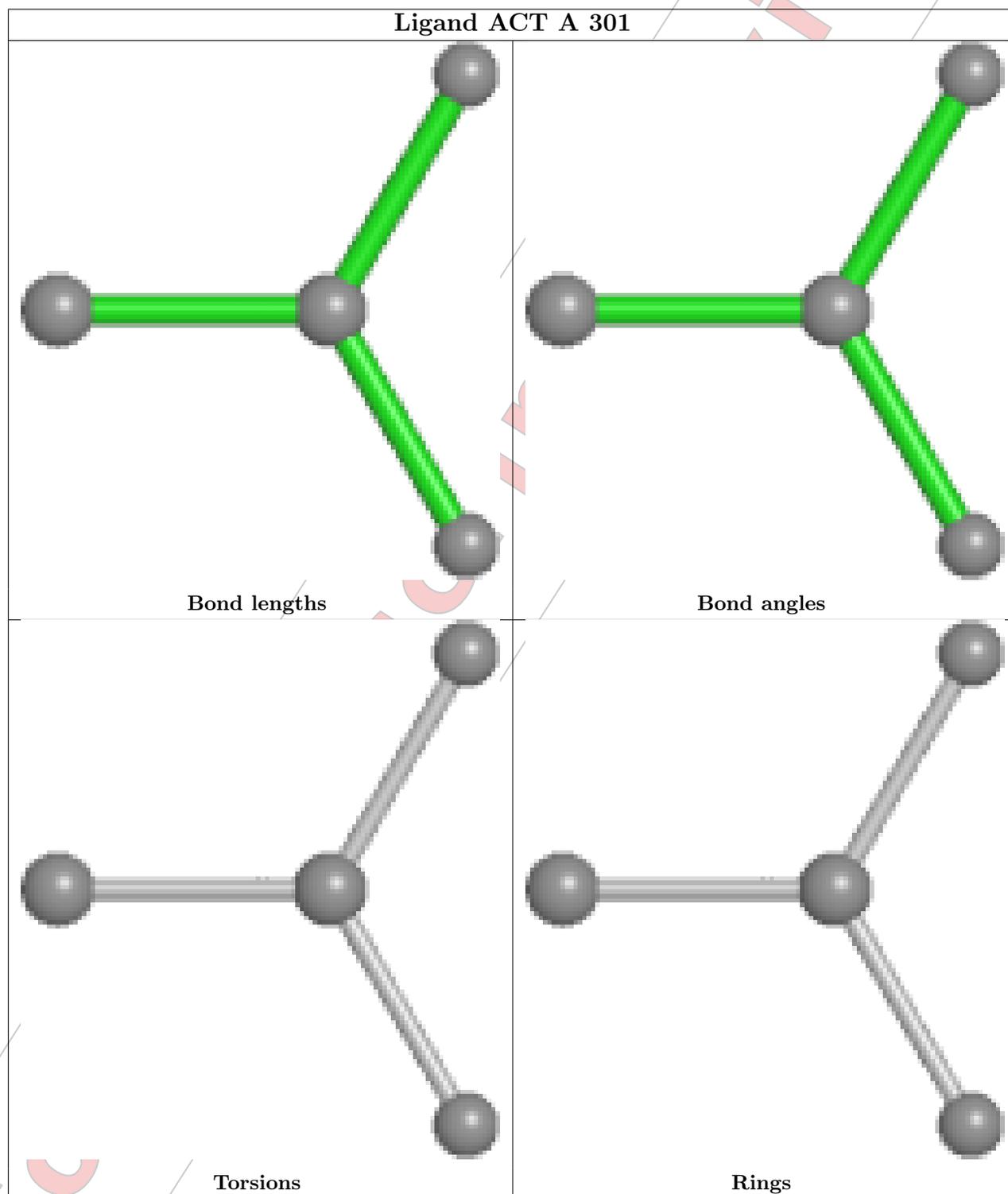
There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring

in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

For Manuscript Review

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2	OWAB(Å ²)	Q < 0.9
1	A	265/270 (98%)	1.03	24 (9%) 9 10	13, 18, 30, 64	18 (6%)

All (24) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	5	MET	3.2
1	A	3	GLY	3.2
1	A	221	ILE	3.0
1	A	109	ALA	3.0
1	A	27	VAL	2.7
1	A	252	ASN	2.7
1	A	222	VAL	2.7
1	A	214	ILE	2.5
1	A	104	VAL	2.5
1	A	29	VAL	2.5
1	A	216	CYS	2.5
1	A	106	ALA	2.4
1	A	169[A]	HIS	2.4
1	A	250	ILE	2.4
1	A	14	GLN	2.3
1	A	200	VAL	2.2
1	A	265	LEU	2.2
1	A	107	ILE	2.2
1	A	125	LEU	2.1
1	A	80	VAL	2.1
1	A	267	GLU	2.1
1	A	4	SER	2.0
1	A	116	ARG	2.0
1	A	97	PHE	2.0

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

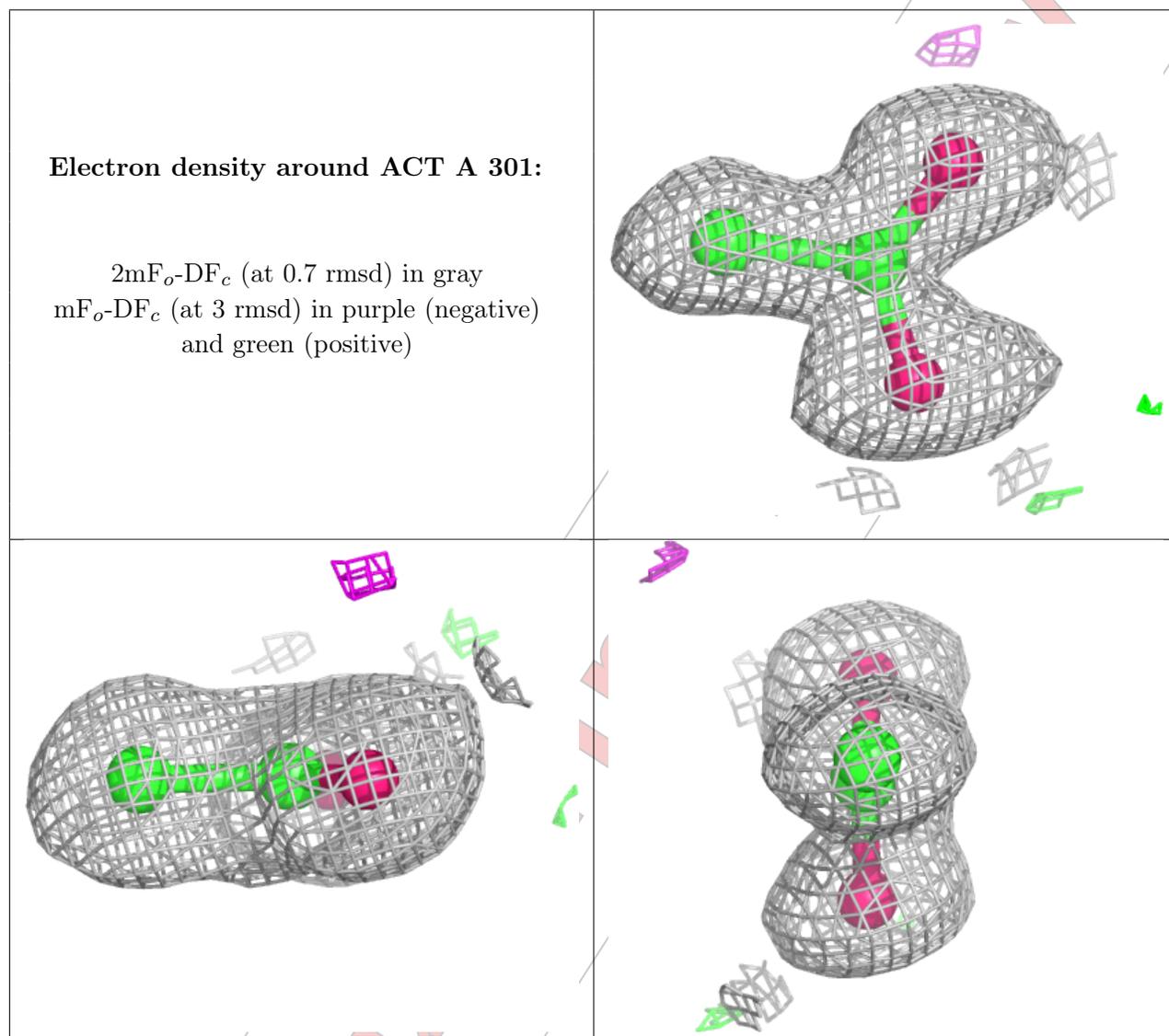
There are no monosaccharides in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q<0.9
2	ACT	A	301	4/4	0.95	0.10	12,14,15,15	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.



6.5 Other polymers [i](#)

There are no such residues in this entry.

For Manuscript



Full wwPDB X-ray Structure Validation Report i

Feb 28, 2024 – 05:10 pm GMT

PDB ID : 8S6L
Title : AzeD dehydratase of Pseudomonas aeruginosa
Deposited on : 2024-02-27
Resolution : 1.55 Å (reported)

This wwPDB validation report is for manuscript review

This is a Full wwPDB X-ray Structure Validation Report.

This report is produced by the wwPDB biocuration pipeline after annotation of the structure.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the i symbol.

The types of validation reports are described at

<https://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references](#) i) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtrriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

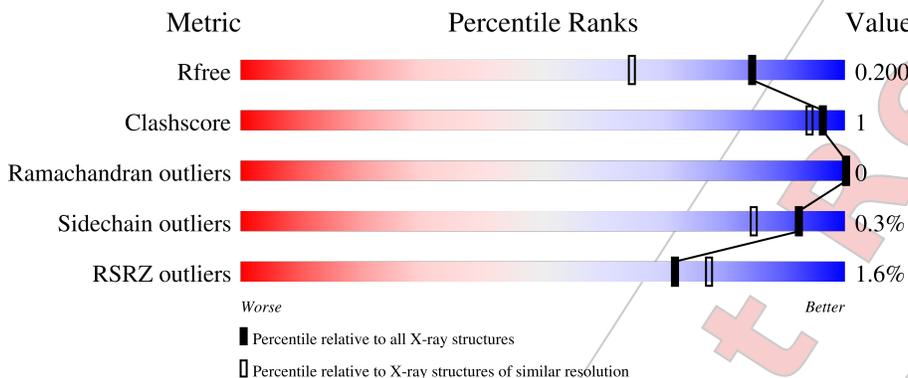
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.55 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1483 (1.56-1.56)
Clashscore	141614	1529 (1.56-1.56)
Ramachandran outliers	138981	1498 (1.56-1.56)
Sidechain outliers	138945	1495 (1.56-1.56)
RSRZ outliers	127900	1465 (1.56-1.56)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	461	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 3763 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

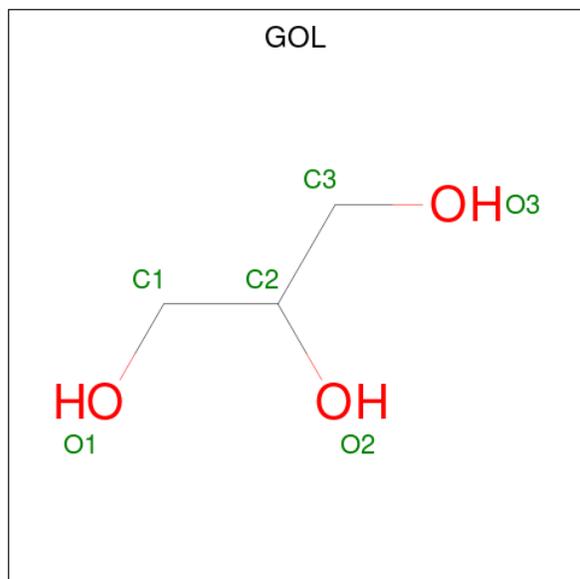
- Molecule 1 is a protein called Phthiocerol/phthiodiolone dimycocerosyl transferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	433	3411	2131	648	625	7	21	2	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	initiating methionine	UNP Q9HYR6
A	-18	GLY	-	expression tag	UNP Q9HYR6
A	-17	SER	-	expression tag	UNP Q9HYR6
A	-16	SER	-	expression tag	UNP Q9HYR6
A	-15	HIS	-	expression tag	UNP Q9HYR6
A	-14	HIS	-	expression tag	UNP Q9HYR6
A	-13	HIS	-	expression tag	UNP Q9HYR6
A	-12	HIS	-	expression tag	UNP Q9HYR6
A	-11	HIS	-	expression tag	UNP Q9HYR6
A	-10	HIS	-	expression tag	UNP Q9HYR6
A	-9	SER	-	expression tag	UNP Q9HYR6
A	-8	SER	-	expression tag	UNP Q9HYR6
A	-7	GLY	-	expression tag	UNP Q9HYR6
A	-6	LEU	-	expression tag	UNP Q9HYR6
A	-5	VAL	-	expression tag	UNP Q9HYR6
A	-4	PRO	-	expression tag	UNP Q9HYR6
A	-3	ARG	-	expression tag	UNP Q9HYR6
A	-2	GLY	-	expression tag	UNP Q9HYR6
A	-1	SER	-	expression tag	UNP Q9HYR6
A	0	HIS	-	expression tag	UNP Q9HYR6

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			6	3	3		
2	A	1	Total	C	O	0	0
			6	3	3		

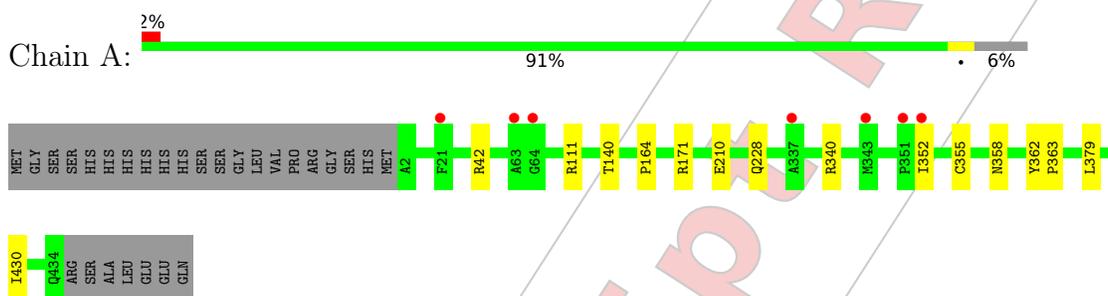
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	340	Total	O	0	0
			340	340		

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ($RSRZ > 2$). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Phthiocerol/phthiodiolone dimycocerosyl transferase



4 Data and refinement statistics 

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	56.46Å 67.59Å 116.68Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.20 – 1.55 44.16 – 1.55	Depositor EDS
% Data completeness (in resolution range)	99.9 (44.20-1.55) 99.9 (44.16-1.55)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.94 (at 1.55Å)	Xtrriage
Refinement program	REFMAC 5.8.0405	Depositor
R, R_{free}	0.146 , 0.194 0.154 , 0.200	Depositor DCC
R_{free} test set	3261 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	24.4	Xtrriage
Anisotropy	0.281	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 43.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	3763	wwPDB-VP
Average B, all atoms (Å ²)	27.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.16% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z >5	RMSZ	# Z >5
1	A	0.46	0/3492	0.77	0/4760

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	111	ARG	Sidechain
1	A	171	ARG	Sidechain
1	A	42	ARG	Sidechain

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3411	0	3361	7	0
2	A	12	0	16	2	0
3	A	340	0	0	0	0
All	All	3763	0	3377	7	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 1.

All (7) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:164:PRO:HA	2:A:502:GOL:H32	1.67	0.76
1:A:352:ILE:HA	1:A:355:CYS:SG	2.27	0.75
1:A:228:GLN:HG2	1:A:430:ILE:HG23	1.71	0.72
1:A:210:GLU:OE1	1:A:340:ARG:NH2	2.34	0.58
1:A:140:THR:HA	1:A:362:TYR:OH	2.11	0.49
1:A:358:ASN:O	2:A:501:GOL:H12	2.20	0.41
1:A:362:TYR:CD2	1:A:363:PRO:HD2	2.57	0.40

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	433/461 (94%)	425 (98%)	8 (2%)	0	100 100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar

resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	351/373 (94%)	350 (100%)	1 (0%)	92 85

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	379	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
2	GOL	A	502	-	5,5,5	0.06	0	5,5,5	0.34	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	GOL	A	501	-	5,5,5	0.09	0	5,5,5	0.33	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	A	502	-	-	3/4/4/4	-
2	GOL	A	501	-	-	3/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (6) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	502	GOL	C1-C2-C3-O3
2	A	502	GOL	O2-C2-C3-O3
2	A	501	GOL	O1-C1-C2-C3
2	A	501	GOL	O2-C2-C3-O3
2	A	501	GOL	O1-C1-C2-O2
2	A	502	GOL	O1-C1-C2-C3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	502	GOL	1	0
2	A	501	GOL	1	0

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled '#RSRZ > 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q < 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ > 2	OWAB(Å ²)	Q < 0.9
1	A	433/461 (93%)	0.03	7 (1%) 72 77	16, 25, 38, 66	19 (4%)

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	352	ILE	3.3
1	A	63	ALA	3.1
1	A	343	MET	2.9
1	A	351	PRO	2.8
1	A	21	PHE	2.5
1	A	64	GLY	2.4
1	A	337	ALA	2.3

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å ²)	Q < 0.9
2	GOL	A	502	6/6	0.80	0.26	24,25,30,30	2

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
2	GOL	A	501	6/6	0.87	0.11	43,44,47,48	2

6.5 Other polymers [i](#)

There are no such residues in this entry.

For Manuscript Review

BIBLIOGRAPHIE

T. S. Agostini-Costa, R. F. Vieira, H. R. Bizzo, D. Silveira, M. A. Gimenes. **Secondary metabolites.** *Chromatography and its application.* **2012**, 8, 131–164. doi: 10.5772/35705.

Y. Ahmed, Y. Rebets, M. Rodríguez Estévez, J. Zapp, M. Myronovskiy, A. Luzhetskyy. **Engineering of *Streptomyces lividans* for heterologous expression of secondary metabolite gene clusters.** *Microb. Cell Fact.* **2020**, 19, 5. doi: 10.1186/s12934-020-1277-8.

F. Alberti, D. J. Leng, I. Wilkening, L. Song, M. Tosin, C. Corre. **Triggering the expression of a silent gene cluster from genetically intractable bacteria results in scleric acide discovery.** *Chem. Sci.* **2019**, 10, 453–463. doi: 10.1039/c8sc03814g.

A. H. Al-Mestarihi, A. Garzan, J. M. Kim, S. Garneau-Tsodikova. **Enzymatic evidence for a revised congocidine biosynthetic pathway.** *ChemBioChem.* **2015**, 16, 1307–1313. doi: 10.1002/cbic.201402711.

J. C. Anderson, J. E. Dueber, M. Leguia, G. C. Wu, J. A. Goler, A. P. Arkin, J. D. Keasling. **BglBricks: a flexible standard for biological part assembly.** *J. Biol. Eng.* **2010**, 4, 1. doi: 10.1186/1754-1611-4-1.

I. Artsimovitch, V. Patlan, S-I. Sekine, M. N. Vassilyeva, T. Hosaka, K. Ochi, S. Yokoyama, D. G. Vassilyev. **Structural basis for transcription regulation by alarmone ppGpp.** *Cell.* **2004**, 117, 299–310. doi: 10.1016/s0092-8674(04)00401-5.

A. G. Atanasov, S. B. Zotchev, V. M. Dirsch, C. T. Supuran. **Natural products in drug discovery: advances and opportunities.** *Nat. Rev. Drug Discov.* **2021**, 20, 200–216. doi: 10.1038/s41573-020-00114-z.

D. Auerbach, F. Yan, Y. Zhang, R. Müller. **Characterization of an unusual glycerate esterification process in vioprolide biosynthesis.** *ACS Chem. Biol.* **2018**, 13, 3123–3130. doi: 10.1021/acscchembio.8b00826.

J. Bains, L. Kaufman, B. Farnell, M. J. Boulanger. **A product analog bound form of 3-oxoadipate-enol-lactonase (PcaD) reveals a multifunctional role for the divergent cap domain.** *J. Mol. Biol.* **2011**, 406, 649–658. doi: 10.1016/j.jmb.2011.01.007.

C. J. Balibar, F. H. Vaillancourt, C. T. Walsh. **Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains.** *Chem. Biol.* **2005**, 12, 1189–1200. doi: 10.1016/j.chembiol.2005.08.010.

- R. H. Baltz. **Function of MbtH homologs in nonribosomal peptide biosynthesis and applications in secondary metabolites discovery.**
J. Ind. Microbiol. Biotechnol. **2011**, *38*, 1747–1760. doi: 10.1007/s10295-011-1022-8.
- R. H. Baltz. **Natural product drug discovery in the genomic era: realities, conjectures, misconceptions, and opportunities.**
J. Ind. Microbiol. Biotechnol. **2019**, *46*, 281–299. doi: 10.1007/s10295-018-2115-4.
- P. Banerjee, J. Erehman, B-O. Gohlke, T. Wilhelm, R. Preissner, M. Dunkel. **Super Natural II: a database of natural products.** *Nucleic Acids Res.* **2015**, *43*, 935–939. doi: 10.1093/nar/gku886.
- J. F. Barajas, R. M. Phelan, A. J. Schaub, J. T. Kliewer, P. J. Kelly, D. R. Jackson, R. Luo, J. D. Keasling, S-C Tsai. **Comprehensive structural and biochemical analysis of the terminal myxalamid reductase domain for the engineered production of primary alcohols.**
Chem. Biol. **2015**, *22*, 1018–1029. doi: 10.1016/j.chembiol.2015.06.022.
- F. M. Barzak, T. M. Ryan, M. V. Kvach, H. M. Kurup, H. Aihara, R. S. Harris, V. V. Filichev, E. Harjes, G. B. Jameson. **Small-angle X-ray scattering models of APOBEC3B catalytic domain in a complex with a single-stranded DNA inhibitor.** *Viruses.* **2021**, *13*, 290. doi: 10.3390/v13020290.
- J. Beld, E. C. Sonnenschein, C. R. Vickery, J. P. Noel, M. D. Burkart. **The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life.**
Nat. Prod. Rep. **2014**, *31*, 61–108. doi: 10.1039/c3np70054b.
- P. J. Belshaw, C. T. Walsh, T. Stachelhaus. **Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis.**
Science. **1999**, *284*, 486–489. doi: 10.1126/science.284.5413.486.
- V. Bergendahl, U. Linne, M. A. Marahiel. **Mutational analysis of the C-domain in nonribosomal peptide synthesis.**
Eur. J. Biochem. **2002**, *269*, 620–629. doi: 10.1046/j.0014-2956.2001.02691.x.
- H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne. **The Protein Data Bank.** *Nucleic Acids Res.* **2000**, *28*, 235–242. doi:10.1093/nar/28.1.235.
- B. Behsaz, E. Bode, A. Gurevich, Y-N. Shi, F. Grundmann, D. Acharya, A. Mauricio Caraballo-Rodriguez, A. Bouslimani, M. Panitchpakdi, A. Linck, C. Guan, J. Oh, P. C. Dorrestein, H. B. Bode, P. A. Pevzner, H. Mohimani. **Integrating genomics and metabolomics for scalable non-ribosomal peptide discovery.** *Nat. Commun.* **2021**, *12*, 4318. doi: 10.1038/s41467-021-24441-w.

- B. E. Blass. **Azabicyclic fused pyrimidine derivatives useful for the treatment of Alzheimer's disease.** *ACS Med. Chem. Lett.* **2018**, *9*, 165–166. doi: 10.1021/acsmchemlett.
- K. Bloudoff, D. Rodionov, T. M. Schmeing. **Crystal structure of the first condensation domain of CDA synthetase suggest conformational changes during synthetic cycle of nonribosomal peptide synthetases.** *J. Mol. Biol.* **2013**, *425*, 3137–3150. doi: 10.1016/j.jmb.2013.06.003.
- K. Bloudoff, T. M. Schmeing. **Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity.** *Biochim. Biophys. Acta.* **2017**, *1865*, 1587–1604. doi: 10.1016/j.bbapap.2017.05.010.
- K. Blin, S. Shaw, H. E. Augustijn, Z. L. Reitz, F. Biermann, M. Alanjary, A. Fetter, B. R. Terlouw, W. W. Metcalf, E. J. N. Helfrich, G. P. van Wezel, M. H. Medema, T. Weber. **antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation.** *Nucleic Acid Res.* **2023**, *51*, 46–50. doi: 10.1093/nar/gkad344.
- H. B. Bode, A. O. Brachmann, K. B. Jadhav, L. Seyfarth, C. Dauth, S. W. Fuchs, M. Kaiser, N. R. Waterfield, H. Sack, S. H. Heinemann, H-D. Arndt. **Structure elucidation and activity of kolossin A, the S-/L-pentadecapeptide product of a giant nonribosomal peptide synthetase.** *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 10352–10355. doi: 10.1002/anie.201502835.
- I. Borodina, J. Siebring, J. Zhang, C. P. Smith, G. van Keulen, L. Dikkhuizen, J. Nielsen. **Antibiotic overproduction in *Streptomyces coelicolor* A3(2) mediated by phosphofructokinase deletion.** *J. Biol. Chem.* **2008**, *283*, 25186–25199. doi: 10.1074/jbc.M803105200.
- M. A. Borovinskaya, S. Shoji, K. Fredrick, J. H. Cate. **Structural basis for hygromycin B inhibition of protein biosynthesis.** *RNA.* **2008**, *14*, 1590–1599. doi: 10.1261/rna.1076908.
- K. A. J. Bozhüyük, J. Watzel, N. Abbood, H. B. Bode. **Synthetic zippers as an enabling tool for engineering of non-ribosomal peptide synthetases.** *Angew. Chem. Int. Ed. Engl.* **2021**, *60*, 17531–17538. doi: 10.1002/anie.202102859.
- G. Brader, S. Compant, B. Mitter, F. Trognitz, A. Sessitsch. **Metabolic potential of endophytic bacteria.** *Curr. Opin. Biotechnol.* **2014**, *27*, 30–37. doi: 10.1016/j.copbio.2013.09.012.
- L. Briand, G. Marcion, A. Kriznik, J. M. Heydel, Y. Artur, C. Garrido, R. Seigneuric, F. Neiers. **A self-inducible heterologous protein expression system in *Escherichia coli*.** *Sci. Rep.* **2016**, *6*, 33037. doi: 10.1038/srep33037.

- H. R. Brielmann, W. N. Setzer, P. B. Kaufman, A. Kirakosyan, L. J. Cseke. **Phytochemicals: the chemical components of plants.**
Natural products from plants. **2006**, 2, 1–50. ISBN: 978-0-8493-2976-0.
- A. Broberg, A. Menkis, R. Vasiliauskas. **Kutznerides 1–4, depsipeptides from the actinomycete *Kutzneria* sp. 744 inhabiting mycorrhizal roots of *Picea abies* seedlings.**
J. Nat. Prod. **2006**, 69, 97–102. doi: 10.1021/np050378g.
- S. D. Bruner, T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh, M. T. Stubbs. **Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioestérase domain SrfTE.** *Structure.* **2002**, 10, 301–310. doi: 10.1016/s0969-2126(02)00716-5.
- D. J. Busby, R. C. Copley, J. A. Hueso, S. A. Readshaw, A. Rivera. J. **SB-253514 and analogues: novel inhibitors of lipoprotein associated phospholipase A2 produced by *Pseudomonas fluorescens* DSM 11579. II. Physico-chemical properties and structure elucidation.**
J. Antibiot. **2000**, 53, 670–676. doi: /
- T. D. H. Bugg. **Diverse catalytic activities in the alphabeta-hydrolase family of enzymes: activation of H₂O, HCN, H₂O₂, and O₂.**
Bioorg. Chem. **2004**, 32, 367–375. doi:10.1016/j.bioorg.2004.05.005.
- X. Cai, L. Zhao, H. B. Bode. **Reprogramming promiscuous nonribosomal peptide synthetases for production of specific peptides.**
Org. Lett. **2019**, 21, 2116–2120. doi:10.1021/acs.orglett.9b00395.
- P. D. Carr, D. L. Ollis. **Alpha/beta hydrolase fold: an update.**
Protein. Pept. Lett. **2009**, 16, 1137–1148. doi: 10.2174/092986609789071298.
- J. Cavarelli. **Biocristallographie - Cristallisation et collecte des données.**
Techniques de l'ingénieur. **2020**, 1110, 1–17.
- J. Cavarelli. **Biocristallographie - Détermination des phases à la structure cristallographique.**
Techniques de l'ingénieur. **2020**, 1111, 1–25.
- V. L. Challinor, H. B. Bode. **Bioactive natural products from novel microbial sources.**
Ann. N. Y. Acad. Sci. **2015**, 1354, 82–97. doi: 10.1111/nyas.12954.
- G. L. Challis. **Mining microbial genomes for new natural products and biosynthetic pathways.**
Microbiology. **2008**, 154, 1555–1569. doi: 10.1099/mic.0.2008/018523-0.

- W. T. Chan, C. S. Verma, D. P. Lane, S. K. Gan. **A comparison and optimization of methods and factors affecting the transformation of *Escherichia coli*.** *Biosci. Rep.* **2013**, *33*, e00086. doi: 10.1042/BSR20130098.
- R. Chao, Y. Yuan, H. Zhao. **Recent advances in DNA assembly technologies.** *FEMS Yeast Res.* **2015**, *15*, 1–9. doi: 10.1111/1567-1364.12171.
- V. B. Chen, W. B. Arendall III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson. **MolProbity: all-atom structure validation for macromolecular crystallography.** *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 12–21. doi: 10.1107/S0907444909042073.
- W-H. Chen, Z-J. Qin, J. Wang, G-P, Zhao. **The MASTER (Methylation-Assisted Tailorable Ends Rational) ligation method for seamless DNA assembly.** *Nucleic Acid Res.* **2013**, *41*, e93. doi: 10.1093/nar/gkt122.
- W-H. Chen, K. Li, N. S. Guntaka, S. D. Bruner. **Interdomain and intermodule organization in epimerization domain containing nonribosomal peptide synthetases.** *ACS. Chem. Biol.* **2016**, *11*, 2293–2303. doi: 10.1021/acscchembio.6b00332.
- W. Chen, B. Wang, J. D. Gruber, Y-M. Zhang, C. Davies. **Acyl carrier protein 3 is involved in oxidative stress response in *Pseudomonas aeruginosa*.** *Front. Microbiol.* **2018**, *9*, 2244. doi: 10.3389/fmicb.2018.02244.
- M. G. Chevrette, F. Aicheler, O. Kohlbacher, C. R. Currie, M. H. Medema. **SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across *Actinobacteria*.** *Bioinformatics.* **2017**, *33*, 3202–3210. doi: 10.1093/bioinformatics/btx400.
- W. G. Choi, M. Prasad, R. Lennon, R. Gulati, A. Prasad, L. O. Lerman, A. Lerman. **Long-term darapladib use does not affect coronary plaque composition assessed using multimodality intravascular imaging modalities : a randomized-controlled study.** *Coron. Artery. Dis.* **2018**, *29*, 104–113. doi: 10.1097/MCA.0000000000000573.
- P. Cimermancic, M. H. Medema, J. Claesen, K. Kurita, L. C. Wieland Brown, K. Mavrommatis, A. Pati, P. A. Godfrey, M. Koehrsen, J. Clardy, B. W. Birren, E. Takano, A. Sali, R. G. Lington, M. A. Fishbach. **Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters.** *Cell.* **2014**, *158*, 412–421. doi: 10.1016/j.cell.2014.06.034.

A. F. Cisneros, F. D. Rouleau, C. Baustista, P. Lemieux, N. Dumont-Leblond. **Toehold: a software for automated design and in silico validation of toehold riboswitches.**

PeerJ Phys. Chem. **2023**, 5, e28. doi: 10.7717/peerj-pchem.28.

C. M. Clark, M. S. Costa, L. M. Sanchez, B. T. Murphy. **Coupling MALDI-TOF mass spectrometry protein and specialized metabolites analyses to rapidly discriminate bacterial function.** *Proc. Natl. Acad. Sci. USA.* **2018**, 115, 4981–4986. doi: 10.1073/pnas.1801247115.

S. D. Colloms, C. A. Merrick, F. J. Olorunniji, W. M. Stark, M. C. M. Smith, A. Osbourn, K. J. Keasling, S. J. Rosser. **Rapid metabolic pathway assembly and modification using serine integrase site specific recombination.** *Nucleic Acid Res.* **2014**, 42, e23. doi: 10.1093/nar/gkt1101.

E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick. **Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S.**

EMBO J. **1997**, 16, 4174–4183. doi: 10.1093/emboj/16.14.4174.

J. C. Corpuz, L. M. Podust, T. D. Davis, M. J. Jaremko, M. D. Burkart. **Dynamic visualization of type II peptidyl carrier protein recognition in pyoluteorin biosynthesis.**

RCS Chem. Biol. **2020**, 1, 8–12. doi: 10.1039/c9cb00015a.

J. C Corpuz, J. O. Sanlley, M. D. Burkart. **Protein-protein interface analysis of the non-ribosomal peptide synthetase peptidyl carrier protein and enzymatic domains.** *Synth. Syst. Biotechnol.* **2022**, 7, 677–688. doi: 10.1016/j.synbio.2022.02.006.

A. Cravens, J. Payne, C. Smolke. **Synthetic biology strategies for microbial biosynthesis of plant natural products.** *Nat. Commun.* **2019**, 10, 2142. doi: 10.1038/s41467-019-09848-w.

J. S. Crawford, R. Kontnik, J. Clardy. **Regulating alternative lifestyles in entomopathogenic bacteria.** *Curr. Biol.* **2010**, 20, 69–74. doi: 10.1016/j.cub.2009.10.059.

E. J. Culp, N. Waglechner, W. Wang, A. A. Fiebig-Comyn, Y-P. Hsu, K. Koteva, D. Sychantha, B. K. Coombes, M. S. Van Nieuwenhze, Y. V. Brun, G. D. Wright. **Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling.**

Nature. **2020**, 578, 582–587. doi: 10.1038/s41586-020-1990-9.

L. Dang, J. Liu, C. Wang, H. Liu, J. Wen. **Enhancement of rapamycin production by metabolic engineering in *Streptomyces hygroscopicus* based on genome-scale metabolic model.**

J. Ind. Microbiol. Biotechnol. **2017**, 44, 259–270. doi: 10.1007/s10295-016-1880-1.

M. J. Daniels. **Studies of the biosynthesis of polymyxin B.**

Biochim. Biophys. Acta. **1968**, 156, 119–127. doi: 10.1016/0304-4165(68)90110-4.

S. Da Vela, D. I. Svergun. **Methods, development and applications of small-angle X-ray scattering to characterize biological macromolecules in solution.**

Curr. Res. Struct. Biol. **2020**, *2*, 164–170. doi: 10.1016/j.crstbi.2020.08.004.

S. de Avila e Silva, D. L. Notari, F. A. Neis, H. G. Ribeiro, S. Echeverrigaray. **BacPP: a web-based tool for Gram-negative bacterial promoter prediction.**

Genet. Mol. Res. **2016**, *15*. doi: 10.4238/gmr.15027973.

E. Dehling, G. Volkmann, J. C. J. Matern, W. Dörner, J. Alfermann, J. Diecker, H. D. Mootz. **Mapping of the communication-mediating interface in non-ribosomal peptide synthetases using a genetically encoded photocrosslinker supports an upside-down helix-hand motif.**

J. Mol. Biol. **2016**, *428*, 4345–4360. doi: 10.1016/j.jmb.2016.09.007.

C. A. Dejong, G. M. Chen, H. Li, C. W. Johnston, M. R. Edwards, P. N. Rees, M. A. Skinnider, A. L. H. Webster, N. A. Magarvey. **Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching.**

Nat. Chem. Biol. **2016**, *12*, 1007–1014. doi: 10.1038/nchembio.2188.

S. Dekimpe, J. Masschelein. **Beyond peptide bond formation: the versatile role of condensation domains in natural product biosynthesis.**

Nat. Prod. Rep. **2021**, *38*, 1910–1937. doi: 10.1039/D0NP00098A.

S. de Kok, L. H. Stanton, T. Salby, M. Durot, V. F. Holmes, K. G. Patel, D. Platt, E. B. Shapland, Z. Serber, J. Dean, J. D. Newman, S. S. Chandran. **Rapid and reliable DNA assembly via ligase cycling reaction.** *ACS Synth. Biol.* **2014**, *3*, 97–106. doi: 10.1021/sb4001992.

M. Dell, K. L. Dunbar, C. Hertweck. **Ribosome-independent peptide biosynthesis: the challenge of a unifying nomenclature.** *Nat. Prod. Rep.* **2022**, *39*, 453–459. doi: 10.1039/D1NP00019E.

J. C. Del Valle, M. L. Buide, J. B. Whittall, F. Valladares, E. Narbona. **UV radiation increases phenolic compound protection but decreases reproduction in *Silene littorea*.**

PLoS One. **2020**, *15*, e0231611. doi: 10.1371/journal.pone.0231611.

A. L. Demain. **Pharmaceutically active secondary metabolites of microorganisms.**

Appl. Microbiol. Biotechnol. **1999**, *52*, 455–463. doi: 10.1007/s002530051546.

A. L. Demain, R. P. Elander. **The beta-lactam antibiotics: past, present, and future.**

Antonie van Leeuwenhoek. **1999**, *75*, 5–19. doi: 10.1023/a:1001738823146.

A. L. Demain, A. Fang. **The natural functions of secondary metabolites.**

Adv. Biochem. Eng. Biotechnol. **2000**, *69*, 1–39. doi: 10.1007/3-540-44964-7_1.

T. de Rond, J. E. Asay, B. S. Moore. **Co-occurrence of enzyme domains guide the discovery of an oxazolone synthetase.** *Nat. Chem. Biol.* **2021**, *17*, 794–799. doi: 10.1038/s41589-021-00808-4.

A. L. Devereaux, S. L. Mercer, C. W. Cunningham. **DARK classics in chemical neuroscience: morphine.** *ACS Chem. Neurosci.* **2018**, *9*, 2395–2407. doi: 10.1021/acscemneuro.8b00150.

Y. Ding, C. M. Rath, K. L. Bolduc, K. Håkansson, D. H. Sherman. **Chemoenzymatic synthesis of cryptophycin anticancer agents by an ester bond-forming non-ribosomal peptide synthetase module.** *J. Am. Chem. Soc.* **2011**, *133*, 14492–14495. doi: 10.1021/ja204716f.

A. D'Onofrio, J. M. Crawford, E. J. Stewart, K. Witt, E. Gavriš, S. Epstein, J. Clardy, K. Lewis. **Siderophores from neighboring organisms promote the growth of uncultured bacteria.** *Chem. Biol.* **2010**, *17*, 254–264. doi: 10.1016/j.chembiol.2010.02.010.

D. P. Dowling, Y. Kung, A. K. Croft, K. Taghizadeh, W. L. Kelly, C. T. Walsh, C. L. Drenna. **Structural elements of an NRPS cyclization domain and its intermodule docking domain.** *Proc. Natl. Acad. Sci. USA.* **2016**, *113*, 12432–12437. doi: 10.1073/pnas.1608615113.

T. W. Doyle, D. E. Nettleton, D. M. Balitz, J. E. Moseley, R. E. J. Grulich. **Isolation and structure of bohemamine (1a β ,2 α ,6a β ,6b β)-3-methyl-N-(1a,6,6a,6b-tetrahydro-2,6a-dimethyl-6-oxo-2H-oxireno[a]pyrrolizin-4-yl)-2-butenamide.** *Org. Chem.* **1980**, *45*, 1324–1326. doi: 10.1021/jo01295a037.

E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. L. Allen, G. Skiniotis, C. C. Aldrich, A. M. Gulick. **Structures of two distinct conformations of holo-non-ribosomal peptide synthetases.** *Nature.* **2016**, *446*, 876–879. doi: 10.1038/nature16163.

L. Du, C. Sánchez, M. Chen, D. J. Edwards, B. Shen. **The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase.** *Chem. Biol.* **2000**, *7*, 623–642. doi: 10.1016/s1074-5521(00)00011-9.

D. Du, L. Wang, Y. Tian, H. Liu, H. Tan, G. Niu. **Genome engineering and direct cloning of antibiotic gene clusters via phage ϕ BT1 integrase-mediated site-specific recombination in *Streptomyces*.** *Sci. Rep.* **2015**, *5*, 8740. doi: 10.1038/srep08740.

- M. Duban, S. Cociancich, V. Leclère. **Nonribosomal peptide synthesis definitely working out the rules.** *Microorganisms*. **2022**, *10*, 577. doi: 10.3390/microorganisms10030577.
- G. Dunn, M. G. Montgomery, F. Mohammed, A. Coker, J. B. Cooper, T. Robertson, J-L. Garcia, T. D. H. Bugg, S. P. Wood. **The structure of the C-C bond hydrolase MhpC provides insights into its catalytic mechanism.** *J. Mol. Biol.* **2005**, *346*, 253–265. doi: 10.1016/j.jmb.2004.11.033.
- M. J. Ek-Ramos, R. Gomez-Flores, A. A. Orozco-Flores, C. Rodríguez-Padilla, G. González-Ochoa, P. Tamez-Guerra. **Bioactive products from plant-endophytic Gram-positive bacteria.** *Front. Microbiol.* **2019**, *10*, 463. doi: 10.3359/fmicb.2019.00463.
- P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan. **Features and development of Coot.** *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486–501. doi: 10.1107/S0907444910007493.
- C. Engler, R. Kandzia, S. Marillonnet. **A one pot, one step, precision cloning method with high throughput capability.** *PLoS One*. **2008**, *3*, e3647. doi: 10.1371/journal.pone.0003647.
- D. J. Ericsson, A. Kasrayan, P. Johansson, T. Bergfors, A. G. Sandström, J-E. Bäckvall, S. L. Mowbray. **X-ray structure of *Candida antartica* lipase A shows a novel lid structure and a likely mode of interfacial activation.** *J. Mol. Biol.* **2008**, *376*, 109–119. doi: 10.1016/j.jmb.2007.10.079.
- S. Ernst, A. N. Volkov, M. Stark, L. Hölscher, K. Steinert, S. Fetzner, U. Hennecke, S. L. Drees. **Azetidomonamide and diazetidomonapyridone metabolites control biofilm formation and pigment synthesis in *Pseudomonas aeruginosa*.** *J. Am. Chem. Soc.* **2022**, *144*, 7676–7685. doi: 10.1021/jacs.1c13653.
- P. R. Evans. **An introduction to data reduction: space-group determination, scaling and intensity statistics.** *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 282–292. doi: 10.1107/S090744491003982X.
- P. R. Evans, G. N. Murshudov. **How good are my data and what is the resolution?** *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69*, 1204–1214. doi: 10.1107/S0907444913000061.
- C. D. Fage, S. Kosol, M. Jenner, C. Öster, A. Gallo, M. Kaniusaite, R. Steinbach, M. Staniforth, V. G. Stavris, M. A. Marahiel, M. J. Cryme, J. R. Lewandowski. **Communication breakdown: dissecting the COM interfaces subunits of nonribosomal peptide synthetases.** *ACS Catal.* **2021**, *11*, 10802–10813. doi: 10.1021/acscatal.1c02113.

Q. Fang, L. Wu, C. Urwald, M. Mugat, S. Wang, K. Kyeremeh, C. Philips, S. Law, Y. Zhou, H. Deng. **Genomic scanning enabling discovery of a new antibacterial bicyclic carbamate-containing alkaloid.** *Synth. Syst. Biotechnol.* **2021**, *6*, 12–19. doi: 10.1016/j.synbio.2021.01.002.

E. A. Felnagle, M. R. Rondon, A. D. Berti, H. A. Crosby, M. G. Thomas. **Identification of the biosynthetic gene cluster and an additional gene for resistance to the antituberculosis drug capreomycin.** *Appl. Environ. Microbiol.* **2007**, *73*, 4162–4170. doi: 10.1128/AEM.00485-07.

E. A. Felnagle, E. E. Jackson, Y. A. Chan, A. M. Podevels, A. D. Berti, M. D. McMahon, M. G. Thomas. **Nonribosomal peptide synthetases involved in the production of medically relevant natural products.** *Mol. Pharm.* **2008**, *5*, 191–211. doi: 10.1021/mp700137g.

A. Fleming. **On the antibacterial action of cultures of a *Penicillium*, with special reference to their use in the isolation of *B. influenzae*.** *Br. J. Exp. Pathol.* **1929**, *10*, 226–236. doi: /

P. D. Fortin, C. T. Walsh, N. A. Magarvey. **A transglutaminase homologue as a condensation catalyst in antibiotic assembly lines.** *Nature.* **2007**, *448*, 824–827. doi: 10.1038/nature06068.

D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, D. I. Svergun. **ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions.** *J. Appl. Cryst.* **2017**, *50*, 1212–1225. doi: 10.1107/S1600576717007786.

J. Fu, X. Bian, S. Hu, H. Wang, F. Huang, P. M. Seibert, A. Plaza, L. Xia, R. Müller, A. F. Stewart, Y. Zhang. **Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting.** *Nat. Biotechnol.* **2012**, *30*, 440–446. doi: 10.1038/nbt.2183.

P. Fu, J. B. MacMillan. **Spithioneines A and B, two new bohemamine derivatives possessing ergothioneine moiety from a marine-derived *Streptomyces spinoverrucosus*.** *Org. Lett.* **2015**, *17*, 3046–3049. doi: 10.1021/acs.orglett.5b01328.

P. Fu, S. La, J. B. MacMillan. **1,3-oxazin-6-one derivatives and bohemamine-type pyrrolizidine alkaloids from a marine-derived *Streptomyces spinoverrucosus*.** *J. Nat. Prod.* **2016**, *79*, 455–462. doi: 10.1021/acs.jnatprod.5b00604.

P. Fu, A. Legako, S. La, J. B. MacMillan. **Discovery, characterization, and analogue Synthesis of bohemamine dimers generated by non-enzymatic biosynthesis.** *Chem.* **2016**, *22*, 3491–3495. doi: 10.1002/chem.201600024.

S. Fushinobu, T. Saku, M. Hidaka, S-Y. Jun, H. Nojiri, H. Yamane, H. Shoun, T. Omori, T. Wakagi. **Crystal structures of a meta-cleavage product hydrolase from *Pseudomonas fluorescens* IP01 (CumD) complexed with cleavage products.**

Protein Sci. **2002**, *11*, 2184–2195. doi: 10.1110/ps.0209602.

D. G. Gamage, A. Gunaratne, G. R. Periyannan, T. G. Russel. **Applicability of instability index for *in vitro* protein stability prediction.**

Protein Pept. Lett. **2019**, *26*, 339–347. doi: 10.2174/0929866526666190228144219.

N. Gaudelli, C. A. Townsend. **Epimerization and substrate gating by a TE domain in β -lactam antibiotic biosynthesis.** *Nat. Chem. Biol.* **2014**, *10*, 251–258. doi: 10.1038/nchembio.1456.

N. M. Gaudelli, D. H. Long, C. A. Townsend. **β -lactam formation by a non-ribosomal peptide synthase during antibiotic biosynthesis.** *Nature.* **2015**, *520*, 383–387. doi: 10.1038/nature14100.

E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch. **Protein identification and analysis tools on the ExPASy server.**

In *The Proteomics Protocols Handbook* (ed. J.M. Walker). **2005**, 571–607.

J. A. Gerlt, J. T. Bouvier, D. B. Davidson, H. J. Imker, B. Sadkhin, D. R. Slater, K. L. Whalen. **Enzyme function initiative-enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks.**

Biochim. Biophys. Acta. **2015**, *1854*, 1019–1037. doi: 10.1016/j.bbapap.2015.04.015.

W. Gevers, H. Kleinkauf, F. Lipmann. **The activation of amino acids for biosynthesis of gramicidin S.** *Proc. Natl. Acad. Sci. USA.* **1968**, *60*, 269–276. doi: 10.1073/pnas.60.1.269.

D. G. Gibson, L. Young, R-Y. Chuang, J. C. Venter, C. A. Hutchinson, H. O. Smith. **Enzymatic assembly of DNA molecules up to several hundred kilobases.**

Nat. Methods. **2009**, *6*, 343–345. doi: 10.1038/nmeth.1318.

C. C. Gilhuus-Moe, T. Kristensen, J. E. Bredesen, T. Zimmer, S. G. Laland. **The presence and possible role of phosphopantothenic acid in gramicidin S synthetase.**

FEBS Lett. **1970**, *7*, 287–290. doi: 10.1016/0014-5793(70)80183-1.

O. Glatter, O. Kratky. **Small-angle X-ray scattering.** *London: Academic Press.* **1982**, *1*, 361–386.

P. Goettig, M. Groll, J-S. Kim, R. Huber, H. Brandstetter. **Structures of the tricorin-interacting aminopeptidase F1 with different ligands explain its catalytic mechanism.**

EMBO J. **2002**, *21*, 5343–5352. doi: 10.1093/emboj/cdf552.

J. P. Gomez-Escribano, M. J. Bibb. **Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters.**

Microb. Biotechnol. **2011**, *4*, 207–215. doi: 10.1111/j.1751-7915.2010.00219.x.

A. Goswami, S. G. Van Lanen. **Enzymatic strategies and biocatalysts for amide bond formation: tricks of the trade outside of the ribosome.**

Mol. Biosyst. **2015**, *11*, 338–353. doi: 10.1039/C4MB00627E.

M. Govindappa, B. Hemashekhar, M-K. Arthikala, R. V. Ravishankar, Y. L. Ramachandra. **Characterization, antibacterial, antioxidant, antidiabetic, anti-inflammatory and antityrosinase activity of green synthesized silver nanoparticles using *Calophyllum tomentosum* leaves extract.** *Results Phys.* **2018**, *9*, 400–408. doi: 10.1016/j.rinp.2018.02.049.

C. Greunke, E. R. Duell, P. M. D'Agostino, A. Glöckle, K. Lamm, T. A. M. Gulder. **Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential.**

Metab. Eng. **2018**, *47*, 334–345. doi: 10.1016/j.ymben.2018.03.010.

A. M. Gulick. **Conformational dynamics in the acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase.**

ACS Chem. Biol. **2009**, *4*, 811–827. doi: 10.1021/cb900156h.

H. Habe, K. Morii, S. Fushinobu, J-W. Nam, Y. Ayabe, T. Yoshida, T. Wagaki, H. Yamane, H. Nojiri, T. Omori. **Crystal structure of a histidine-tagged serine hydrolase involved in the carbazole degradation (CarC enzyme).**

Biochem. Biophys. Res. Commun. **2003**, *303*, 631–639. doi: 10.1016/s0006-291x(03)00375-9.

C. Hacker, X. Cai, C. Kegler, L. Zhao, A. K. Weickhmann, J. P. Wurm, H. B. Bode, J. Wöhnert. **Structure-based redesign of docking domain interactions modulates the product spectrum of a rhabdopeptide-synthesizing NRPS.**

Nat. Commun. **2018**, *9*, 4366. doi: 10.1038/s41467-018-06712-1.

N. Hadadi, V. Hatzimanikatis. **Design of computational retrobiosynthesis tools for the design of de novo synthetic pathways.**

Curr. Opin. Chem. Biol. **2015**, *28*, 99–104. doi: 10.1016/j.cbpa.2015.06.025.

M. Hahn, T. Stachelhaus. **Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains.**

Proc. Natl. Acad. Sci. USA. **2004**, *101*, 15585–15590. doi: 10.1073/pnas.0404932101.

M. Hahn, T. Stachelhaus. **Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides.**

Proc. Natl. Acad. Sci. USA. **2006**, *103*, 275–280. doi: 10.1073/pnas.0508409103.

G. D. Hannigan, D. Prihoda, A. Palicka, J. Soukup, O. Klempir, L. Rampula, J. Durcak, M. Wurst, J. Kotowski, D. Chang, R. Wang, G. Piizzi, G. Temesi, D. J. Hazuda, C. H. Woelk, D. A. Bitton. **A deep learning genome-mining strategy for biosynthetic gene cluster prediction.**

Nucleic Acid Res. **2019**, *47*, e110. doi: 10.1093/nar/gkz654.

D. B. Hansen, S. B. Bumpus, Z. D. Aron, N. L. Kelleher, C. T. Walsh. **The loading module of mycosubtilin: an adenylation domain with fatty acid selectivity.**

J. Am. Chem. Soc. **2007**, *129*, 6366–6367. doi: 10.1021/ja070890j.

K. Haslinger, C. Brieke, S. Uhlmann, L. Sieverling, R. D. Süssmuth, M. J. Cryle. **The structure of a transient complex of a nonribosomal peptide synthetase and a CYP450 monooxygenase.**

Angew. Chem. Int. Ed. Engl. **2014**, *53*, 8518–8522. doi: 10.1002/anie.201404977.

K. Haslinger, M. Peschke, C. Brieke, E. Maximowitsch, M. J. Cryle. **X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis.**

Nature. **2015**, *521*, 105–109. doi: 10.1038/nature14141.

G. W. Heberlig, J. T. C. Brown, R. D. Simard, M. Wirz, W. Zhang, L. I. Susser, M. E. Horsman, C. N. Boddy. **Chemoenzymatic macrocycle synthesis using resorcylic acid lactone thioesterase domains.** *Org. Biomol. Chem.* **2018**, *16*, 5771–5779. doi: 10.1039/C8OB01512K.

L. Hedström. **Serine protease mechanism and specificity.**

Chem. Rev. **2002**, *102*, 4501–4524. doi: 10.1021/cr000033x.

F. Hemmerling, J. Piel. **Strategies to access biosynthetic novelty in bacterial genomes for drug discovery.** *Nat. Rev. Drug. Discov.* **2022**, *21*, 359–378. doi: 10.1038/s41573-022-00414-6.

D. A. Herbst, B. Boll, G. Zocher, T. Stehle, L. Heide. **Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylation enzymes.** *J. Biol. Chem.* **2013**, *288*, 1991–2003. doi: 10.1074/jbc.M112.420182.

L. Holm, A. Laiho, P. Törönen, M. Salgado. **DALI shines a light on remote homologs: one hundred discoveries.** *Protein Sci.* **2023**, *32*, e4519. doi: 10.1002/pro.4519.

Z. Hong, A. Bolard, C. Giraud, S. Prévost, G. Genta-Jouve, C. Deregnacourt, S. Häussler, K. Jeannot, Y. Li. **Azetidine-containing alkaloids produced by a quorum-sensing regulated nonribosomal peptide synthetase pathway in *Pseudomonas aeruginosa*.**

Angew. Chem. Int. Ed. Engl. **2019**, *58*, 3178–3182. doi: 10.1002/anie.201809981.

Y. Horiuchi, S. Kondo, T. Ikeda, D. Ikeda, K. Miura, M. Hamada, T. Takeuchi, H. Umezawa. **New antibiotics clazamycins A and B.** *J. Antibiot.* **1979**, *32*, 762–764. doi: 10.7164/antibiotics.32.762.

R. M. Horton, H. D. Hunt, S. N. Ho, J. K. Pullen, L. R. Pease. **Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension.**

Gene. **1989**, *77*, 61–68. doi: 10.1016/0378-1119(89)90359-4.

J. Hou, L. Robbel, M. A. Marahiel. **Identification and characterization of the lysobactin biosynthetic gene cluster reveals mechanistic insights into an unusual termination module architecture.** *Chem. Biol.* **2011**, *18*, 655–664. doi: 10.1016/j.chembiol.2011.02.012.

J-F. Hu, D. Wunderlich, R. Thiericke, H-M. Dahse, S. Grabley, X-Z. Feng, I. Sattler. **Jenamidines A to C: unusual alkaloids from *Streptomyces* sp. with specific antiproliferative properties obtained by chemical screening.** *J. Antibiot.* **2003**, *56*, 747–754. doi: 10.7164/antibiotics.56.747.

S. Huang, J. Tabudravu, S. S. Elsayed, J. Travert, D. Peace, M. Him Tong. K. Kyeremeh, S. M. Kelly, L. Trembleau, R. Ebel, M. Jaspars, Y. Yu, H. Deng. **Discovery of a single monooxygenase that catalyzes carbamate formation and ring contraction in the biosynthesis of the legonmycins.** *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 12697–12701. doi: 10.1002/anie.201502902.

G. H. Hur, C. R. Vickery, M. D. Burkart. **Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology.** *Nat. Prod. Rep.* **2012**, *29*, 1074–1098. doi: 10.1039/c2np20025b.

A. Isogai, S. Sakuda, K. Shindo, S. Watanabe, A. Suzuki, S. Fujita, T. Furuya. **Structures of cyclocarbamides A and B, new plant growth regulators from *Streptoverticillium* sp.** *Tetrahydron Lett.* **1986**, *27*, 1161–1164. doi: 10.1016/S0040-4039(00)84205-7.

T. Izoré, M. J. Cryle. **The many faces and important roles of protein–protein interactions during non-ribosomal peptide synthesis.**

Nat. Prod. Rep. **2018**, *35*, 1120–1139. doi: 10.1039/C8NP00038G.

T. Izoré, Y. T. Candace Ho, J. A. Kaczmarek, A. Gavriilidou, K. H. Chow, D. L. Steer, R. J. A. Goode, R. B. Schittenhelm, J. Tailhades, M. Tosin, G. L. Challis, E. H. Krenske, N. Ziemert, C. J. Jackson, M. J. Cryle. **Structures of a non-ribosomal peptide synthetase condensation domain suggest the basis of substrate selectivity.**

Nat. Commun. **2021**, *12*, 2511. doi: 10.1038/s41467-021-22623-0.

G. Jahanshah, Q. Yan, H. Gerhardt, Z. Pataj, M. Lämmerhofer, I. Pianet, M. Josten, H-G. Sahl, M. W. Silby, J. E. Loper, H. Gross. **Discovery of the cyclic lipopeptide gacamide A by genome mining and repair of defective GacA regulator in *Pseudomonas fluorescens* Pf0-1.**

J. Nat. Prod. **2019**, *82*, 301–308. doi: 10.1021/acs.jnatprod.8b00747.

M. J. Jaremko, D. John Lee, S. J. Opella, M. D. Burkart. **Structure and substrate sequestration in the pyoluteorin type II peptidyl carrier protein PltL.**

J. Am. Chem. Soc. **2015**, *137*, 11546–11549. doi: 10.1021/jacs.5b04525.

W. Jiang, X. Zhao, T. Gabrieli, C. Lou, Y. Ebenstein, T. F. Zhu. **Cas9-Assisted Targeting of Chromosome segments CATCH enables one-step targeted cloning of large gene clusters.**

Nat. Commun. **2015**, *6*, 8101. doi: 10.1038/ncomms9101.

S. K. Johansen, C. E. Maus, B. B. Plikaytis, S. Douthwaite. **Capreomycin binds across the ribosomal subunit interface using tlyA-encoded 2'-O-methylations in 16S and 23S rRNAs.**

Mol. Cell. **2006**, *23*, 173–182. doi: 10.1016/j.molcel.2006.05.044.

C. H. Johnston, R. Zvanych, N. Khyzha, N. A. Magarvey. **Nonribosomal assembly of natural lipocyclocarbamates lipoprotein-associated phospholipase inhibitors.**

ChemBioChem. **2013**, *14*, 431–435. doi: 10.1002/cbic.201200598.

C. W. Johnston, M. A. Skinnider, M. A. Wyatt, W. Li, M. R. M. Ranieri, L. Yang, D. L. Zechel, B. Ma, N. A. Magarvey. **An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products.** *Nat. Commun.* **2015**, *6*, 8421. doi: 10.1038/ncomms9421.

J. Jumper et al. **Highly accurate protein structure prediction with AlphaFold.**

Nature. **2021**, *596*, 583–589. doi: 10.1038/s41586-021-03819-2.

W. Kabsch. **XDS.**

Acta Crystallogr. D Biol. Crystallogr. **2010**, *66*, 125–132. doi: 10.1107/S0907444909047337.

D. Kahne, C. Leimkuhler, W. Lu, C. Walsh. **Glycopeptide and lipoglycopeptide antibiotics.**

Chem. Rev. **2005**, *105*, 425–448. doi: 10.1021/cr030103a.

- S. A. Kautsar, *et al.* **MIBiG 2.0: a repository for biosynthetic gene clusters of known function.** *Nucleic Acids Res.* **2020**, *48*, 454–458. doi: 10.1093/nar/gkz882.
- S. A. Kautsar, K. Blin, S. Shaw, T. Weber, M. H. Medema. **BiG-FAM: the biosynthetic gene cluster families database.** *Nucleic Acids Res.* **2021**, *49*, 490–497. doi: 10.1093/nar/gkaa812.
- T. A. Keating, C. G. Marshall, C. T. Walsh, A. E. Keating. **The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains.** *Nat. Struct. Biol.* **2002**, *9*, 522–526. doi: 10.1038/nsb810.
- C. Kegler, H. B. Bode. **Artificial splitting of a non-ribosomal peptide synthetase by inserting natural docking domains.** *Angew. Chem. Int. Ed. Engl.* **2020**, *59*, 13463–13467. doi: 10.1002/anie.201915989.
- P. Khandel, S. K. Shahi. **Mycogenic nanoparticles and their bio-prospective applications: current status and future challenges.** *J. Nanostruct. Chem.* **2018**, *8*, 369–391. doi: 10.1007/s40097-018-0285-2.
- M. C. Kim, J. H. Lee, B. Shin, L. Subedi, J. W. Cha, J-S. Park, D-C. Oh, S. Y. Kim, H. C. Kwon. **Salinazinones A and B: pyrrolidinyl-oxazinones from solar saltern-derived *Streptomyces* sp. KMF-004.** *Org. Lett.* **2015**, *17*, 5024–5027. doi:10.1021/acs.orglett.5b02495.
- S-H. Kim, W. Lu, M. Kamal Ahmadi, D. Montiel, M. A. Ternei, S. F. Brady. **Atolypenes, ticyclic bacterial sesterterpenes discovered using a multiplexed *in vitro* Cas9-TAR gene cluster refactoring approach.** *ACS Synth. Biol.* **2019**, *8*, 109–118. doi: 10.1021/acssynbio.8b00361.
- M. Klaus, A. D. D'Souza, A. Nivina, C. Khosla, M. Grninger. **Engineering of chimeric polyketide synthases using SYNZIP docking domains.** *ACS Chem. Biol.* **2019**, *14*, 426–433. doi: 10.1021/acscchembio.8b01060.
- H. Kleinkauf, W. Gevers, R. Roskoski Jr, F. Lipmann. **Enzyme-bound phosphopantetheine in tyrocidine biosynthesis.** *Biochem. Biophys. Res. Commun.* **1970**, *41*, 1218–1222. doi: 10.1016/0006-291x(70)90216-0.
- M. Kobayashi, K. Fujita, K. Matsuda, T. Wakimoto. **Chemo-enzymatic synthesis of non-ribosomal macrolactams by a penicillin-binding protein-type thioesterase.** *Methods Mol. Biol.* **2023**, *2670*, 127–144. doi: 10.1007/978-1-0716-3214-7_6.
- A. Koglin, M. R. Mofid, F. Löhr, B. Schäfer, V. V. Rogov, M-M. Blum, T. Mittag, M. A. Marahiel, F. Bernhard, V. Dötsch. **Conformational switches modulate protein interactions in peptide antibiotic synthetases.** *Science.* **2006**, *312*, 273–276. doi: 10.1126/science.1122928.

R. M. Kohli, J. W. Trauger, D. Schwarzer, M. A. Marahiel, C. T. Walsh. **Generality of peptide cyclization catalyzed by isolated thioesterase domains nonribosomal peptides synthetases.** *Biochemistry*. **2001**, *40*, 7099–7108. doi: 10.1021/bi010036j.

K. Koketsu, K. Watanabe, H. Suda, H. Oguri, H. Oikawa. **Reconstruction of the saframycin core scaffold defines dual Pictet-Spengler mechanisms.**

Nat. Chem. Biol. **2010**, *6*, 408–410. doi: 10.1038/nchembio.365.

D. N. Konanov, D. V. Krivonos, E. N. Ilana, V. V. Babenko. **BioCAT: search for biosynthetic gene clusters producing nonribosomal peptides with known structure.**

Comput. Struct. Biotechnol. J. **2022**, *20*, 1218–1226. doi: 10.1016/j.csbj.2022.02.013.

S. Kosol, A. Gallo, D. Griffiths, T. R. Valentic, J. Masschelein, M. Jenner, E. L. C. de Los Santos, L. Manzi, P. K. Sydor, D. Rea, S. Zhou, V. Fülöp, N. J. Oldham, S-C. Tsai, G. L. Challis, J. R. Lewandowski. **Structural basis for chain release from the enacyloxin polyketide synthase.**

Nat. Chem. **2019**, *11*, 913–923. doi: 10.1038/s41557-019-0335-5.

M. Kotowska, K. Pawlik. **Roles of type II thioesterases and their application for secondary metabolite yield improvement.**

Appl. Microbiol. Biotechnol. **2014**, *98*, 7735–7746. doi: 10.1007/s00253-014-5952-8.

M. B. Kozin, D. I. Svergun. **Automated matching of high- and low-resolution models.**

J. Appl. Cryst. **2001**, *34*, 33–41. doi: 10.1107/S0021889800014126.

F. I. Kraas, V. Helmetag, M. Wittmann, M. Strieker, M. A. Marahiel. **Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation.**

Chem. Biol. **2010**, *17*, 872–880. doi: 10.1016/j.chembiol.2010.06.015.

F. I. Kraas, T. W. Giessen, M. A. Marahiel. **Exploring the mechanism of lipid transfer during biosynthesis of the acidic lipopeptide antibiotic CDA.**

FEBS Lett. **2012**, *586*, 283–288. doi: 10.1016/j.febslet.2012.01.003.

T. L. LaFleur, A. Hossain, H. M. Salis. **Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria.**

Nat. Commun. **2022**, *13*, 5159. doi: 10.1038/s41467-022-32829-5.

M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins. **Clustal W and Clustal X version 2.0.**

Bioinformatics. **2007**, *23*, 2947–2948. doi: 10.1093/bioinformatics/btm404.

- R. A. Laskowski, J. Jabłońska, L. Pravda, R. Svobodová Vařeková, J. M. Thornton. **PDBsum: structural summaries of PDB entries.** *Protein Sci.* **2017**, *27*, 129–134. doi: 10.1002/pro.3289.
- O. Lazos, M. Tosin, A. L. Slusarczyk, S. Boakes, J. Cortés, P. J. Sidebottom, P. F. Leadlay. **Biosynthesis of the putative siderophore erythrochelin requires unprecedented crosstalk between separate nonribosomal peptide gene clusters.** *Chem. Biol.* **2010**, *17*, 160–173. doi: 10.1016/j.chembiol.2010.01.011.
- S. G. Lee, R. Roskoski Jr, K. Bauer, F. Lipmann. **Purification of the polyenzymes responsible for tyrocidine synthesis and their dissociation into subunits.** *Biochemistry.* **1973**, *12*, 398–405. doi: 10.1021/bi00727a006.
- A. G. Leslie, P. C. Moody, W. V. Shaw. **Structure of chloramphenicol acetyltransferase at 1.75 Å resolution.** *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 4133–4137. doi: 10.1073/pnas.85.12.4133.
- M. Z. Li, S. J. Elledge. **Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC.** *Nat. Methods.* **2007**, *4*, 251–256. doi: 10.1038/nmeth1010.
- Y-X. Li, Z. Zhong, W-P. Zhang, P-Y. Qian. **Discovery of cationic nonribosomal peptides as Gram-negative antibiotics through global genome mining.** *Nat. Commun.* **2018**, *9*, 3273. doi: 10.1038/s41467-018-05781-6.
- H. Li, C. L. M. Gilchrist, C-S. Phan, H. J. Lacey, D. Vuong, S. A. Moggach, E. Lacey, A. M. Piggott, Y-H. Chooi. **Biosynthesis of a new benzazepine alkaloid nanangelinin A from *Aspergillus nanangensis* involves an unusual l-kynurenine-incorporating NRPS catalyzing regioselective lactamization.** *J. Am. Chem. Soc.* **2020**, *142*, 7145–7152. doi: 10.1021/jacs.0c01605.
- L. Li, L. W. MacIntyre, S. F. Brady. **Refactoring biosynthetic gene clusters for heterologous production of microbial natural products.** *Curr. Opin. Biotechnol.* **2021**, *69*, 145–162. doi: 10.1016/j.copbio.2020.12.011.
- E. Lilkova, P. Petkov, N. Ilieva, L. Litov. **The PyMOL Molecular Graphics System.** Version 2.0 Schrodinger, LLC. **2015**, 10919–10925.
- C. Lin, Z. Zeng, Y. Lin, P. Wang, D. Cao, K. Xie, Y. Luo, H. Yang, J. Yang, W. Wang, L. Luo, H. Lin, H. Chen, Y. Zhao, Y. Shi, Z. Gao, H. Liu, S-L. Liu. **Naringenin suppresses epithelial ovarian cancer by inhibiting proliferation and modulating gut microbiota.** *Phytomedicine.* **2022**, *106*, 154401. doi: 10.1016/j.phymed.2022.154401.

- T. Lincke, S. Behnken, K. Ishida, M. Roth, C. Hertweck. **Closthioamide: an unprecedented polythioamide antibiotic from the strictly anaerobic bacterium *Clostridium cellulolyticum*.** *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 2011–2013. doi: 10.1002/anie.200906114.
- L. L. Ling et al. **A new antibiotic kills pathogens without detectable resistance.** *Nature.* **2015**, *517*, 455–459. doi: 10.1038/nature14098.
- F. Lipmann. **Nonribosomal polypeptide synthesis on polyenzyme templates.** *Acc. Chem. Res.* **1973**, *6*, 361–367. doi: 10.1021/ar50071a001.
- J. Liu, B. Wang, H. Li, Y. Xie, Q. Li, X. Qin, X. Zhang, J. Ju. **Biosynthesis of the anti-infective marformycins featuring pre-NRPS assembly line *N*-formylation and *O*-methylation and post-assembly line *C*-hydroxylation chemistries.** *Org. Lett.* **2015**, *17*, 1509–1512. doi: 10.1021/acs.orglett.5b00389.
- Z. Liu, Y. Zhao, C. Huang, Y. Luo. **Recent advances in silent gene cluster activation in *Streptomyces*.** *Front. Bioeng. Biotechnol.* **2021**, *9*, 632230. doi: 10.3389/fbioe.2021.632230.
- Y. Liu, X. Yang, J. Gan, S. Chen, Z-X. Xiao, Y. Cao. **CB-Dock2: improved protein-ligand blind docking integrating cavity detection, docking and homologous template fitting.** *Nucleic Acids Res.* **2022**, *50*, 159–164. doi: 10.1093/nar/gkac394.
- A. Loeschcke, S. Thies. ***Pseudomonas putida*-a versatile host for the production of natural products.** *Appl. Microbiol. Biotechnol.* **2015**, *99*, 6197–6214. doi: 10.1007/s00253-015-6745-4.
- J. R. Lohman, M. Ma, M. E. Cuff, L. Bigelow, J. Bearden, G. Babnigg, A. Joachimiak, G. N. Phillips. B. Shen. **The crystal structure of BlmI as a model for nonribosomal peptide synthetase peptidyl carrier proteins.** *Proteins*, **2014**, *82*, 1210–1218. doi: 10.1002/prot.24485.
- N. K. Lokanath, I. Shiromizu, M. Miyano, S. Yokoyama, S. Kuramitsu, N. Kunishima. **Structure of beta-glucosidase at atomic resolution from *Thermus thermophilus*.** En attente de publication, **2003**.
- T. A. Lundy, S. Mori, S. Garneau-Tsodikova. **Engineering bifunctional enzymes capable of adenylating and selectively methylating the side chain or core of amino acids.** *ACS Synth. Biol.* **2018**, *7*, 399–404. doi: 10.1021/np050378g.
- Y. Luo, L. Zhang, K. W. Barton, H. Zhao. **Systematic identification of a panel of string constitutive promoters from *Streptomyces albus*.** *ACS Synth. Biol.* **2015**, *4*, 1001–1010. doi: 10.1021/acssynbio.5b00016.

B. Mach, E. Reich, E. L. Tatum. **Separation of the biosynthesis of the antibiotic polypeptide tyrocidine from protein biosynthesis.** *Proc. Natl. Acad. Sci. USA.* **1963**, *50*, 175–181.

S. S. Mahmoud, R. B. Croteau. **Strategies for transgenic manipulation of monoterpene biosynthesis in plants.**

Trends Plant Sci. **2002**, *7*, 366–373. doi: 10.1016/s1360-1385(02)02303-8.

K. Manalastas-Cantos, P. V. Konarev, N. R. Hajizadeh, A. G. Kikhney, M. V. Petoukhov, D. S. Molodenskiy, A. Panjkovich, H. D. T. Mertens, A. Gruzinov, C. Borges, C. M. Jeffries, D. I. Svergun, D. Franke. **ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis.**

J. Appl. Crystallogr. **2021**, *54*, 343–355. doi: 10.1107/S1600576720013412.

G. L. Marcone, E. Binda, F. Berini, F. Marinelli. **Old and new glycopeptide antibiotics: from products to gene and back in the post-genomic era.**

Biotechnol. Adv. **2018**, *36*, 534–554. doi: 10.1016/j.biotechadv.2018.02.009.

M. A. Martínez-Núñez, V. E. L. López. **Nonribosomal peptides synthetases and their applications in industry.** *Sustain. Chem. Process.* **2016**, *4*, 13. doi: 10.1186/s40508-016-0057-6.

M. Martin-Pascual, C. Batianis, L. Bruinsma, E. Asin-Garcia, L. Garcia-Morales, R. A. Weusthuis, R. van Kranenburg, V. A. P. Martins Dos Santos. **A navigation guide of synthetic biology tools for *Pseudomonas putida*.**

Biotechnol. Adv. **2021**, *49*, 107732. doi: 10.1016/j.biotechadv.2021.107732.

J. J. May, N. Kessler, M. A. Marahiel, M. T. Stubbs. **Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases.**

Proc. Natl. Acad. Sci. USA. **2002**, *99*, 12120–12125. doi: 10.1073/pnas.182156699.

E. K. McCranie, B. O. Bachmann. **Bioactive oligosaccharide natural products.**

Nat. Prod. Rep. **2014**, *31*, 1026–1042. doi: 10.1039/c3np70128j.

M. McErlean, J. Overbay, S. Van Lanen. **Refining and expanding nonribosomal peptide synthetase function and mechanism.**

J. Ind. Microbiol. Biotechnol. **2019**, *46*, 493–513. doi: 10.1007/s10295-018-02130-w.

A. McPherson, J. A. Gavira. **Introduction to protein crystallization.**

Acta Crystallogr. F Struct. Biol. Commun. **2014**, *70*, 2–20. doi: 10.1107/S2053230X13033141.

A. McPherson. **Protein crystallization.**

Methods Mol. Biol. **2017**, *1607*, 17–50. doi: 10.1007/978-1-4939-7000-1_2.

M. H. Medema, M. A. Fischbach. **Computational approaches to natural product discovery.** *Nat. Chem. Biol.* **2015**, *11*, 639–648. doi: 10.1038/nchembio.1884.

M. W. Medema et al. **Artificial intelligence for natural product drug discovery.**

Nat. Rev. Drug Discov. **2023**, *22*, 895–916. doi: 10.1038/s41573-023-00774-7.

A. C. Mercer, M. D. Burkart. **The ubiquitous carrier protein – a window to metabolite biosynthesis.** *Nat. Prod. Rep.* **2007**, *24*, 750–773. doi: 10.1039/b603921a.

S. Meyer, J.-C. Kehr, A. Mainz, D. Dehm, D. Petras, R. D. Süßmuth, E. Dittman. **Biochemical dissection of the natural diversification of microcystin provides lessons for synthetic biology of NRPS.** *Cell Chem. Biol.* **2016**, *23*, 462–471. doi: 10.1016/j.chembiol.2016.03.011.

S. Meyer, A. Mainz, J.-C. Kehr, R. D. Süßmuth, E. Dittman. **Prerequisites of isopeptide bond formation in microcystin biosynthesis.**

ChemBioChem. **2017**, *18*, 2376–2379. doi: 10.1002/cbic.201700389.

B. R. Miller, J. A. Sundlov, E. J. Drake, T. A. Makin, A. M. Gulick. **Analysis of the linker region joining the adenylation and carrier protein domains of the modular nonribosomal peptide synthetases.** *Proteins.* **2014**, *82*, 2691–2702. doi: 10.1002/prot.24635.

B. R. Miller, A. M. Gulick. **Structural Biology of Nonribosomal Peptide Synthetases.**

Methods Mol. Biol. **2016**, *1401*, 3–29. doi: 10.1007/978-1-4939-3375-4_1.

G. Minasov, X. Wang, B. K. Shoichet. **An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation.**

J. Am. Chem. Soc. **2002**, *124*, 5333–5340. doi: 10.1021/ja0259640.

Y. Minowa, M. Araki, M. Kanehisa. **Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes.**

J. Mol. Biol. **2007**, *368*, 1500–1517. doi: 10.1016/j.jmb.2007.02.099.

A. Mitra, A. K. Kesarwani, D. Pal, V. Nagaraja. **WebGeSTer DB: a transcription terminator database.** *Nucleic Acids Res.* **2011**, *39*, 129–135. doi: 10.1093/nar/gkq971.

S. Mohammadi, B. Jafari, P. Asgharian, M. Martorell, J. Sharifi-Rad. **Medicinal plants used in the treatment of malaria: a key emphasis to *Artemisia*, *Cinchona*, *Cryptolepis*, and *Tabebuia* genera.** *Phytother. Res.* **2020**, *34*, 1556–1569. doi: 10.1002/ptr.6628.

H. Mohimani, A. Gurevich, A. Shlemov, A. Mikheenko, A. Korobeynikov, L. Cao, E. Shcherbin, L-F. Nothias, P. C. Dorrestein, P. A. Pevzner. **Dereplication of microbial metabolites through database search of mass spectra.** *Nat. Commun.* **2018**, *9*, 4035. doi: 10.1038/s41467-018-06082-8.

S. Mori, A. H. Pang, T. A. Lundy, A. Garzan, O. V. Tsodikov, S. Garneau-Tsodikova. **Structural basis for backbone N-methylation by an interrupted adenylation domain.**

Nat. Chem. Biol. **2018**, *14*, 428–430. doi: 10.1038/s41589-018-0014-7.

M. Moussa, W. Ebrahim, M. Bonus, H. Gohlke, A. Mándi, T. Kurtán, R. Hartmann, R. Kalscheuer, W. Lin, Z. Liu, P. Proksch. **Co-culture of the fungus *Fusarium tricinctum* with *Streptomyces lividans* induces production of the cryptic naphthoquinone dimers.**

RSC Adv. **2019**, *9*, 1491–1500. doi: 10.1039/c8ra09067j.

A. M. Mulichak, W. Lu, H. C. Losey, C. T. Walsh, R. M. Garavito. **Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interaction with acceptor and nucleotide ligands.** *Biochemistry.* **2004**, *43*, 5170–5180. doi: 10.1021/bi036130c.

S. Müller, E. Garcia-Gonzalez, A. Mainz, G. Hertlein, N. C. Heid, E. Mösker, H. van den Elst, H. S. Overkleeft, E. Genersch, R. D. Süßmuth. **Paenilamicin: structure and biosynthesis of a hybrid nonribosomal peptide/polyketide antibiotic from the bee pathogen *Paenibacillus larvae*.**

Angew. Chem. Int. Ed. Engl. **2014**, *53*, 10821–10825. doi: 10.1002/anie.201404572.

M. W. Mullooney et al. **Artificial intelligence for natural product drug discovery.**

Nat. Rev. Drug Discov. **2023**, *22*, 895–916. doi: 10.1038/s41573-023-00774-7.

U. Munawer, V. B. Raghavendra, S. Ningaraju, K. L. Krishna, A. R. Ghosh, G. Meleppa, A. Pugazhendhi. **Biofabrication of gold nanoparticles mediated by the endophytic *Cladosporium* species: photodegradation, *in vitro* anticancer activity and *in vivo* antitumor studies.**

Int. J. Pharm. **2020**, *88*, 119729. doi: 10.1016/j.ijpharm.2020.119729.

G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. Vagin. **REFMAC5 for the refinement of macromolecular crystal structures.**

Acta Crystallogr. D Biol. Crystallogr. **2011**, *67*, 355–367. doi: 10.1107/S0907444911001314.

C. U. Mussagy, L. Duffosé. **A review of natural astaxanthin production in a circular bioeconomy context using *Paracoccus carotinifaciens*.**

Bioresour. Technol. **2023**, *369*, 128499. doi: 10.1016/j.biortech.2022.128499.

J. C. Navarro-Muñoz et al. **A computational framework to explore large-scale biosynthetic diversity.** *Nat. Chem. Biol.* **2020**, *16*, 60–68. doi: 10.1038/s41589-019-0400-9.

D. Nichols, N. Cahoon, E. M. Trakhtenberg, L. Pham, A. Mehta, A. Belanger, T. Kanigan, K. Lewis, S. S. Epstein. **Use of ichip for high-throughput in situ cultivation of uncultivable microbial species.** *Appl. Environ. Microbiol.* **2010**, *76*, 2445–2450. doi: 10.1128/AEM.01754-09.

K. Ochi, T. Hosaka. **New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters.**

Appl. Microbiol. Biotechnol. **2013**, *97*, 87–98. doi: 10.1007/s00253-012-4551-9.

A. Oladipo, V. Enwemiwe, O. Ejeromedoghene, A. Adebayo, O. Ogunyemi, F. Fu. **Production and functionalities of specialized metabolites from different organic sources.**

Metabolites. **2022**, *12*, 534. doi: 10.3390/metabo12060534.

D. L. Ollis, E. Cheah, M. Cygler, B. W. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. D. Schrag, J. Sussman, K. H. G. Verschueren, A. Goldman. **The alpha/beta hydrolase fold.** *Protein. Eng. Des. Sel.* **1992**, *5*, 197–211. doi: 10.1093/protein/5.3.197.

T. L. Orr-Weaver, J. W. Szostak, R. J. Rothstein. **Yeast transformation: a model system for the study of recombination.**

Proc. Natl. Acad. Sci. USA. **1981**, *78*, 6354–6358. doi: 10.1073/pnas.78.10.6354.

M. Otyepka, J. Damborský. **Functionally relevant motions of haloalkane dehalogenases occur in the specificity-modulating cap domains.**

Protein Sci. **2002**, *11*, 1206–1217. doi: 10.1110/ps3830102.

K. Palaniappan, I-M. A. Chen, K. Chu, A. Ratner, R. Seshadri, N. C. Kyrpides, N. N. Ivanova, N. J. Mouncey. **IMG-ABC v.5.0: an update to the IMG/Atlas of biosynthetic gene clusters knowledgebase.** *Nucleic Acid Res.* **2020**, *48*, D422–D430. doi: 10.1093/nar/gkz932.

A. Panjkovich, D. I. Svergun. **Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis.**

Phys. Chem. Chem. Phys. **2016**, *18*, 5707–5719. doi: 10.1039/C5CP04540A.

H. M. Patel, C. T. Walsh. **In vitro reconstitution of the *Pseudomonas aeruginosa* nonribosomal peptide synthesis of pyocelin: characterization of backbone tailoring thiazoline reductase and N-methyltransferase activities.** *Biochemistry.* **2001**, *40*, 9023–9031. doi: 10.1021/bi010519n.

H. M. Patel, J. Tao, C. T. Walsh. **Epimerization of an L-cysteinyll to a D-cysteinyll residue during thiazoline ring formation in siderophore chain elongation by pyochelin synthetase from *Pseudomonas aeruginosa*.** *Biochemistry.* **2003**, *42*, 10514–10527. doi: 10.1021/bi034840c.

- J. B. Patteson, A. R. Lescallete, B. Li. **Discovery and biosynthesis of azabicyclene, a conserved nonribosomal peptide in *Pseudomonas aeruginosa*.**
Org. Lett. **2019**, *21*, 4955–4959. doi: 10.1021/acs.orglett.9b01383.
- J. B. Patteson, C. M. Fortinez, A. T. Putz, J. Rodriguez-Rivas, L. Henry Bryant III, K. Adhikari, M. Weigt, T. M. Schmeing, B. Li. **Structure and function of dehydrating condensation domain in nonribosomal peptide biosynthesis.**
J. Am. Chem. Soc. **2022**, *144*, 14057–14070. doi: 10.1021/jacs.1c13404.
- C. Paulitz, R. R. Bélanger. **Biological control in greenhouse systems.**
Annu. Rev. Phytopathol. **2001**, *39*, 103–133. doi: 10.1146/annurev.phyto.39.1.103.
- A. E. Payne, M. Schoppet, M. H. Hansen, M. J. Cryle. **Diversity of nature's assembly lines - recent discoveries in non-ribosomal peptide synthesis.**
Mol. Biosyst. **2017**, *13*, 9–22. doi: 10.1039/c6mb00675b.
- M. Peschke, M. Gonsior, R. D. Süßmuth, M. J. Cryle. **Understanding the crucial interactions between cytochrome P450s and non-ribosomal peptide synthetases during glycopeptide antibiotic biosynthesis.** *Curr. Opin. Struct. Biol.* **2016**, *41*, 46–53. doi: 10.1016/j.sbi.2016.05.018.
- M. V. Petoukhov, D. I. Svergun. **Ambiguity assessment of small-angle scattering curves from monodisperse systems.**
Acta Crystallogr. D Biol. Crystallogr. **2015**, *71*, 1051–1058. doi: 10.1107/S1399004715002576.
- V. H. T. Pham, J. Kim. **Cultivation of unculturable soil bacteria.**
Trends Biotechnol. **2012**, *30*, 475–484. doi: 10.1016/j.tibtech.2012.05.007.
- S. J. Pidot, S. Coyne, F. Kloss, C. Hertweck. **Antibiotics from neglected bacterial sources.**
Int. J. Med. Microbiol. **2014**, *340*, 14–22. doi: 10.1016/j.ijmm.2013.08.011.
- J. Quan, J. Tian. **Circular polymerase extension cloning of complex gene libraries and pathway.**
PLoS One. **2009**, *4*, e6441. doi: 10.1371/journal.pone.0006441.
- J. Ratelade, M-C. Miot, E. Johnson, J-M. Betton, P. Mazodier, N. Benaroudi. **Production of recombinant proteins in the lon-deficient BL21(DE3) strain of *Escherichia coli* in the absence of the DnaK chaperone.**
Appl. Environ. Microbiol. **2007**, *75*, 3803–3807. doi: 10.1128/AEM.00255-09.
- C. Rausch, I. Hoof, T. Weber, W. Wohlleben, D. H. Huson. **Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution.**
BMC. Evol. Biol. **2007**, *16*, 78. doi: 10.1186/1471-2148-7-78.

- L. Ray, K. Yamanaka, B. S. Moore. **A peptidyl-transesterifying type I thioesterase in salinamide biosynthesis.** *Angew. Chem. Int. Ed. Engl.* **2016**, *55*, 364–367. doi: 10.1002/anie.201508576.
- R. Reher et al. **A convolutional neural network-based approach for the rapid annotation of molecularly diverse natural products.**
J. Am. Chem. Soc. **2020**, *142*, 4114–4120. doi: 10.1021/jacs.9b13786.
- J. M. Reimer, M. N. Aloise, P. M. Harrison, T. M. Schmeing. **Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase.**
Nature. **2016**, *529*, 239–242. doi: 10.1038/nature16503.
- J. M. Reimer, M. Eivaskhani, I. Harb, A. Guarné, M. Weigt, T. M. Schmeing. **Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility.**
Science. **2019**, *366*, eaaw4388. doi: 10.1126/science.aaw4388.
- A. W. Reinke, R. A. Grant, A. E. Keating. **A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering.**
J. Am. Chem. Soc. **2010**, *132*, 6025–6031. doi: 10.1021/ja907617a.
- H. Ren, C. Shi, H. Zhao. **Computational tools for discovering and engineering natural product biosynthetic pathways.** *iScience*, **2020**, *23*, 100795. doi: 10.1016/j.isci.2019.100795.
- E. Ricart, V. Leclère, A. Flissi, M. Mueller, M. Pupin, F. Lisacek. **rBAN: retro-biosynthetic analysis of nonribosomal peptides.**
J. Cheminform. **2019**, *11*, 13. doi: 10.1186/s13321-019-0335-x.
- C. D. Richter, D. Nietlispach, R. W. Broadhurst, K. J. Weissman. **Multienzyme docking in hybrid megasynthases.** *Nat. Chem. Biol.* **2008**, *4*, 75–81. doi: 10.1038/nchembio.2007.61.
- L. Robbel, M. A. Marahiel. **Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery.** *J. Biol. Chem.* **2010**, *285*, 27501–27508. doi: 10.1074/jbc.R110.128181.
- X. Robert, P. Gouet. **Deciphering key features in protein structures with the new ENDscript server.** *Nucl. Acids Res.* **2014**, *42*, 320–324. doi: 10.1093/nar/gku316.
- F. Roberts, M. Wink. **Biochemistry, ecology and medicinal applications.**
Alkaloids. **1998**, *1*, 1–6. ISBN: 0-306-45465-3.
- J. Robertson, K. Stevens. **Pyrrolizidine alkaloids: occurrence, biology, and chemical synthesis.**
Nat. Prod. Rep. **2017**, *34*, 62–89. doi: 10.1039/c5np00076a.

- D. Roche, C. T. Walsh. **Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis.**
Biochemistry. **2003**, *42*, 1334–1344. doi: 10.1021/bi026867m.
- R. Roskoski Jr, H. Kleinkauf, W. Gevers, F. Lipmann. **Isolation of enzyme-bound peptide intermediates in tyrocidine biosynthesis.**
Biochemistry. **1970**, *9*, 4846–4851. doi: 10.1021/bi00827a003.
- M. G. Rossman, D. M. Blow. **The detection of sub-units within the crystallographic asymmetric unit.** *Acta Cryst.* **1962**, *15*, 24–31. doi: 10.1107/S0365110X62000067.
- M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch, O. Kohlbacher. **NRPSpredictor2: a web server for predicting NRPS adenylation domain specificity.**
Nucleic Acid Res. **2011**, *39*, 362–367. doi: 10.1093/nar/gkr323.
- Z. Rui, W. Huang, F. Xu, M. Han, X. Liu, S. Lin, W. Zhang. **Sparsomycine biosynthesis highlights unusual module architecture and processing mechanism in non-ribosomal peptide synthase.** *ACS. Chem. Biol.* **2015**, *10*, 1765–1769. doi: 10.1021/acscchembio.5b00284.
- P. J. Rutledge, G. L. Challis. **Discovery of microbial natural products by activation of silent biosynthetic gene clusters.** *Nat. Rev. Microbiol.* **2015**, *13*, 509–523. doi: 10.1038/nrmicro3496.
- Z. Saati-Santamaría, N. Selem-Mojica, E. Peral-Aranega, R. Rivas, P. García-Fraile. **Unveiling the genomic potential of *Pseudomonas* type strains for discovering new natural products.**
Microb. Genom. **2022**, *8*, 000758. doi: 10.1099/mgen.0.000758.
- H. Salis. **The ribosome binding site calculator.**
Methods. Enzymol. **2011**, *498*, 19–42. doi: 10.1016/B978-0-12-385120-8.00002-4.
- S. A. Samel, B. Wagner, M. A. Marahiel, L-O. Essen. **The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide.**
J. Mol. Biol. **2006**, *359*, 876–889. doi: 10.1016/j.jmb.2006.03.062.
- S. A. Samel, G. Schoenafinger, T. A. Knappe, M. A. Marahiel, L-O. Essen. **Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase.** *Structure*. **2007**, *15*, 781–792. doi: 10.1016/j.str.2007.05.008.
- S. A. Samel, P. Czodrowski, L-O. Essen. **Structure of the epimerization domain of tyrocidine synthetase A.**
Acta Crystallogr. D Biol. Crystallogr. **2014**, *70*, 1442–1452. doi: 10.1107/S1399004714004398.

C. Sayer, M. N. Isupov, E. Bonch-Osmolovskaya, J. A. Littlechild. **Structural studies of a thermophilic esterase from a new Planctomycetes species, *Thermogutta terrifontis*.**

FEBS J. **2015**, *282*, 2846–2857. doi:10.1111/febs.13326.

A. Scaglione, M. R. Fullone, L. C. Montemiglio, G. Parisi, C. Zamparelli, B. Vallone, C. Savino, I. Grgurina. **Structure of the adenylation domain Thr1 involved in the biosynthesis of 4-chlorothreonine in *Streptomyces* sp. OH-5093 - Protein flexibility and molecular basis of substrate specificity.** *FEBS J.* **2017**, *284*, 2981–2999. doi: 10.1111/febs.14163.

T. H. Scheuermann, C. A. Brautigam. **High-precision, automated integration of multiple isothermal titration calorimetric thermograms: new features of NITPIC.**

Methods. **2015**, *76*, 87–98. doi: 10.1016/j.ymeth.2014.11.024.

O. Schimming, V. L. Challinor, N. J. Tobias, H. Adihou, P. Grün, L. Pöschel, C. Richter, H. Schwalbe, H. B. Bode. **Structure, biosynthesis, and occurrence of bacterial pyrrolizidine alkaloids.** *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 12702–12705. doi: 10.1002/anie.201504877.

Y. Schmidt, M. van der Voort, M. Crüsemann, J. Piel, M. Josten, H-G. Sahl, H. Miess, J. M. Raaijmakers, H. Gross. **Biosynthetic origin of the antibiotic cyclocarbamate brabantamide A (SB-253514) in plant-associated *Pseudomonas*.**

Chembiochem. **2014**, *15*, 259–266. doi: 10.1002/cbic.201300527.

T. Schneider, B. Shen, C. T. Walsh. **Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation.**

Biochemistry. **2003**, *42*, 9722–9730. doi: 10.1021/bi034792w.

N. Schnell, K. D. Entian, U. Schneider, F. Götz, H. Zähner, R. Kellner, G. Jung. **Prepeptide sequence of epidermin, a ribosomally synthesized antibiotic with four sulphide-rings.**

Nature. **1988**, *333*, 276–288. doi: 10.1038/333276a0.

B. K. Scholz-Schroeder, J. D. Soule, D. C. Gross, **The *sypA*, *sypS*, and *sypC* synthetase genes encode twenty-two modules involved in the nonribosomal peptide synthesis of syringopeptine by *Pseudomonas syringae* pv. *Syringar* B301D.**

Mol. Plant. Microbe. Interact. **2003**, *16*, 271–280. doi: 10.1094/MPMI.2003.16.4.271.

M. R. Seyedsayamdost. **High-throughput platform for the discovery of elicitors of silent bacterial gene clusters.**

Proc. Natl. Acad. Sci. USA. **2014**, *111*, 7266–7271. doi: 10.1073/pnas.1400019111.

- Z. Shao, H. Zhao, H. Zhao. **DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways.** *Nucleic Acid Res.* **2009**, *37*, e16. doi: 10.1093/nar/gkn991.
- E. Shao, S. Zhao, Y. Dong, Y. Wang, Y. Fei, S. Li, L. Wang, T. Bashir, T. Luan, L. Lin, Y. Wang, W. Zhao, Z. Zhong. **Anisomycin inhibits Coxsackievirus B replication by promoting the lysosomal degradation of eEF1A1.** *Antiviral. Res.* **2023**, *215*, 105621. doi: 10.1016/j.antiviral.2023.105621.
- T. D. Sharkey, S. Yeh. **Isoprene emission from plants.** *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **2001**, *52*, 407–436. doi: 10.1146/annurev.arplant.52.1.407.
- R. Shi, A. Proteau, M. Villarroja, I. Moukadiri, L. Zhang, J-F. Trempe, A. Matte, M. E. Armengod, M. Cygler. **Structural basis for Fe-S cluster assembly and tRNA thiolation mediated by IscS protein-protein interactions.** *PloS Biol.* **2010**, *13*, e1000354. doi: 10.1371/journal.pbio.1000354.
- Y-M. Shi, A. O. Brachmann, M. A. Westphalen, N. Neubacher, N. J. Tobias, H. B. Bode. **Dual phenazine gene clusters enable diversification during biosynthesis.** *Nat. Chem. Biol.* **2019**, *15*, 331–339. doi: 10.1038/s41589-019-0246-1.
- J. Shima, A. Hesketh, S. Okamoto, S. Kawamoto, K. Ochi. **Induction of actinorhodine production by *rpsL* (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2).** *J. Bacteriol.* **1996**, *178*, 7276–7284. doi: 10.1128/jb.178.24.7276-7284.1996.
- M. Singh, S. Chaudhary, D. Sareen. **Non-ribosomal peptide synthetases: identifying the cryptic gene clusters and decoding the natural product.** *J. Biosci.* **2017**, *42*, 175–187. doi: 10.1007/s12038-017-9663-z.
- M. A. Skinnider, N. J. Merwin, C. W. Johnston, N. A. Magarvey. **PRISM 3: expanded prediction of natural product chemical structures from microbial genomes.** *Nucleic Acid Res.* **2017**, *45*, 49–54. doi: 10.1093/nar/gkx320.
- H. G. Smith, M. J. Beech, J. R. Lewandowski, G. L. Challis, M. Jenner. **Docking domain-mediated subunit interactions in natural product megasynth(et)ases.** *J. Ind. Microbiol. Biotechnol.* **2021**, *48*, kuab018. doi: 10.1093/jimb/kuab018.
- V. Solovyev, A. Salamov. **Automatic annotation of microbial genomes and metagenomic sequences.** *Metagenomics and its applications in agriculture, biomedicine and environmental studies.* **2011**, *1*, 61–78.

- M. Sorokina, C. Steinbeck. **Review on natural products databases: where to find data in 2020.** *J. Cheminform.* **2020**, *12*, 20. doi: 10.1186/s13321-020-00424-9.
- J. Staunton, B. Wilkinson. **Biosynthesis of erythromycin and rapamycin.** *Chem. Rev.* **1997**, *97*, 2611–2630. doi: 10.1021/cr9600316.
- T. Stein, J. Vater, V. Kruff, A. Otto, B. Wittmann-Liebold, P. Franke, M. Panico, R. McDoweel, H. R. Morris. **The multiple carrier model of nonribosomal peptide biosynthesis at modular multienzymatic templates.** *J. Biol. Chem.* **1996**, *271*, 15428–15435. doi: 10.1074/jbc.271.26.15428.
- J. Stetefeld, S. A. McKenna, T. R. Patel. **Dynamic light scattering: a practical guide and applications in biomedical sciences.** *Biophys. Rev.* **2016**, *8*, 409–427. doi: 10.1007/s12551-016-0218-6.
- Y. Sugimoto, F. R. Camacho, S. Wang, P. Chankhamjon, A. Odabas, A. Biswas, P. D. Jeffrey, M. S. Donia. **A metagenomic strategy for harnessing the chemical repertoire of the human microbiome.** *Science.* **2019**, *366*, eaax9176. doi: 10.1126/science.aax9176.
- J. A. Sundlov, C. Shi, D. J. Wilson, C. C. Aldrich, A. M. Gulick. **Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains.** *Chem. Biol.* **2012**, *19*, 188–198. doi: 10.1016/j.chembiol.2011.11.013.
- R. D. Süßmuth, A. Mainz. **Nonribosomal Peptide Synthesis - Principles and Prospects.** *Angew. Chem. Int. Ed. Engl.* **2017**, *56*, 3770–3821. doi: 10.1002/anie.201609079.
- L. Tallorin, J. Wang, W. E. Kim, S. Sahu, N. M. Kosa, P. Yang, M. Thompson, M. K. Gilson, P. I. Frazier, M. D. Burkart, N. C. Gianneschi. **Discovering *de novo* peptide substrates for enzymes using machine learning.** *Nat. Commun.* **2018**, *9*, 5253. doi: 10.1038/s41467-018-07717-6.
- K. Tan, M. Zhou, R. P. Jedrzejczak, R. Wu, R. A. Higuera, D. Borek, G. Babnigg, A. Joachimiak. **Structures of teixobactin-producing nonribosomal peptide synthetase condensation and adenylation domains.** *Curr. Res. Struct. Biol.* **2020**, *2*, 14–24. doi: 10.1016/j.crstbi.2020.01.002.
- A. Tanovic, S. A. Samel, L-O. Essen, M. A. Marahiel. **Crystal structure of the termination modul of a nonribosomal peptide synthetase.** *Science.* **2008**, *321*, 659–663. doi: 10.1126/science.1159850.
- H. R. Thapa, J. M. Robbins, B. S. Moore, V. Agarwal. **Insights onto thiotemplated pyrrole biosynthesis gained from the crystal structure of flavin-dependent oxidase in complex with carrier protein.** *Biochemistry.* **2019**, *58*, 918–929. doi: 10.1021/acs.biochem.8b01177.

- D. Thirumurugan, A. Cholarajan, S. S. S. Raja, R. Ramasamy. **An introductory chapter: secondary metabolites.** *Secondary metabolites*. **2018**. doi: 10.5772/intechopen.71955.
- M. G. Thomas, M. D. Burkart, C. T. Walsh. **Conversion of L-proline to pyrrolyl-2-carboxyl-S-PCP during undecylprodigiosin and pyoluteorin biosynthesis.** *Chem. Biol.* **2002**, *9*, 171–184. doi: 10.1016/s1074-5521(02)00100-x.
- K. E. Thompson, C. J. Bashor, W. A. Lim, A. E. Keating. **SYNZIP protein interaction toolbox: *in vitro* and *in vivo* specifications of heterospecific coiled-coil interaction domains.** *ACS Synth. Biol.* **2012**, *1*, 118–129. doi: 10.1021/sb200015u.
- A. Thureau, P. Roblin, J. Pérez. **BioSAXS on the SWING beamline at synchrotron SOLEIL.** *J. Appl. Cryst.* **2021**, *54*, 1698–1710. doi: 10.1107/S1600576721008736.
- J. Tian, Q. Li, X. Chu, N. Wu. **Presyncodon, a web server for gene design with the evolutionary information of the expression hosts.** *Int. J. Mol. Sci.* **2018**, *19*, 3872. doi: 10.3390/ijms19123872.
- N. J. Tobias, H. Wolff, B. Djahanschiri, F. Grundmann, M. Kronenwerth, Y-M. Shi, S. Simonyi, P. Grün, D. Shapiro-Ilan, S. J. Pidot, T. P. Stinear, I. Ebersberger, H. B. Bode. **Natural product diversity associated with the nematode symbionts *Photorhabdus* and *Xenorhabdus*.** *Nat. Microbiol.* **2017**, *2*, 1676–1685. doi: 10.1038/s41564-017-0039-9.
- C. A. Townsend. **Convergent biosynthetic pathways to β -lactam antibiotics.** *Curr. Opin. Chem. Biol.* **2016**, *35*, 97–108. doi: 10.1016/j.cbpa.2016.09.013.
- P. A. Trinidad-Calderón, C. D. Varela-Chinchilla, S. García-Lara. **Despsipeptides targeting tumor cells: milestones from *in vitro* to clinical trials.** *Molecules*. **2023**, *9*, 670. doi: 10.3390/molecules28020670.
- P. Tufar, S. Rahighi, F. I. Kraas, D. K. Kirchner, F. Löhr, E. Henrich, J. Köpke, I. Dikic, P. Günter, M. A. Marahiel, V. Dötsch. **Crystal structure of a PCP/Sfp complex reveals the structural basis for carrier protein posttranslational modification.** *Chem. Biol.* **2014**, *21*, 552–562. doi: 10.1016/j.chembiol.2014.02.014.
- A. Vagin, A. Teplyakov. **MOLREP: an automated program for molecular replacement.** *J. Appl. Cryst.* **1997**, *30*, 1022–1025. doi: 10.1107/S0021889897006766.
- J. A van Santen et al. **The Natural Products Atlas: an open access knowledge base for microbial natural products discovery.** *ACS Cent. Sci.* **2019**, *5*, 1824–1833. doi: 10.1021/acscentsci.9b00806.

- V. Volkov, D. I. Svergun. **Uniqueness of *ab initio* shape determination in small-angle scattering.** *J. Appl. Cryst.* **2003**, *36*, 860–864. doi: 10.1107/S0021889803000268.
- C. T. Walsh, H. Chen, T. A. Keating, B. K. Hubbard, H. C. Losey, L. Luo, C. G. Marshall, D. A. Miller, H. M. Patel. **Tailoring enzymes that modify non-ribosomal peptides during and after chain elongation on NRPS assembly lines.** *Curr. Opin. Chem Biol.* **2001**, *5*, 525–534. doi: 10.1016/s1367-5931(00)00235-0.
- C. T. Walsh. **Where will new antibiotics come from ?** *Nat. Rev. Microbiol.* **2003**, *1*, 65–70. doi: 10.1038/nrmicro727.
- C. T. Walsh, R. V. O'Brien, C. Khosla. **Nonproteinogenic amino acid building blocks for nonribosomal peptide and hybrid polyketide scaffolds.** *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 7098–7124. doi: 10.1002/anie.201208344.
- R-Y. Wang, Z-Y. Shi, Y-Y. Guo, J-C. Chen, G-Q. Chen. **DNA fragments assembly based on nicking enzyme system.** *PLoS One.* **2013**, *8*, e57943. doi: 10.1371/journal.pone.0057943.
- H. Wang, D. P. Fewer, L. Holm, L. Rouhiainen, K. Sivonen. **Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes.** *Proc. Natl. Acad. Sci. USA.* **2014**, *111*, 9259–9264. doi: 10.1073/pnas.1401734111.
- M. Wang et al. **Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking.** *Nat. Biotechnol.* **2016**, *34*, 828–837. doi: 10.1038/nbt.3597.
- H. Wang, Z. Li, R. Jia, J. Yin, A. Li, L. Xia, Y. Yin, R. Müller, J. Fu, A. F. Stewart, Y. Zhang. **ExoCET: exonuclease *in vitro* assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes.** *Nucleic Acid Res.* **2018**, *46*, e28. doi: 10.1093/nar/gkx1249.
- G. Wang et al. **CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria.** *Nat. Microbiol.* **2019**, *4*, 2498–2510. doi: 10.1038/s41564-019-0573-8.
- K. Watanabe, K. Hotta, A. P. Praseuth, K. Koketsu, A. Migita, C. N. Boddy, C. C. C. Wang, H. Oguri, H. Oikawa. **Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*.** *Nat. Chem. Biol.* **2006**, *2*, 423–428. doi: 10.1038/nchembio803.
- D. Watts, E. A. Palombo, A. J. Castillo, B. Zaferanloo. **Endophytes in agriculture: potential to improve yields and tolerance of agricultural crops.** *Microorganisms.* **2023**, *11*, 1276. doi: 10.3390/microorganisms11051276.

- M. G. Watve, R. Tickoo, M. M. Jog, B. D. Whole. **How many antibiotics are produced by the genus *Streptomyces*.** *Arch. Microbiol.* **2001**, *176*, 386–390. doi: 10.1007/s002030100345.
- T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel, T. A. Holak. **Solution structure of a PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases.** *Structure.* **2000**, *8*, 407–418. doi: 10.1016/s0969-2126(00)00120-9.
- K. J. Weissman. **The structural biology of biosynthetic megaenzymes.** *Nat. Chem. Biol.* **2015**, *11*, 660–670. doi: 10.1038/nchembio.1883.
- J. Watzel, C. Hacker, E. Duchardt-Ferner, H. B. Bode, J. Wöhnert. **A new docking domain type in the peptide-antimicrobial-*Xenorhabdus* peptide producing nonribosomal peptide synthetase from *Xenorhabdus bovienii*.** *ACS Chem. Biol.* **2020**, *15*, 982–989. doi: 10.1021/acscchembio.9b01022.
- M. D. Winn *et al.* **Overview of the CCP4 suite and current developments.** *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235–242. doi: 10.1107/S09074444910045749.
- D. S. Wishart *et al.* **NP-MRD: the Natural Products Magnetic Resonance Database.** *Nucleic Acid Res.* **2022**, *50*, 665–677. doi: 10.1093/nar/gkab1052.
- C. Wu, Z. Shang, C. Lemetre, M. A. Ternei, S. F. Brady. **Cadasides, calcium-dependent acidic lipopeptides from the soil metagenome that are active against multidrug-resistant bacteria.** *J. Am. Chem. Soc.* **2019**, *141*, 3910–3919. doi: 10.1021/jacs.8b12087.
- J. Xu, J. Zhang, J. Zhuon Y. Li, Y. Tian, H. Tan. **Activation and mechanism of a cryptic ovidomycin gene cluster via the disruption of a global regulatory gene, *adpA*, in *Streptomyces ansiochromogenes*.** *J. Biol. Chem.* **2017**, *292*, 19708–19720. doi: 10.1074/jbc.M117.809145.
- R-X. Yang, S-W. Zhang, D. Xue, J-H. Xuan, Y-B. Zhang, B-B. Peng. **Culturable endophytes diversity isolated from *Paeonia ostii* and the genetic basis for their bioactivity.** *Pol. J. Microbiol.* **2018**, *67*, 441–454. doi: 10.21307/pjm-2018-052.
- J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter, C. T. Walsh. **Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase.** *Proc. Natl. Acad. Sci. USA.* **2005**, *102*, 15815–15820. doi: 10.1073/pnas.0507705102.
- Y. Yoshikuni, T. E. Ferrin, J. D. Keasling. **Designed divergent evolution of enzyme function.** *Nature.* **2006**, *440*, 1078–1082. doi: 10.1038/nature04607.

J. Y. Young *et al.* **OneDep: unified wwPDB system for deposition, biocuration, and validation of macromolecular structures in the PDB archive.**

Structure. **2017**, *25*, 536–545. doi: 10.1016/j.str.2017.01.004.

Q. Zhang, K. K. Schrader, H. N. ElSohly, S. Takamatsu. **New-cell-cell adhesion from *Streptomyces* sp. UMA-044.** *J. Antibiot.* **2003**, *56*, 673–681. doi: 10.7164/antibiotics.56.673.

Y. Zhang, U. Werling, W. Edelmann. **SLiCE: a novel bacterial cell-extract based DNA cloning method.** *Nucleic Acid Res.* **2012**, *40*, e55. doi: 10.1093/nar/gkr1288.

C. Zhang, L. Kong, Q. Liu, X. Lei, T. Zhu, J. Yin, B. Lin, Z. Deng, D. You. ***In vitro* characterization of echinomycin of L-tryptophanyl-S-enzyme and oxidation of (2S, 3S) β -hydroxytryptophan.** *PLoS One*. **2013**, *8*, e56772. doi: 10.1371/journal.pone.0056772.

J. Zhang, N. Liu, R. A. Cacho, Z. Gong, Z. Liu, W. Qin, C. Tang, Y. Tang, J. Zhou. **Structural basis of non-ribosomal peptide macrocyclization in fungi.**

Nat. Chem. Biol. **2016**, *12*, 1001–1003. doi: 10.1038/nchembio.2202.

Q. Zhang, J-W. Ren, W. Wang, J. Zhai, J. Yang, N. Liu, Y. Huang, Y. Chen, G. Pan, K. Fan. **A versatile Transcription-Translation in One approach for activation of cryptic biosynthetic gene clusters.** *ACS Chem. Biol.* **2020**, *15*, 2551–2557. doi: 10.1021/acscchembio.0c00581.

F. Zhao, Z. Liu, S. Yang, N. Ding, X. Gao. **Quinolactacin biosynthesis involves non-ribosomal-peptide-synthetase-catalyzed Dieckmann condensation to form the quinolone- γ -lactam hybrid.** *Angew. Chem. Int. Ed. Engl.* **2020**, *59*, 19108–19114. doi: 10.1002/anie.202005770.

Z. Zhou, P. Cironi, A. J. Lin, Y. Xu, S. Hrvatin, D. E. Golan, P.A. Silver, C. T. Walsh, J. Yin. **Genetically encoded short peptide tags for orthogonal protein labeling by Sfp.**

ACS Chem. Biol. **2007**, *2*, 337–346. doi: 10.1021/cb700054k.

J. Zhou, O. G. Troyanskaïa. **Predicting effects of noncoding variants with deep learning-based sequence model.** *Nat. Methods*. **2015**, *12*, 931–934. doi: 10.1038/nmeth.3547.

Les alcaloïdes pyrrolizidiniques (AP) bactériens sont des métabolites secondaires d'une grande diversité chimique à l'origine de leurs propriétés pharmacologiques prometteuses (antibiotiques, anticancéreux, *etc.*). Bien qu'il ait été déterminé que le noyau hétérocyclique soit assemblé par des méga-synthétases de peptides non-ribosomiques (NRPS) bimodulaires, les mécanismes moléculaires qui conditionnent la variété structurale des produits et sous-tendent leurs activités biologiques restent inconnus. Compte tenu de l'intérêt majeur de ces systèmes biosynthétiques, ce projet vise à établir les relations structure-fonction de l'hydrolase XhpG et de la déshydratase AzeD impliquées dans le processus de maturation post-NRPS des AP intermédiaires chez les γ -protéobactéries pathogènes *Xenorhabdus hominickii* et *Pseudomonas aeruginosa*.

Dans cet objectif, les enzymes recombinantes ont fait l'objet d'une stratégie expérimentale intégrative, de leur expression hétérologue au sein de l'hôte *E. coli* BL21 (DE3), dans une forme native ou mutante, jusqu'à leur purification par chromatographie en phase liquide et analyse subséquente par diffusion dynamique de la lumière. Ces étapes ont permis de confirmer l'homogénéité et la monodispersité des échantillons obtenus, des conditions nécessaires à la caractérisation structurale approfondie des *apo*-protéines par cristallographie aux rayons X. Dès lors, les efforts entrepris dans le criblage et l'affinement des conditions de cristallisation, ainsi que dans l'acquisition de données de diffraction à résolution atomique (1,2–1,6 Å), ont permis de résoudre la structure tridimensionnelle des enzymes biosynthétiques d'intérêt. À la lumière d'analyses complémentaires, ces résultats ont fourni une vue globale des états oligomériques et conformationnels plausibles en solution, et également un aperçu du mécanisme catalytique avec un accent sur la spécificité de substrat.

À terme, les connaissances obtenues contribueront aux travaux d'ingénierie biosynthétique afin de générer des analogues azacycliques avec un potentiel thérapeutique intéressant, en l'occurrence dans la lutte contre le germe opportuniste *P. aeruginosa* responsable d'infections nosocomiales à l'échelle mondiale.

Mots-clefs : cristallographie des protéines, NRPS, enzymes biosynthétiques, alcaloïdes pyrrolizidiniques.

Bacterial pyrrolizidine alkaloids (PAs) are secondary metabolites of wide chemical diversity at the origin of their promising pharmacological properties (antibiotics, anticancer, *etc.*). Although it has been determined that the heterocyclic core is assembled by modular non-ribosomal peptide synthetases (NRPS), the molecular mechanisms which govern the structural variety of the products and underlie their biological activities remain unknown. Given the major interest of these biosynthetic systems, this project aims to establish the structure-function relationships of the XhpG hydrolase and the AzeD dehydratase involved in the post-NRPS maturation of intermediate PAs in the pathogenic γ -proteobacteria *Xenorhabdus hominickii* and *Pseudomonas aeruginosa*.

In pursuit of this objective, the recombinant enzymes have been subjected to an integrative experimental strategy, from their heterologous expression within the host *E. coli* BL21 (DE3), in either a native or mutant form, up to their purification by liquid chromatography and subsequent analysis *via* dynamic light scattering. These steps have allowed to confirm the homogeneity and monodispersity of the obtained samples, conditions necessary for the in-depth structural characterization of the *apo*-proteins by X-ray crystallography. Therefore, the efforts undertaken in screening and refining crystallization conditions, as well as in acquiring diffraction data at atomic resolution (1.2–1.6 Å), have permitted to elucidate the three-dimensional structure of the target biosynthetic enzymes. In light of complementary analyses, these results have provided a general overview of plausible oligomeric and conformational states in solution, and also insights into the catalytic mechanism with a focus on substrate specificity.

In due course, the obtained knowledge will contribute to biosynthetic engineering endeavors to generate azacyclic analogues with interesting therapeutic potential, specifically in fighting the opportunistic pathogen *P. aeruginosa* responsible for nosocomial infections on a global scale.

Keywords: protein crystallography, NRPS, biosynthetic enzymes, pyrrolizidine alkaloids.