



**HAL**  
open science

# Caractérisation structurale d'une méga-enzyme bactérienne participant à la biosynthèse du carcinogène colibactine

Sarah Bonhomme

► **To cite this version:**

Sarah Bonhomme. Caractérisation structurale d'une méga-enzyme bactérienne participant à la biosynthèse du carcinogène colibactine. Biologie structurale [q-bio.BM]. Université Grenoble Alpes [2020-..], 2022. Français. NNT : 2022GRALV050 . tel-03853505

**HAL Id: tel-03853505**

**<https://theses.hal.science/tel-03853505>**

Submitted on 15 Nov 2022

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

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

## THÈSE

Pour obtenir le grade de

### DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Biologie Structurale et Nanobiologie

Arrêté ministériel : 25 mai 2016

Présentée par

**Sarah BONHOMME**

Thèse dirigée par **Andréa DESSEN**  
et co-encadrée par **Pauline MACHEBOEUF**, CNRS

préparée au sein du **Laboratoire Institut de Biologie Structurale**  
dans l'**École Doctorale Chimie et Sciences du Vivant**

**Caractérisation structurale d'une méga-enzyme  
bactérienne participant à la biosynthèse du  
carcinogène colibactine**

**Structural characterization of a bacterial  
megaenzyme involved in the biosynthesis of the  
colibactin carcinogen**

Thèse soutenue publiquement le **12 juillet 2022**,  
devant le jury composé de :

**Madame Andréa DESSEN**

DIRECTRICE DE RECHERCHE, CNRS délégation Alpes, Directrice de thèse

**Madame Kira WEISSMAN**

PROFESSEUR DES UNIVERSITES, Université de Lorraine, Rapporteur

**Monsieur Richard BONNET**

PROFESSEUR DES UNIVERSITES - PRATICIEN HOSPITALIER,  
Université Clermont Auvergne, Rapporteur

**Monsieur Patrice GOUET**

PROFESSEUR DES UNIVERSITES, Université Lyon 1 - Claude Bernard,  
Examineur

**Monsieur Guy SCHOEHN**

DIRECTEUR DE RECHERCHE, CNRS délégation Alpes, Examineur

**Monsieur Winfried WEISSENHORN**

PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes,  
Président





## Remerciements

J'aimerais tout d'abord remercier l'Ecole Normale Supérieure de Lyon pour m'avoir accordé une bourse de thèse.

Je tiens ensuite à remercier Andréa, pour m'avoir acceptée dans son équipe et m'avoir donné la chance de travailler sur un sujet passionnant. Merci également pour ton enthousiasme à tout épreuve.

Merci évidemment à Pauline, pour m'avoir accompagnée tout au long de ma thèse, depuis ma formation à la paillasse, la mise en place des expériences et l'aide à l'écriture. En plus de tes talents de formatrice, merci aussi de m'avoir fait profiter de tes talents de cristallographe en m'aidant à pêcher de (très très) nombreux cristaux et en résolvant la structure de la PKS. Merci aussi d'avoir partagé ton bureau, que j'ai largement envahi.

Je remercie sincèrement Madame Kira Weissman et Monsieur Richard Bonnet de me faire l'honneur d'être rapporteurs de ma thèse.

Je remercie également les autres membres du jury, Patrice Gouet, Guy Schoehn et Winfried Weissenhorn d'avoir accepté d'évaluer mon travail.

J'aimerais remercier les membres de mon comité de suivi de thèse, Mickaël Cherrier, Patrice Gouet et Cécile Morlot, pour leurs conseils et nos discussions, qui m'ont remotivée.

Merci à tous ceux qui ont participé aux expériences sur ClbK :

Merci à Caroline Mas de m'avoir formée patiemment aux techniques de biophysique.

Je remercie Luca Signor pour les expériences de spectrométrie de masse.

Je tiens à remercier Guy Schoehn et la plate-forme de microscopie pour les observations de ClbK. Merci aussi à Alexandre Dos Santos et Mayara Miyachiro pour les nombreuses sessions de coloration négative.

Merci à Sylvain Engilberge pour son aide dans la génération des modèles Alphafold. Merci aussi à Eric Girard pour sa collaboration sur les cristallographes.

Merci à Petra Pernot et Mark Tully pour leur aide à la collecte des données SAXS. J'aimerais aussi remercier chaleureusement Frank Gabel, pour ses conseils sur l'analyse des données SAXS.

Un grand merci à tous les membres, passés et présents, de l'équipe PATBAC, pour leur accueil chaleureux et les bons moments passés ensemble. Merci à Carlos pour tes conseils qui m'ont bien souvent aidée, au laboratoire et en dehors. Merci à Alex, pour ton aide au laboratoire et les moments de détente nécessaires aux pauses café. Merci aussi à Kévin, pour ta bonne humeur infaillible. Merci à Mayara pour ta gentillesse et ta générosité.

Merci à Jérôme, Mickaël et Jean-Marie de m'avoir intégrée dans votre équipe pédagogique et de m'avoir donné l'opportunité d'enseigner à l'Université.

Merci à ma famille, qui m'a soutenue pendant que je travaillais sur « mes petites bêtes ». Merci à mes parents, à qui je dois mon goût pour l'étude du vivant. Merci aussi à mes sœurs, pour les moments de détente et de repos partagés ensemble, bien nécessaires.

Merci à Maxime, pour ton oreille attentive et ton soutien au quotidien.



# Table des matières

<b>Table des matières</b>	<b>5</b>
<b>Abréviations</b>	<b>9</b>
<b>Introduction</b>	<b>15</b>
<b>Préambule</b>	<b>17</b>
<b>I. Activités biologiques des souches <i>pks+</i> d'<i>Escherichia coli</i> du microbiote intestinal</b>	<b>19</b>
I.1. Le microbiote intestinal et son rôle	19
I.2. Diversité des souches d' <i>Escherichia coli</i>	19
I.3. Génotoxicité des souches <i>pks+</i> d' <i>E. coli</i>	20
I.4. Implication des souches <i>pks+</i> d' <i>E. coli</i> dans le cancer colorectal	23
I.5. Implication des souches <i>pks+</i> d' <i>E. coli</i> dans d'autres types de cancers	28
I.6. Autres effets des souches <i>pks+</i> d' <i>E. coli</i>	29
<b>II. La biosynthèse de la colibactine par l'îlot génomique <i>pks</i></b>	<b>31</b>
II.1. Distribution et acquisition de l'îlot <i>pks</i> chez les bactéries	31
II.2. Organisation et régulation de l'îlot <i>pks</i>	32
II.3. Elucidation de la structure de la molécule génotoxique, la colibactine	33
II.4. Production de la colibactine par l'îlot <i>pks</i>	35
<b>III. La biosynthèse des peptides non ribosomiaux et des polycétides par des machineries enzymatiques multi-domaines</b>	<b>39</b>
III.1. Importance des métabolites secondaires synthétisés par les NRPS/PKS	39
III.2. Les NRPS/PKS partagent une organisation commune	41
III.3. Le domaine « carrier protein »	42
<b>IV. Structures et fonctions des méga-enzymes NRPS</b>	<b>44</b>
IV.1. Les domaines enzymatiques des NRPS	44
IV.2. Les structures des assemblages NRPS monomériques	49
IV.3. Les structures des assemblages NRPS dimériques	50
<b>V. Structures et fonctions des méga-enzymes PKS</b>	<b>52</b>
V.1. Les domaines enzymatiques des PKS	52
V.2. Les structures des assemblages PKS de type <i>cis</i> -AT	55
V.3. Les structures des assemblages PKS de type <i>trans</i> -AT	62

<b>VI. Les hybrides NRPS/PKS</b>	<b>67</b>
VI.1. Classification des hybrides	67
VI.2. Interfaces NRPS/PKS	68
VI.3. Interfaces PKS/NRPS	69
VI.4. Etat oligomérique des protéines hybrides de type fusionné	69
<b>Choix du modèle d'étude et objectifs</b>	<b>71</b>
<b>Matériels et méthodes</b>	<b>77</b>
<b>I. Matériel biologique</b>	<b>79</b>
I.1. Vecteurs d'expression	79
I.2. Souches bactériennes	79
I.3. Milieux de culture	80
<b>II. Bio-informatique</b>	<b>81</b>
II.1. Logiciels utilisés pour l'alignement de séquences et l'analyse des structures	81
II.2. Analyse de la séquence de ClbK	81
II.3. Génération de modèles par le logiciel Alphafold2	82
<b>III. Biologie moléculaire</b>	<b>84</b>
III.1. Clonage de ClbK et de ses fragments	84
III.2. Mutagénèse dirigée	86
<b>IV. Expression et purification de ClbK et de ses fragments</b>	<b>88</b>
IV.1. Tests d'expression, de solubilité et d'attache à la résine d'affinité	88
IV.2. Expression et lyse des cultures de ClbK	89
IV.3. Purification de ClbK	90
IV.4. Préparation de ClbK pour la microscopie électronique	91
IV.5. Expression et lyse des cultures de fragments de ClbK	92
IV.6. Expression de la forme sélénométhionée de la protéine ClbK-PKS	92
IV.7. Purification de ClbK-PKS	93
IV.8. Purification de ClbK-PKS-Cy et ClbK-PKS-A <sub>core</sub>	93
<b>V. Caractérisation biochimique et biophysique de ClbK et de ses fragments</b>	<b>95</b>
V.1. Caractérisation par spectrométrie de masse en conditions dénaturantes	95
V.2. Caractérisation par diffusion de lumière laser multi-angles couplée à une chromatographie d'exclusion de taille (SEC-MALLS)	95
V.3. Caractérisation par photométrie de masse	97
<b>VI. Caractérisation de l'activité enzymologique du domaine d'adénylation de ClbK</b>	<b>99</b>
<b>VII. Introduction à la diffusion des rayons X en vue d'études structurales</b>	<b>101</b>

<b>VIII.</b>	<b>Caractérisation structurale par cristallographie aux rayons X</b>	<b>103</b>
VIII.1.	Tests de stabilité de la protéine dans différents tampons par essai Thermofluor	103
VIII.2.	Cristallogenèse	104
VIII.3.	Congélation des cristaux et enregistrement des données de diffraction	105
VIII.4.	Traitement des données de diffraction	105
VIII.5.	Phasage et affinement de la structure	106
<b>IX.</b>	<b>Caractérisation structurale par diffusion des rayons X aux petits angles (SAXS)</b>	<b>107</b>
IX.1.	Dispositif expérimental et acquisition des données	107
IX.2.	Soustraction du signal de diffusion du tampon	107
IX.3.	Analyse de Guinier	108
IX.4.	Fonction de distribution des distances	109
IX.5.	Graphique de Kratky	110
IX.6.	Modélisation <i>ab initio</i>	110
<b>X.</b>	<b>Caractérisation structurale par microscopie électronique à transmission</b>	<b>113</b>
X.1.	Principe de la méthode	113
X.2.	Coloration négative	114
X.3.	Cryo-microscopie électronique	115
<b>Résultats</b>		<b>117</b>
<b>I.</b>	<b>Etudes biochimiques et structurales de ClbK entière</b>	<b>119</b>
I.1.	Objectifs de l'étude de ClbK entière	119
I.2.	Mise en place d'un protocole d'expression et de purification de ClbK	119
I.3.	Détermination de l'état oligomérique de ClbK	123
I.4.	Détermination de la préférence de substrat du domaine d'adénylation de ClbK	124
I.5.	Eléments suggérant la présence d'un cofacteur FMN associé à ClbK	127
I.6.	Essais de cristallisation de ClbK	128
I.7.	Analyse structurale de ClbK en solution par diffusion des rayons X aux petits angles (SAXS)	131
I.8.	Observation de ClbK en microscopie électronique	137
I.9.	Bilan de l'étude de ClbK entière et perspectives	142
I.10.	Choix d'une stratégie de « dissection-reconstruction » de ClbK	143
<b>II.</b>	<b>Etudes biochimiques et structurales de ClbK-PKS</b>	<b>145</b>
II.1.	Objectifs de l'étude de ClbK-PKS	145
II.2.	Purification de ClbK-PKS	145
II.3.	Détermination de l'état de modification post-traductionnelle de ClbK-PKS	146
II.4.	Détermination de l'état oligomérique de ClbK-PKS	147
II.5.	Cristallisation de ClbK-PKS	147
II.6.	Résolution de la structure de ClbK-PKS*	151



II.7. Analyse de la structure cristallographique de ClbK-PKS* _____	153
<b>III. Etudes biochimiques et structurales d'autres fragments de ClbK _____</b>	<b>174</b>
III.1. Objectifs de l'étude d'autres fragments de ClbK _____	174
III.2. Purification des fragments ClbK-PKS-LK, ClbK-PKS-Cy et ClbK-PKS-A <sub>core</sub> _____	175
III.3. Détermination de l'état oligomérique des fragments ClbK-PKS-LK, ClbK-PKS-Cy et ClbK-PKS-A <sub>core</sub> _____	176
III.4. Tentatives d'analyse structurale de ClbK-PKS-Cy _____	177
III.5. Bilan de l'études des fragments ClbK-PKS-LK, ClbK-PKS-Cy et ClbK-PKS-A <sub>core</sub> et perspectives _____	178
<b><i>Discussion et perspectives</i> _____</b>	<b>181</b>
<b>Résumé des informations acquises sur ClbK _____</b>	<b>183</b>
<b>Modèle possible de l'architecture de ClbK _____</b>	<b>184</b>
<b>Hypothèses sur l'interaction de ClbK avec ses partenaires enzymatiques _____</b>	<b>186</b>
Interaction avec l'enzyme ClbG _____	186
Interactions avec les autres méga-enzymes Clb _____	188
<b>Intérêt des travaux pour l'ingénierie des NRPS/PKS _____</b>	<b>190</b>
<b><i>Références</i> _____</b>	<b>193</b>
<b><i>Annexe 1</i> _____</b>	<b>213</b>
<b><i>Annexe 2 : publication 1</i> _____</b>	<b>219</b>

# Abréviations



Les termes entre guillemets sont en anglais.

A (domaine)	domaine d'adénylation
Å	ångström
A <sub>core</sub> (sous-domaine)	grand sous-domaine A
A <sub>sub</sub> (sous-domaine)	petit sous-domaine A
ACP (domaine)	domaine « acyl carrier protein »
ADN	acide désoxyribonucléique
AMP	adénosine monophosphate
AT (domaine)	domaine acyltransférase
AT* (domaine)	domaine acyltransférase non fonctionnel
AT <sub>i</sub> (sous-domaine)	grand sous-domaine AT
AT <sub>s</sub> (sous-domaine)	petit sous-domaine AT
ATM	« ataxia telangiectasia mutated »
ATP	adénosine triphosphate
AU	unité d'absorbance
AVS	adénosine vinyl-sulfonamide
BAC	chromosome artificiel bactérien
BS <sup>3</sup>	bis(sulfosuccinimidyl) subérate
C (domaine)	domaine de condensation
CCR	cancer colorectal
ClbK-A-Ox-A	protéine contenant les résidus 1268 à 2023 de ClbK entière
ClbK-NRPS <sub>court</sub>	protéine contenant les résidus 812 à 2115 de ClbK entière
ClbK-NRPS <sub>long</sub>	protéine contenant les résidus 812 à 2154 de ClbK entière
ClbK-PKS	protéine contenant les résidus 1 à 787 de ClbK entière
ClbK-PKS*	protéine ClbK-PKS après clivage de l'étiquette poly-histidines
ClbK-PKS-A <sub>core</sub>	protéine contenant les résidus 1 à 1659 de ClbK entière
ClbK-PKS-A <sub>sub</sub>	protéine contenant les résidus 1 à 2023 de ClbK entière
ClbK-PKS-Cy	protéine contenant les résidus 1 à 1241 de ClbK entière
ClbK-PKS-LK	protéine contenant les résidus 1 à 812 de ClbK entière
ClbK-SER	forme de ClbK contenant les mutations K1718A, E1719A, E1721A
CoA	coenzyme A

Cy (domaine)	domaine de cyclisation
DEBS	« 6-deoxyerythronolide B synthase »
DMSO	diméthylsulfoxyde
D <sub>max</sub>	dimension maximale
DO <sub>600</sub>	densité optique mesurée à 600 nm
DTT	dithiothréitol
EDTA	acide éthylène diamine tétraacétique
EGS	éthylène glycol bis(sulfosuccinimidyl succinate)
ESRF	« European Synchrotron Radiation Facility »
ExPEC	<i>Escherichia coli</i> pathogène extra-intestinal
FAS	synthase d'acides gras, de l'anglais « fatty acid synthase »
F <sub>ab</sub>	« fragment antigen binding »
FMN	flavine mononucléotide
GA	glutaraldéhyde
GST	glutathion-S-transférase
Hepes	acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
HTX	« High Throughput Crystallization »
IPTG	isopropyl β-D-1-thiogalactopyranoside
ISBG	« Integrated Structural Biology Grenoble »
KS (domaine)	domaine cétosynthase, de l'anglais « ketosynthase »
LB	« lysogeny broth »
LD	domaine de liaison, de l'anglais « linker domain »
<i>linker</i>	peptide reliant deux domaines repliés
LINKS	« laterally-interacting ketosynthase sequence »
LS	lumière diffusée, de l'anglais « light scattering »
MES	acide 2-éthanosulfonique
MesG	7-méthyl-6-thioguanosine
MMT	acide malique, MES, Tris
NaCl	chlorure de sodium
NRPS	synthétase de peptide non ribosomaux, de l'anglais « non-ribosomal peptide synthetase »

Ox (domaine)	domaine oxydase
PAL	« post-AT linker »
PCP (domaine)	domaine « peptidyl carrier protein »
PCR	réaction en chaîne par polymérase
PDB	banque de données des structures protéiques, de l'anglais « protein data bank »
PEG	polyéthylène glycol
<i>pks</i>	îlot génomique « polyketide synthase »
PKS	synthase de polycétide, de l'anglais « polyketide synthase »
pLDDT	« predicted local-distance difference test »
PPant (bras)	bras phosphopantéthéine
PP <sub>i</sub>	pyrophosphate inorganique
PPTase	phosphopanthéthéinyl-transférase
R <sub>g</sub>	rayon de giration
RMSD	déviation de la racine de la moyenne des carrés
rpm	rotation par minute
SAXS	diffusion des rayons X aux petits angles
SDS	dodécylsulfate de sodium
SEC-MALLS	diffusion de lumière laser multi-angles couplée à la chromatographie d'exclusion de taille
SER	réduction de l'entropie de surface
SST	sodium silico tungstate
TB	« terrific broth »
TEV (protéase)	« tobacco etch virus »
T <sub>m</sub>	« melting temperature »
Tris	2-amino-2-hydroxyméthylpropane-1,3-diol
u.a.	unité arbitraire
UV	ultraviolet



# Introduction





## Préambule

*Escherichia coli* est une bactérie commensale de notre intestin mais aussi un pathogène humain majeur (Kaper *et al.*, 2004). La nature de l'interaction entre *E. coli* et son hôte repose notamment sur l'utilisation de facteurs de virulence. L'un d'eux est la colibactine, une génotoxine produite par l'îlot génomique « *polyketide synthase* » (*pkc*) (Wernke *et al.*, 2020). Les souches *pkc+* d'*E. coli* ont été décrites pour la première fois en 2006 (Nougayrède *et al.*, 2006) et les preuves de leur génotoxicité et de leur implication dans le cancer colorectal se sont accumulées depuis (Dougherty & Jobin, 2021). L'îlot *pkc* code 19 protéines permettant la synthèse d'un métabolite secondaire, la colibactine. Des efforts exceptionnels ont été fournis depuis 2006 pour élucider la structure de cette toxine et ont abouti en 2019 puisque deux groupes indépendants ont déterminé la structure d'un adduit entre la colibactine et l'ADN (Xue *et al.*, 2019 ; Jiang *et al.*, 2019). L'introduction débutera d'abord par un état de l'art des activités biologiques des souches *pkc+* d'*E. coli* (partie I). La détermination de la structure de la colibactine et de sa biosynthèse grâce à l'îlot *pkc* seront ensuite résumées (partie II).

Les protéines au cœur de la synthèse de la colibactine sont un assemblage de synthétases de peptides non ribosomiaux (NRPS) et de synthèses de polycétides (PKS). Ces méga-enzymes synthétisent de nombreux métabolites secondaires de grand intérêt médical (Katz & Baltz, 2016). La partie III explicitera l'importance de ces métabolites ainsi que la logique biochimique des NRPS/PKS permettant la synthèse de métabolites d'une extraordinaire diversité (Weissman, 2015b). Les NRPS/PKS fonctionnent selon le même principe qu'une chaîne d'assemblage : les domaines enzymatiques interviennent successivement selon un ordre prédéfini afin de synthétiser un métabolite. Le métabolite en construction est ainsi transporté d'un domaine à l'autre dans la chaîne d'assemblage. Dans les NRPS/PKS de type I, les domaines sont fusionnés entre eux dans un seul polypeptide. L'étude structurale de ces multi-enzymes, flexibles par nature, est donc délicate. Des avancées majeures ont eu lieu ces dernières années et les parties IV et V rendront compte des connaissances structurales des NRPS/PKS (Weissman, 2015a ; Bonhomme *et al.*, 2021 ; Bagde *et al.*, 2021 ; Cogan *et al.*, 2021 ; Fortinez *et al.*, 2022 ; Wang *et al.*, 2022). Une attention particulière sera portée aux domaines enzymatiques constitutifs de ClbK, le modèle d'étude de ma thèse, une méga-enzyme participant à la biosynthèse de la colibactine. La dernière partie s'attachera à introduire les protéines hybrides NRPS/PKS, dont ClbK fait partie. Mon

sujet d'étude est rattaché à deux sujets de la biologie en pleine expansion, la colibactine et les NRPS/PKS. Les connaissances sur ces deux sujets ont considérablement augmenté entre le début de ma thèse et aujourd'hui et une attention particulière sera donnée dans l'introduction aux avancées très récentes sur ces deux sujets.

## I. Activités biologiques des souches *pks+* d'*Escherichia coli* du microbiote intestinal

### I.1. Le microbiote intestinal et son rôle

Le microbiote désigne l'ensemble des micro-organismes associés à un individu, il est propre à chacun. Ces micro-organismes, en majorité des bactéries, sont localisés à la surface et à l'intérieur du corps humain et l'organe le plus riche en bactéries est le gros intestin, appelé ici côlon (Savage, 1977). On estime qu'un individu humain est composé d'au moins autant de cellules bactériennes que de cellules humaines, de l'ordre de  $10^{13}$  cellules (Sender *et al.*, 2016). Ce ratio suggère un rôle essentiel du microbiote dans la physiologie du corps humain.

Le développement des méthodes d'analyse génomique a permis des progrès considérables dans la caractérisation des microbiomes, notamment grâce au projet microbiote humain (Turnbaugh *et al.*, 2007). L'implication du microbiote a pu être démontrée dans plusieurs pathologies telles que l'obésité (Gérard, 2016) ou les maladies chroniques inflammatoires de l'intestin (Lavelle & Sokol, 2020). Des dysbioses, c'est-à-dire des déséquilibres du microbiote, sont associées à ces maladies (Levy *et al.*, 2017). La grande complexité du microbiote intestinal entrave généralement la capacité à associer une souche bactérienne unique à un effet spécifique sur la santé humaine. Une rare exception est la souche d'*Escherichia coli* possédant l'îlot génomique *polyketide synthase (pks)*, dont les preuves de l'implication dans la pathogenèse du cancer colorectal se sont accumulées ces dernières années.

### I.2. Diversité des souches d'*Escherichia coli*

*Escherichia coli* est une espèce de bactérie en forme de bacille de la famille des *Enterobacteriaceae*. C'est un organisme modèle en génétique, biologie moléculaire et microbiologie. Cette espèce versatile peut être aussi bien commensale de l'homme que pathogène (Kaper *et al.*, 2004). Plusieurs classifications rendent compte de la diversité de cette espèce. En particulier, l'espèce est divisée en groupes phylogénétiques ou phylogroupes et la plupart des souches appartiennent aux phylogroupes A, B1, B2 et D (Escobar-Páramo *et al.*, 2004). Le mode de vie d'une souche est relié à son histoire évolutive et donc à son phylogroupe. Ainsi, les souches commensales appartiennent généralement au phylogroupe A. Les souches pathogènes sont également divisées en différents pathotypes selon la maladie

qu'elles génèrent et la combinaison de facteurs de virulence qu'elles possèdent (Kaper *et al.*, 2004 ; Gatsios *et al.*, 2021). Par exemple, il existe six pathotypes d'*E. coli* causant des diarrhées. Les souches pathogènes extra-intestinales d'*E. coli* (ExPEC) causent des infections urinaires et des méningites et sont généralement associées au phylogroupe B2 (Russo & Johnson, 2000).

### I.3. Génotoxicité des souches *pks+* d'*E. coli*

#### I.3.a. Les souches *pks+* d'*E. coli* génèrent des cassures double-brin de l'ADN des cellules eucaryotes

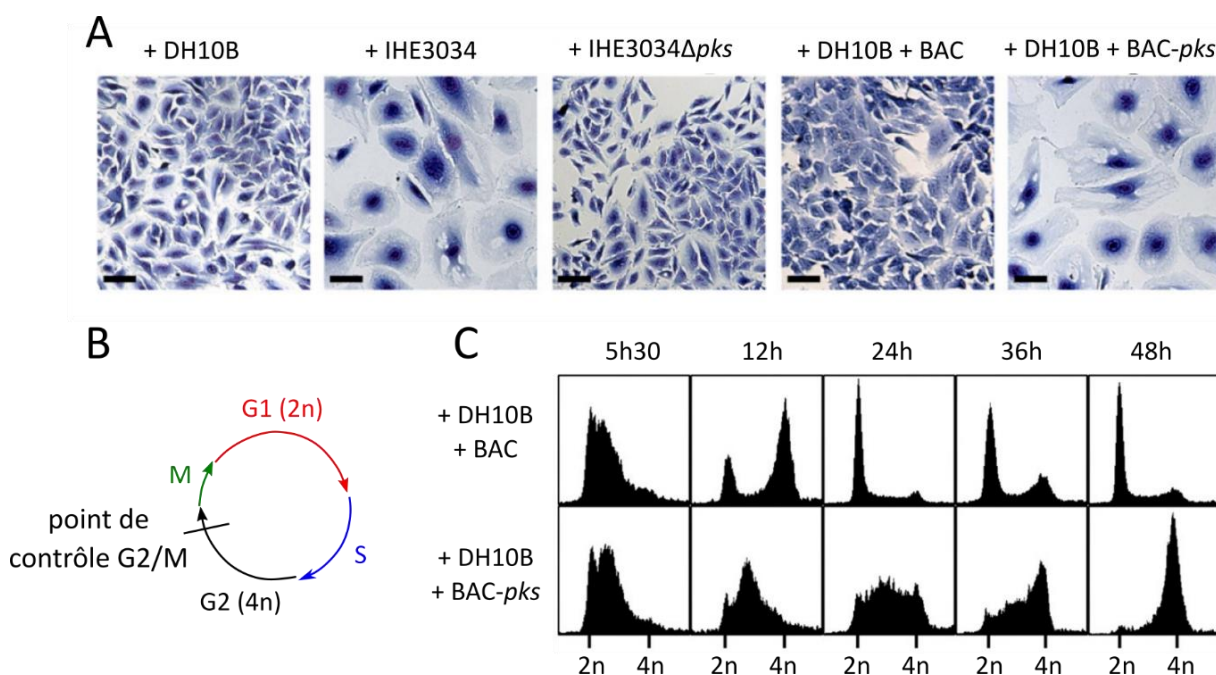


Figure 1. L'infection transitoire de cellules HeLa par des souches *pks+* d'*E. coli* induit un phénotype de mégalocytose et un arrêt du cycle cellulaire. Adapté de Nougayrède *et al.*, 2006. A) Des cellules HeLa ont été infectées pendant 4 heures par des bactéries *E. coli* à une multiplicité d'infection de 100 puis lavées et examinées par coloration Giemsa 72 heures après infection. La souche DH10B est une souche de laboratoire alors que la souche IHE3034 est une souche ExPEC possédant l'îlot *pks*. BAC : chromosome artificiel bactérien vide, BAC-*pks* : BAC contenant l'îlot *pks*. Barre d'échelle : 50  $\mu$ m. B) Schéma du cycle cellulaire. C) Des cellules HeLa synchronisées en phase G1/S ont été incubées pendant 4 heures avec des souches DH10B-BAC ou DH10B-BAC-*pks* et la progression du cycle cellulaire a été analysée par cytométrie en flux plusieurs heures après l'infection.

L'effet génotoxique de certaines souches d'*E. coli* sur des cellules de mammifères a été démontré pour la première fois en 2006 par Nougayrède *et al.* (Nougayrède *et al.*, 2006). L'équipe a remarqué que certaines souches d'*E. coli* du phylogroupe B2, telle que la souche ExPEC IHE3034 isolée d'un patient souffrant d'une méningite, induisent chez des cellules de mammifères en co-culture un phénotype de mégalocytose, c'est-à-dire un accroissement du volume du cytoplasme et du noyau associé à une absence de mitose (Figure 1A). Des

mutagénèses ont permis l'identification de l'îlot génomique *pks*, d'une taille de 54 kilobases, comme étant responsable du phénotype de mégalytose. Un îlot génomique est une portion de génome acquise par transfert horizontal et l'acquisition de l'îlot *pks* chez *E. coli* sera décrite dans la partie II.1 ci-dessous. L'îlot *pks* a été cloné dans un chromosome bactérien artificiel (BAC-*pks*) et des expériences de complémentation ont démontré que cet îlot est suffisant pour induire le phénotype de mégalytose chez les cellules en co-culture (Figure 1A). Ce phénotype n'est pas induit par l'incubation avec un lysat ou un surnageant de cultures de souches *pks+* d'*E. coli* et nécessite un contact entre la cellule eucaryote et la bactérie. L'inhibition de l'internalisation de la bactérie dans la cellule eucaryote n'abolit pas le phénotype induit de mégalytose. Ces résultats suggèrent que les bactéries possédant l'îlot *pks* produisent une toxine, la colibactine, responsable du phénotype de mégalytose observé chez les cellules eucaryotes avec lesquelles elles ont été co-cultivées.

Le phénotype de mégalytose rappelle l'effet des cyclomodulines, qui sont des toxines bactériennes perturbant le cycle cellulaire des cellules eucaryotes (Nougayrède *et al.*, 2005). Par exemple, les cyclomodulines Cif (« cycle inhibiting factor ») et CDT (« cytolethal distending toxin ») inhibent le cycle cellulaire. Au cours du cycle cellulaire, les cellules eucaryotes répliquent leur ADN pendant la phase S et se divisent pendant la mitose (Figure 1B). Le passage d'une phase à l'autre est régulé par des points de contrôle. Des cellules HeLa contrôles suivent le cycle classique et répliquent leur ADN puis se divisent en moins de 24 heures alors que des cellules HeLa infectées par des souches *pks+* répliquent leur ADN plus lentement et ne se divisent plus, ce qui indique qu'elles n'ont pas passé avec succès le point de contrôle G2/M (Figure 1C). Certaines cellules entrent en apoptose.

Le passage de la phase G2 à la mitose est régulé par le point de contrôle G2/M, qui est déclenché notamment par des cassures double-brin (Figure 1B). L'analyse de l'ADN des cellules eucaryotes par électrophorèse en conditions alcalines a montré que les souches *pks+* induisent des cassures double-brin dans l'ADN des cellules HeLa (Nougayrède *et al.*, 2006). Des analyses biochimiques ont montré que les souches *pks+* d'*E. coli* induisent l'autophosphorylation de la kinase ATM (de l'anglais « ataxia telangiectasia mutated ») (Bakkenist & Kastan, 2003). La phosphorylation détectée des protéines H2AX et Chk2 indique que toute la voie de signalisation initiée par l'autophosphorylation d'ATM est déclenchée par les souches *pks+*. La déphosphorylation de Chk2 est nécessaire au passage de la phase G2 à la

mitose. En résumé, les bactéries *pks+* induisent des cassures double-brin de l'ADN des cellules HeLa, ce qui initie la voie de signalisation ATM et *in fine* bloque l'entrée en mitose des cellules HeLa. L'induction de cassures double-brin et l'activation de la voie de signalisation ATM sont confirmées *in vivo* sur des portions d'intestin de souris incubées avec des souches *pks+* d'*E. coli* (Cuevas-Ramos *et al.*, 2010).

### **I.3.b. Les souches *pks+* d'*E. coli* induisent des alkylations de l'ADN et des ponts inter-brin**

Des cassures double-brin de l'ADN peuvent être causées par un mécanisme de réparation de certaines lésions à l'ADN, les ponts inter-brin (de l'anglais « interstrand crosslink ») (Deans & West, 2011). L'activation de la voie de signalisation ATM pourrait donc être une conséquence d'autres dommages infligés à l'ADN par les bactéries *pks+* d'*E. coli*. L'équipe de Nougayrède a montré que des cellules HeLa co-cultivées avec des souches *pks+* d'*E. coli* sont protégées de la génotoxicité bactérienne par l'ajout d'ADN extracellulaire de façon dose-dépendante (Bossuet-Greif *et al.*, 2018). L'analyse biochimique de l'ADN extracellulaire a révélé que cet ADN contient des ponts inter-brin, suggérant que la colibactine produite par les souches *pks+* est responsable de la formation de ponts inter-brin sur l'ADN. Les ponts inter-brin sont formés par alkylation de deux nucléotides localisés chacun sur un brin. Ils empêchent l'ouverture de l'ADN par l'hélicase au moment de la réplication, la bloquant. Les *E. coli pks+* induisent la voie de l'anémie de Fanconi, une voie de réparation de l'ADN utilisée par les cellules eucaryotes en réponse à des ponts inter-brin, qui est en lien avec d'autres voies de réparation (Bossuet-Greif *et al.*, 2018).

### **I.3.c. Les souches *pks+* d'*E. coli* génèrent des aberrations chromosomiques**

Les souches *pks+* induisent des ponts inter-brin et des cassures double-brin mais également des dommages plus importants de l'ADN, qui pourraient provenir d'une réparation incomplète des dommages initiaux. Des aberrations chromosomiques sont observées sur des portions d'intestin de souris exposées aux bactéries *pks+* (Cuevas-Ramos *et al.*, 2010) et sont également observées sur des organoïdes humains infectés par les souches *pks+* d'*E. coli* (Iftekhar *et al.*, 2021). Les dommages à l'ADN générés par les souches *pks+* d'*E. coli* ainsi que leur localisation intestinale suggèrent un lien entre ces souches et le cancer colorectal.

## I.4. Implication des souches *pks+* d'*E. coli* dans le cancer colorectal

### I.4.a. Le cancer colorectal

Le côlon est la portion du tube digestif après l'intestin grêle. Il possède une fonction de récupération d'eau et de propulsion des fèces vers le rectum. Le côlon et le rectum ont une histologie semblable. Le cancer est une maladie causée par la prolifération anarchique de cellules qui forment *in fine* une masse nommée tumeur. Le cancer colorectal (CCR) est le deuxième cancer le plus meurtrier et 916 000 décès ont été attribués au CCR dans le monde en 2020 (Organisation mondiale de la santé). Il peut être guéri s'il est détecté précocement. Seulement 5 % des CCR sont héréditaires (Kwak & Chung, 2007) et les 95 % restants sont sporadiques, c'est-à-dire causés par une combinaison de facteurs génétiques et environnementaux.

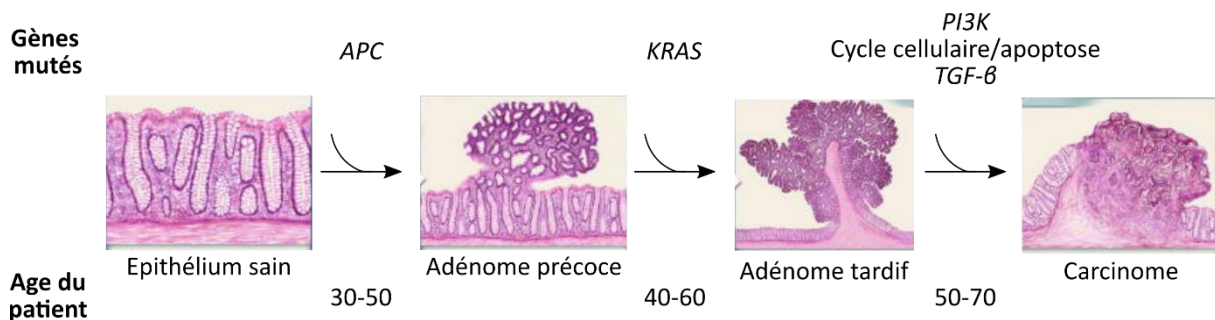


Figure 2. Stades du cancer colorectal (CCR) et mutations génétiques associées, modèle valable pour la plupart des cas de CCR. Adapté de Vogelstein *et al.*, 2013. L'âge qu'ont généralement les patients lors de la mutation des gènes clés dans la carcinogenèse est indiqué. APC : « adenomatous polyposis coli », KRAS : « kristen rat sarcoma viral oncogen homolog », PI3K : phosphoinositide 3-kinase, TGF-β : facteur de croissance transformant β.

La transformation d'un épithélium sain en une tumeur est divisée en étapes et les mutations génétiques clés initiant le passage d'une étape à l'autre sont bien connues (Vogelstein *et al.*, 2013) (Figure 2). La plupart de ces mutations sont des substitutions d'un seul nucléotide. La première mutation est généralement celle du gène suppresseur de tumeur APC (de l'anglais « adenomatous polyposis coli ») dans une cellule épithéliale. Ce gène doit son nom à la polypose adénomateuse familiale, une forme héréditaire du CCR causée par une mutation du gène APC. Cette mutation participe à l'activation constitutive de la voie de signalisation Wnt, normalement activée constitutivement uniquement dans les cellules souches et induit donc la prolifération de la cellule mutée, générant un adénome précoce. Une mutation d'un deuxième gène, tel que le gène KRAS (de l'anglais « kristen rat sarcoma viral oncogene homolog ») augmente encore la capacité proliférative des cellules mutées. Une



troisième mutation dans des gènes tels que *P53* génère une tumeur maligne ou carcinome, qui peut envahir d'autres parties du corps en formant des métastases.

Plusieurs facteurs de risque sont identifiés pour le cancer colorectal tels que l'alimentation trop riche, le surpoids ou l'inflammation chronique de l'intestin. Les maladies inflammatoires de l'intestin, comprenant la maladie de Crohn et la rectocolite hémorragique, sont d'ailleurs un facteur de risque pour le CCR (Stidham & Higgins, 2018). Ces dernières années, les preuves du rôle du microbiote intestinal dans l'initiation et le développement du CCR se sont accumulées (Tilg *et al.*, 2018). Des dysbioses sont associées au CCR et certaines espèces bactériennes sont particulièrement retrouvées chez les patients ayant un CCR, notamment *Fusobacterium nucleatum*, *Bacteroides fragilis* et certaines souches d'*Escherichia coli*.

#### **I.4.b. Etudes épidémiologiques**

Plusieurs études épidémiologiques démontrent que les souches *pks+* d'*E. coli* sont plus associées à des patients ayant un CCR qu'à des patients contrôles. Une première étude montre que les souches *pks+* d'*E. coli* sont retrouvées dans 55 % des biopsies de patients avec CCR et dans seulement 19 % des patients contrôles ayant des diverticules sur la surface du côlon (Buc *et al.*, 2013). Des valeurs très similaires sont retrouvées par étude des selles de patients avec CCR ou patients contrôles (Eklöf *et al.*, 2017). Les souches *pks+* sont également significativement plus retrouvées chez les patients souffrant de maladie inflammatoire de l'intestin et de polypose adénomateuse familiale (Dejea *et al.*, 2018).

Ces chiffres suggèrent un lien entre les souches *pks+* d'*E. coli* et le CCR mais ne démontrent pas une causalité. On pourrait par exemple supposer que l'environnement tumoral confère un avantage sélectif aux souches *pks+* d'*E. coli*.

#### **I.4.c. Effet des souches *pks+* d'*E. coli* sur des modèles murins**

Afin de démontrer une causalité des souches *pks+* dans l'initiation ou la progression du CCR, plusieurs expériences ont été menées sur des modèles murins. Une première étude a investigué l'effet des souches *pks+* d'*E. coli* sur un modèle de souris mimant le contexte du CCR (souris IL10<sup>-/-</sup> traitées par l'azoxyméthane) (Arthur *et al.*, 2012). Les souches *pks+* d'*E. coli* ont un effet pro-tumoral dans ce contexte pro-inflammatoire et en présence d'un carcinogène. L'effet pro-tumoral des souches *pks+* d'*E. coli* a également été démontré chez

des souris Min, un autre modèle animal de CCR (Bonnet *et al.*, 2014). Ces études démontrent que les souches *pks+* d'*E. coli* favorisent la progression du CCR dans les modèles murins.

Afin de tester l'effet des souches *pks+* dans un contexte non pro-tumoral, des souris aseptisées (« germ free ») ont été infectées par des *E. coli pks+* ou *pks-* puis l'ADN des cellules épithéliales du côlon de ces souris a été extrait (Wilson *et al.*, 2019). Des adduits de l'ADN avec la molécule de colibactine sont détectés uniquement dans les cellules des souris infectées par les souches *pks+*, ce qui montre que l'effet génotoxique des souches *pks+* se produit en l'absence de conditions pro-tumorales et suggère que les souches *pks+* jouent également un rôle dans l'initiation du CCR.

#### I.4.d. Signature mutationnelle induite par les souches *pks+* d'*E. coli*

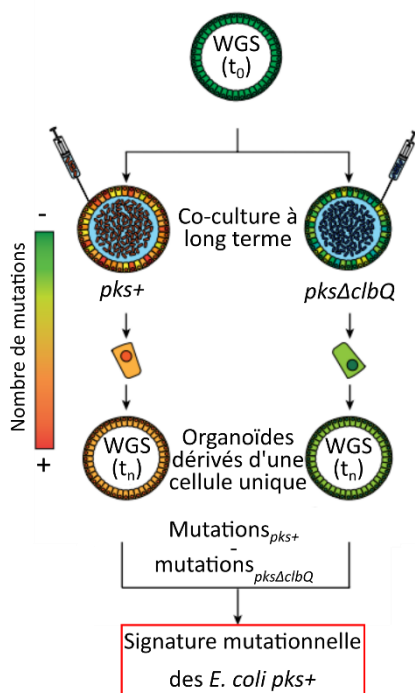


Figure 3. Schéma de l'expérimentation utilisée pour la détermination de la signature mutationnelle des souches *pks+* d'*E. coli*. Adapté de Pleguezuelos-Manzano *et al.*, 2020. Des souches *pks+* ou *pksΔclbQ* d'*E. coli* ont été micro-injectées dans des organoïdes, qui ont été co-cultivés avec ces bactéries pendant plusieurs mois. Des nouveaux organoïdes dérivés d'une cellule unique issue des premiers organoïdes ont été générés et le génome complet de ces organoïdes a été séquencé (WGS : « whole genome sequencing »). La comparaison de ces génomes a permis l'identification de la signature mutationnelle des *E. coli pks+*.

Deux études publiées en 2020 ont déterminé la signature mutationnelle des souches *pks+* d'*E. coli* et ont identifié cette signature dans une portion des CCR, démontrant ainsi le rôle des souches *pks+* dans l'initiation de certains CCR (Pleguezuelos-Manzano *et al.*, 2020 ; Dziubańska-Kusibab *et al.*, 2020). La première étude a utilisé un outil novateur, les organoïdes, qui sont des structures tri-dimensionnelles mimant un organe, un intestin dérivé de cellules épithéliales humaines dans cette étude (V. S. W. Li, 2020) (Figure 3). Des *E. coli pks+* ou délétées d'un gène essentiel à la génotoxicité, *clbQ* (*pksΔclbQ*) ont été injectées dans le lumen de deux organoïdes. Après seulement un jour d'exposition, les dommages à l'ADN décrits

précédemment sont détectés uniquement chez les organoïdes exposés à la souche *pks+* d'*E. coli*. Des injections d'*E. coli pks+* ou *pksΔclbQ* ont alors été répétées pendant cinq mois afin de mimer une exposition à long terme par les souches bactériennes. Plusieurs organoïdes ont été dérivés des organoïdes initiaux et leurs génomes ont été séquencés entièrement. Des mutations spécifiques sont alors identifiées uniquement dans les organoïdes exposés à la souche *pks+* d'*E. coli*.

Deux sortes de mutations sont détectées : les substitutions d'une seule base (« single-base substitution », SBS) et des délétions dans des séquences riches en base thymine. Les SBS sont principalement retrouvées sur la base thymine localisée dans des motifs ATA, ATT et TTT (la base mutée est soulignée) présentant une base adénine 3 paires de bases en aval de la base thymine. Ces signatures mutationnelles ont été recherchées dans les séquences de tumeurs. Un enrichissement significatif (5-10 %) est retrouvé dans les tumeurs de CCR, ainsi que dans certains autres types de tumeurs. Les mutations connues des gènes causant le CCR ont été analysées afin d'établir un lien entre cette signature et l'initiation du CCR. La signature est retrouvée dans 2.4 % (112/4712) des mutations de gènes clés dans l'initiation et la progression du CCR. 52 mutations touchent le gène *APC*, gène clé dans la carcinogénèse du CCR. Ces résultats, confortés par l'étude de Dziubańska-Kusibab *et al.*, suggèrent la causalité des souches *pks+* *E. coli* dans certains CCR (Dziubańska-Kusibab *et al.*, 2020). Cette signature mutationnelle est compatible avec la structure et le mécanisme d'action de la colibactine (voir II.3 ci-dessous).

#### **I.4.e. Les souches *pks+* d'*E. coli* induisent la transformation tumorale de cellules épithéliales de côlon**

Des organoïdes ont également été utilisés pour démontrer la capacité de transformation de cellules saines en cellules tumorales par les souches *pks+* d'*E. coli* (Iftekhhar *et al.*, 2021). Une exposition courte (quelques heures) d'organoïdes dérivés de cellules épithéliales primaires de côlon de souris par des *E. coli pks+* confère à ces organoïdes des caractéristiques de cellules cancéreuses du CCR. Notamment, la voie de signalisation Wnt est constitutivement active dans ces cellules. De plus, ces cellules prolifèrent anormalement et présentent des signes de dédifférenciation. Le séquençage du génome des organoïdes ayant la voie Wnt activée constitutivement a également révélé plusieurs aberrations chromosomiques, qui sont retrouvés dans une majorité des CCR sporadiques (Cancer Genome

Atlas Network, 2012). La signature mutationnelle n'a pas été identifiée dans ces organoïdes, suggérant que la signature mutationnelle et les aberrations chromosomiques résultent de mécanismes différents.

#### I.4.f. Modèle du rôle des souches *pks+* d'*E. coli* sur le cancer colorectal

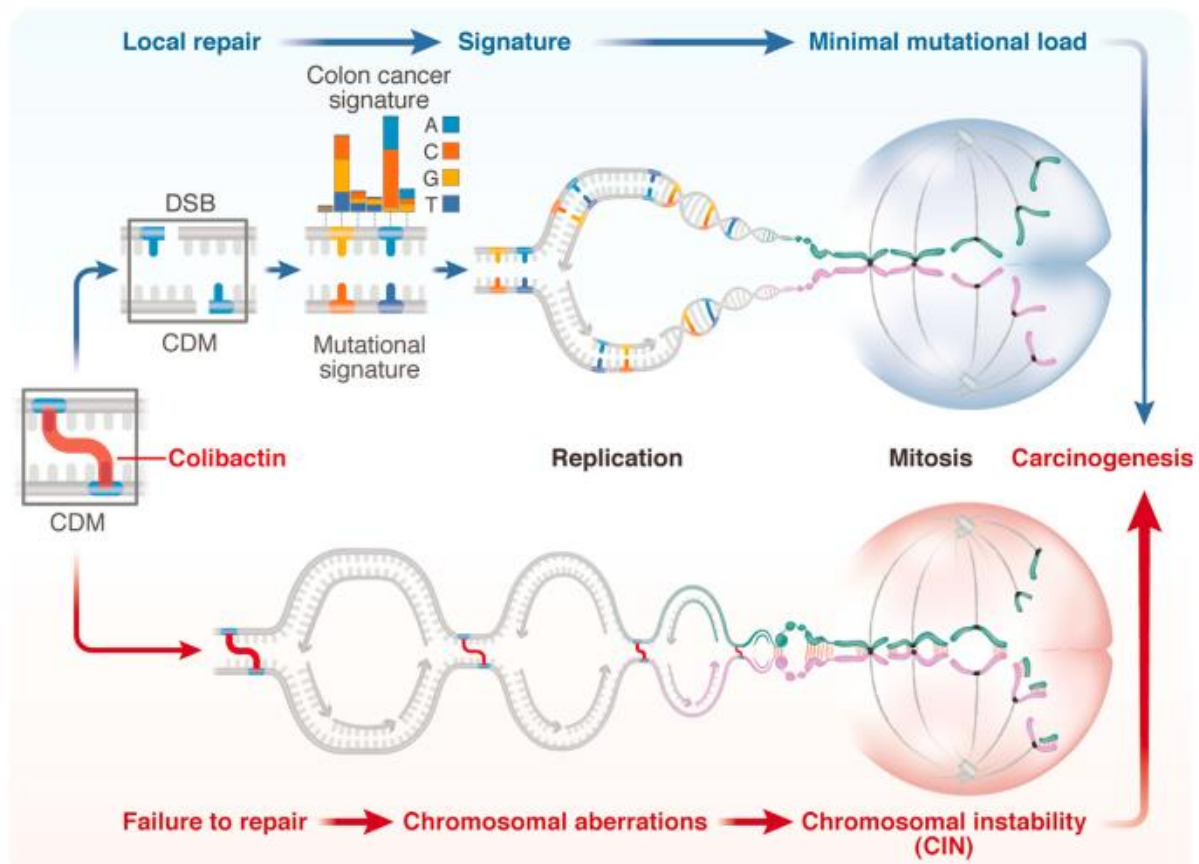


Figure 4. Conséquences sur la cellule des ponts inter-brin générés par la colibactine produite par les souches *pks+* d'*E. coli*. Modèle proposé par Berger & Meyer, 2021. CDM : motif cible de la colibactine, de l'anglais « colibactin damage motif », DSB : cassure double-brin, de l'anglais « double-strand break ».

La co-culture de souches *pks+* d'*E. coli* avec des cellules eucaryotes génère plusieurs types de dommages à l'ADN dans ces cellules : ponts inter-brin, cassures double-brin, mutations et aberrations chromosomiques. Le modèle présenté à la Figure 4 résume l'impact des souches *pks+* d'*E. coli* produisant la colibactine sur l'ADN des cellules eucaryotes (Berger & Meyer, 2021). L'effet génotoxique direct de la colibactine est le pontage inter-brin de l'ADN (Xue *et al.*, 2019). Selon ce modèle, deux cas de figure se présentent alors. Dans le premier cas, les ponts sont suffisamment peu nombreux pour être réparés par les mécanismes de réparation de la cellule, ce qui génère des cassures double-brin et des mutations. La signature mutationnelle détectée dans une faible proportion de cancers serait la marque des dommages

dus à la colibactine imparfaitement réparés par la cellule. Dans le deuxième cas, les ponts sont tellement nombreux que la cellule ne peut pas tous les réparer, ce qui bloque la réplication et entraîne des aberrations chromosomiques. Ce modèle suggère que la signature mutationnelle présente dans 5-10 % des CCR ne serait que la partie émergée de l'iceberg et que la colibactine produite par les souches *pks+* d'*E. coli* pourrait jouer un rôle dans une bien plus large proportion des CCR.

On peut signaler ici que les souches *pks+* d'*E. coli* sont acquises très tôt au cours de la vie, par des contacts mère-enfant à la naissance ou lors de l'allaitement (Tsunematsu *et al.*, 2021). Une détection précoce de cette souche pourrait être envisagée en prévention du cancer colorectal.

#### 1.5. Implication des souches *pks+* d'*E. coli* dans d'autres types de cancers

La signature mutationnelle des souches *pks+* d'*E. coli* est également retrouvée dans des tumeurs du côlon mais aussi dans une faible fraction de tumeurs de la tête et du cou, ainsi que des tumeurs des voies urinaires (vessie ou reins) (Pleguezuelos-Manzano *et al.*, 2020). Ces deux sites sont des sites d'infection par *E. coli* donc les souches génotoxiques *pks+* d'*E. coli* pourraient jouer un rôle dans certains de ces cancers.

Deux études récentes ont commencé à établir un lien de causalité entre les souches *pks+* d'*E. coli* et un cancer autre que le CCR. Une première étude a d'abord montré que les souches *pks+* d'*E. coli* sont particulièrement retrouvées chez les patients atteints d'infection urinaire (Chagneau *et al.*, 2021). Un modèle murin d'infection urinaire a permis de démontrer la génotoxicité d'une souche isolée d'un patient ayant une infection urinaire sur les cellules urothéliales. Une deuxième étude a suggéré une association entre les souches *pks+* d'*E. coli* et l'initiation du cancer de la prostate (Shrestha *et al.*, 2021). Environ la moitié des cancers de la prostate sont caractérisés par une fusion entre les gènes *TMPRSS2* et *ERG* conduisant à une surexpression de l'oncogène *ERG* (Tomlins *et al.*, 2005). Les souches *pks+* d'*E. coli* sont détectées dans certains spécimens de prostatectomie et semblent capables d'induire des cassures chromosomiques dans les gènes *TMPRSS2* et *ERG*, suggérant que certains cancers de la prostate pourraient être causés par une infection précoce par des souches *pks+* d'*E. coli*.

### 1.6. Autres effets des souches *pks+* d'*E. coli*

L'intérêt biologique de l'îlot *pks* pour la bactérie le possédant est une question centrale. En effet, l'induction ou la promotion de cancer chez son hôte n'apporte pas de bénéfices évidents à une bactérie, il semble donc peu probable que l'effet carcinogène soit la seule fonction physiologique de l'îlot *pks*. Plusieurs études suggèrent l'importance des souches possédant l'îlot *pks* dans l'écologie du microbiote, qui est un environnement complexe composé de nombreuses espèces microbiennes. Une première étude a démontré un effet antibiotique modéré des souches *pks+* d'*E. coli* contre plusieurs souches de *Staphylococcus aureus*, en particulier contre des souches multi-résistantes aux antibiotiques (Faïs *et al.*, 2016). Le mécanisme de l'effet antibiotique n'a pas été déterminé à l'époque mais une étude récente suggère fortement que cet effet reposerait sur l'induction du cycle lytique de prophages intégrés dans le génome de *S. aureus* (Silpe *et al.*, 2022).

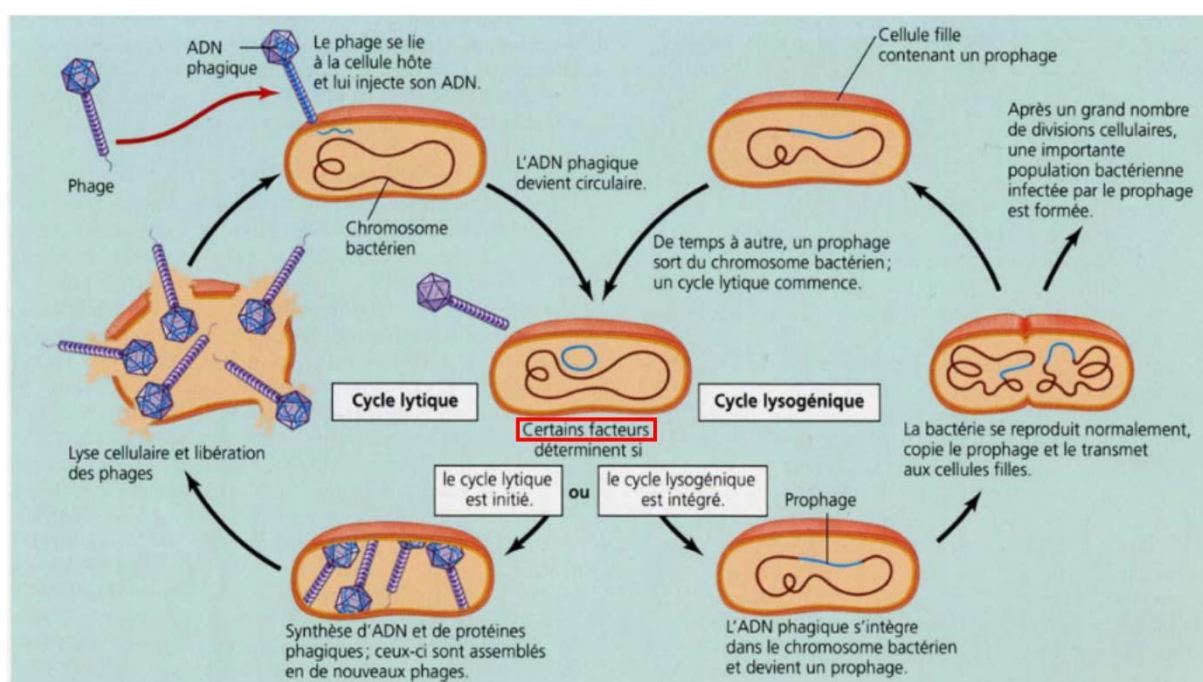


Figure 5. Le cycle lytique et le cycle lysogénique du phage lambda. D'après l'ouvrage de biologie Campbell, 2012. La colibactine produite par les souches *pks+* d'*E. coli* est un des facteurs déterminant le passage du cycle lysogénique au cycle lytique.

Les bactériophages sont des virus infectant les bactéries (Campbell, 2012) (Figure 5). Après infection, le virus peut entrer dans un cycle lysogénique, c'est-à-dire que l'ADN du phage s'intègre dans le chromosome bactérien et devient un prophage. Le virus peut également entrer dans un cycle lytique, c'est-à-dire détourner la machinerie bactérienne pour synthétiser de nouveaux phages et finalement lyser la cellule hôte. Les souches *pks+* d'*E. coli*

induisent l'initiation du cycle lytique chez plusieurs espèces de bactéries contenant des prophages intégrés dans leur génome, en particulier chez des souches d'*E. coli*, de *S. aureus*, de *Salmonella enterica* ou encore d'*Enterococcus faecium*. L'ajout d'ADN extracellulaire diminue l'induction du cycle lytique du phage, aussi appelée induction du prophage, suggérant que les dommages à l'ADN générés par la colibactine déclenchent cette induction du prophage. Par ailleurs, certaines bactéries possèdent uniquement un gène de l'îlot *pks*, le gène *clbS* codant la protéine de résistance à la colibactine ClbS (Bossuet-Greif *et al.*, 2016 ; Tripathi *et al.*, 2017). Silpe *et al.* ont démontré l'effet protecteur de ClbS chez des souches *pks*- contre l'induction du prophage par les souches *pks*+. L'effet d'induction du prophage par des souches *pks*+ pourrait expliquer les changements de composition du microbiote observés chez des souris après introduction d'une souche *pks*+ d'*E. coli* (Tronnet *et al.*, 2020).

Une autre étude a montré un effet bactéricide des souches *pks*+ contre des souches de *Vibrio cholerae*, l'entéropathogène responsable de la maladie du choléra (Chen *et al.*, 2022). Comme pour la génotoxicité des cellules eucaryotes et l'induction du prophage des cellules bactériennes, un contact est nécessaire entre la bactérie *pks*+ et l'autre cellule. L'effet bactéricide repose sur les cassures double-brin dans le génome des *Vibrio cholerae* mais le rôle des prophages n'a pas été étudié dans cette étude.

Enfin, l'îlot *pks* confère un effet anti-inflammatoire à certaines souches le possédant, notamment la souche d'*E. coli* Nissle 1917 utilisée comme probiotique (Olier *et al.*, 2012). Commercialisé sous le nom de Mutaflor®, ce probiotique est utilisé depuis plus d'un siècle dans le traitement de certaines maladies intestinales dont la rectocolite hémorragique (Schultz, 2008). Il a été démontré qu'un lipopeptide produit par l'îlot *pks* confère à la souche ses propriétés analgésiques (Pérez-Berezo *et al.*, 2017). Toutefois, la souche Nissle 1917, comme les autres souches *pks*+, est génotoxique *in vitro* et *in vivo* et la sûreté de ce traitement est aujourd'hui questionnée (Nougayrède *et al.*, 2021). Il semblerait donc que, chez la souche Nissle 1917, l'îlot *pks* soit responsable de la biosynthèse de plusieurs métabolites ayant une variété d'actions biologiques.

## II. La biosynthèse de la colibactine par l'îlot génomique *pks*

### II.1. Distribution et acquisition de l'îlot *pks* chez les bactéries

L'îlot *pks* a été décrit principalement chez des Entérobactéries commensales de l'homme. Il a été découvert chez *E. coli* mais a été reporté chez d'autres Entérobactéries telles que *Klebsiella pneumoniae*, *Citrobacter koseri* et *Enterobacter aerogenes* (Putze *et al.*, 2009). L'organisation et l'identité génétique de l'îlot *pks* sont très conservées parmi ces Entérobactéries (Figure 6). En particulier, l'îlot *pks* du pathogène opportuniste *Klebsiella pneumoniae* présente une identité de séquence de 100 % avec celui d'*E. coli*, suggérant que les souches *pks+* de *K. pneumoniae* pourraient également jouer un rôle dans le CCR (Strakova *et al.*, 2021).

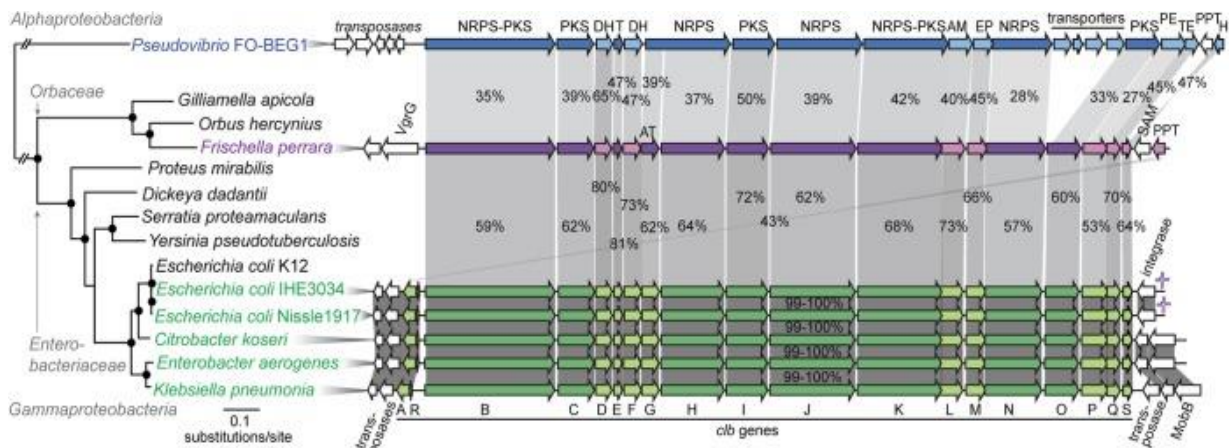


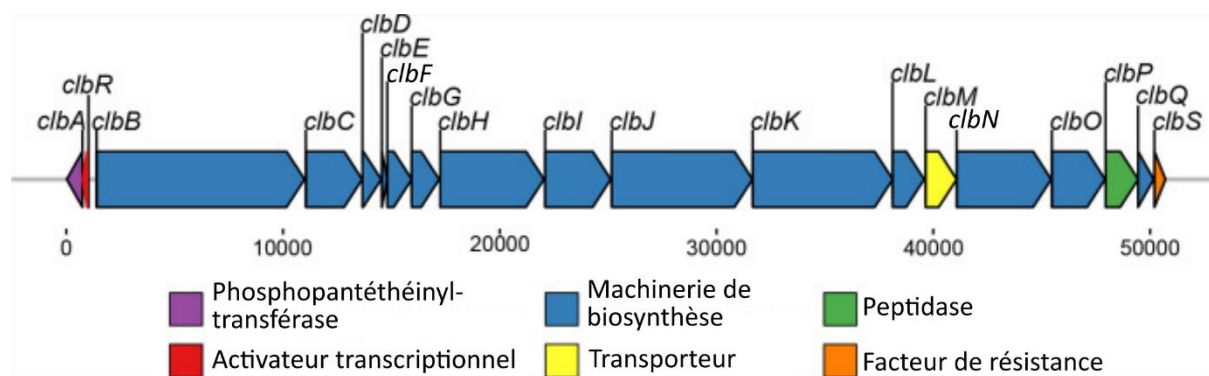
Figure 6. Relations phylogénétiques entre les bactéries possédant l'îlot *pks* et comparaison de l'organisation génétique de l'îlot chez ces bactéries. D'après Engel *et al.*, 2015. Les souches en bleu (*Pseudovibrio* FO-BEG1), violet (*Frischella perrara* PEB0191) et vert (*Enterobacteriaceae*) possèdent l'îlot *pks*. Les gènes orthologues sont connectés entre eux par des blocs gris. Les gènes sans homologues sont représentés en blanc. AM : amidase, DH : déshydrogénase, EP : protéine d'efflux, H : hydrolase, MobB : protéine de mobilisation B, PE : peptidase, PPT : phosphopantéthéinyl-transférase, SAM : protéine liant la S-adénosylméthionine, T : « carrier protein », TE : thioestérase.

L'îlot *pks* a également été décrit chez deux symbiontes bactériens n'appartenant pas à la famille des Entérobactéries, une souche de *Frischella perrara* isolée d'une abeille (Engel *et al.*, 2015) et une souche de *Pseudovibrio* isolée d'un corail (Bondarev *et al.*, 2013). Malgré une faible identité de séquence avec l'îlot *pks* des Entérobactéries, l'organisation génétique de l'îlot chez ces souches est globalement conservée avec celle des Entérobactéries (Figure 6). La génotoxicité de la souche *pks+* de *Frischella perrara* a même été démontrée *in vitro* (Engel *et al.*, 2015).



A l'exception de *Frischella perrara*, l'îlot *pks* des autres souches est bordé par des éléments génétiques mobiles, en particulier des gènes qui codent des transposases et des intégrases, témoignant d'une acquisition par transfert horizontal de gènes (Engel *et al.*, 2015) (Figure 6). Chez la plupart des souches d'*E. coli* du phylogroupe B2, cet îlot est inséré au niveau du locus de l'ARNt *asnW* (Auvray *et al.*, 2021). Ce locus est situé à proximité de l'îlot génomique codant la biosynthèse du sidérophore yersiniabactine, indiquant que l'îlot *pks* a suivi une trajectoire évolutive similaire à cet îlot (Wami *et al.*, 2021).

## II.2. Organisation et régulation de l'îlot *pks*



L'îlot *pks* code 19 protéines, toutes essentielles à la génotoxicité de la souche d'*E. coli* le possédant, à l'exception de la protéine de transport ClbM, pouvant probablement être complétée par une autre pompe à efflux (Nougayrède *et al.*, 2006 ; Mousa *et al.*, 2016). La majorité des protéines codées par l'îlot sont des enzymes qui participent directement à la biosynthèse de la colibactine (Figure 7). La peptidase ClbP (Dubois *et al.*, 2011 ; Cougnoux *et al.*, 2012) et l'hydrolase ClbS (Bossuet-Greif *et al.*, 2016 ; Tripathi *et al.*, 2017) sont deux protéines permettant à la bactérie productrice de se protéger contre l'effet génotoxique de la colibactine. ClbQ est une thioestérase hydrolysant les intermédiaires attachés aux protéines de biosynthèse de la colibactine (Guntaka *et al.*, 2017). Deux protéines, ClbA et ClbR, sont codées sur le brin opposé du brin où sont codées toutes les autres protéines Clb. ClbA est une enzyme qui modifie post-traductionnellement certaines enzymes de biosynthèse de la colibactine et ClbR est l'activateur transcriptionnel de l'îlot *pks* (Wallenstein *et al.*, 2020).

Les 19 protéines codées par l'îlot sont traduites à partir de sept transcrits dont quatre polycistroniques (Homburg *et al.*, 2007). Les protéines ClbB, ClbH et ClbQ sont codées à partir d'ARNm monocistroniques. Les protéines codées par des ARNm polycistroniques sont ClbC à

ClbG, ClbI à ClbN, ClbO-ClbP et ClbA-ClbR. L'unité génétique codant ClbA et ClbR a été acquise postérieurement au reste de l'îlot génomique et apparaît essentielle pour la régulation de la biosynthèse de la colibactine (Wami *et al.*, 2021).

Un paramètre important pour la régulation de l'expression de l'îlot *pks* est la disponibilité en fer. En effet, l'expression de ClbA et de ClbR est diminuée dans un milieu riche en fer (Tronnet *et al.*, 2017 ; Wallenstein *et al.*, 2020). Par conséquent, la production de colibactine est réduite dans un milieu riche en fer (Tronnet *et al.*, 2017). Les voies de biosynthèse de la colibactine et du sidérophore yersiniabactine sont connectées par la phosphopantéthéinyl-transférase ClbA, capable de modifier post-traductionnellement les enzymes de la biosynthèse de la yersiniabactine dont la voie de biosynthèse ne possède pas d'enzyme équivalente (Martin *et al.*, 2013). L'effet synergique de la colibactine et de la yersiniabactine pourrait augmenter la virulence des souches possédant les îlots génomiques permettant la biosynthèse de ces deux métabolites.

D'autres facteurs physiologiques impactent l'expression des gènes de l'îlot *pks* et/ou la production de colibactine. Par exemple, la spermidine, une polyamine abondante dans les tissus cancéreux, favorise la génotoxicité des souches *pks+* (Chagneau *et al.*, 2019). L'inflammation active la transcription des gènes de l'îlot *pks* dans des souris modèles du CCR (Arthur *et al.*, 2014). La régulation de la biosynthèse de la colibactine par de nombreux facteurs physiologiques souligne l'importance de ce métabolite pour la bactérie le produisant (Dougherty & Jobin, 2021).

### II.3. Elucidation de la structure de la molécule génotoxique, la colibactine

Malgré des efforts considérables, la faible production de colibactine par les souches *pks+* et son instabilité n'ont pas encore permis d'isoler la molécule de colibactine afin d'en élucider directement sa structure (Williams *et al.*, 2020). Plusieurs stratégies ont toutefois permis la détermination indirecte de la structure de la colibactine. Il a rapidement été établi que les bactéries *pks+* ne synthétisent pas directement la génotoxine mais une molécule non toxique, la pré-colibactine contenant un motif protecteur N-myristoyl-D-asparagine (Brotherton & Balskus, 2013). L'export de la pré-colibactine dans le périplasme par ClbM (Mousa *et al.*, 2016) et le clivage du motif protecteur par la peptidase ClbP (Dubois *et al.*, 2011) génèrent la molécule de colibactine. Plusieurs molécules de pré-colibactines ont été isolées à

partir de cultures à grande échelle de souches *pks+ΔclbP* d'*E. coli* mais il a ensuite été démontré que certaines de ces molécules étaient des artefacts causés par la mutation de *clbP* (Williams *et al.*, 2020).

Les recherches sur la structure de la colibactine ont alors reposé sur une stratégie différente, l'isolation et l'élucidation de fragments d'ADN endommagés par la colibactine (Williams *et al.*, 2020). La structure d'un adduit entre la colibactine et une adénine a d'abord été déterminée par deux groupes indépendants (Xue *et al.*, 2018 ; Wilson *et al.*, 2019). Peu après, c'est la structure du pont inter-brin généré par la colibactine entre deux adénines qui a été caractérisée par les deux groupes (Xue *et al.*, 2019 ; Jiang *et al.*, 2019). Ces caractérisations structurales ont été permises par une multitude d'expériences, notamment des marquages isotopiques, de la spectrométrie de masse en tandem et de la synthèse chimique.

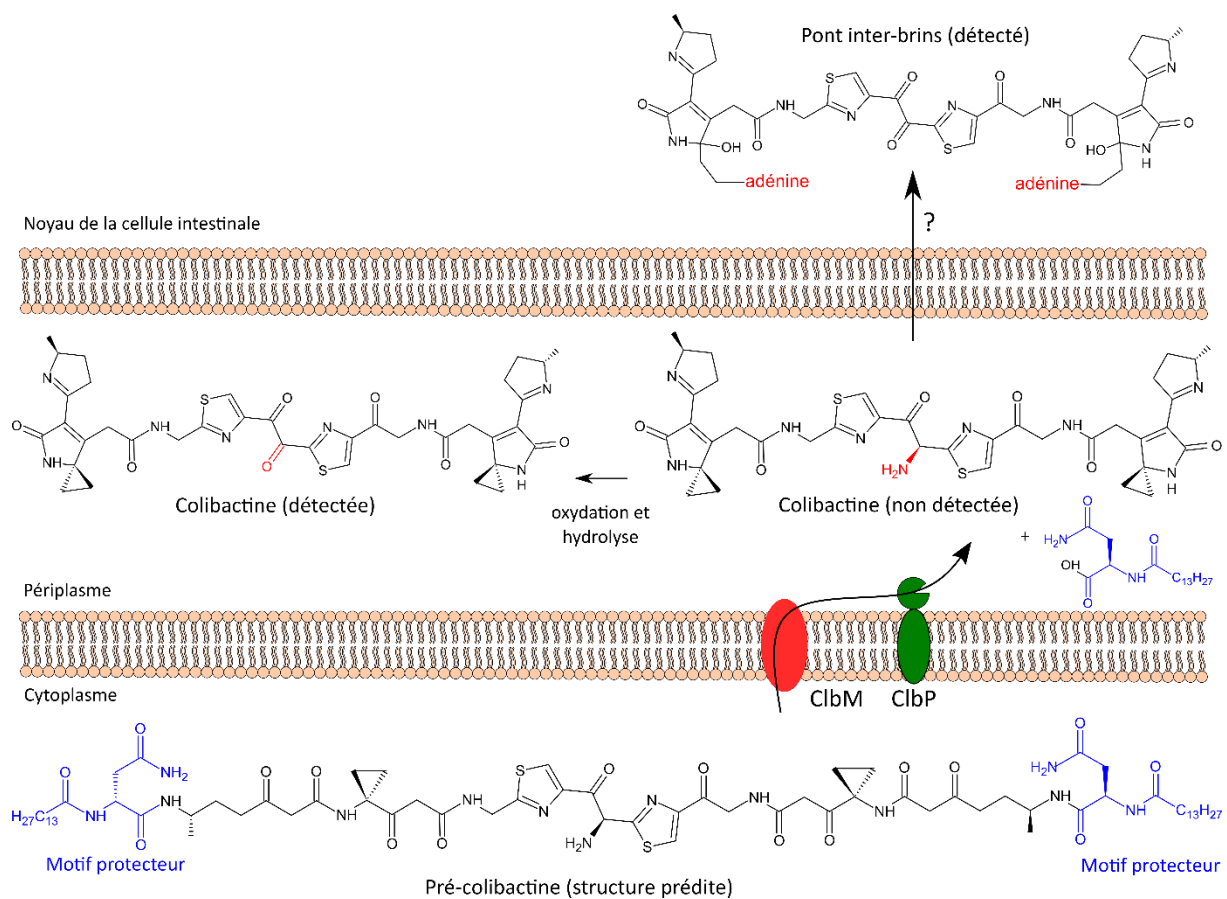


Figure 8. Modèle de la structure de la colibactine, et des dernières étapes de sa biosynthèse. Adapté de Gatsios *et al.*, 2021. La pré-colibactine est exportée dans le périplasm par la protéine d'export ClbM puis le motif protecteur, en bleu, est clivé par la peptidase périplasmique ClbP, générant la génotoxine colibactine. Le mode de transport de la colibactine depuis le périplasm jusqu'au noyau de la cellule intestinale n'est pas connu.

La colibactine est une molécule hétérodimérique contenant deux cycles cyclopropanes responsables de l'alkylation des adénines de l'ADN générant les ponts inter-brin (Figure 8). La structure de la colibactine est compatible avec sa signature mutationnelle décrite à la section I.4.d ci-dessus (Pleguezuelos-Manzano *et al.*, 2020 ; Dziubańska-Kusibab *et al.*, 2020). En effet, l'alkylation d'une adénine suivie de la réparation incorrecte par la cellule génèrerait des mutations A>N sur un brin et donc T>N sur le brin complémentaire (Wernke *et al.*, 2020). La colibactine contient également deux cycles thiazoles qui sont des motifs fréquemment retrouvés comme intercalants de l'ADN (Roy *et al.*, 1999). Le groupement amino-cétone entre les deux cycles thiazoles n'a jamais été identifié, suggérant qu'il est instable et qu'il est oxydé puis hydrolysé en un groupement di-cétone (Healy *et al.*, 2019) (Figure 8).

Une question majeure non résolue de la biosynthèse de la colibactine est son transport depuis le périplasme jusqu'au noyau de la cellule hôte. La génotoxicité des souches *pks+* nécessite un contact cellulaire entre la bactérie et la cellule hôte (Nougayrède *et al.*, 2006). L'instabilité de la colibactine pourrait expliquer la nécessité de ce contact. Un système de sécrétion pourrait permettre le transport dans la cellule hôte mais aucun système n'a été identifié à ce jour (Figure 8).

Bien que la structure présentée à la Figure 8 soit relativement consensuelle, d'autres structures de colibactine ont été proposées. Notamment, la molécule colibactine-645 induit directement des cassures double-brin par un mécanisme d'oxydation dépendant du cuivre (Z.-R. Li *et al.*, 2019). Cependant, la voie de biosynthèse proposée pour la colibactine-645 ne fait pas intervenir l'ensemble des protéines Clb démontrées comme nécessaires à la génotoxicité de la colibactine (Wernke *et al.*, 2020). La molécule de colibactine présentée à la Figure 8 est la seule molécule décrite actuellement dont la biosynthèse fait intervenir toutes les protéines de la machinerie de biosynthèse (Figure 7).

#### II.4. Production de la colibactine par l'îlot *pks*

La machinerie de biosynthèse de la colibactine est constituée de synthétases de peptides non ribosomiaux (NRPS), de synthases de polycétides (PKS) et de protéines hybrides NRPS/PKS. Ces méga-enzymes construisent la colibactine par addition d'acides aminés par les NRPS et de groupements acyles par les PKS. Un module est l'unité protéique responsable de l'ajout d'un monomère (acide aminé ou acyle) et il se divise en plusieurs domaines. Chaque

domaine possède une activité enzymatique, à l'exception du domaine « carrier protein », qui attache les substrats et la colibactine en construction grâce à un groupement phosphopantéthéinyl (PPant) ajouté post-traductionnellement et les présente aux domaines enzymatiques. Ce domaine est nommé ACP (« acyl carrier protein ») dans les modules PKS et PCP (« peptidyl carrier protein ») dans les modules NRPS.

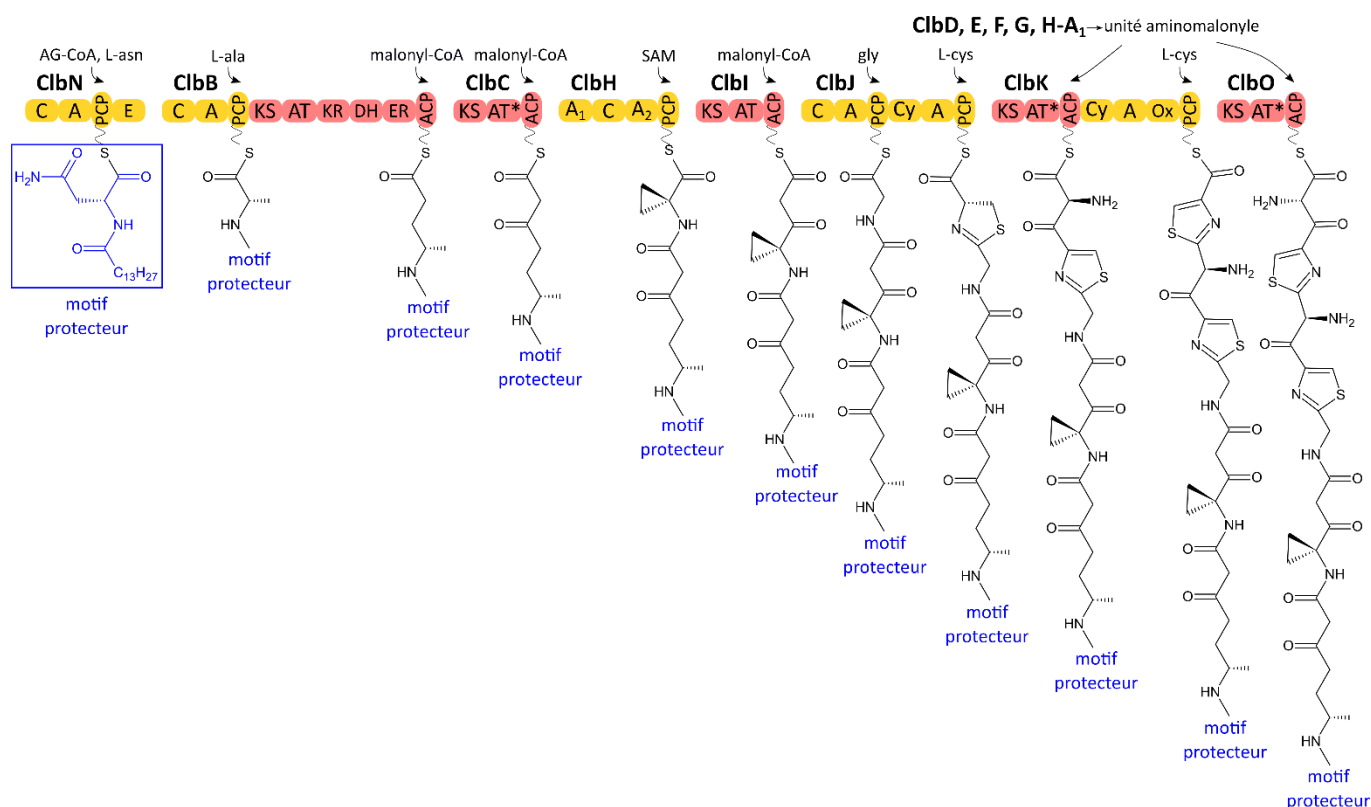


Figure 9. Etapes initiales de la biosynthèse de la colibactine. Les modules NRPS sont représentés en jaune et les modules PKS sont représentés en saumon. Les modules NRPS sont composés des domaines A (adénylation), C (condensation), Cy (cyclisation), E (épimérisation), Ox (oxydase) et PCP (« peptidyl carrier protein »). Les modules PKS sont composés des domaines AT (acyltransférase), AT\* (AT tronqué), ACP (« acyl carrier protein »), DH (déshydratase), ER (énoylréductase), KS (cétosynthase) et KR (cétoréductase). Les substrats incorporés par chaque module sont indiqués. AG : acide gras, CoA : coenzyme A, SAM : S-adenosylméthionine.

La première partie de la biosynthèse de la colibactine est bien établie (Faïs *et al.*, 2018 ; Wernke *et al.*, 2020) (Figure 9). Elle débute par la synthèse du motif protecteur N-myristoyl-D-asparagine par la NRPS ClbN. L'hybride NRPS-PKS ClbB ajoute ensuite une alanine puis une unité malonyle qui est entièrement réduite par les domaines réducteurs de ClbB. Une unité malonyle est ensuite incorporée par la PKS ClbC. ClbH n'utilise pas un acide aminé protéinogène mais incorpore le substrat S-adenosylméthionine, ensuite converti en cycle cyclopropane par la PKS ClbI (Zha *et al.*, 2017), qui rajoute également une unité malonyle. La NRPS ClbJ incorpore deux acides aminés, une glycine et une cystéine. Cette dernière formera

le premier cycle thiazole de la colibactine après cyclisation et oxydation. Les deux derniers modules PKS, le premier module de ClbK et ClbO, ne sont pas entièrement fonctionnels. Ces modules ne sélectionnent pas leur substrat directement mais utilisent un substrat aminomalonyl fourni par ClbG (Brachmann *et al.*, 2015). La construction de ce substrat fait intervenir les enzymes ClbD, ClbE, ClbF et ClbH. Le deuxième module de ClbK est responsable de la synthèse du deuxième cycle thiazole.

Le schéma présenté à la Figure 9 fait intervenir toutes les protéines essentielles à la génotoxicité des souches *pks+* d'*E. coli* à l'exception de ClbL et ClbQ, il est donc incomplet. Il a récemment été démontré que ClbL, bien qu'appartenant à la famille enzymatique des amidases, est capable de catalyser la formation de liaisons peptidiques entre substrats amino-cétone et céto-thioesters (Jiang *et al.*, 2019). ClbQ est une thioestérase (TE) de type II hydrolysant des métabolites chargés sur les domaines « carrier protein » des NRPS/PKS (Guntaka *et al.*, 2017). Les deux groupes ayant élucidé la structure du pont inter-brin généré par la colibactine ont proposé une voie de biosynthèse complète de la colibactine (J.-W. Tang *et al.*, 2022) (Figure 10).

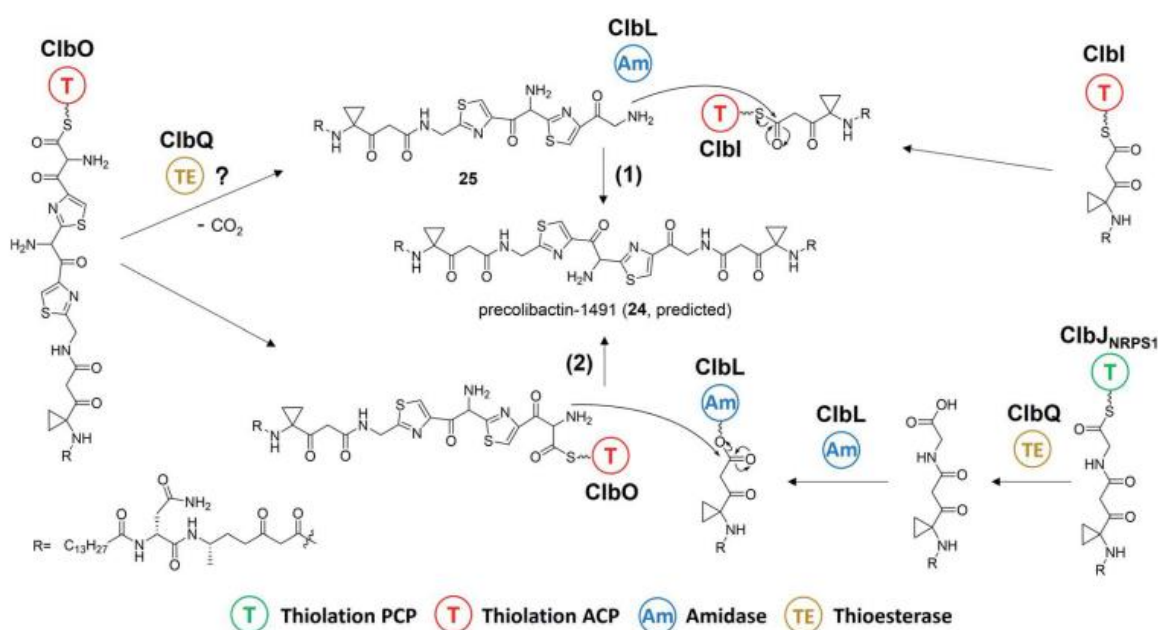


Figure 10. Modèles des dernières étapes de la biosynthèse de la pré-colibactine, proposés par les groupes de Balskus (1) et d'Herzon et Crawford (2). D'après J.-W. Tang *et al.*, 2022.

La différence principale entre ces deux modèles est le rôle joué par la protéine ClbL. Dans le modèle proposé par le groupe de Balskus (Figure 10, (1)), ClbL catalyserait la formation d'une liaison peptidique entre le produit décarboxylé du métabolite attaché à ClbO et le

métabolite chargé sur ClbI (Jiang *et al.*, 2019). Le rôle de la thioestérase ClbQ, essentielle à la génotoxicité des souches *pks+*, n'est pas explicité dans ce modèle (Nougayrède *et al.*, 2006).

ClbQ intervient à la première étape du modèle proposé par Herzon et Crawford (Figure 10, (2)) en hydrolysant le métabolite chargé sur le premier module de ClbJ (Xue *et al.*, 2019). L'amidase ClbL catalyserait ensuite le clivage d'une liaison amide libérant un résidu glycine, ce qui favoriserait la formation d'une liaison amide entre le métabolite ainsi généré et le métabolite chargé sur la PKS ClbO.

Ces deux modèles de la biosynthèse totale de la pré-colibactine n'ont pas été prouvés expérimentalement. En particulier, ClbQ est une enzyme capable d'hydrolyser plusieurs substrats (Guntaka *et al.*, 2017) donc d'autres voies de biosynthèse à partir d'autres précurseurs sont envisageables. De plus, d'autres enzymes d'autres voies de biosynthèse pourraient intervenir. En particulier, la biosynthèse de métabolites récemment isolés fait intervenir des enzymes de la voie de biosynthèse des acides gras en plus des enzymes de l'îlot *pks* (Kim *et al.*, 2021).

En résumé, l'îlot *pks* permet la biosynthèse de métabolites variés dont la génotoxine colibactine. La construction de ce métabolite hybride peptide-polycétide est permise par les NRPS/PKS, des méga-enzymes générant des métabolites secondaires d'une incroyable diversité.

### III. La biosynthèse des peptides non ribosomaux et des polycétides par des machineries enzymatiques multi-domaines

#### III.1. Importance des métabolites secondaires synthétisés par les NRPS/PKS

Les peptides non ribosomaux (NRP), les polycétides (PK) et leurs hybrides (NRP/PK) sont des métabolites secondaires qui, par opposition aux métabolites primaires, ne sont pas dédiés à la croissance et au développement de l'organisme les produisant mais jouent généralement un rôle d'adaptation dans des conditions environnementales particulières. Comme les autres métabolites secondaires, les NRP/PK ont majoritairement été décrits chez des micro-organismes. Un seul exemple de métabolite NRP/PK animal est connu à ce jour, les némamides produits par *Caenorhabditis elegans*, permettant la survie de l'animal en l'absence de nourriture (Shou *et al.*, 2016 ; Feng *et al.*, 2021).

Les NRP/PK sont construits à partir d'unités chimiques diverses qui sont ensuite assemblées et modifiées chimiquement par des domaines enzymatiques dédiés. Les métabolites finaux présentent donc une très grande diversité chimique. Ces produits naturels sont généralement solubles dans un milieu aqueux, capables de traverser les membranes et relativement stables chimiquement, ce qui n'est pas toujours le cas des molécules synthétisées par les chimistes (Gulick, 2017). Par conséquent, ces métabolites ont été et sont encore largement exploités en médecine (Katz & Baltz, 2016) (Figure 11).

Dès la fin du XIX<sup>ème</sup> siècle, la capacité des moisissures à produire des molécules antimicrobiennes a été étudiée par le médecin français Ernest Duchesne (*La pénicilline I - Découverte d'un antibiotique*, s. d.). Le premier antibiotique isolé par Alexandre Flemming en 1928, la pénicilline, est synthétisé à partir d'un NRP. De nombreux autres antibiotiques utilisés en médecine sont synthétisés par des NRPS/PKS, en particulier l'érythromycine synthétisée par les protéines PKS DEBS (« 6-deoxyerythronolide B synthase ») de *Saccharopolyspora erythraea*. L'immunosuppresseur rapamycine et les anticancéreux bléomycine et épothilone sont d'autres exemples de molécules NRP/PK d'importance cruciale en médecine (Katz & Baltz, 2016).



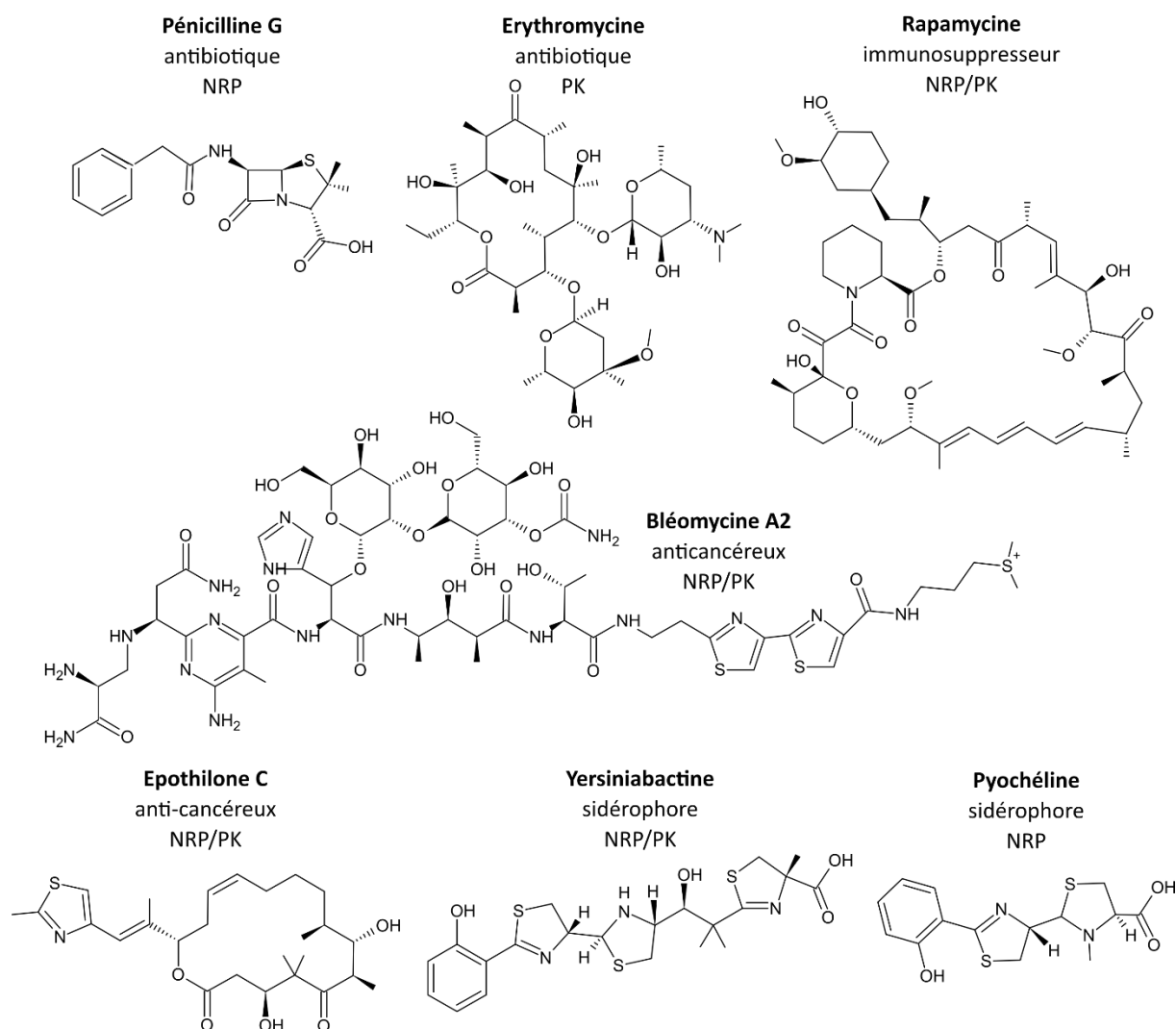


Figure 11. Peptides non ribosomaux (NRP), polycétides (PK) et leurs hybrides (NRP/PK) présentant un intérêt médical. D'après Katz & Baltz, 2016 et Miyanaga, Kudo, *et al.*, 2018.

L'étude des NRP/PK produits par les micro-organismes présente un intérêt majeur dans la recherche de nouveaux antimicrobiens. La résistance aux antimicrobiens est un problème de santé public majeur (Iskandar *et al.*, 2022). 700 000 personnes meurent chaque année d'une infection causée par un agent pathogène résistant aux anti-infectieux et ce nombre pourrait s'élever à 10 millions en 2050 (Organisation mondiale de la santé). Une cause de cette crise est l'absence de nouveaux antimicrobiens novateurs. Les facteurs de virulence synthétisés par les NRPS/PKS sont une nouvelle cible potentielle d'antimicrobiens. Ainsi, de nombreux pathogènes synthétisent des sidérophores afin d'acquérir le fer nécessaire au fonctionnement de certaines de leurs enzymes (Gulick, 2017). Par exemple, le sidérophore yersiniabactine est synthétisé entre autres par la bactérie *Yersinia pestis* à partir d'un îlot NRPS/PKS (Figure 11) (Miller *et al.*, 2002). Le sidérophore pyochéline est quant à lui produit par *Pseudomonas aeruginosa* à partir d'une unité d'assemblage exclusivement NRPS (Braud

*et al.*, 2009). Des molécules inhibitrices de la synthèse de ces sidérophores sont de potentielles sources de nouveaux antibiotiques. La découverte ou la synthèse de telles molécules nécessite une compréhension poussée des enzymes synthétisant les NRP/PK, les NRPS/PKS.

### III.2. Les NRPS/PKS partagent une organisation commune

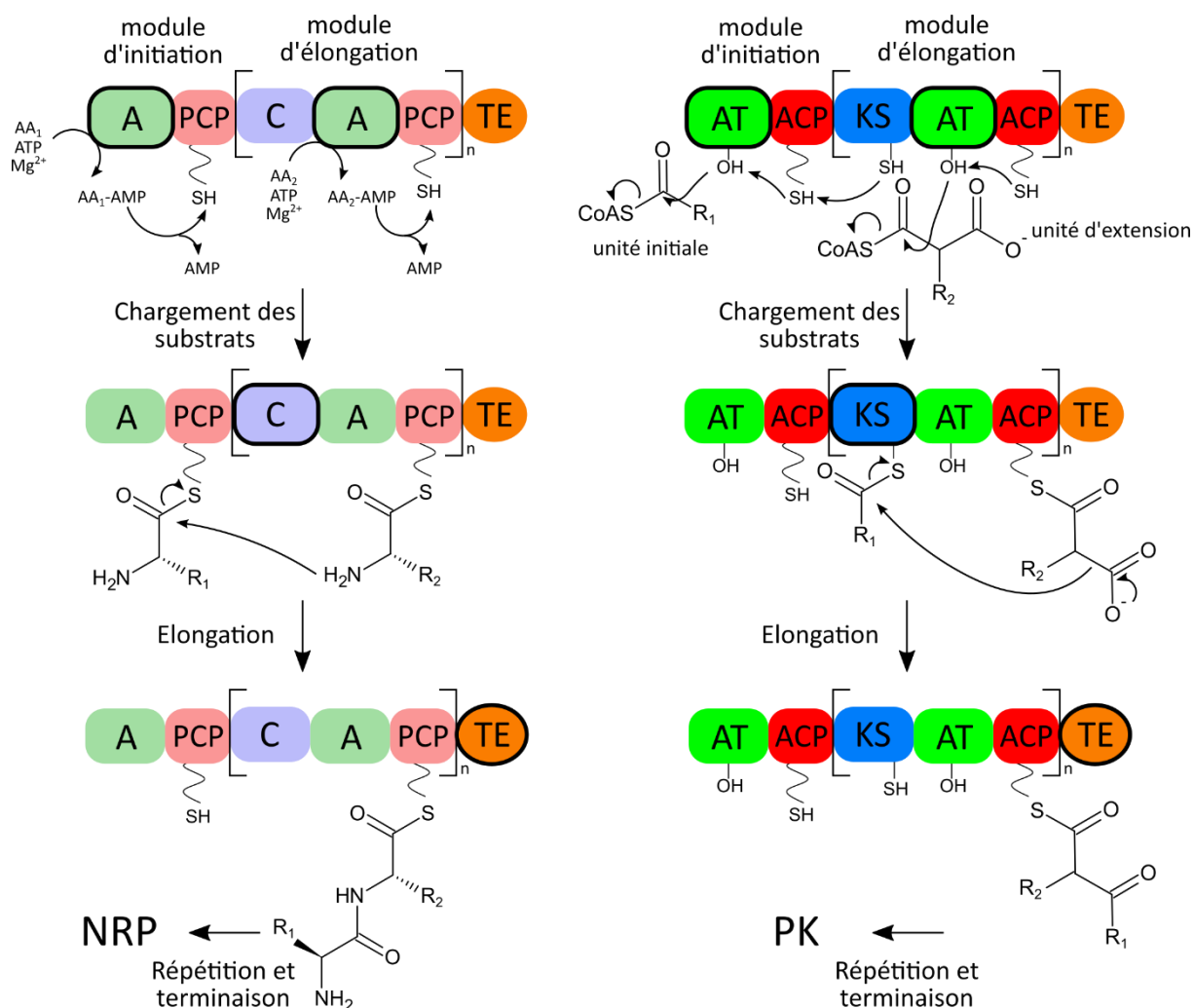


Figure 12. Fonctionnement des NRPS/PKS. Adapté de Weissman, 2015b. Les domaines essentiels à la biosynthèse des NRP/PK sont représentés et le domaine catalysant la réaction chimique décrite est entouré en noir. La première étape de la biosynthèse des NRP/PK est la sélection du substrat par le domaine d'adénylation (A) ou le domaine acyltransférase (AT) puis le transfert du substrat au domaine « carrier protein » (PCP/ACP). Les substrats des NRPS sont des acides, notamment des acides aminés (AA) alors que les substrats des PKS sont des thioesters contenant le coenzyme A (CoA). Les domaines de condensation (C) et cétyosynthase (KS) catalysent respectivement la formation d'une liaison peptidique et d'une liaison carbone-carbone entre deux substrats. Après plusieurs cycles d'élongation, le métabolite final est libéré par le domaine thioestérase (TE).

La biosynthèse des métabolites NRP/PK est une succession de réactions chimiques catalysées par des domaines enzymatiques. Les structures des métabolites synthétisés par les NRPS/PKS sont généralement bien corrélées à l'architecture des méga-enzymes les produisant

(Hertweck, 2009). Cette règle de colinéarité n'est pas toujours valable et sera discutée dans la section V.3. Pour rappel, un module est l'unité multi-domaines responsable de l'ajout d'un monomère au métabolite en construction. Les domaines enzymatiques sont soit portés par des protéines distinctes (NRPS/PKS de type II) soit fusionnés en un seul polypeptide (NRPS/PKS de type I). Parmi les protéines NRPS/PKS de type I, on distingue les méga-enzymes modulaires des méga-enzymes itératives. Les modules des NRPS/PKS modulaires incorporent un seul substrat dans le métabolite final alors que les modules de NRPS/PKS itératives incorporent plusieurs fois le même substrat dans le métabolite final. Seules les NRPS/PKS de type I modulaires seront décrites ici.

Les NRPS/PKS utilisent la même logique de biosynthèse, celle retrouvée dans les chaînes d'assemblage. Cette chaîne d'assemblage est constituée de plusieurs modules. La chaîne est constituée d'un module d'initiation, d'un ou plusieurs modules d'élongation et enfin d'un module de terminaison. La reconnaissance entre les modules est directe lorsque les modules sont codés par le même polypeptide ou indirecte lorsque les modules sont codés par différents polypeptides. Dans ce cas-là, elle est permise par des régions nommées « docking domains », généralement localisées aux extrémités des polypeptides (Smith *et al.*, 2021). Un module d'élongation est composé au minimum de trois domaines : un domaine de sélection du substrat, un domaine responsable de l'élongation et un domaine « carrier protein » délivrant les substrats aux domaines catalytiques (Figure 12). Les réactions catalysées par ces différents domaines seront présentées dans les sections IV.1 et V.1 ci-dessous.

### III.3. Le domaine « carrier protein »

Chaque module NRPS/PKS possède un domaine non catalytique dénommé « carrier protein » (CP), responsable du transport du substrat dans la chaîne d'assemblage. Ce domaine est modifié post-traductionnellement par l'enzyme phosphopantéthéinyl-transférase (PPTase) qui ajoute un bras phosphopantéthéine (PPant) sur une sérine conservée du domaine ACP/PCP. Les substrats sont chargés directement sur la fonction thiol terminale du bras PPant par formation d'une liaison thioester (Figure 13). Certaines PPTases sont capables de modifier des domaines CP n'appartenant pas à la même voie de biosynthèse. Par exemple, la PPTase Sfp de *Bacillus subtilis*, est fréquemment utilisée pour modifier post-traductionnellement les domaines CP et ainsi exprimer les NRPS/PKS sous leur forme *holo*

(Tufar *et al.*, 2014). La souche d'*E. coli* BL21(DE3) Bap1 a été modifiée génétiquement à partir de la souche BL21(DE3) afin que cette dernière exprime de façon inducible la PPTase Sfp (Pfeifer *et al.*, 2001).

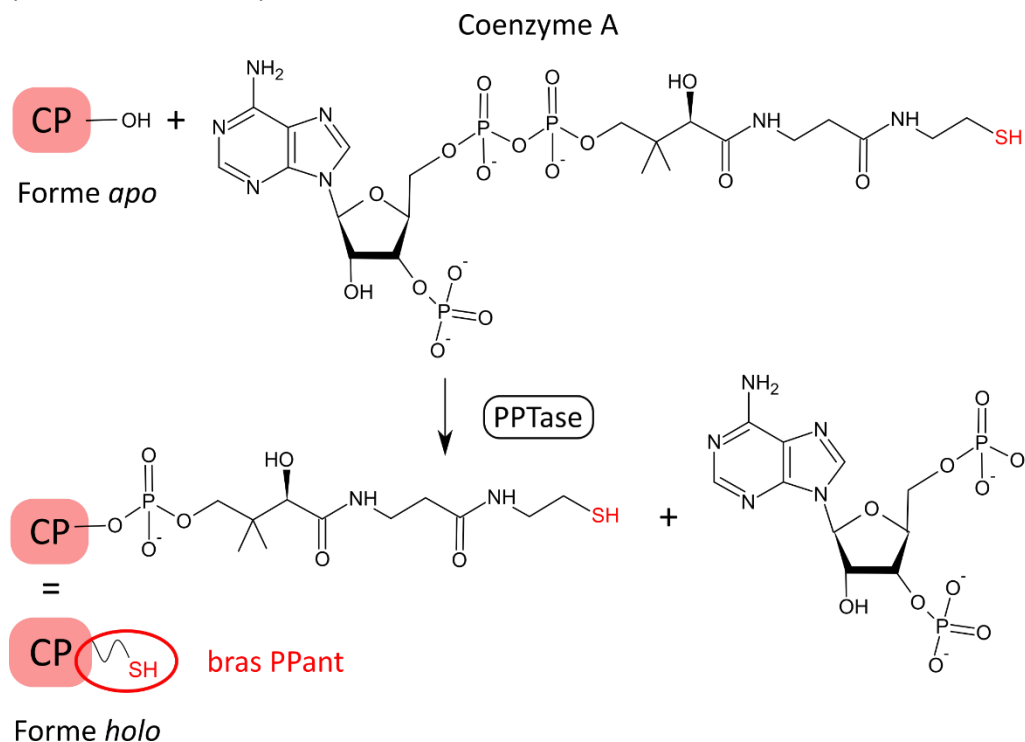


Figure 13. Addition d'un bras phosphopantéthéine (PPant) sur les domaines « carrier protein » des NRPS/PKS par les phosphopantéthéinyl-transférases (PPTases).

Les domaines CP sont des petits domaines de moins de 100 résidus et ils sont repliés en hélices (Crosby & Crump, 2012). Les domaines CP sont constitués d'au moins quatre hélices, les hélices I et II étant généralement plus longues que les hélices III et IV. La sérine modifiée post-traductionnellement est localisée au début de l'hélice II. Certains domaines CP, généralement les domaines ACP des PKS de type *cis*-AT, possèdent une hélice supplémentaire en N-terminal nommée H0 (Moretto *et al.*, 2019). Le domaine CP est central dans les NRPS/PKS car il présente les substrats à tous les autres domaines enzymatiques.

## IV. Structures et fonctions des méga-enzymes NRPS

La publication en Annexe 2 est complémentaire à cette partie.

Durant les deux parties suivantes (IV et V), une attention particulière sera portée aux domaines enzymatiques constitutifs de la méga-enzyme ClbK, le modèle d'étude de ma thèse. Notamment, deux études structurales de NRPS publiées cette année sont de grand intérêt pour l'étude de ClbK. La première structure d'une enzyme oxydase isolée, BmdC, ainsi que la structure de ce domaine en complexe avec le module 2 de BmdB (BmdB<sub>M2</sub>, Cy-A-PCP) ont été dévoilées cette année (Fortinez *et al.*, 2022). En plus d'avoir la même composition en domaines que le module NRPS de ClbK, ce complexe BmdB<sub>M2</sub>-BmdC a la même fonction puisqu'il synthétise lui aussi un cycle thiazole, celui de la bacillamide D. La deuxième étude a capturé la dynamique de PchE, un module à cinq domaines synthétisant un des deux cycles thiazolines de la pyochéline (Wang *et al.*, 2022).

### IV.1. Les domaines enzymatiques des NRPS

#### IV.1.a. **Le domaine d'adénylation**

Les NRPS sont des synthétases alors que les PKS sont des synthases. Cette distinction s'explique car les NRPS utilisent une molécule d'adénosine triphosphate (ATP) afin d'activer leurs substrats alors que les PKS utilisent des substrats contenant une liaison riche énergétiquement, la liaison thioester formée entre une fonction acide carboxylique et la fonction thiol du coenzyme A. La réaction d'activation du substrat est une réaction d'adénylation catalysée par le domaine d'adénylation (A).

Le domaine A est responsable de la sélection du substrat, de son adénylation et de son transfert au domaine PCP (Figure 14A). Le domaine A est divisé en deux sous-domaines, A<sub>core</sub> (≈ 400 résidus), et A<sub>sub</sub> (≈ 100 résidus), séparés par un *linker* flexible (Izoré & Cryle, 2018). Les conformations adoptées par le domaine A à chaque étape ont été déterminées grâce aux structures du module d'initiation de la NRPS LgrA participant à la synthèse de l'antibiotique gramicidine linéaire (Reimer, Aloise, Harrison, *et al.*, 2016) (Figure 14B). Ce module est constitué d'un domaine de formylation (F), d'un domaine A et d'un domaine PCP. Le passage d'une conformation à l'autre est permis par la rotation du sous-domaine A<sub>sub</sub> (Gulick, 2009) (Figure 14C).

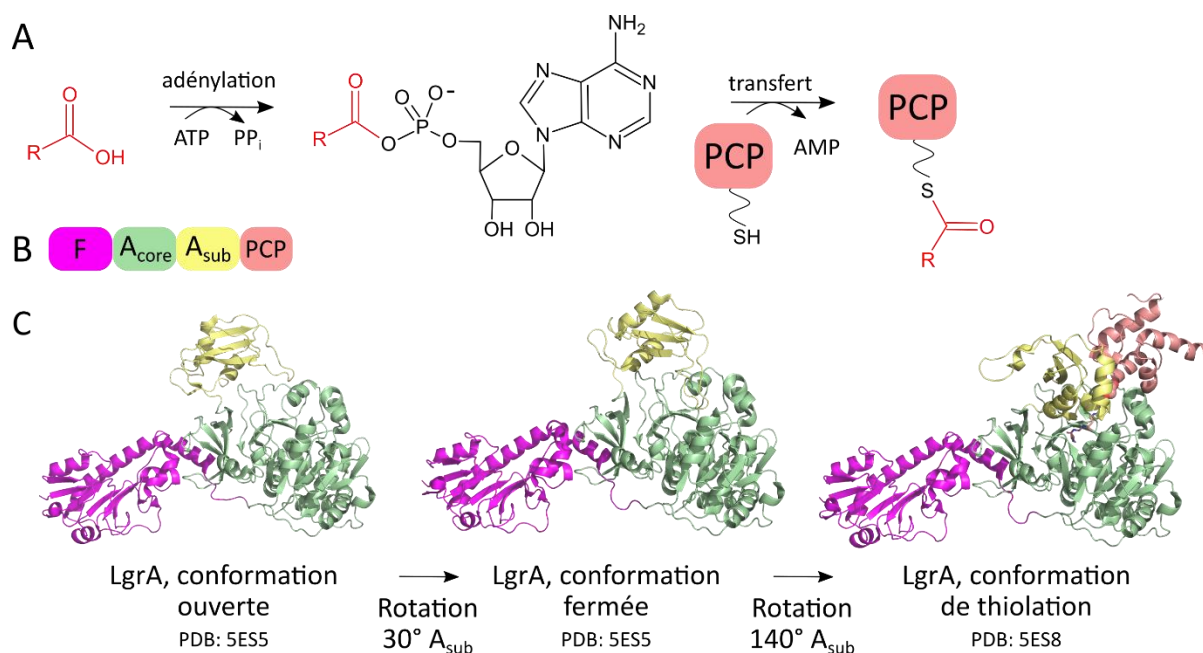


Figure 14. Le domaine d'adénylation (A) sélectionne un substrat, l'active et le transfère au domaine PCP. A) Réactions catalysées par le domaine A. B) Organisation du module d'initiation de LgrA, composé d'un domaine de formylation (F), d'un domaine A constitué d'un grand (A<sub>core</sub>) et d'un petit (A<sub>sub</sub>) sous-domaine, et d'un domaine « peptidyl carrier protein » (PCP). C) Changements conformationnels du module d'initiation de LgrA. La conformation ouverte permet la fixation des substrats valine et ATP ; la conformation fermée est compétente pour la réaction d'adénylation ; la conformation de thiolation est adoptée pour le transfert de la valine sur le domaine PCP (Reimer, Aloise, Harrison, *et al.*, 2016). Le domaine PCP n'a pu être reconstruit dans les conformations ouverte et fermée. Le fragment F-A<sub>core</sub> adopte la même conformation dans chaque structure alors que le sous-domaine A<sub>sub</sub> est mobile.

Les domaines A sélectionnent des substrats très divers tels que des acides aminés protéinogènes ou non protéinogènes et mêmes des groupes aryles. L'analyse de séquences et de structures de domaines A sélectionnant des acides aminés protéinogènes a permis de définir le code des domaines A, c'est-à-dire de définir les résidus conférant sa spécificité de substrat au domaine A (Stachelhaus *et al.*, 1999). Ce code a permis la création d'outils de prédiction du substrat d'un domaine A à partir de sa séquence (Strieker *et al.*, 2010). La sélection d'acides aminés non protéinogènes par le domaine A repose également sur des bases structurales (Kudo *et al.*, 2019). La fixation des substrats par le sous-domaine A<sub>core</sub> est permise lorsque le domaine A est en conformation ouverte (Figure 14C).

Le domaine A, comme les autres enzymes de la famille ANL (« adenylate forming enzyme ») catalyse deux réactions, d'abord une réaction d'adénylation, puis une réaction de thiolation (Figure 14A) (Gulick, 2009). La réaction d'adénylation est la réaction d'activation du substrat, c'est-à-dire le transfert d'un groupement adénosine monophosphate (AMP) sur le substrat acide après hydrolyse d'une molécule d'adénosine triphosphate (ATP). Lorsque

les deux substrats sont fixés, le domaine A<sub>sub</sub> effectue une rotation de 30°, refermant ainsi le site actif du domaine A : le domaine A adopte alors la conformation fermée compétente pour la réaction d'adénylation (Figure 14C).

La réaction de thiolation est la réaction de transfert de l'acide aminé sur le groupement thiol du bras PPant. La conformation de thiolation permet l'interaction productive entre les domaines A et PCP et a été capturée dans plusieurs structures, souvent à l'aide d'analogues non hydrolysables du produit de la réaction, généralement des molécules adénosine vinyl-sulfonamides (AVS) (Gulick & Aldrich, 2018). Le passage de la conformation fermée à la conformation de thiolation est permis par une rotation de 140° du sous-domaine A<sub>sub</sub> (Reimer, Aloise, Harrison, *et al.*, 2016) (Figure 14C).

#### IV.1.b. Les domaines de condensation et de cyclisation

Le domaine de condensation (C) catalyse la formation d'une liaison peptidique entre deux substrats attachés à deux domaines PCP. Le substrat donneur est apporté par le domaine PCP appartenant au module en amont du domaine C et le substrat accepteur est apporté par le domaine PCP localisé en aval du domaine C. La réaction de condensation nécessite que les domaines PCP donneur et accepteur se fixent tous les deux sur le domaine C. Dans certains modules, le domaine de cyclisation (Cy), également nommé domaine d'hétérocyclisation, remplace le domaine C (Bloudoff & Schmeing, 2017). En plus de catalyser la formation de la liaison peptidique, le domaine Cy catalyse la réaction de cyclo-déshydratation du substrat accepteur lorsque celui-ci est une sérine, une thréonine ou une cystéine (Figure 15A). Cette réaction génère un cycle à cinq atomes, nommé respectivement cycle oxazoline, méthyl-oxazoline ou thiazoline.

Les domaines C possèdent un motif HHxxxD conservé alors que les domaines Cy possèdent le motif DxxxxD à la position équivalente (Bloudoff & Schmeing, 2017). Ces motifs ne joueraient pas un rôle catalytique direct mais un rôle structural dans le positionnement adéquat des substrats, facilitant ainsi les réactions de condensation/cyclisation. Pour la réaction de cyclisation catalysée par les domaines Cy, une dyade catalytique aspartate-thréonine a été identifiée et est retrouvée dans les motifs SQTPQVxLD et PVVFIS (Bloudoff & Schmeing, 2017). Le résidu aspartate agirait d'abord comme une base en déprotonant la fonction thiol de la cystéine (Figure 15A). La thréonine catalytique donnerait ensuite un

proton, permettant la formation de l'intermédiaire hydroxy-thiazolidine. La restitution du proton par le résidu aspartate permettrait la formation du cycle thiazoline en libérant une molécule d'eau.

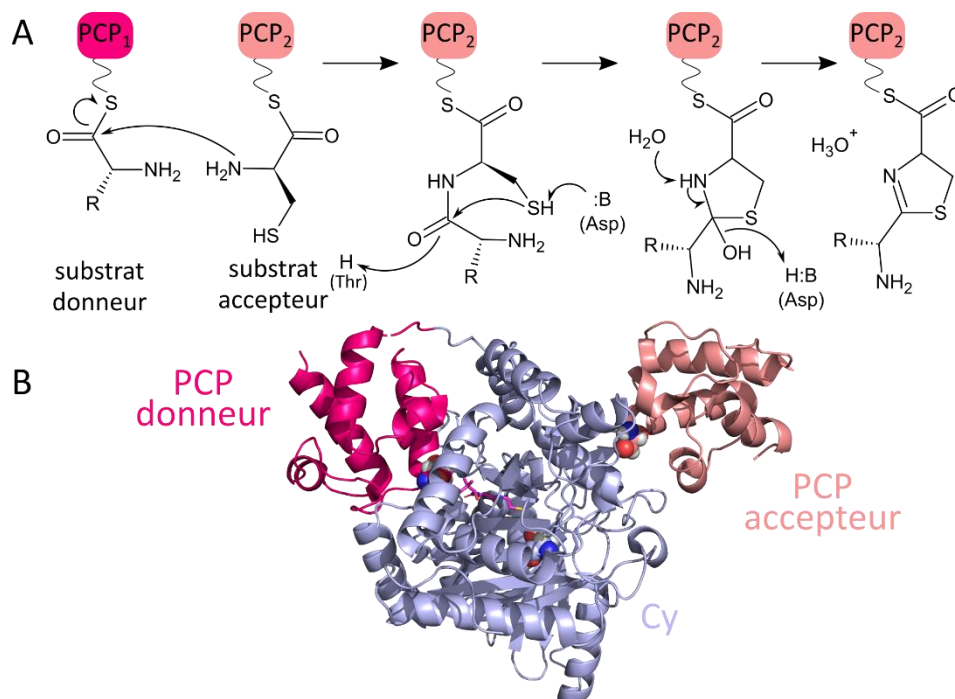


Figure 15. Le domaine de cyclisation (Cy) catalyse les réactions de condensation et de cyclo-déshydratation. A) Schéma des réactions catalysées par un domaine Cy sur un substrat accepteur cystéine, permettant la synthèse d'un cycle thiazoline. D'après Bloudoff & Schmeing, 2017. B) Structure du domaine Cy de PchE dans la conformation de condensation (PDB : 7EN2) (Wang *et al.*, 2022). Le résidu Asp critique pour la cyclisation est représenté par une sphère, ainsi que les résidus sérines modifiés des domaines PCP. Le bras PPant du domaine PCP donneur est représenté par des baguettes.

Les domaines C et Cy ( $\approx 450$  résidus) appartiennent à la super-famille des domaines de condensation (Bloudoff & Schmeing, 2017). Les structures de ces domaines excisés révèlent qu'ils adoptent un repliement pseudo-dimérique en V, chaque lobe adoptant le repliement typique de l'enzyme chloramphénicol acyltransférase (Figure 15B). La conformation de condensation visualisant les PCP donneurs et accepteurs fixés au domaine C/Cy a été observée chez LgrA et chez PchE (Reimer *et al.*, 2019 ; Wang *et al.*, 2022). Les positions des domaines PCP sont équivalentes dans les deux structures : les deux PCP se fixent à des sites opposés du domaine C/Cy et dirigent leur sérine modifiée en direction du résidu aspartate catalytique (Figure 15B).



### IV.1.c. Les domaines de modification

Une diversification importante des métabolites NRP est permise par les domaines de modification. Ces domaines optionnels peuvent agir avant l'incorporation du substrat par le domaine A, sur le substrat attaché au domaine PCP ou après la libération du NRP. Ces domaines sont soit codés par des protéines indépendantes, soit inclus dans les NRPS, généralement dans les domaines A, qui sont alors nommés domaines A interrompus (Labby *et al.*, 2015). Parmi les domaines caractérisés structuralement, on trouve les domaines cétoréductase (Alonzo *et al.*, 2020), d'épimérisation (Wang *et al.*, 2022), de formylation (Reimer, Aloise, Harrison, *et al.*, 2016), de méthylation (Mori *et al.*, 2018), oxydase (Fortinez *et al.*, 2022),...

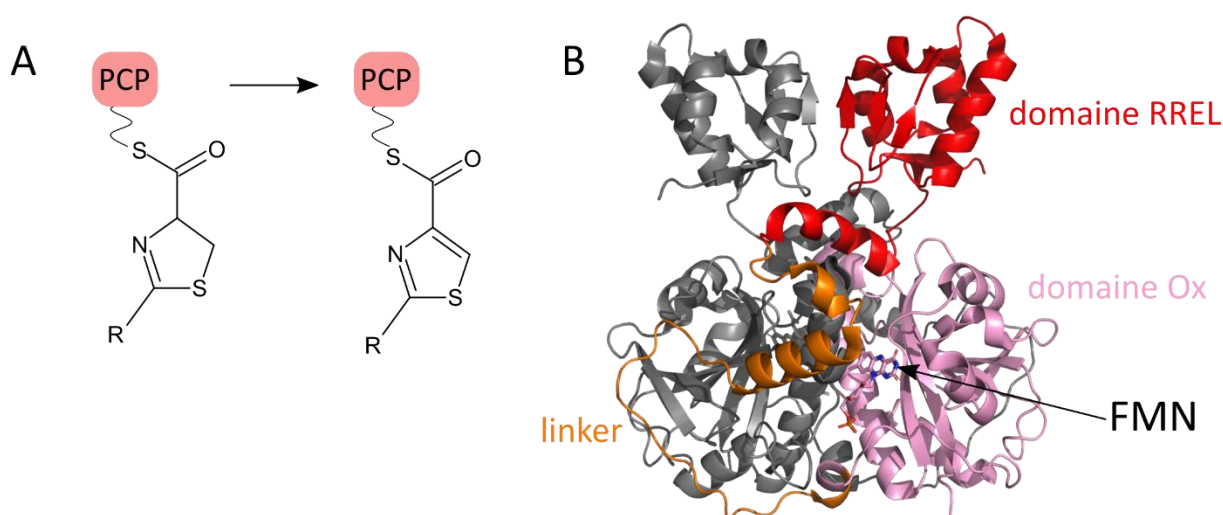


Figure 16. Le domaine oxydase (Ox) catalyse l'oxydation d'un cycle thiazoline en cycle thiazole. A) Réaction catalysée par le domaine Ox. B) Structure du dimère de BmdC (PDB : 7LY6), une enzyme à activité oxydase agissant en *trans* (Fortinez *et al.*, 2022). Un monomère est coloré en gris et l'autre monomère est coloré par domaines. RREL : « RiPP recognition element-like ».

Les cycles thiazoles confèrent souvent un rôle biologique important aux molécules qui en contiennent, telles que les anticancéreux épothilone et bléomycine (Figure 11). Le domaine oxydase (Ox) catalyse l'oxydation d'un cycle thiazoline en cycle thiazole grâce à un cofacteur flavine mononucléotide (FMN) (Schneider *et al.*, 2003) (Figure 16A). La structure de l'enzyme BmdC possédant une activité oxydase a été révélée cette année (Fortinez *et al.*, 2022) (Figure 16B). Cette protéine dimérique est constituée d'un domaine « RiPP recognition element-like » (RREL) en N-terminal ( $\approx 100$  résidus), un *linker* ( $\approx 40$  résidus), et le domaine Ox proprement dit ( $\approx 180$  résidus). Les deux cofacteurs FMN se fixent dans le site actif à l'interface entre les

deux monomères. L'oxydation serait catalysée grâce à l'action combinée d'une tyrosine catalytique qui jouerait le rôle de base et du facteur FMN (Pang *et al.*, 2020).

#### IV.2. Les structures des assemblages NRPS monomériques

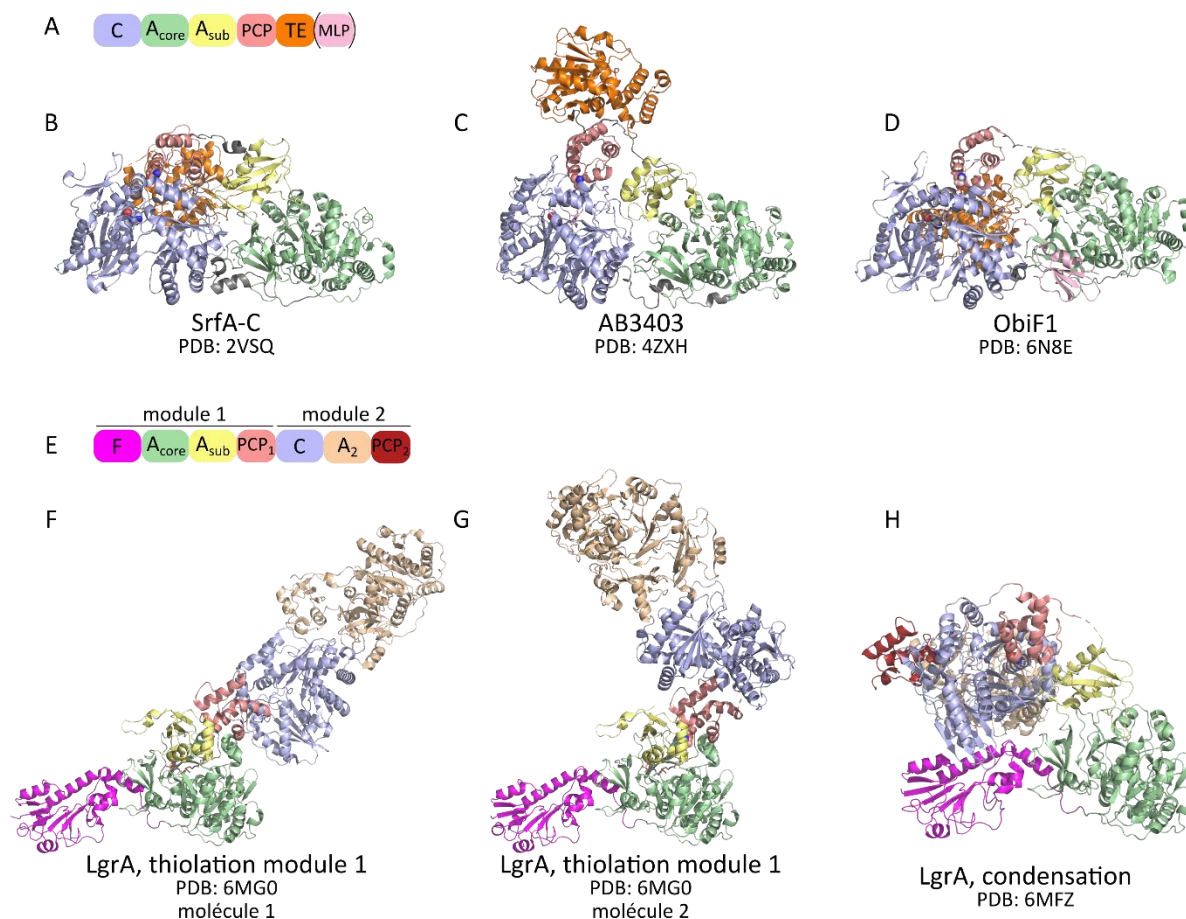


Figure 17. Structures de NRPS monomériques. A) Composition en domaines d'un module de terminaison générique. MLP : « MtbH-like protein », un activateur du domaine A. B-D) Structures de trois modules de terminaison (Tanovic *et al.*, 2008 ; Drake *et al.*, 2016 ; Kreitler *et al.*, 2019). Le résidu histidine catalytique du domaine C est représenté par des sphères. E) Composition en domaines de la NRPS di-modulaire LgrA. F-H) Structures de plusieurs fragments de LgrA dans différentes conformations (Reimer *et al.*, 2019). Le domaine PCP<sub>2</sub> n'est présent qu'à la structure H. Dans toutes les structures, la sérine modifiée du domaine PCP est représentée par des sphères et le bras PPant par des bâtons s'il est présent.

Les structures de modules complets de NRPS se sont accumulées ces dernières années et, jusqu'à récemment, toutes les NRPS décrites étaient monomériques (Tanovic *et al.*, 2008 ; Drake *et al.*, 2016 ; Reimer, Aloise, Harrison, *et al.*, 2016 ; Kreitler *et al.*, 2019 ; Reimer *et al.*, 2019). Ces structures ont révélé qu'un module NRPS est généralement composé d'un di-domaine C-A<sub>core</sub> stable et d'un fragment A<sub>sub</sub>-PCP mobile (Figure 17). Le module d'initiation de LgrA ne possède pas de domaine de condensation mais possède à la place un domaine de formylation, qui joue le même rôle structural, c'est-à-dire qu'il forme une plate-forme stable

avec le sous-domaine  $A_{\text{core}}$ . Le domaine TE étant lui aussi mobile, des modules de terminaison ayant la même composition en domaines adoptent des structures variées (Figure 17A-D).

Deux études structurales se sont intéressées à l'architecture supra-modulaire des NRPS. La première étude portait sur la NRPS di-modulaire DhbF participant à la biosynthèse du sidérophore bacillibactine (Tarry *et al.*, 2017). Malgré la stabilisation de cette protéine par deux inhibiteurs AVS, l'observation de DhbF en microscopie électronique a révélé que DhbF adopte des conformations très variées. La flexibilité inter-modulaire des NRPS a été confirmée grâce aux premières structures hautes résolutions d'une NRPS di-modulaire, celles de LgrA (Reimer *et al.*, 2019). Plusieurs structures de fragments multi-domaines contenant le module d'initiation entier et un ou plusieurs domaines du module d'élongation de LgrA ont mis en évidence la flexibilité du *linker* inter-modulaire (Figure 17E-H). L'illustration la plus probante de cette flexibilité est la structure cristallographique du fragment F-A<sub>1</sub>-PCP<sub>1</sub>-C-A<sub>2</sub> de LgrA dont le premier module a été stabilisé dans la conformation de thiolation par un inhibiteur AVS (Figure 17F-G). Les deux molécules contenues dans l'unité asymétrique sont superposables au niveau du module d'initiation mais adoptent des positions radicalement différentes au niveau du module d'élongation.

#### IV.3. Les structures des assemblages NRPS dimériques

Jusqu'en 2021, le paradigme était que, contrairement aux PKS dimériques, les NRPS étaient monomériques. Trois exemples de NRPS dimériques sont venus contredire cette affirmation : FmoA3, PchE et BmdB<sub>M2</sub>-BmdC (Katsuyama *et al.*, 2021 ; Wang *et al.*, 2022 ; Fortinez *et al.*, 2022). FmoA3 est une NRPS constituée de trois domaines (Cy-A-PCP) intervenant dans la biosynthèse de JBIR-34 et -35, des métabolites synthétisés par une espèce de *Streptomyces* isolée d'une éponge (Katsuyama *et al.*, 2021). FmoA3 et PchE dimérisent en tête-à-queue par le fragment Cy-A<sub>core</sub> (Katsuyama *et al.*, 2021 ; Wang *et al.*, 2022) (Figure 18A-B). Le mode de dimérisation est similaire mais pas strictement identique chez ces deux protéines. En plus du di-domaine Cy-A<sub>core</sub>, des domaines mobiles « carrier protein » et du sous-domaine A<sub>sub</sub>, la NRPS PchE contient également un domaine d'épimérisation (E). Chez PchE, la dimérisation par le fragment Cy-A<sub>core</sub> conduit à une forme en H de l'assemblage NRPS. Les trois structures obtenues par cryo-microscopie électronique (cryo-ME), ainsi que la reconstitution de la dynamique de la protéine à partir des images de microscopie grâce à l'outil CryoDRGN

(Zhong *et al.*, 2021), ont même permis aux auteurs de proposer un cycle catalytique complet pour la NRPS PchE (Wang *et al.*, 2022).

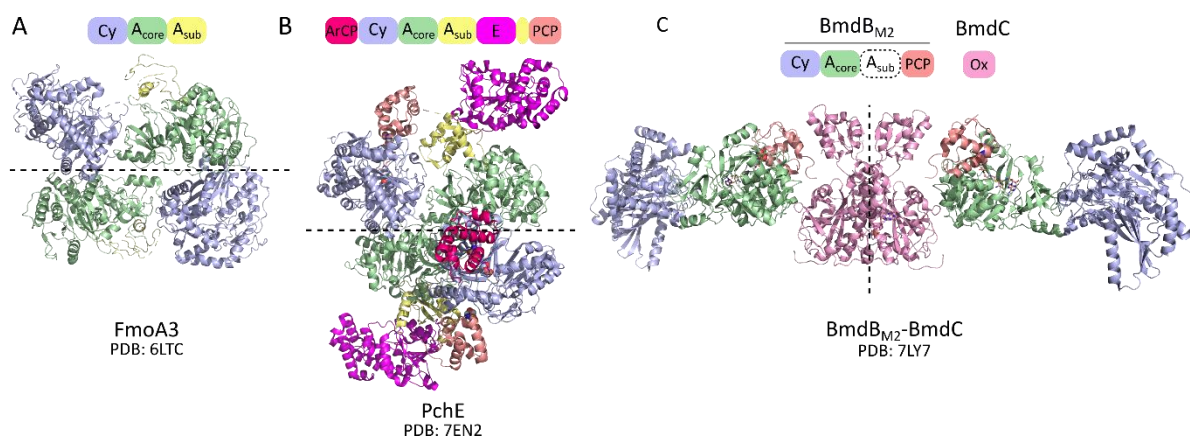


Figure 18. Structures de NRPS dimériques. L'axe de dimérisation est représenté en pointillés. La sérine conservée des domaines CP est représentée par des sphères, et le bras PPant, lorsqu'il est présent, est représenté par des baguettes. A) Composition en domaines et structure cristallographique de FmoA3 (Katsuyama *et al.*, 2021). B) Composition en domaines et structure obtenue par cryo-microscopie électronique de PchE (Wang *et al.*, 2022). ArCP : domaine « aryl carrier protein », E : domaine d'épimérisation. Le résidu aspartate catalytique du domaine Cy de PchE est représenté par des sphères. Un domaine ArCP de PchE n'a pas pu être reconstruit. C) Composition en domaines et structure cristallographique du complexe entre BmdB<sub>M2</sub>, bloqué dans la conformation de thiolation, et BmdC (Fortinez *et al.*, 2022). Le sous-domaine A<sub>sub</sub> de Bmd<sub>M2</sub> n'a pas pu être reconstruit.

Le deuxième mode de dimérisation identifié chez les NRPS est la dimérisation par une enzyme à activité oxydase, BmdC, qui induit la dimérisation de l'ensemble du second module NRPS (Cy<sub>2</sub>-A<sub>2</sub>-PCP<sub>2</sub>) de BmdB (Fortinez *et al.*, 2022) (Figure 18C). Plusieurs structures du complexe BmdB<sub>M2</sub>-BmdC ont été déterminées, notamment une structure cristallographique du complexe stabilisé dans la conformation de thiolation, et une structure par cryo-microscopie électronique. Seul le domaine BmdC contribue à la dimérisation du complexe BmdB<sub>M2</sub>-BmdC, qui présente une forme très allongée, dont les dimensions sont 6.5 x 6.5 x 28 nm (Figure 18C). Le domaine PCP<sub>2</sub> doit visiter, dans l'ordre, les sites actifs des domaines Cy<sub>2</sub>, A<sub>2</sub> et Ox. De façon intéressante, ces trois sites actifs sont alignés et la forme légèrement courbée du dimère BmdB<sub>M2</sub>-BmdC diminue la distance à parcourir pour le domaine PCP<sub>2</sub>.

## V. Structures et fonctions des méga-enzymes PKS

### V.1. Les domaines enzymatiques des PKS

#### V.1.a. Le domaine acyltransférase

On distingue deux classes de PKS de type I. Chez les PKS *cis*-AT, le domaine acyltransférase (AT) est inséré dans chaque module (Dodge *et al.*, 2018) alors que chez les PKS *trans*-AT, le domaine AT est codé par une protéine indépendante (Kosol *et al.*, 2018). Le domaine AT sélectionne un substrat, malonyl-CoA ou dérivé, puis le transfère au domaine ACP. Contrairement au domaine A des NRPS, le domaine AT établit une liaison covalente avec le substrat sélectionné puisqu'il charge le substrat sur un résidu sérine conservé.

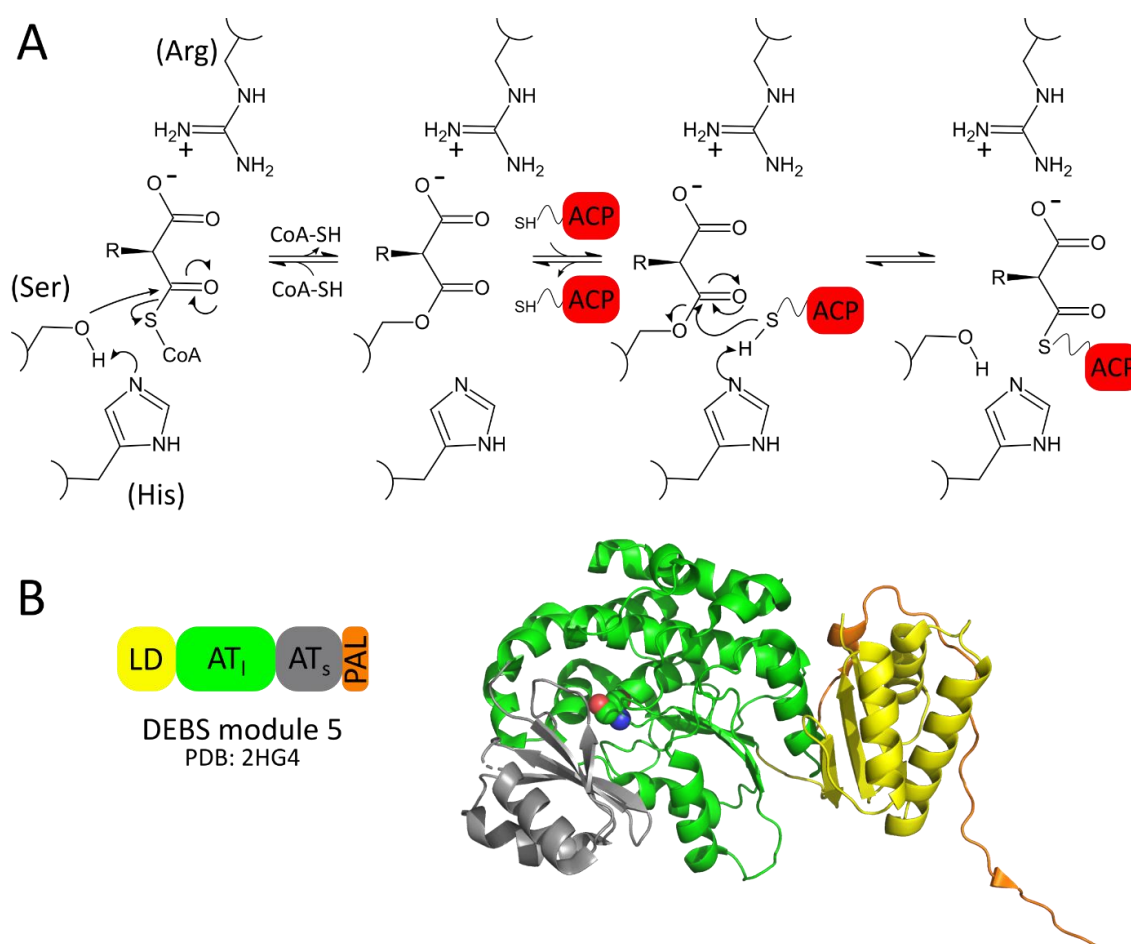


Figure 19. Le domaine acyltransférase (AT) sélectionne un substrat de type malonyl-CoA et le transfère sur le bras PPant du domaine ACP. A) Mécanisme réactionnel du domaine AT, d'après A. T. Keatinge-Clay, 2012. B) Organisation et structure du domaine AT de DEBS module 5. La sérine catalytique est représentée par des sphères (Y. Tang *et al.*, 2006). AT<sub>1</sub> : grand sous-domaine AT, AT<sub>s</sub> : petit sous-domaine AT.

Le mécanisme réactionnel du domaine AT est de type ping-pong (A. T. Keatinge-Clay, 2012) (Figure 19A). Dans la première demi-réaction, la sérine catalytique, préalablement

activée par une histidine, attaque la liaison thioester du substrat. La charge négative du substrat est stabilisée par une arginine conservée, ce qui favorise le départ du coenzyme A et la formation du complexe enzyme-acyl. Dans la deuxième demi-réaction, la fonction thiol du bras PPant porté par l'ACP attaque la liaison ester, permettant ainsi le transfert du substrat sur le domaine ACP. Le domaine AT discrimine différents substrats, notamment le malonyl-CoA du méthylmalonyl-CoA. Ainsi, l'histidine des domaines AT sélectionnant spécifiquement du malonyl-CoA est localisé dans un motif HAFH alors que ce motif est YASHH pour les domaines AT spécifiques du méthylmalonyl-CoA (A. T. Keatinge-Clay, 2012). La phénylalanine (en gras) de la poche pour le substrat malonyl-CoA empêche stériquement la sélection d'un substrat méthylmalonyl-CoA.

Les domaines AT des PKS *cis*-AT et *trans*-AT ont la même structure. Ils sont constitués d'un grand sous-domaine AT<sub>I</sub> (AT<sub>large</sub>, ≈ 240 résidus) à repliement type α,β hydrolase et d'un petit sous-domaine AT<sub>s</sub> (AT<sub>small</sub>, ≈ 60 résidus) à repliement type ferrédoxine (A. T. Keatinge-Clay, 2012) (Figure 19B). La sérine catalytique est localisée à l'interface entre les deux sous-domaines. Les AT de type *cis*-AT sont entourés par des régions *linker* en N-terminal et C-terminal, nommées respectivement domaine de liaison (LD, ≈ 90 résidus) et « post-AT linker » (PAL, ≈ 40 résidus). Étonnamment, les régions LD et PAL sont retrouvées en C-terminal des domaines KS de PKS de type *trans*-AT bien que ces domaines KS ne soient pas suivis d'un domaine AT (Lohman *et al.*, 2015).

### V.1.b. Le domaine cétoynthase

Le domaine cétoynthase (KS) des PKS joue le même rôle que le domaine C des NRPS, il catalyse la formation d'une liaison covalente entre deux substrats. Contrairement au domaine C, le domaine KS établit une liaison covalente avec un des deux substrats. En effet, le domaine KS catalyse d'abord une réaction de trans-thioestérification, c'est-à-dire l'attaque de la liaison thioester du substrat chargé sur l'ACP donneur par la fonction thiol d'une cystéine conservée. Le domaine KS catalyse ensuite une condensation de Claisen, c'est-à-dire la formation d'une liaison carbone-carbone par une condensation décarboxylative.

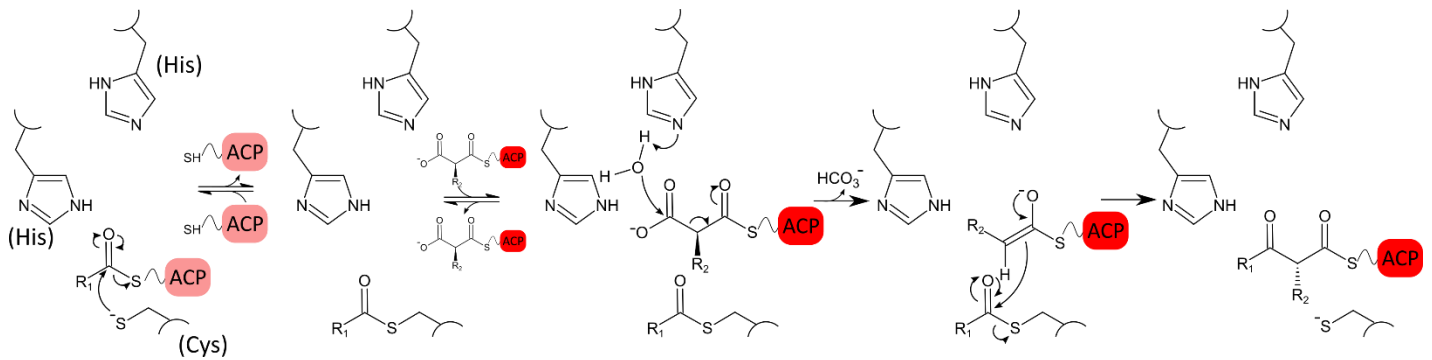


Figure 20. Mécanisme réactionnel hypothétique du domaine cétoynthase. Adapté de A. T. Keatinge-Clay, 2012. Le domaine KS catalyse d'abord la trans-thioestérification du substrat apporté par l'ACP donneur (en saumon) sur un résidu cystéine conservé puis la condensation décarboxylative entre ce substrat et l'unité d'extension apportée par l'ACP accepteur (en rouge). Deux histidines catalytiques jouent un rôle important pour la condensation.

Le mécanisme réactionnel du domaine KS repose sur une triade Cys-His-His et a été déduit du mécanisme du domaine KS des synthèses d'acides gras, évolutivement et structuralement reliées aux PKS *cis*-AT (A. T. Keatinge-Clay, 2012) (Figure 20). La réaction de trans-thioestérification est permise par l'attaque de la liaison thioester du polycétide donneur par le groupement thiol déprotoné du résidu cystéine catalytique. La première histidine catalytique favoriserait la fixation du groupe carboxylate de l'unité d'extension apportée par le domaine ACP accepteur alors que la seconde histidine catalytique agirait comme une base activant une molécule d'eau qui attaquerait alors ce groupe carboxylate. L'énolate généré après décarboxylation attaquerait alors le groupement thioester du substrat donneur. La réaction de condensation génère une inversion de stéréochimie sur le C2 de l'unité d'extension (Weissman *et al.*, 1997).

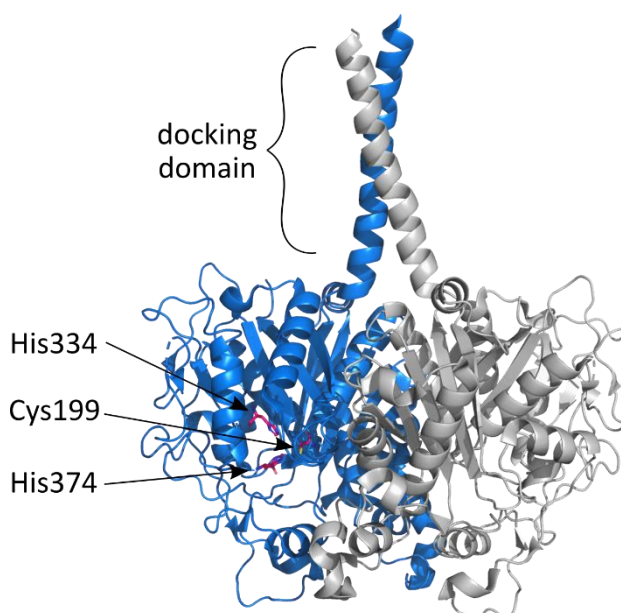


Figure 21. Structure du domaine KS de DEBS module 5 (PDB : 2HG4) (Y. Tang *et al.*, 2006). Les deux monomères sont représentés en bleu et en gris. Les résidus catalytiques du monomère en bleu sont représentés en rose par des bâtons.

Le domaine KS ( $\approx 430$  résidus) adopte un repliement thiolase constitué d'une architecture  $\alpha\beta\alpha\beta\alpha$  (Figure 21). La grande interface de dimérisation du domaine KS, quasiment  $2000 \text{ \AA}^2$ , fait du domaine KS un contributeur majeur à l'architecture dimérique des PKS (Dodge *et al.*, 2018). Certains domaines KS sont précédés d'une hélice N-terminale participant elle-aussi à l'interface de dimérisation des PKS. Cette hélice est un « docking domain » (DD) permettant la reconnaissance du module précédent, localisé sur un autre polypeptide (Broadhurst *et al.*, 2003).

### V.1.c. Les domaines de modification optionnels

Un module d'élongation minimal d'une PKS *cis*-AT est constitué des domaines KS, AT et ACP. De nombreux modules d'élongation contiennent un ou plusieurs domaine(s) qui modifie(nt) chimiquement le groupement cétone incorporé par le module précédent. Le domaine cétoréductase (KR) réduit le groupement  $\beta$ -cétone en groupement hydroxyle, qui peut éventuellement être éliminé pour former une double liaison carbone-carbone par le domaine déshydratase (DH). Finalement, cette double liaison peut être réduite en une simple liaison par le domaine énoylréductase (ER). En plus de diversifier les polycétides générés par les PKS, certains de ces domaines, notamment le domaine DH, ont un rôle architectural majeur puisqu'ils contribuent à la dimérisation du module PKS complet (Dodge *et al.*, 2018).

## V.2. Les structures des assemblages PKS de type *cis*-AT

### V.2.a. L'apport des méga-enzymes itératives

Les PKS *cis*-AT sont reliées évolutivement et structuralement aux synthèses d'acides gras (« fatty acid synthases », FAS) (Herbst, Townsend, *et al.*, 2018). La FAS peut être considérée comme une PKS itérative complètement réductrice puisqu'elle comprend les trois domaines de modification du groupement  $\beta$ -cétone décrits auparavant. La structure à  $3.2 \text{ \AA}$  de la FAS porcine (mFAS pour FAS mammalienne) obtenue par cristallographie aux rayons X a longtemps servi de modèle pour les structures de PKS (Maier *et al.*, 2008) (Figure 22). La méga-enzyme mFAS adopte une structure dimérique en X qui sépare physiquement le di-domaine KS-malonyl-acyltransférase (MAT) constituant la portion de condensation du tri-domaine DH-ER-KR formant la portion de modification. L'interface de dimérisation est composée des domaines KS, DH et ER et de l'extrémité C-terminale du *linker* MAT-DH. Cette structure en X, de dimensions  $20 \times 17 \times 11 \text{ nm}$ , délimite deux chambres catalytiques non fermées, chaque



chambre étant bordée par l'ensemble des sites actifs des domaines catalytiques. L'étude d'une autre mFAS en microscopie électronique a mis en évidence la grande flexibilité de cette méga-enzyme, qui adopte un continuum de conformations (Brignole *et al.*, 2009). En particulier, des mouvements de balancier et de rotation de la portion de condensation par rapport à la portion de modification ont été observés.

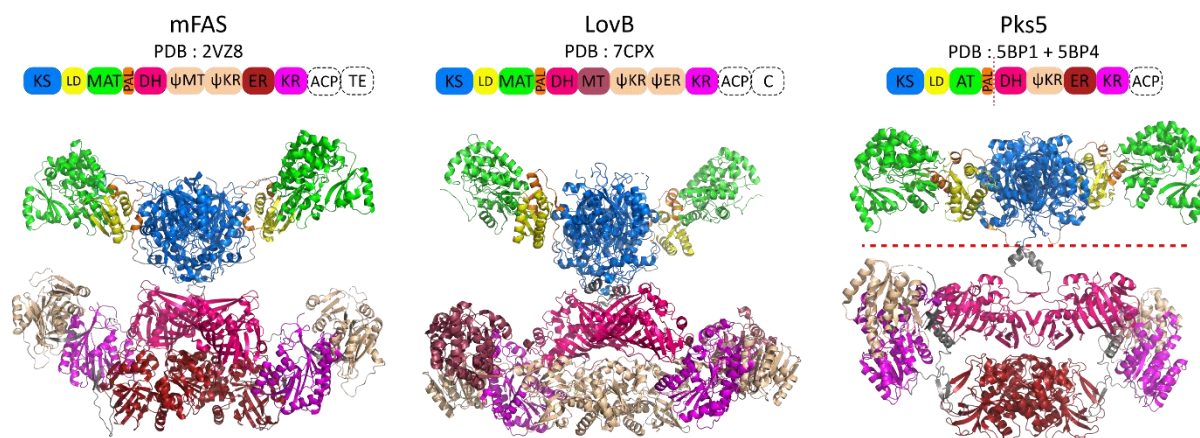


Figure 22. Structures en X des méga-enzymes itératives (Maier *et al.*, 2008 ; Wang *et al.*, 2021 ; Herbst *et al.*, 2016). Les domaines en beige sont des pseudo-domaines non fonctionnels. Les noms des domaines non visibles dans les structures sont entourés en pointillés. La méga-enzyme Pks5 a été cristallisée en deux fragments indépendants. Abréviations des domaines : KS : cétyosynthase, LD : domaine de liaison, MAT : malonyl-acyltransférase, PAL : *linker* post-AT, DH : déshydratase, MT : méthyltransférase, KR : cétoréductase, ER : énoylréductase, ACP : « acyl carrier protein », TE : thioestérase, C : condensation.

Deux structures récentes de PKS itératives ont montré que la forme en X identifiée chez la mFAS est conservée chez les PKS itératives. Le champignon *Aspergillus terreus* synthétise la lovastatine, le précurseur des statines utilisées en traitement contre l'hypercholestérolémie, grâce à la PKS itérative LovB (Wang *et al.*, 2021). La structure de LovB, résolue par cryo-ME à 3.6 Å, est très ressemblante à l'architecture en X de la mFAS (Figure 22). De leur côté, s'inspirant de la structure de la mFAS, Herbst *et al.* ont cristallisé indépendamment les portions de condensation et de modification de Pks5, une synthèse de lipo-oligosaccharides de *Mycobacterium segmentatis* (Herbst *et al.*, 2016 ; Herbst, Townsend, *et al.*, 2018). Étonnamment, malgré une structure très conservée avec celle de mFAS, la portion de condensation a cristallisé sous forme monomérique, indiquant que l'excision d'un fragment d'une multi-enzyme ne maintient pas toujours son état oligomérique. La structure reconstituée du dimère de Pks5 adopte une conformation similaire à celle de mFAS malgré quelques différences dans la portion de modification. Ces différences, telles que l'absence du

domaine pseudo-méthyltransférase, font de Pks5 un modèle des PKS modulaires comprenant les trois domaines de modification.

Ces structures de PKS itératives sont des modèles utiles pour la compréhension des PKS modulaires, même si leur fonctionnement est différent puisque les PKS modulaires ne catalysent qu'un seul cycle d'incorporation/modification du substrat. Malheureusement, le domaine ACP, d'importance cruciale pour le fonctionnement des méga-enzymes, n'a pu être modélisé dans les structures décrites auparavant. Comme le domaine PCP des NRPS, l'adoption de positions multiples par le domaine ACP au cours du cycle catalytique des PKS est certainement responsable de sa difficulté à le visualiser.

### V.2.b. Le cycle catalytique de la PKS modulaire PikAIII

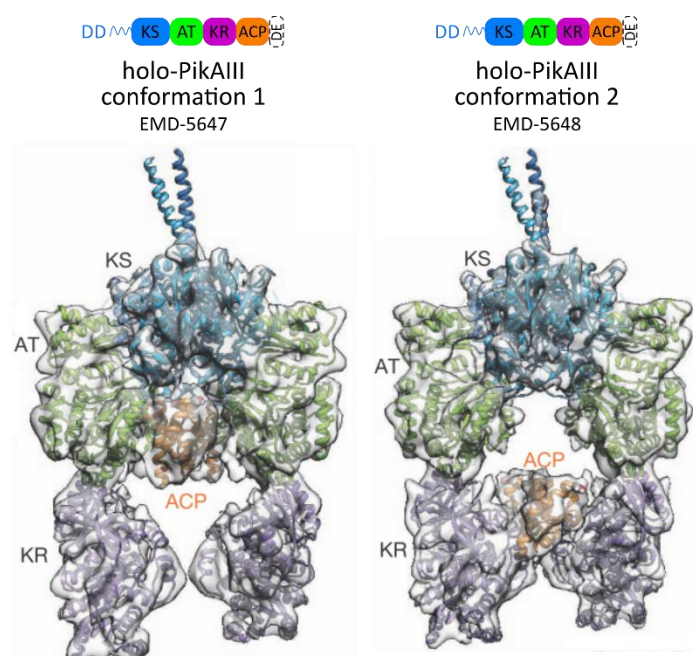


Figure 23. Structures de PikAIII sous forme *holo*, obtenues par cryo-ME. D'après Dutta *et al.*, 2014. Les structures de domaines homologues de DEBS ont été placées dans les cartes de densité ( $\approx 9 \text{ \AA}$ ). DD : « docking domain », DE : élément de dimérisation.

En 2014, les connaissances structurales sur les PKS modulaires ont bondi grâce à l'étude de PikAIII, le cinquième module de la chaîne d'assemblage synthétisant l'antibiotique pikromycine de *Streptomyces venezuelae* (Dutta *et al.*, 2014 ; Whicher *et al.*, 2014). La protéine PikAIII ( $\text{KS}_5\text{-AT}_5\text{-KR}_5\text{-ACP}_5$ ) a été visualisée par cryo-microscopie électronique à moyenne résolution ( $7\text{-}12 \text{ \AA}$ ) à chaque étape de son cycle catalytique (Figure 23 et Figure 24). Les structures des domaines individuels des protéines homologues de DEBS ont été positionnées dans les cartes de densité. La conformation globale de PikAIII résultante est très différente de celle de la mFAS (Figure 22). En particulier, le di-domaine  $\text{KS}_5\text{-AT}_5$  de PikAIII adopte une conformation en arche alors que tous les autres di-domaines KS-AT de PKS,

itératives comme modulaires, adoptent une conformation étendue (Y. Tang *et al.*, 2006, 2007 ; Maier *et al.*, 2008) (Figure 22). La conformation en arche du di-domaine  $KS_5$ - $AT_5$  ainsi que la présence d'éléments de dimérisation post-ACP conduit à la formation d'une chambre catalytique unique et close, contrairement aux deux chambres ouvertes observées chez les PKS itératives. Deux conformations ont été détectées pour la forme *holo* de PikAIII, c'est-à-dire modifiée post-traductionnellement avec un bras PPant (Figure 23). Les deux domaines  $ACP_5$  sont localisés à l'intérieur de la chambre catalytique mais positionnés différemment selon la conformation.

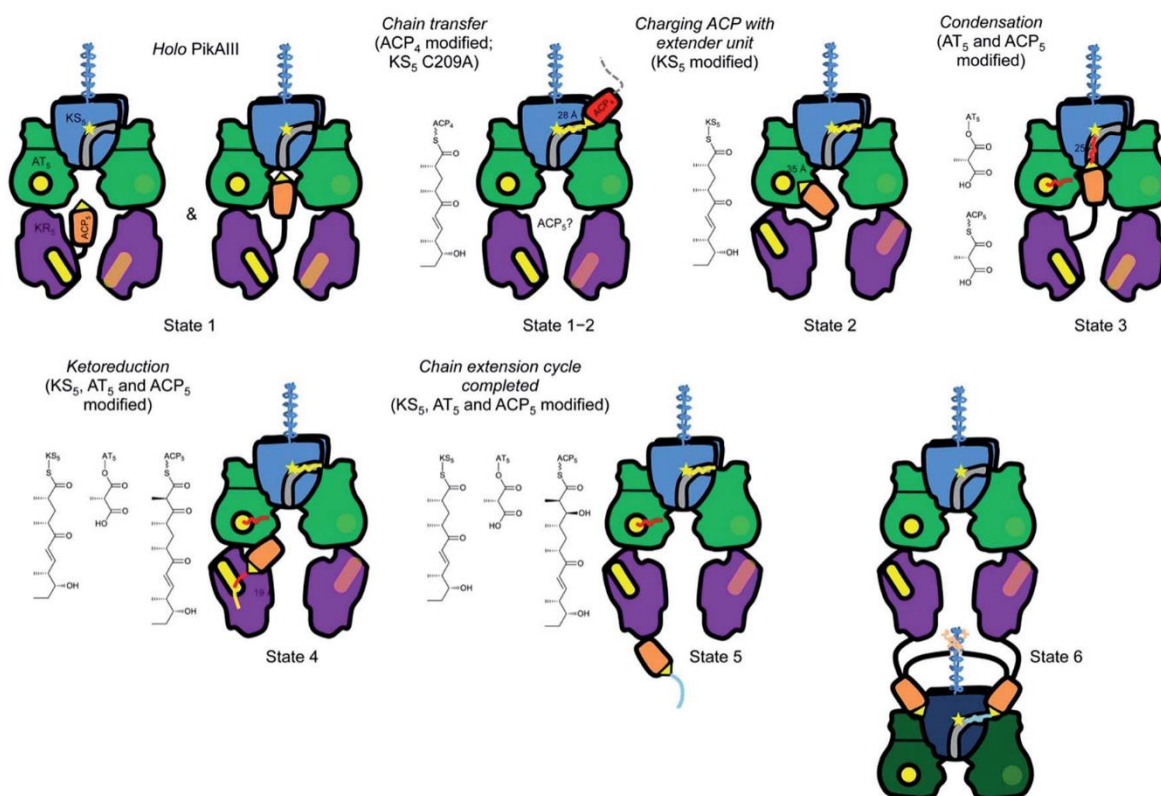


Figure 24. Schéma du cycle catalytique de PikAIII, construit à partir des structures de la PKS PikAIII chargée avec différents substrats, obtenues en cryo-ME (Dutta *et al.*, 2014 ; Whicher *et al.*, 2014). D'après Weissman, 2015a. Les domaines sont de la même couleur qu'à la Figure 23. Les sites actifs des domaines sont en jaune et la distance estimée entre le résidu catalytique et la sérine modifiée du domaine  $ACP_5$  est indiquée. Les substrats chargés sur chaque domaine sont indiqués à côté de chaque état. Pour les états 1 à 5, un seul des deux domaines  $ACP_5$  est représenté. Pour visualiser l'interaction productive entre le  $KS_5$  et l' $ACP_4$  chargé avec son substrat penta-cétide (état 1-2), une délétion de l' $ACP_5$  a été générée. L'état 2 représente l'étape suivante, lorsque l' $ACP_4$  a transféré son substrat au domaine  $KS_5$ . A l'état 3, les domaines  $AT_5$  et  $ACP_5$  sont chargés avec un substrat méthylmalonyl et l' $ACP_5$  se positionne pour la condensation. A l'état 4, le domaine  $ACP_5$  est chargé avec un substrat  $\beta$ -céto-hexa-cétide donc cet état reflète l'état post-condensation. Enfin, à l'état 5, le domaine  $ACP_5$  est chargé avec un substrat  $\beta$ -hydroxy-hexa-cétide donc cet état est représentatif d'un état post-réduction par le domaine  $KR_5$ . L'état 6 est déduit de l'état 1-2 et de l'état 5 et représente le transfert du substrat à PikAIV.

L'ensemble du cycle catalytique de PikAIII a pu être reconstitué grâce aux structures de la PKS chargée avec différents substrats (Figure 24). La nature du substrat chargé sur le domaine ACP<sub>5</sub> apparaît comme cruciale pour la position du domaine ACP<sub>5</sub> et même pour la conformation de l'ensemble du module. En effet, le chargement de certains substrats induit un retournement du domaine KR<sub>5</sub> (Figure 24, états 2-4-5). Dans toutes les structures de PikAIII, les deux domaines ACP<sub>5</sub> ont un mouvement concerté, probablement à cause de l'élément de dimérisation localisé en C-terminal du domaine ACP<sub>5</sub>.

La première étape du cycle de PikAIII est le chargement par le domaine KS<sub>5</sub> du penta-cétide généré par le module 4 de la chaîne d'assemblage. La structure d'une fusion PikAIII( $\Delta$ ACP<sub>5</sub>)-C209A+ACP<sub>4</sub> a montré que l'ACP<sub>4</sub> donneur utilise l'entrée de côté (« side entrance ») du domaine KS<sub>5</sub>, et est donc exclu de la chambre catalytique (Figure 24, état 1-2). Pour que le domaine KS<sub>5</sub> catalyse la réaction de condensation, le domaine ACP<sub>5</sub> accepteur doit être chargé avec son substrat. Ainsi, la présence du substrat donneur sur le KS<sub>5</sub> combinée à l'absence du substrat accepteur sur l'ACP<sub>5</sub> induit un repositionnement de l'ACP<sub>5</sub>, qui se réoriente en direction de l'AT<sub>5</sub> (Figure 24, état 2). Une fois chargé avec le méthylmalonyl, l'ACP<sub>5</sub> accepteur se positionne à l'entrée interne (« bottom entrance ») du KS<sub>5</sub>, nouvellement identifiée chez PikAIII (Figure 24, état 3). Ces résultats suggèrent que la discrimination des ACP donneur et accepteur est permise par des entrées dédiées, ce qui contredit les données antérieures qui suggéraient que les deux ACP utilisaient l'entrée de côté mais n'utilisaient pas les mêmes interfaces d'interaction KS-ACP (Kapur *et al.*, 2010). Le chargement du domaine ACP<sub>5</sub> avec la molécule  $\beta$ -céto-hexa-cétide, produit du domaine KS<sub>5</sub> et substrat du domaine KR<sub>5</sub>, induit un repositionnement de l'ACP<sub>5</sub> à proximité du KR<sub>5</sub> (Figure 24, état 4). Le domaine KR<sub>5</sub> a subi une large rotation par rapport à la structure initiale dans cet état. Enfin, le chargement du domaine ACP<sub>5</sub> avec la molécule  $\beta$ -hydroxy-hexa-cétide, produit de la réduction par le KR<sub>5</sub>, conduit à la sortie du domaine ACP<sub>5</sub> de la chambre catalytique (Figure 24, états 5 et 6).

Bien que des réserves quant à la représentativité du cycle de PikAIII pour les PKS modulaires aient été émises (Rittner & Grninger, 2014 ; Weissman, 2015a), les structures à moyenne résolution de PikAIII ont fourni les premières captures des complexes transitoires entre le domaine ACP et les domaines catalytiques des PKS. Il a ensuite fallu attendre 2021, soit sept ans après les structures de PikAIII pour obtenir d'autres structures de PKS modulaires

visualisant le domaine ACP. Grâce à la révolution de la cryo-ME (Callaway, 2020), les deux modules complets de PKS modulaires caractérisés ont été visualisés par cryo-ME à haute résolution (3-4 Å) (Bagde *et al.*, 2021 ; Cogan *et al.*, 2021).

### V.2.c. Structures des PKS modulaires asymétriques

Les deux PKS modulaires caractérisées en 2021 sont la protéine Lsd14 de *Streptomyces lasalocidi* participant à la biosynthèse de l'antibiotique lasalocide A (Bagde *et al.*, 2021) et le module 1 de DEBS (DEBS-M1) (Cogan *et al.*, 2021) (Figure 25). A l'exception de la structure de la forme *apo* de Lsd14, élucidée par cristallographie aux rayons X, les autres structures ont été obtenues par cryo-ME en présence de fragments d'anticorps reconnaissant un antigène (F<sub>ab</sub>, « fragment antigen-binding »). Le F<sub>ab</sub> 1B2 reconnaît spécifiquement le di-domaine KS-AT du module 3 de DEBS, et en particulier le « docking domain » (DD) de ce module, sans entraver l'activité catalytique du module (X. Li *et al.*, 2018 ; Guzman & Khosla, 2022). Ainsi, les DD de Lsd14 et de DEBS-M1 ont été remplacés par le DD du module 3 de DEBS (DD\*), générant des protéines chimériques reconnues et stabilisées par le F<sub>ab</sub> 1B2. Le F<sub>ab</sub> 1B2 stabilise en particulier la forme dimérique des PKS Lsd14-DD\* et DEBS-M1-DD\*, et empêche la dissociation de ces PKS en monomères.

Les structures de DEBS-M1 et Lsd14 ont révélé pour la première fois une asymétrie dans l'architecture des PKS. Les structures montrent que, similairement à la mFAS, ces PKS possèdent deux chambres catalytiques ouvertes, et non une seule comme PikAIII, mais les deux chambres n'ont pas le même volume. Aucune structure ne visualise les deux domaines ACP, indiquant que les deux ACP du dimère n'ont pas un mouvement concerté bien que leurs mouvements soient probablement contraints à cause d'un élément de dimérisation (DE) post-ACP. L'asymétrie visualisée implique une asynchronie des deux chambres de réaction, c'est-à-dire qu'à un temps donné, une seule chambre serait active. Similairement à mFAS, des mouvements de balancier et de rotation de la portion DE-ψKR-KR-ACP par rapport à la portion de condensation ont été envisagés pour Lsd14 (Bagde *et al.*, 2021).

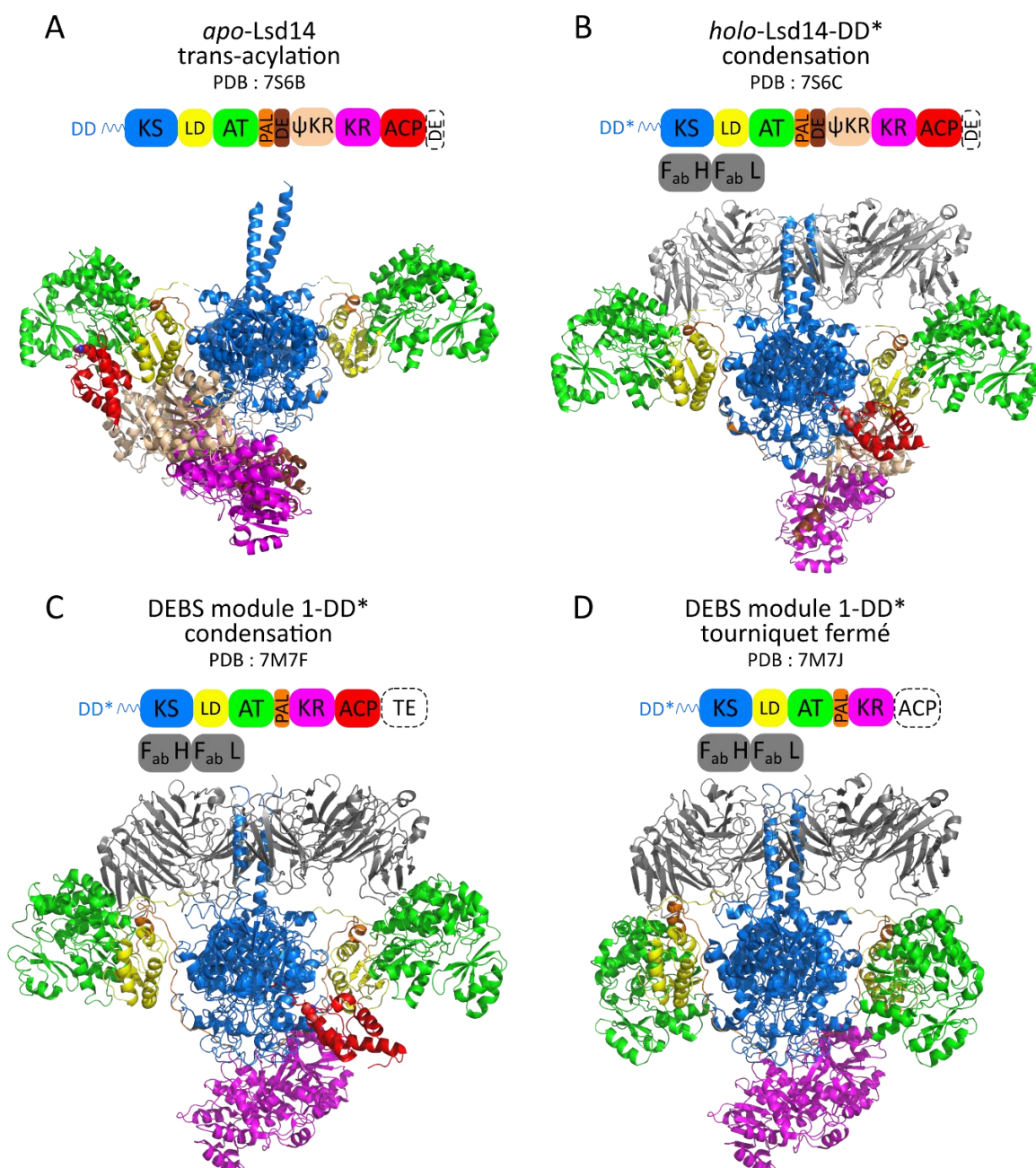


Figure 25. Structures des PKS modulaires Lsd14 et DEBS module 1. La structure A a été obtenue par cristallographie aux rayons X et les autres structures ont été obtenues par cryo-ME (Bagde *et al.*, 2021 ; Cogan *et al.*, 2021). Dans les structures B, C et D, les « docking domains » (DD) natifs des protéines ont été remplacés par le DD de DEBS module 3 (DD\*) et les protéines chimériques ont été stabilisées par des fragments d'anticorps dirigés contre le DD\*. Les domaines en beige sont des pseudo-domaines non fonctionnels. Les noms des domaines non visibles dans les structures sont entourés en pointillés. Dans la structure A, un seul domaine ACP a pu être modélisé et les domaines  $\Psi$ KR-KR et ACP n'ont pu être assignés à une chaîne. Dans la structure B, un seul tri-domaine  $\Psi$ KR-KR-ACP a pu être modélisé. Dans la structure C, un seul ACP et un seul KR ont pu être modélisés. Dans la structure D, un seul KR a pu être modélisé. DE : élément de dimérisation. F<sub>ab</sub> : fragment d'anticorps reconnaissant un antigène.

L'interaction du domaine ACP avec les deux domaines essentiels d'un module PKS a été visualisée grâce à ces études. En effet, dans la structure de Lsd14 sous forme *apo*, le domaine ACP s'oriente en direction du site actif du domaine AT et 22 Å séparent les deux sérines conservées, suggérant que le module est dans une conformation compatible avec la trans-acylation, c'est-à-dire le transfert du substrat sélectionné par le domaine AT sur le bras PPant du domaine ACP (Figure 25A). L'ACP accepteur en position de condensation a été visualisé à la fois chez *holo*-Lsd14-DD\* et DEBS-M1-DD\* (Figure 25B et C). Les deux structures montrent que l'ACP accepteur se fixe sur l'entrée de côté du domaine KS, dans une fente entre le dimère de KS et le domaine de liaison. Ces structures supportent le modèle selon lequel les ACP donneur et accepteur utilisent l'entrée de côté du KS, mais adoptent une orientation différente (Kapur *et al.*, 2010).

Les structures de DEBS-M1-DD\* montrent que l'accessibilité de l'entrée de côté du domaine KS pour le domaine ACP est contrôlée par des mouvements du domaine AT, démontrant ainsi structuralement le mécanisme du « tourniquet » postulé précédemment (Lowry *et al.*, 2016). Une personne ne peut pas rentrer dans un tourniquet tant que la personne précédente n'est pas sortie du tourniquet. Selon ce modèle, un nouveau substrat ne peut être chargé sur la cystéine du domaine KS tant que le substrat chargé sur l'ACP n'a pas été transloqué au module suivant. La structure de DEBS-M1-DD\* avec le produit d'élongation du KS chargé sur l'ACP montre que les domaines AT se sont repliés contre le domaine KS, fermant ainsi l'accès à l'entrée de côté (Figure 25D). Cette conformation « tourniquet fermé » repose sur un mouvement du domaine AT lui-même et non du domaine de liaison conduisant à la conformation en arche chez PikAIII (Figure 23). Le mouvement du domaine AT ainsi que les mouvements déduits de l'asymétrie des structures de Lsd14-DD\* et DEBS-M1-DD\* soulignent la flexibilité des méga-enzymes PKS.

### V.3. Les structures des assemblages PKS de type *trans*-AT

Environ 40 % des PKS modulaires chez les bactéries sont de type *trans*-AT (Kosol *et al.*, 2018). Contrairement aux métabolites produits par les PKS *cis*-AT, les structures des métabolites produits par les PKS *trans*-AT sont difficilement prédictibles à partir de l'organisation de la chaîne d'assemblage car la règle de colinéarité est bien moins vraie chez les PKS *cis*-AT que chez les PKS *trans*-AT. En effet, les chaînes d'assemblage des PKS *trans*-AT possèdent de nombreuses particularités, telles que la division d'un module sur plusieurs

polypeptides, les domaines KS non compétents pour la condensation ou l'existence de nombreux domaines catalytiques agissant en *trans*. Ainsi, plus de 160 types de modules de PKS *trans*-AT ont été décrits, ce qui souligne la complexité de ce type de méga-enzymes (Helfrich *et al.*, 2019). Malgré cette grande diversité et la large répartition des PKS *trans*-AT, ces méga-enzymes et les produits qu'elles synthétisent sont très peu caractérisés.

Plusieurs domaines KS excisés de modules PKS de type *trans*-AT ont été caractérisés structurellement (D. C. Gay *et al.*, 2014 ; Lohman *et al.*, 2015 ; D. C. Gay *et al.*, 2016). Malgré l'absence d'un domaine AT fonctionnel, ces domaines KS possèdent le domaine de liaison (LD) connectant les domaines KS et AT chez les PKS *cis*-AT, ainsi que le *linker* post-AT (PAL). La région LD pourrait servir de plate-forme d'ancrage pour le domaine *trans*-AT mais cela n'a pas été prouvé (Kosol *et al.*, 2018). Alors que les fragments LD et PAL sont conservés, la région entre les deux est très variable chez les PKS *trans*-AT. Parfois, des reliquats de domaine AT sont retrouvés à cette position, comme observé chez la PKS *trans*-AT OzmQ de *Streptomyces albus* participant à la synthèse des antibiotiques oxazolomycines (Lohman *et al.*, 2015) (Figure 26A). Une insertion constituée de trois hélices, nommée « Laterally-INteracting Ketosynthase Sequence » (LINKS) est fréquemment retrouvée à cette position et l'interaction homotypique entre deux insertions LINKS a été visualisée dans plusieurs structures cristallographiques (D. C. Gay *et al.*, 2016). Ces hélices pourraient jouer un rôle dans l'organisation tri-dimensionnelle de méga-complexes de PKS.

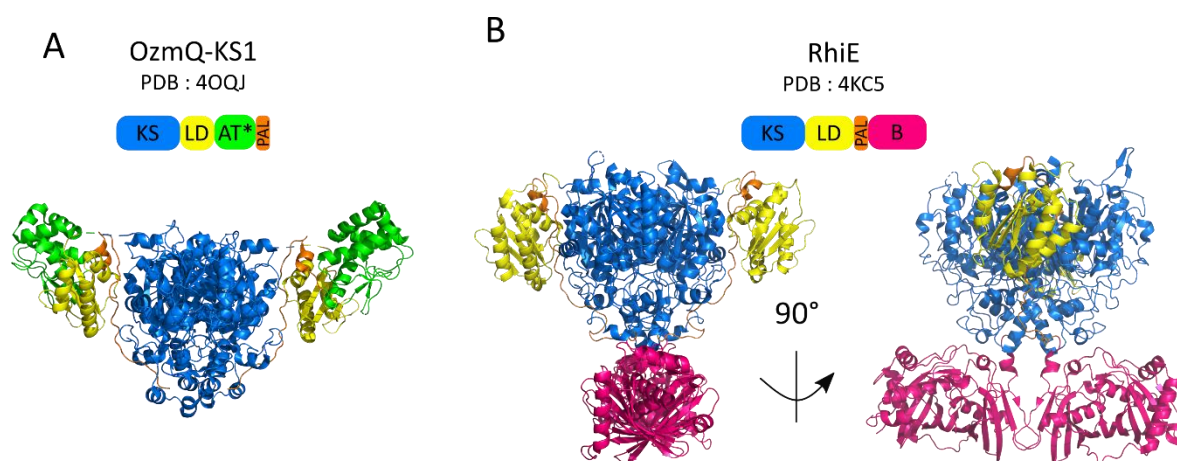


Figure 26. Structures cristallographiques de fragments de PKS *trans*-AT. A) Structure du dimère d'OzmQ-KS1, reconstituée à partir de la forme cristalline monomérique (Lohman *et al.*, 2015). B) Structure cristallographique de la PKS RhiE constituée d'un domaine KS et d'un domaine de branchement (B) non fonctionnel (Bretschneider *et al.*, 2013).



Une structure importante pour nos connaissances sur les PKS *trans*-AT est celle du domaine KS-domaine de branchement de RhiE (Bretschneider *et al.*, 2013) (Figure 26B). RhiE participe à la biosynthèse de l'antimitotique rhizoxine, produite par la bactérie *Burkholderia rhizoxinica*, symbionte d'un champignon pathogène. Le domaine KS de RhiE, en plus de l'activité de condensation, possède une activité de branchement de la chaîne carbonée et le domaine B est non fonctionnel. Ce domaine B adopte un repliement en double « hot-dog » similaire à celui du domaine DH et se positionne sous le domaine KS. Cette structure nous renseigne donc sur la position d'un domaine optionnel par rapport au domaine KS mais n'est pas représentative d'un module complet car elle ne contient pas le domaine ACP essentiel à chaque module.

Le seul modèle d'un module complet d'une PKS *trans*-AT est celui du cinquième module de VirA (VirA<sub>M5</sub>) participant à la biosynthèse de l'antibiotique virginiamycine chez *Streptomyces virginiae* (Davison *et al.*, 2014). Ce modèle a été obtenu grâce à une stratégie de « dissection-reconstruction », c'est-à-dire que la reconstruction de la structure et de la dynamique du module complet a été permise par la caractérisation de fragments multi-domaines de VirA<sub>M5</sub> (Weissman, 2015a). Le module 5 de VirA, est un module minimal composé d'un domaine KS et d'un tandem de domaines ACP. Les structures des domaines ACP et d'un modèle du domaine KS<sub>5</sub> ont été positionnées dans le modèle *ab initio* du module entier obtenu à partir des données de diffusion des rayons X aux petits angles (SAXS) (Figure 27A). Similairement à PikAIII, VirA<sub>M5</sub> adopte une conformation archée avec le dimère de KS au sommet de l'arche, mais cette arche est plus ouverte et ne délimite pas une chambre catalytique close (Figure 27A). Les frontières de l'arche sont les deux ACP séparés par un *linker* de 28 résidus. L'inclusion du « docking domain » C-terminal (°DD) du module 5 n'induit pas la fermeture de l'arche mais au contraire une ouverture plus importante, comme en témoigne l'augmentation de la dimension maximale ( $D_{max}$ ) de cette protéine, ainsi que son modèle *ab initio* (Figure 27B). Il semblerait donc que le *linker* entre les deux ACP soit flexible et contrôle le degré d'ouverture de l'arche.

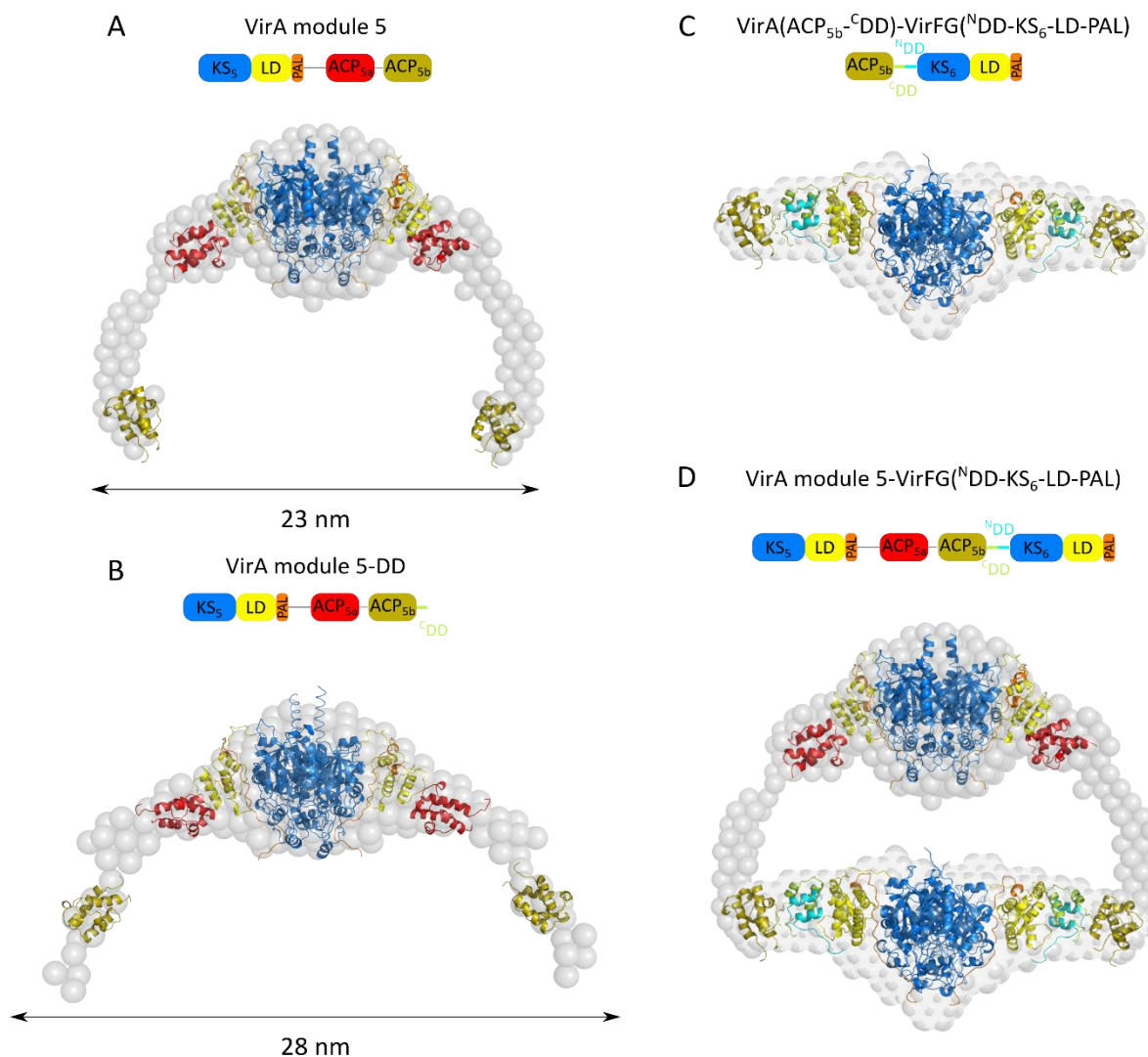


Figure 27. Structures multi-domaines des PKS *trans*-AT VirA/VirFG obtenues par diffusion des rayons X aux petits angles (SAXS). Modifié d'après Davison *et al.*, 2014 et Dorival *et al.*, 2016. Les structures des domaines individuels ont été positionnées dans les modèles *ab initio* de VirA module 5 (A), VirA module 5 incluant le « docking domain » C-terminal (<sup>cDD</sup>) (B) et la fusion entre l'ACP<sub>5b</sub>-<sup>cDD</sup> du module 5 et le <sup>NDD</sup>-KS du module 6 (C). Les dimensions maximales ( $D_{max}$ ) des protéines sont indiquées pour les modèles A et B. D) Juxtaposition des structures A et C.

Le module 6 de la chaîne d'assemblage de la virginiamycine est codé par un autre polypeptide et la reconnaissance entre les modules 5 et 6 est permise grâce à l'interaction entre le <sup>cDD</sup> à l'extrémité C-terminale du module 5 et le <sup>NDD</sup> à l'extrémité N-terminale du module 6 (Dorival *et al.*, 2016). Les interactions de « docking » étant transitoires, une protéine hybride composée de l'ACP<sub>5b</sub> et du <sup>cDD</sup> du module 5 fusionnés au <sup>NDD</sup> du module 6 et au KS<sub>6</sub> a été étudiée par SAXS (Dorival *et al.*, 2016) (Figure 27C). Le modèle résultant, plutôt allongé, suggère que l'interaction entre les DD positionne l'ACP<sub>5b</sub> à côté du LD. Cette structure est compatible avec la structure de VirA module 5 et suggère que l'interaction entre les modules 5 et 6 induit la fermeture de l'arche formée par le module 5 (Figure 27D). Ces données

suggèrent que les PKS *trans*-AT, malgré une organisation différente des PKS *cis*-AT, pourraient aussi utiliser des chambres réactionnelles.

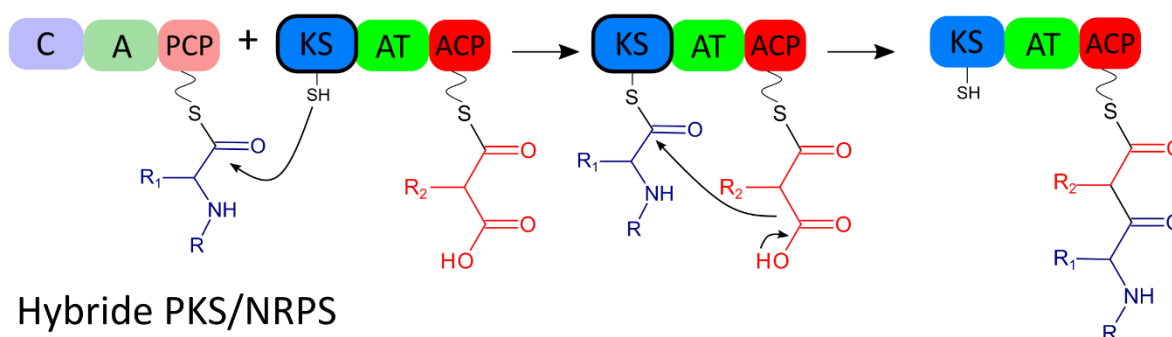
En résumé, les connaissances structurales des PKS *trans*-AT sont encore parcellaires. De nombreux points importants restent à déterminer, tels que les modalités de recrutement des domaines agissant en *trans*, en particulier le domaine AT. Le domaine ACP ayant seulement été visualisé dans des modèles SAXS, des structures à haute résolution de modules *trans*-AT complets sont nécessaires pour visualiser les surfaces d'interaction entre ce domaine et les domaines catalytiques, et les comparer à celles déterminées pour les PKS *cis*-AT.

## VI. Les hybrides NRPS/PKS

### VI.1. Classification des hybrides

De nombreux métabolites d'intérêt biologique sont des hybrides NRP/PK (Figure 11). En effet, la logique commune des NRPS et des PKS, en particulier l'utilisation du domaine « carrier protein » pour présenter le métabolite en construction aux domaines catalytiques, rend possible l'existence de chaînes d'assemblage contenant les deux types de mégaszymes. Au moins un tiers des îlots génomiques NRPS ou PKS décrits sont des îlots hybrides NRPS/PKS (Miyana, Kudo, *et al.*, 2018).

#### A Hybride NRPS/PKS



#### B Hybride PKS/NRPS

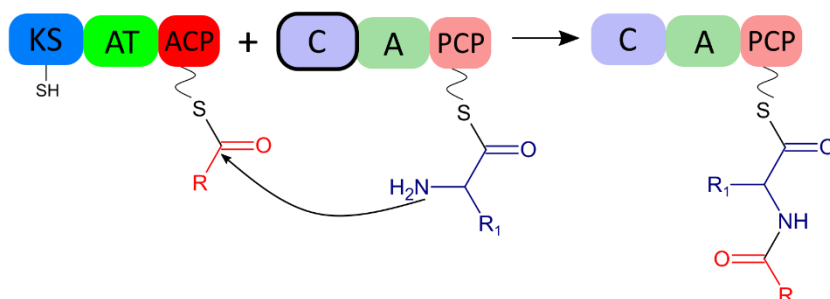


Figure 28. Réactions catalysées par les domaines cétyosynthase (KS) d'hybrides NRPS/PKS et les domaines de condensation (C) d'hybrides PKS/NRPS. Adapté de Miyana, Kudo, *et al.*, 2018. Le domaine catalysant la réaction chimique décrite est entouré en noir. A) Chez les hybrides NRPS/PKS, le domaine KS accepte d'abord le substrat donneur du domaine PCP puis catalyse la formation d'une liaison carbone-carbone entre ce substrat et le substrat accepteur présenté par l'ACP. B) Chez les hybrides PKS/NRPS, le domaine C catalyse la formation d'une liaison peptidique entre le substrat donneur présenté par le domaine ACP et le substrat accepteur présenté par le domaine PCP.

Deux types d'interface sont possibles entre un module NRPS et PKS : une interface séparée si les deux modules sont codés par différents polypeptides ou une interface fusionnée si les deux modules sont fusionnés au sein du même polypeptide (Miyana, Kudo, *et al.*, 2018). Dans le reste de cette thèse, une protéine hybride NRPS-PKS désignera une protéine hybride à interface fusionnée composée d'un module NRPS en N-terminal et d'un module PKS en C-terminal. Quel que soit le type d'interface, une interaction entre un domaine NRPS et un

domaine PKS est nécessaire à l'interface hybride. Chez les hybrides NRPS/PKS, le domaine KS catalyse la trans-thioestérification d'un substrat donneur en provenance du module NRPS, puis la condensation avec le substrat polycétide chargé sur le domaine ACP (Figure 28A). Chez les hybrides PKS/NRPS, le domaine C catalyse la formation d'une liaison peptidique entre un substrat donneur PK et un substrat accepteur NRP (Figure 28B). L'hybridation fonctionnelle des NRPS et PKS repose donc sur l'interaction PCP:KS chez les hybrides NRPS/PKS et sur l'interaction ACP:C chez les hybrides PKS/NRPS.

## VI.2. Interfaces NRPS/PKS

Plusieurs domaines KS acceptant un substrat provenant d'un module NRPS ont été caractérisés structurellement. Le domaine OzmQ-KS1 présente une interface séparée avec un module NRPS en amont (Lohman *et al.*, 2015) (Figure 26A). Le domaine KS de TAS1, qui participe à la biosynthèse de l'acide tenuazonique, une mycotoxine produite par *Pyricularia oryzae*, présente lui une interface fusionnée avec un module NRPS en amont (Yun *et al.*, 2020). Ces structures ont montré que les domaines KS acceptant des substrats en provenance des NRPS présentent des adaptations, notamment l'absence d'une hélice recouvrant la poche catalytique, permettant probablement l'acceptation de substrats contenant des acides aminés, plus gros que les chaînes polycétides (Yun *et al.*, 2020). Ces adaptations sont cohérentes avec le fait que les domaines KS d'hybrides NRPS/PKS forment un groupe phylogénétique distinct des autres domaines KS (Du *et al.*, 2001).

Un type d'interface NRPS/PKS particulier est retrouvé chez les PKS *trans*-AT : les domaines KS non fonctionnels (KS<sup>0</sup>) (Kosol *et al.*, 2018). L'importance du domaine KS<sup>0</sup> à l'interface NRPS/PKS a été démontrée chez Lnml, une hybride NRPS-PKS de *Streptomyces atrolivaceus* participant à la biosynthèse de la leinamycine, un métabolite aux propriétés antimicrobiennes et antitumorales (Huang *et al.*, 2016). Ce domaine KS<sup>0</sup> est capable de catalyser la trans-thioestérification d'un substrat mais pas son élongation car la première de ses histidines catalytiques est mutée en alanine. Un substrat NRPS présenté par un domaine PCP ne serait pas reconnu par un domaine KS donc le domaine KS<sup>0</sup> permettrait le transfert de ce substrat à un domaine ACP, qui le délivrerait ensuite au domaine KS (Miyanaga, Kudo, *et al.*, 2018). Le di-domaine KS<sup>0</sup>-ACP serait donc une interface de transition entre un module NRPS et un module PKS.

### VI.3. Interfaces PKS/NRPS

Certaines interfaces PKS/NRPS font également intervenir des domaines KS<sup>0</sup>. En particulier, un domaine KS<sup>0</sup> intervient dans la biosynthèse de l'antibiotique énacyloxine (Masschelein *et al.*, 2019). Comme chez Lnml, le domaine KS<sup>0</sup> de Bamb\_5919 possède une activité de trans-thioestérification mais pas de condensation et transfère donc le substrat PK d'un domaine ACP à un domaine PCP. Les auteurs de l'étude ont montré que le transfert du substrat au domaine PCP est nécessaire à la prise en charge par le domaine NRPS en aval, qui est incapable de reconnaître le substrat PK chargé sur le domaine ACP. Les domaines KS<sup>0</sup> semblent donc être des adaptations répandues chez les PKS *trans*-AT présentant une interface hybride avec un module NRPS en amont ou en aval.

La reconnaissance entre deux modules NRPS et PKS appartenant à différents polypeptides fait intervenir des « docking domains » (DD) (Smith *et al.*, 2021). Par exemple, dans la chaîne d'assemblage de l'anticancéreux épothilone, la reconnaissance entre le module PKS d'EpoA en amont et le module NRPS EpoB en aval fait intervenir des DD (Miyanaga, Kudo, *et al.*, 2018). La structure du di-domaine <sup>N</sup>DD-Cy d'EpoB a été résolue par cristallographie aux rayons X (Dowling *et al.*, 2016). Le DD adopte une structure monomérique en épingle à cheveux. Ce DD est connecté par un *linker* flexible au domaine Cy et ne contacte pas le domaine Cy. Le domaine Cy ne semble pas présenter d'adaptation à la reconnaissance d'un substrat polycétide.

### VI.4. Etat oligomérique des protéines hybrides de type fusionné

Les PKS sont des protéines dimériques et la plupart des NRPS caractérisées sont monomériques, l'état oligomérique des hybrides n'est donc pas facilement prévisible. L'état oligomérique de l'hybride PKS-NRPS HMWP1 participant à la biosynthèse de la yersiniabactine, a été étudié par filtration sur gel et a révélé que HMWP1 forme à la fois des monomères et des dimères (Sieber *et al.*, 2002). La dimérisation uniquement par le domaine KS pourrait ne pas être suffisante pour maintenir l'état dimérique de la protéine entière (Miyanaga, Kudo, *et al.*, 2018).



# Choix du modèle d'étude et objectifs





L'effet génotoxique de la colibactine s'explique par sa structure, comprenant deux cycles cyclopropanes générant des ponts inter-brin dans l'ADN et deux cycles thiazoles, suspectés d'agir comme agent intercalant. Cette structure complexe est produite par l'îlot génomique *pks*, codant plusieurs NRPS/PKS, des machineries macromoléculaires microbiennes spécialisées dans la biosynthèse de métabolites secondaires.

Les connaissances structurales sur les NRPS/PKS se sont accumulées ces dernières années. En revanche, celles sur les hybrides NRPS/PKS, en particulier les hybrides fusionnés, sont très rares. Aucune structure ou aucun modèle d'hybride fusionné n'est disponible, et même l'état oligomérique de ces méga-enzymes est incertain. L'îlot *pks* comprend deux hybrides fusionnés, ClbB et ClbK (Figure 9). L'hybride NRPS-PKS ClbB participe au début de la biosynthèse de la colibactine puisque qu'elle intervient juste après ClbN générant le motif protecteur clivé après export de la pré-colibactine. L'hybride PKS-NRPS ClbK intervient quant à elle plus tardivement puisqu'elle agit après la NRPS di-modulaire ClbJ, générant un cycle thiazoline. ClbK oxyde le cycle thiazoline généré par ClbJ, incorpore une unité aminomalonyle ainsi que le deuxième cycle thiazole de la colibactine (Trautman *et al.*, 2017).

Le cycle catalytique de ClbK est déduit de la structure de la colibactine et du mode de fonctionnement des protéines NRPS/PKS (Figure 29) (Wernke *et al.*, 2020). Le module PKS de ClbK (ClbK-PKS) est un module *trans*-AT possédant un domaine AT atypique non fonctionnel (AT\*), complété par le domaine AT isolé ClbG, qui transfère une unité aminomalonyle sur le domaine ACP de ClbK (Brachmann *et al.*, 2015) (Figure 29, réaction 2). Le domaine KS de ClbK catalyse d'abord la *trans*-thioestérification du substrat en provenance de ClbJ (Figure 29, réaction 1) puis la condensation de Claisen de ce substrat avec l'unité aminomalonyle attachée au domaine ACP (Figure 29, réaction 3). En parallèle, le domaine A active et transfère la L-cystéine sur le domaine PCP (Figure 29, réaction 4). Le domaine Cy, à l'interface hybride PKS-NRPS, catalyse la condensation du produit du KS avec la L-cystéine, ainsi que la cyclisation de la L-cystéine en cycle thiazoline (Figure 29, réaction 5). Le cycle thiazoline généré est ensuite oxydé en cycle thiazole par le domaine Ox, inséré dans le domaine A (Figure 29, réaction 6).

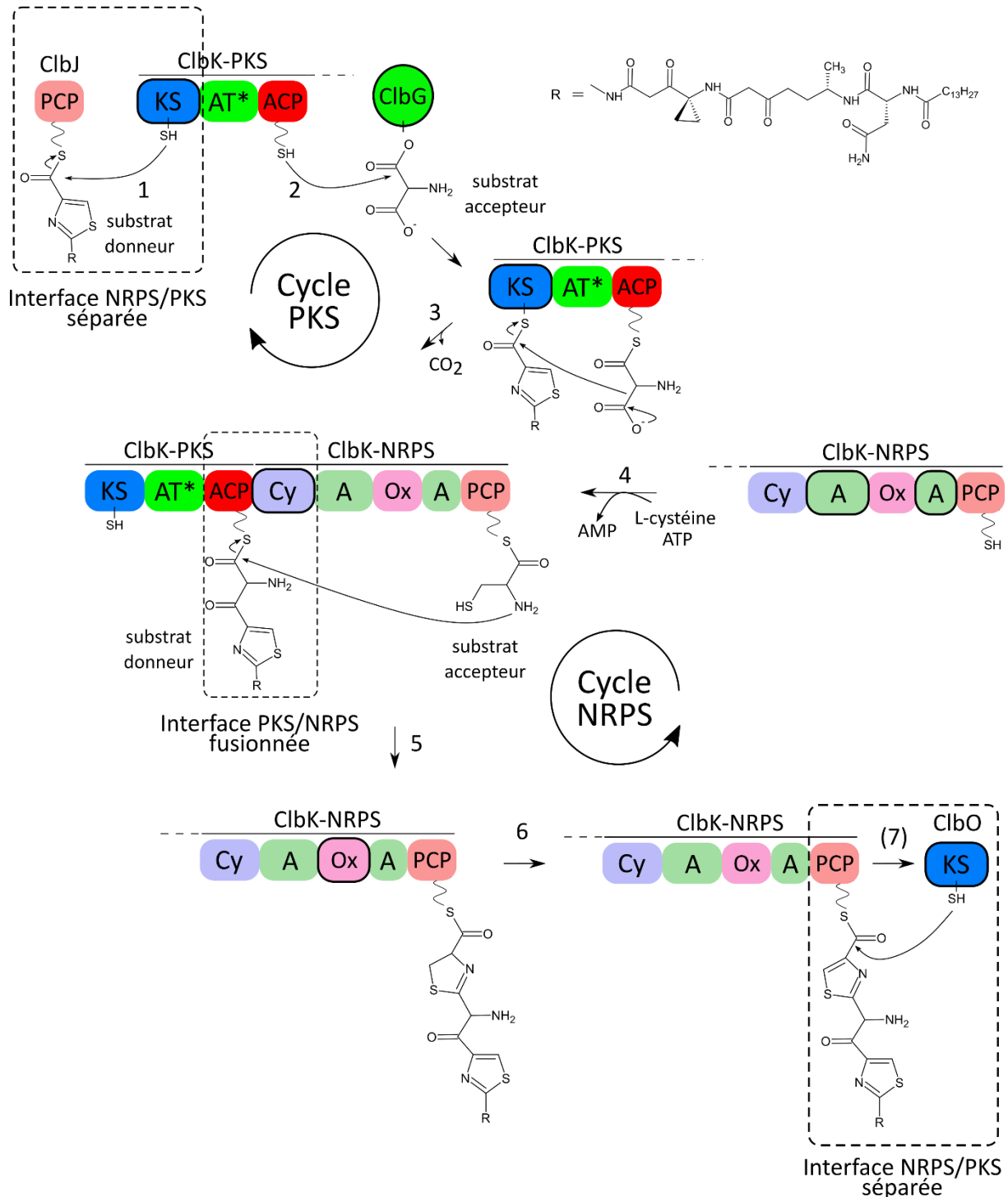


Figure 29. Réactions catalysées par la protéine hybride PKS-NRPS ClbK. Le domaine catalysant la réaction décrite est entouré en noir. 1) Le domaine KS catalyse la trans-thioestérification du substrat apporté par la NRPS ClbJ. 2) Le domaine AT isolé ClbG transfère une unité aminomalonyl sur le domaine ACP de ClbK. 3) Le domaine KS catalyse la condensation de Claisen entre les deux substrats. 4) Le domaine A catalyse l'adénylation et le transfert de la L-cystéine sur le domaine PCP. 5) Le domaine Cy catalyse la condensation des substrats et la cyclisation de la cystéine en cycle thiazoline. 6) Le domaine Ox catalyse l'oxydation du cycle thiazoline en cycle thiazole. 7) Le produit généré est transféré au domaine KS de ClbO (réaction catalysée par ClbO). L'activité oxydase du domaine Ox sur le cycle thiazoline généré par ClbJ n'a pas été représentée sur ce schéma. ClbK comprend une interface hybride PKS-NRPS fusionnée, ainsi que deux interfaces hybrides NRPS/PKS séparées, une avec ClbJ en amont et une avec ClbO en aval.

ClbK est un modèle d'étude de choix pour la caractérisation des interfaces hybrides PKS/NRPS (Figure 29). En effet, ClbK est non seulement une protéine hybride PKS-NRPS fusionnée, mais elle forme aussi deux interfaces hybrides séparées NRPS/PKS, une en amont avec ClbJ et une en aval avec ClbO. Ainsi, le module ClbK-PKS est entouré par deux modules NRPS et le module ClbK-NRPS est entouré par deux modules PKS.

L'objectif global de cette thèse était d'augmenter les connaissances structurales sur les hybrides NRPS/PKS. Avant mon arrivée dans l'équipe Pathogénie Bactérienne, le gène codant ClbK avait été cloné afin de permettre son expression et les premiers tests d'expression et de purification avaient été menés. Le premier objectif de ma thèse était d'établir un protocole de purification reproductible de ClbK. Le deuxième objectif était la détermination de l'état d'oligomérisation de l'hybride NRPS-PKS fusionné.

Le troisième objectif était la caractérisation structurale de cette méga-enzyme. L'ensemble des données structurales sur les NRPS/PKS démontrent que ces protéines adoptent plusieurs conformations au cours de leur cycle catalytique. La caractérisation structurale à haute résolution par cristallographie aux rayons X et microscopie électronique est empêchée par une trop grande flexibilité. Ainsi, l'étude de la protéine entière (240 kDa) a été complétée par l'étude structurale de fragments multi-domaines, supposés moins flexibles que la protéine entière. L'objectif final était l'établissement du premier modèle d'une protéine hybride NRPS/PKS fusionnée.



# Matériels et méthodes



## I. Matériel biologique

### I.1. Vecteurs d'expression

Nom	Gène de résistance	Etiquette d'affinité	Site de clivage de l'étiquette	Promoteur
pET-15b	Ampicilline	N-His	Thrombine	Lactose
pET-21a	Ampicilline	C-His		Lactose
pBADM-10	Ampicilline	N-His	TEV	Arabinose
pBADM-30	Ampicilline	N-His, N-GST	TEV	Arabinose
pETM-11	Kanamycine	N-His	TEV	Lactose
pETM-30	Kanamycine	N-His, N-GST	TEV	Lactose

Tableau 1. Vecteurs d'expression utilisés durant cette thèse. Pour l'étiquette d'affinité, N ou C désigne la position N-terminale ou C-terminale et His désigne l'étiquette poly-histidines alors que GST désigne l'étiquette glutathion-S-transférase. L'expression des protéines est induite par le lactose ou l'IPTG dans les vecteurs contenant un promoteur lactose et par l'arabinose dans les vecteurs contenant un promoteur arabinose.

### I.2. Souches bactériennes

Plusieurs souches d'*E. coli* ont été utilisées pour la préparation de plasmide ou l'expression protéique. Toutes les souches sont des souches commerciales déjà utilisées dans le laboratoire, à l'exception de la souche BL21(DE3) Bap1 (Pfeifer *et al.*, 2001), fournie par Chaitan Khosla.

Nom de la souche	Génotype	Intérêt
NEB 5-alpha	<i>fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>	Préparation de plasmide
XL10-Gold	<i>Tet<sup>R</sup> Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F' proAB lacI<sup>q</sup> ZΔM15 Tn10 (Tet<sup>R</sup>) Amy Cam<sup>r</sup>]</i>	Cellules ultra-compétentes utilisées pour la mutagenèse dirigée
BL21(DE3)	<i>F<sup>-</sup> ompT gal dcm lon hsdS<sub>B</sub>(r<sub>B</sub>,m<sub>B</sub>), λ (DE3)</i>	Souche d'expression



BL21(DE3) Bap1	BL21(DE3) $\Delta prpRBCD::T7prom-sfp$ , T7prom- <i>prpE</i> (Pfeifer <i>et al.</i> , 2001)	Modification post-traductionnelle des domaines « carrier protein »
BL21(DE3) pLysS	BL21(DE3), pLysS ; Cm <sup>R</sup>	Réduction de l'expression avant induction
BL21(DE3) STAR	BL21(DE3), <i>rne131</i>	Haute stabilité des ARNm
C41(DE3)	F', <i>ompT, gal, dcm, hsdS<sub>B</sub>(r<sub>B</sub>,m<sub>B</sub>)</i> , $\lambda$ (DE3)	Expression de protéines toxiques

Tableau 2. Souches d'*Escherichia coli* utilisées dans cette thèse.

### 1.3. Milieux de culture

- Le milieu LB (Lennox) a été utilisé pour les transformations, pour les cultures en vue d'extraction d'ADN par mini-préparation, pour toutes les précultures et pour l'expression de certaines protéines.
- Le milieu TB (Sigma, supplémenté en glycérol) a été utilisé pour les cultures d'expression de certaines protéines.
- Le milieu d'auto-induction a été utilisé pour l'expression de certaines protéines.  
Composition pour 500 mL : 5 g Tryptone (US Biological), 2.5 g d'extrait de levure (Euromedex) dans 460 mL H<sub>2</sub>O milliQ, à autoclaver. Rajouter 0.5 mL 1 M MgSO<sub>4</sub> filtré, 25 mL 20 X NPS filtré, 4.2 mL glycérol 60 % autoclavé, 2.5 mL glucose 10 % filtré, 10 mL  $\alpha$ -lactose 10 % filtré.  
Préparation de la solution 20 X NPS (100 mL) : 6.6 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 13.6 g KH<sub>2</sub>PO<sub>4</sub> et 14.2 g Na<sub>2</sub>HPO<sub>4</sub>. Ajuster le pH à 6.75.
- Le milieu minimum M9 a été utilisé pour l'expression de protéine sélénométhionée.  
Composition pour 1 L : 780 mL H<sub>2</sub>O autoclavée, 200 mL 5X sels M9, 20 mL glucose 20 % filtré, 2 mL 1 M MgSO<sub>4</sub> filtré, 100  $\mu$ L 1 M CaCl<sub>2</sub>.  
Préparation de la solution 5X sels M9 (1L) : 64 g Na<sub>2</sub>HPO<sub>4</sub>-7 H<sub>2</sub>O, 15 g KH<sub>2</sub>PO<sub>4</sub>, 2.5 g NaCl, 5 g NH<sub>4</sub>Cl. Ajuster le pH à 7.2.

## II. Bio-informatique

### II.1. Logiciels utilisés pour l'alignement de séquences et l'analyse des structures

Les alignements de séquences ont été effectués grâce au logiciel Clustal Omega (Sievers *et al.*, 2011). Ces alignements ont ensuite été analysés et représentés grâce au serveur ESPript (<https://esprpt.ibcp.fr>), qui permet aussi de représenter les structures secondaires associées aux séquences lorsque la structure de la protéine est connue (Robert & Gouet, 2014).

Toutes les structures de protéines présentées dans cette thèse ont été représentées grâce au logiciel Pymol (The PyMOL Molecular Graphics System, Schrödinger, LLC., <https://www.pymol.org>). Les modèles *ab initio* de SAXS ont été représentés soit avec Pymol, soit avec Chimera (Pettersen *et al.*, 2004). L'ajustement des structures cristallographiques dans les cartes de densité de cryo-ME a été effectué avec l'outil « Fit in map » de Chimera. Les interfaces protéiques des structures ont été analysées grâce au programme PISA ([http://www.ebi.ac.uk/pdbe/prot\\_int/pistart.html](http://www.ebi.ac.uk/pdbe/prot_int/pistart.html)) (Krissinel & Henrick, 2007).

### II.2. Analyse de la séquence de ClbK

Les analyses de contenu en acides aminés, de masse molaire et de point isoélectrique (pHi) de la protéine ClbK ont été effectuées avec l'outil ProtParam du serveur ExPASy (Gasteiger *et al.*, 2003).

La technique de réduction d'entropie de surface (SER) a été appliquée à ClbK afin de muter les résidus flexibles en surface qui pourraient gêner la cristallisation (Longenecker *et al.*, 2001). Les résidus à muter pour le mutant ClbK-SER ont été déterminés grâce au serveur SERp (Goldschmidt *et al.*, 2007). Ce serveur suggère plusieurs triplets de résidus à muter, chacun avec un score associé à la prédiction.

Plusieurs fragments multi-domaines de ClbK ont été générés durant cette thèse et ont nécessité une analyse précise de sa séquence afin d'en délimiter les domaines (Figure 30). L'alignement de séquences avec le programme Clustal Omega associé à l'analyse de structures existantes ont guidé cette délimitation. Le logiciel PONDR (<http://www.pondr.com>) a également été utilisé afin de définir les régions désordonnées de ClbK, qui ont été exclues de certaines constructions.

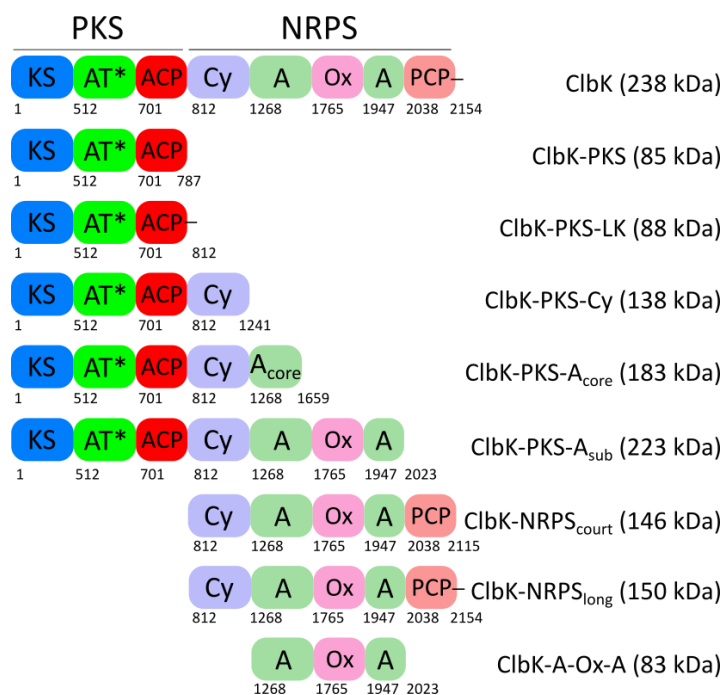


Figure 30. Fragments de ClbK générés et étudiés durant cette thèse. Les masses molaires indiquées et la numérotation des résidus ne tiennent pas compte de l'étiquette d'affinité. Pour chaque construction, le premier résidu de chaque domaine est indiqué, ainsi que le dernier résidu de la construction.

La délimitation C-terminale du fragment ClbK-PKS a été choisie par homologie avec le domaine ACP du module 2 de DEBS (PDB : 2JU2) (Alekseyev *et al.*, 2007). Les structures des domaines Cy d'EpoB (PDB : 5T7Z) (Dowling *et al.*, 2016) et de BdmB (PDB : 5T3E) (Bloudoff *et al.*, 2017) ont servi aux délimitations C-terminales des fragments ClbK-PKS-LK et ClbK-PKS-Cy, ainsi qu'à la délimitation N-terminale des fragments ClbK-NRPS<sub>court</sub> et ClbK-NRPS<sub>long</sub>. Les délimitations C-terminales des fragments ClbK-PKS-A<sub>core</sub> et ClbK-PKS-A<sub>sub</sub> ont bénéficié des comparaisons avec les structures du di-domaine F-A<sub>core</sub> de LgrA (PDB : 5JNF) (Reimer, Aloise, Powell, *et al.*, 2016) et du tridomaine F-A-PCP de LgrA (PDB : 5ES8) (Reimer, Aloise, Harrison, *et al.*, 2016). L'extrémité C-terminale de ClbK-NRPS<sub>court</sub> a été choisie pour ne pas inclure la région prédite désordonnée en aval du domaine PCP. De même, l'extrémité N-terminale de ClbK-A-Ox-A a été choisie pour ne pas inclure la région désordonnée en amont du domaine A.

### II.3. Génération de modèles par le logiciel AlphaFold2

Des modèles de ClbK-PKS et ClbG ont été générés par le logiciel de prédiction de structure AlphaFold2 à l'aide de Sylvain Engilberge (ESRF). Ce logiciel d'intelligence artificielle prédit avec une précision remarquable la structure de protéines à partir de leurs séquences primaires. En 2020, le logiciel AlphaFold2 a largement remporté la quatorzième édition du concours CASP consacré à la prédiction de structure protéique en aveugle (Jumper *et al.*, 2021). La prédiction des structures par AlphaFold2 repose sur des contraintes évolutives, physiques et géométriques. Ce logiciel repose sur un apprentissage profond (« deep

learning ») des quelques 100 000 structures déposées sur la PDB. Une force de ce programme est que la confiance dans la prédiction est estimée, notamment grâce au score pLDDT (en anglais « predicted local distance difference test »).

### III. Biologie moléculaire

#### III.1. Clonage de ClbK et de ses fragments

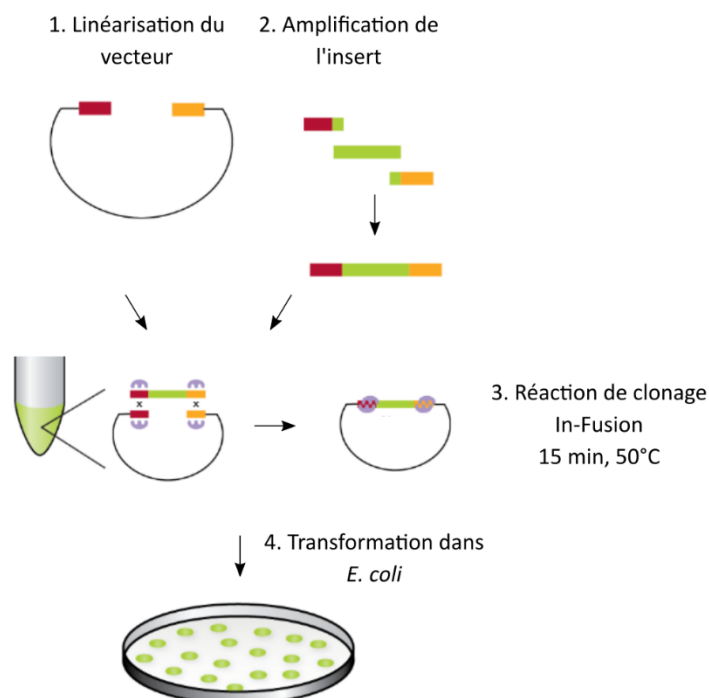


Figure 31. Principe du protocole In-Fusion (simplifié d'après le manuel de Takara Bio). Les zones en rouge et jaune du vecteur désignent les sites de restriction, qui sont inclus dans les amorces servant à amplifier l'insert.

Au vu de la grande taille des séquences ADN à cloner, contenant souvent des sites de restriction, tous les clonages ont été effectués grâce au kit In-Fusion (Takara Bio) ne nécessitant pas de digérer l'insert (Figure 31). L'ADN a d'abord été amplifié par réaction en chaîne par polymérase (PCR) grâce à un couple d'amorces contenant des séquences complémentaires au vecteur digéré (Tableau 3), synthétisées chimiquement par Eurofins Genomics. L'ADN amplifié a été purifié sur gel d'agarose (Nucleospin Gel and PCR clean up, Macherey-Nagel), puis introduit par recombinaison homologue dans un vecteur prédigéré (Tableau 1). Le plasmide recombinant a été introduit par transformation dans la souche NEB-5-alpha. L'ADN de plusieurs colonies a été extrait par mini-préparation (Nucleospin, Macherey-Nagel), digéré avec les enzymes appropriées et analysées par électrophorèse sur gel d'agarose. Les séquences des plasmides recombinants ayant un insert à la taille attendue ont été vérifiées par séquençage (Eurofins Genomics).

L'ADN utilisé pour le clonage du gène codant la protéine ClbK entière provient d'un chromosome bactérien artificiel contenant l'îlot génomique *pks* d'*E. coli* (BAC-*pks*), donné à l'équipe Pathogénie Bactérienne par Jean-Philippe Nougayrède (Nougayrède *et al.*, 2006). Le

clonage du gène codant ClbK dans le plasmide pET-15b a été effectué par Viviana Job en amplifiant l'ADN de BAC-*pks* par PCR grâce aux amorces *clbKf\_NdeI\_pET-15b* et *clbKr\_BamHI* (Tableau 3). Le plasmide pET-15b-*clbK* généré a servi de matrice pour tous les autres clonages. Les amorces *clbKf\_NcoI* et *clbKr\_HindIII* ont permis le clonage du gène codant ClbK dans les plasmides pBADM-10 et pBADM-30 (Tableau 3). Ce clonage, ainsi que celui du gène codant ClbK dans le plasmide pET-21a grâce aux amorces *clbKf\_NdeI\_pET-21a* et *clbKr\_XhoI* ont été réalisés par Pauline Macheboeuf (Tableau 3).

Nom de l'amorce	Séquence
<i>clbKf_NdeI_pET-15b</i>	5'-CGCGCGGCAGCC <u>CATATG</u> ACTTACAGTGAAAGCGATATTGC-3'
<i>clbKr_BamHI</i>	5'-GTTAGCAGCC <u>GATCCT</u> CACTCATCGCCCTTTTCTCCC-3'
<i>clbKf_NcoI</i>	5'-TTCAGGGCGC <u>CCATGGG</u> TACTTACAGTGAAAGCGATATTGCCA-3'
<i>clbKr_HindIII</i>	5'-TGCGGCCGCA <u>AAGCTTT</u> CACTCATCGCCCTTTTCTCC-3'
<i>clbKf_NdeI_pET-21a</i>	5'-GGAGATATACATATGACTTACAGTGAAAGCGATATTG-3'
<i>clbKr_XhoI</i>	5'-GTGGTGGT <u>GCTCGAG</u> CTCATCGCCCTTTTCTCCCT-3'
<i>clbK-PKSr_XhoI</i>	5'-GTGGTGGT <u>GCTCGAG</u> TCAGCCTTGTTGAGCAGATAGC-3'
<i>clbK-PKS-LKr_XhoI</i>	5'-GTGGTGGT <u>GCTCGAG</u> TCACAGCGGAAAAGGTAGGTG-3'
<i>clbK-PKS-Cyr_XhoI</i>	5'-GTGGTGGT <u>GCTCGAG</u> TCAAGTCTCATGCCAGAGCTGT-3'
<i>clbK-PKS-A<sub>coref</sub>_XhoI</i>	5'-GTGGTGGT <u>GCTCGAG</u> TCAATCGTTGCGCCCCAAAAG-3'
<i>clbK-PKS-A<sub>subf</sub>_XhoI</i>	5'-GTGGTGGT <u>GCTCGAG</u> TCCTGCAGCTGTAATTTTGGCGAT-3'
<i>clbK-NRPS<sub>f</sub>_NcoI</i>	5'-TTCAGGGCGC <u>CCATGGG</u> GCTGACTGACGTGCAACAG-3'
<i>clbK-NRPS<sub>court</sub>_XhoI</i>	5'-GTGGTGGT <u>GCTCGAG</u> TCAATTGTGCAGTTGCTCATCAAGGT-3'
<i>clbK-NRPS<sub>long</sub>_XhoI</i>	5'-GTGGTGGT <u>GCTCGAG</u> TCCTCATCGCCCTTTTCTCC-3'
<i>clbK-A<sub>coref</sub>_NcoI</i>	5'-TTCAGGGCGC <u>CCATGGG</u> GATTACTGCGTCAGGCAGCA-3'
<i>clbK-SERf</i>	5'-GCATTGCAAGCT <u>GCGGCAGCGGCGG</u> CGGCGGCGGCACTG-3'
<i>clbK-SERr</i>	5'-CAGTGCCGCCG <u>CCGCGCTGCCGC</u> AGCTTGCAATGC-3'
<i>clbK-s747af</i>	5'-CTGGGTGGGCAT <u>GCG</u> CTGTTAGCGTC-3'
<i>clbK-s747ar</i>	5'-GACGCTAACAG <u>GCG</u> CATGCCACCCAG-3'

Tableau 3. Amorces d'ADN utilisées durant cette thèse. Les lettres f et r dans le nom de l'amorce désignent respectivement les amorces en amont et en aval du gène à amplifier. Les quinze premières amorces ont servi au clonage de *clbK* et de ses fragments alors que les quatre dernières amorces ont servi à la mutagenèse dirigée de *clbK*. Pour les amorces ayant servi au clonage, le nom de l'enzyme de restriction est indiqué dans le nom de l'amorce et le site de restriction est souligné dans la séquence. Pour les quatre dernières amorces, les nucléotides mutés sont surlignés en gras et les triplets de nucléotides impactés sont soulignés.

J'ai effectué les clonages des différents fragments multi-domaines de ClbK. Les séquences d'ADN codant les protéines ClbK-PKS, ClbK-PKS-LK, ClbK-PKS-Cy, ClbK-PKS-A<sub>core</sub> et ClbK-PKS-A<sub>sub</sub> ont été amplifiées par PCR grâce à l'amorce *clbKf\_NcoI* en amont et aux amorces *clbK-PKSr\_XhoI*, *clbK-PKS-LKr\_XhoI*, *clbK-PKS-Cyr\_XhoI*, *clbK-PKS-A<sub>core</sub>r\_XhoI* et *clbK-PKS-A<sub>sub</sub>r\_XhoI* et en aval (Tableau 3). Les séquences d'ADN codant les fragments ClbK-NRPS<sub>court</sub> et ClbK-NRPS<sub>long</sub> ont été amplifiées grâce aux amorces *clbK-NRPSf\_NcoI* en amont et en aval *clbK-NRPS<sub>court</sub>r\_XhoI* ou *clbK-NRPS<sub>long</sub>r\_XhoI* (Tableau 3). L'ADN codant le fragment ClbK-A-Ox-A a été amplifié grâce aux amorces *clbK-A<sub>core</sub>f\_NcoI* et *clbK-PKS-A<sub>sub</sub>r\_XhoI* (Tableau 3).

La numérotation des résidus ne tient pas compte de l'étiquette d'affinité introduite grâce aux vecteurs. Les fragments ayant été clonés avec les amorces *clbKf\_NcoI*, *clbK-NRPSf\_NcoI* et *clbK-A<sub>core</sub>f\_NcoI* ont un résidu glycine supplémentaire après le premier résidu méthionine. Par souci d'homogénéité, pour ces fragments, le résidu méthionine est numéroté 0, le résidu glycine 1. Ainsi, toutes les constructions de ClbK et des fragments incluant le module PKS ont un résidu thréonine en position 2 puis tyrosine en position 3, ... (voir Annexe).

### III.2. Mutagenèse dirigée

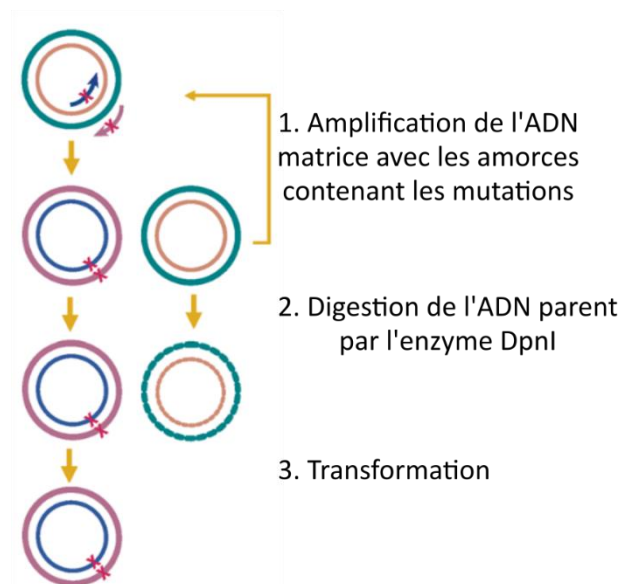


Figure 32. Principe du protocole de mutagenèse dirigée QuikChange Lightning Site-Directed Mutagenesis Kit. D'après le manuel d'instruction d'Agilent.

Les plasmides mutants pET-15b-*clbK-SER* et pET-15b-*clbK-s747a* ont été générés par mutagenèse dirigée grâce au kit QuikChange Lightning Site-Directed Mutagenesis (Agilent) (Figure 32). Quatre nucléotides ont été mutés dans le plasmide pET-15b-*clbK-SER*, générant trois résidus mutés dans la protéine ClbK-SER, les résidus Lys1718, Glu1719 et Glu1721. Un seul nucléotide a été muté dans le plasmide pET-15b-*clbK-s747a*, générant la mutation du

résidu Ser747 en alanine. L'ADN de pET15b-*clbK* a été amplifié avec les amorces contenant la/les mutations(s) souhaitée(s) (Tableau 3). L'ADN parental a ensuite été digéré par l'endonucléase DpnI puis l'ADN a été transformé dans les cellules XL10-Gold d'*E. coli*. L'ADN a ensuite été extrait par mini-préparation et séquencé (Eurofins Genomics). Le plasmide pET15b-*clbK-s747a* été utilisé comme matrice pour la génération du mutant pETM-11-*clbK-PKS-s747a* grâce au protocole In-Fusion décrit précédemment.



## IV. Expression et purification de ClbK et de ses fragments

Toutes les incubations ont été réalisées avec une rotation de 180 rpm. Toutes les précultures et cultures avant induction ont été incubées à 37°C. Tous les milieux solides et liquides utilisés pour la multiplication des bactéries transformées avec les plasmides pET-15b, pET-21a, pBADM-10 et pBADM-30 ont été supplémentés avec de l'ampicilline (100 µg.mL<sup>-1</sup>). De même 50 µg.mL<sup>-1</sup> de kanamycine ont été ajoutés dans tous les milieux utilisés pour la multiplication des bactéries transformées avec les plasmides pETM-11 et pETM-30.

### IV.1. Tests d'expression, de solubilité et d'attache à la résine d'affinité

Le protocole décrit dans cette section a été utilisé pour tester l'expression, la solubilité et la première étape de purification des fragments de ClbK. Un protocole équivalent a été utilisé pour les mêmes tests effectués sur la protéine ClbK entière, en faisant varier la souche d'expression, la température d'induction, la quantité d'inducteur et la composition du tampon de lyse (Skiba *et al.*, 2018).

Le plasmide recombinant codant le fragment de ClbK d'intérêt a été introduit par transformation dans la souche BL21(DE3) Bap1. Une préculture a été lancée à partir d'une colonie isolée. Le lendemain matin, des cultures de 50 mL en milieu LB, TB et milieu d'auto-induction ont été initiées par l'ajout de 0.5 mL de préculture dans le milieu. La croissance a été poursuivie à 37°C jusqu'à ce que la densité optique à 600 nm (DO<sub>600</sub>) atteigne 0.5. Les cultures ont alors été transférées dans un incubateur à 18°C et divisées en deux cultures de 10 mL, chacune dans un tube à centrifuger de 50 mL (Greiner). Dans le milieu LB et TB, l'expression a été induite par l'ajout de 0.2 mM ou 0.5 mM d'IPTG après 1h30 d'incubation à 18°C alors que dans le milieu d'auto-induction, l'induction ne nécessite pas l'ajout d'inducteur. L'efficacité de l'expression dans les différents milieux a été analysée sur gel SDS après une nuit d'expression à 18°C. Les cultures ont été centrifugées (10 minutes, 4200 g) et stockées à -20°C.

Les étapes suivantes ont été effectuées à 4°C. Les cultures positives pour l'expression de la protéine d'intérêt ont été dissoutes dans 1 mL de tampon de lyse (50 mM Hepes pH 7.5, 300 mM NaCl, 5 mM DTT) puis lysées par sonication. Une fraction des protéines totales a été conservée pour un gel SDS. Les protéines solubles ont été isolées par centrifugation (10 minutes, 12000 g). Les fractions solubles contenant des protéines à étiquette poly-histidines ont été incubées avec environ 50 mg de résine Protino-Ni-IDA (Macherey-Nagel). Les fractions

solubles contenant des étiquettes GST ont été incubées avec environ 100  $\mu$ L de résine glutathion (Glutathione Sepharose High performance, Cytiva) sur la nuit. Le lendemain, le surnageant contenant les protéines non retenues a été retiré puis les résines d'affinité ont été lavées avec 3 x 1 mL de tampon de lyse. Les protéines fusionnées à une étiquette poly-histidines ont été éluées dans 150  $\mu$ L de tampon de lyse supplémenté avec 250 mM d'imidazole alors que les protéines fusionnées à une étiquette GST ont été éluées dans 150  $\mu$ L de tampon de lyse supplémenté avec 10 mM de glutathion réduit. L'analyse sur gel SDS des protéines totales, solubles, non retenues et éluées dans les différentes conditions d'expression a permis de choisir la meilleure condition d'expression pour des cultures à plus grande échelle. L'analyse par Western Blot révélé avec des anticorps anti-histidines ou anti-GST a permis de confirmer l'identité de la protéine exprimée et partiellement purifiée.

#### IV.2. Expression et lyse des cultures de ClbK

Le protocole d'expression et de purification de ClbK a été largement optimisé et est sensible aux variations, il sera donc détaillé précisément. Ce même protocole a été employé pour l'expression et la purification du mutant ClbK-SER. Le plasmide pET-15b-*clbK* a été transformé dans la souche BL21(DE3) Bap1. Une préculture a été lancée à partir d'une colonie isolée et le lendemain, une culture a été lancée en milieu LB à partir de la préculture diluée au 1/100<sup>ème</sup>. Typiquement, quatre litres de culture ont été répartis entre quatre erlenmeyers de trois litres. La culture a été maintenue à 37°C jusqu'à ce que la  $DO_{600}$  ait atteint 0.5 puis la culture a été transférée dans un incubateur à 18°C. Après 1h30 à 18°C, la culture a été induite avec 0.2 mM d'IPTG. Le lendemain matin, après 16-20 heures d'incubation à 18°C, les cellules ont été centrifugées (20 minutes, 7800 g) et les culots ont été congelés à -20°C ou lysés directement.

La protéine ClbK contient 33 cystéines et est encline à l'agrégation donc l'agent réducteur dithiothréitol (DTT) a été inclus dans tous les tampons afin d'éviter la formation aberrante de ponts disulfures inter-chaînes pouvant conduire à l'agrégation. Le DTT étant instable, il a été rajouté juste avant l'utilisation des tampons.

Les culots ont été dissous dans du tampon de lyse (50 mM Hepes pH 7.5, 500 mM NaCl, 5 mM DTT). Un volume de tampon de lyse trop faible ayant conduit à une forte agrégation de la protéine ClbK, le volume de tampon de lyse a été fixé à 15 mL pour 5 grammes de culot.

1  $\mu$ L de nucléase Benzonase (Sigma) par litre de culture a été ajouté, ainsi que trois inhibiteurs de protéase (cOMplete EDTA Free, Roche) pour quatre litres de culture. Les cellules ont ensuite été lysées par quatre passages dans un broyeur cellulaire Constant Systems, à 20-22-25-25 kpsi.

#### IV.3. Purification de ClbK

Le protocole décrit ici est celui pour quatre litres de cultures. Pour certaines expériences ayant nécessité des quantités de ClbK de l'ordre du milligramme, huit litres de cultures ont alors été utilisés mais ils ont été traités comme deux fois quatre litres de cultures jusqu'à l'étape de filtration sur gel.

Toutes les étapes suivantes ont été réalisées à 4°C. Le surnageant de centrifugation (40 minutes, 39000 *g*) contenant les protéines solubles a été réparti entre quatre tubes à centrifuger de 50 mL, chacun contenant 1.5 g de résine Protino-Ni-IDA (Macherey-Nagel). L'incubation avec la résine a été effectuée pendant au moins 1h30. Le surnageant a été retiré et trois lavages avec du tampon de lyse ont été réalisés. La résine a été regroupée en deux tubes de 50 mL contenant chacun 3 g de résine et les protéines ont été éluées avec le tampon de lyse supplémenté avec 250 mM d'imidazole en trois étapes (12, 8, 6 mL dans chaque tube). 1 mM EDTA (concentration finale) a été ajouté directement dans chaque fraction d'éluion afin de chélater les cations divalents nécessaires à certaines protéases et ainsi limiter la protéolyse enzymatique de ClbK.

Les éluions de chromatographie d'affinité ont été dialysées une première fois contre 1.5 L de tampon 50 mM Hepes pH 7.5, 200 mM NaCl, 2 mM DTT, 1 mM EDTA sur la nuit puis une deuxième fois contre 1.5 L de tampon 50 mM Hepes pH 7.5, 100 mM NaCl, 2 mM DTT, 1 mM EDTA pendant 1h30. Les fractions dialysées ont été centrifugées (10 minutes, 12000 *g*) puis filtrées afin d'éliminer les agrégats. Le filtrat a été injecté sur une colonne échangeuse d'anions Mono Q 5/50 GL (GE Healthcare), préalablement équilibrée dans le tampon 50 mM Hepes pH 7.5, 100 mM NaCl, 5 mM DTT, 1 mM EDTA et les protéines ont été éluées par un gradient linéaire de chlorure de sodium.

Une dernière étape de chromatographie par exclusion de taille ou filtration sur gel a permis d'obtenir un échantillon pur. Les fractions contenant ClbK ont été identifiées par électrophorèse sur gel SDS 7.5%. Ces fractions ont été concentrées jusqu'à 500  $\mu$ L (Amicon

Ultra-4, 100 kDa, Merckmillipore) et injectées sur une colonne d'exclusion de taille Superdex 200 Increase 10/300 GL (GE Healthcare), préalablement équilibrée dans le tampon 50 mM Hepes pH 7.5, 200 mM NaCl, 5 mM DTT.

#### IV.4. Préparation de ClbK pour la microscopie électronique

L'effet de la concentration en chlorure de sodium sur l'état oligomérique de ClbK a été étudié en microscopie électronique. Pour cela, des fractions pures de ClbK après filtration sur gel ont été dialysées dans un tampon sans sel (50 mM Hepes pH 7.5, 5 mM DTT) grâce à des cassettes de dialyse Slide-A-Lyzer 3500 MWCO, 0.1-0.5 mL (ThermoFisher). L'échantillon ainsi préparé a été observé en microscopie électronique et analysé en photométrie de masse.

L'effet d'agents de réticulation sur l'état oligomérique de ClbK a également été étudié. Lors d'une réaction de réticulation, des liaisons covalentes sont formées entre les groupements chimiques de l'agent de réticulation et les groupements chimiques de la protéine. Les trois agents de réticulation testés ici, le glutaraldéhyde, le bis(sulfosuccinimidyl) subérate (BS<sup>3</sup>) et l'éthylène glycol bis(sulfosuccinimidyl succinate) (EGS) établissent des liaisons covalentes entre les groupements amines primaires des protéines, présents sur les chaînes latérales des lysines. La concentration de l'agent de réticulation doit être optimisée afin d'être suffisamment élevée pour réagir avec une quantité importante de protéines mais pas trop élevée pour ne pas générer d'agrégats.

Le protocole détaillé ici est celui pour tester l'impact du glutaraldéhyde mais le même protocole a été suivi pour tester l'impact du BS<sup>3</sup> (ThermoFisher) et de l'EGS (Sigma). Des solutions stocks de glutaraldéhyde (Fluka) à 2 %, 0.5 % et 0.1 % ont été préparées et 1  $\mu$ L de chaque solution a été incubé avec 9  $\mu$ L de ClbK ( $\approx 0.15 \text{ mg.mL}^{-1}$ ) pendant 15 minutes à température ambiante. La réaction a été arrêtée par ajout de bleu de Laemmli et par dénaturation thermique à 95°C. L'effet du glutaraldéhyde sur ClbK avec ou sans sel a été visualisé sur gel NuPage 3-8 % Tris-Acétate (ThermoFisher). Une fois la condition optimale de glutaraldéhyde définie, l'expérience a été répétée et modifiée dans le but de visualiser l'échantillon réticulé en microscopie électronique. Pour cela, les volumes ont été augmentés et la réaction de réticulation a été arrêtée par ajout de Tris pH 7.5, à une concentration finale de 50 mM.

#### IV.5. Expression et lyse des cultures de fragments de ClbK

Le protocole décrit ici pour l'expression et la lyse de ClbK-PKS est valable aussi pour les protéines ClbK-PKS-LK, ClbK-PKS-Cy et ClbK-PKS-A<sub>core</sub>. Il tolère certaines variations, concernant notamment le milieu d'expression et les volumes de lyse. Le plasmide pETM-11-*clbK-PKS* a été transformé dans la souche BL21(DE3) Bap1. Une préculture a été lancée à partir d'une colonie isolée et le lendemain, une culture a été lancée en milieu d'auto-induction à partir de la préculture diluée à une DO<sub>600</sub> d'environ 0.05. Généralement, un erlenmeyer de deux litres à baffles a été utilisé pour une culture de 500 mL. La culture a été maintenue à 37°C jusqu'à ce que la DO<sub>600</sub> ait atteint 0.5 puis la culture a été transférée dans un incubateur à 18°C. Le lendemain, après 20-24 heures d'incubation à 18°C, les cellules ont été centrifugées (20 minutes, 7800 g).

Les culots ont été dissous dans du tampon de lyse (50 mM Hepes pH 7.5, 300 mM NaCl, 5 mM DTT). Le volume de tampon de lyse a été fixé à 20 mL pour 5 grammes de culot. 1 µL de nucléase Benzonase (Sigma) et un inhibiteur de protéase (cOmplete EDTA Free, Roche) ont été ajoutés par demi-litre de culture. Les cellules ont ensuite été lysées par quatre passages dans un broyeur cellulaire Constant Systems, à 20-22-25-25 kpsi.

#### IV.6. Expression de la forme sélénométhionée de la protéine ClbK-PKS

Le plasmide pETM-11-*clbK-PKS* a été transformé dans la souche BL21(DE3) Bap1. Une préculture en LB a été lancée à partir d'une colonie isolée et après huit heures, 2 mL de cette culture ont été dilués dans 50 mL de milieu minimum M9. Le lendemain, un litre de milieu minimum M9, pré-aéré et pré-incubé à 37°C sur la nuit, a été inoculé avec 20 mL de la préculture en milieu minimum. Lorsqu'une DO<sub>600</sub> de 0.5 a été atteinte, la synthèse de méthionine a été inhibée par un mélange d'acides aminés (100 mg lysine, 100 mg phénylalanine, 100 mg thréonine, 50 mg leucine, 50 mg isoleucine, 50 mg valine) et 60 mg de sélénométhionine ont été introduits dans la culture pour compléter la méthionine par de la sélénométhionine. Après 10 minutes supplémentaires d'incubation à 37°C, la culture a été transférée dans un incubateur à 18°C et poursuivie pendant 30 minutes avant l'induction de la protéine par 0.2 mM d'IPTG, à 18°C sur la nuit.

#### IV.7. Purification de ClbK-PKS

Le protocole décrit ici est celui pour 500 mL de culture et toutes les étapes ont été réalisées à 4°C. Ce protocole a été utilisé pour la purification de ClbK-PKS, ClbK-PKS sous forme séléénométhionée et ClbK-PKS-LK. Les protéines solubles ont été isolées par centrifugation (40 minutes, 39000 *g*) et incubées avec 1.5 g de résine Protino-Ni-IDA (Macherey-Nagel) pendant au moins 1h30. Le surnageant a été retiré et la résine a été lavée d'abord avec du tampon de lyse puis du tampon de lyse supplémenté avec 10 mM d'imidazole et du tampon de lyse supplémenté avec 20 mM d'imidazole. Les protéines ont été éluées avec du tampon de lyse supplémenté avec 250 mM d'imidazole en trois étapes (8, 7 et 6 mL). La protéine ClbK-PKS a ensuite été purifiée par chromatographie d'exclusion de taille avec la colonne HiLoad 16/600 Superdex 200 pg (GE Healthcare) pré-équilibrée dans le tampon 25 mM Hepes pH 7.5, 150 mM NaCl, 5 mM DTT.

Pour la cristallographie, l'étiquette poly-histidines de la protéine ClbK-PKS a été clivée, générant ainsi la protéine ClbK-PKS\*. La protéine ClbK-PKS après chromatographie d'affinité a été soumise à un clivage par la protéase TEV. La quantité de protéines dans l'élution de chromatographie d'affinité a été estimée par la méthode du Bradford puis l'élution a été incubée sur la nuit avec la protéase TEV dans un ratio 1 : 100 en masse, typiquement 0.5 mg de protéase TEV pour 50 mg de protéine à digérer. Le mélange a été dialysé contre le tampon de lyse sur la nuit puis incubé avec 2 g de résine Ni-IDA pendant deux heures. La résine a ensuite été transférée dans une colonne de gravité et la fraction non retenue a été récupérée. La résine a été lavée par du tampon de lyse, du tampon de lyse contenant 10 mM imidazole (12 mL) et du tampon de lyse contenant 20 mM Imidazole (2 x 12 mL). L'analyse par Western Blot révélé avec un anticorps anti-histidines a permis l'identification des fractions contenant ClbK-PKS\*. Ces fractions ont été concentrées à 5 mL et injectées sur une colonne HiLoad 16/600 Superdex 200 pg (GE Healthcare) pré-équilibrée dans le tampon 25 mM Hepes pH 7.5, 50 mM NaCl, 5 mM DTT.

#### IV.8. Purification de ClbK-PKS-Cy et ClbK-PKS-A<sub>core</sub>

1.5 litres de culture ont été lysés puis soumis à une chromatographie d'affinité selon le même protocole que celui décrit pour ClbK-PKS. Les fractions d'élution ont ensuite été dialysées contre un tampon Hepes 50 mM pH 7.5, 5 mM DTT puis soumises à une étape

supplémentaire de chromatographie échangeuse d'anions par injection sur une colonne Mono Q 5/50 GL (GE Healthcare) pré-équilibrée dans le tampon de dialyse. Les protéines ont été éluées par un gradient de chlorure de sodium. Les fractions contenant la protéine d'intérêt sur gel SDS ont été injectées sur une colonne d'exclusion de taille Superdex 200 Increase 10/300 GL (GE Healthcare) ou HiLoad 16/600 Superdex 200 pg (GE Healthcare) pré-équilibrée dans le tampon 25 mM Hepes pH 7.5, 150 mM NaCl, 5 mM DTT.

## V. Caractérisation biochimique et biophysique de ClbK et de ses fragments

### V.1. Caractérisation par spectrométrie de masse en conditions dénaturantes

La spectrométrie de masse est une méthode de caractérisation donnant accès à la masse exacte d'une protéine. Elle permet notamment de vérifier l'identité de la protéine et son état de modification post-traductionnelle. L'échantillon est d'abord ionisé puis les ions générés sont séparés selon leur rapport masse/charge ( $m/z$ ). La déconvolution du spectre de masse obtenu permet d'obtenir la masse exacte de la protéine dénaturée. Les analyses ont été effectuées à la plate-forme de spectrométrie de masse de l'IBS par Luca Signor.

L'instrumentation utilisée pour les analyses est composée d'un instrument TOF LS/MS (6210, Agilent) couplé à un capillaire, un système nano-HPLC (1110, Agilent) et d'une source d'ionisation électrospray (HPLC-Chip, Agilent). Le tampon de purification de la protéine contenant des composés non volatils, l'échantillon protéique est donc d'abord dessalé dans un mélange eau/acétonitrile (95/5) contenant 0.03 % d'acide trifluoroacétique. L'échantillon est ionisé par électro-nébulisation (de l'anglais « electropray ionisation », ESI). Au cours de cette étape, l'échantillon passe d'une phase liquide à une phase gazeuse. Les ions sont ensuite accélérés par un champ électrique et séparés en fonction de leur rapport masse/charge, qui est mesuré par un analyseur à temps de vol (de l'anglais « time of flight », TOF). En effet, la mesure du temps mis par les ions pour atteindre le détecteur donne accès à leur vitesse, qui dépend de leur rapport  $m/z$ . Le logiciel MassHunter Acquisition (Agilent) permet l'acquisition des spectres de masse, qui sont ensuite déconvolués par le logiciel MassHunter Qualitative Analysis (Agilent).

### V.2. Caractérisation par diffusion de lumière laser multi-angles couplée à une chromatographie d'exclusion de taille (SEC-MALLS)

La spectrométrie de masse en conditions dénaturantes ne renseigne pas sur l'état oligomérique d'une protéine. Le volume d'élution d'une protéine sur colonne d'exclusion de taille est indicatif de la masse molaire de la protéine et donc de son état oligomérique mais ce n'est pas un indicateur absolu car il dépend notamment de la forme de la protéine. Une méthode pour obtenir la masse molaire absolue d'un échantillon est la diffusion de lumière laser multi-angles couplée à la chromatographie d'exclusion de taille (SEC-MALLS). En effet,



l'intensité de la lumière diffusée par une protéine est directement proportionnelle à sa masse molaire et à sa concentration.

La diffusion de la lumière par une solution protéique est décrite par la relation de Rayleigh :

$$\Delta LS = \left(\frac{I_{\theta}}{I_0}\right)_{solution} - \left(\frac{I_{\theta}}{I_0}\right)_{tampon} = K \left(\frac{dn}{dc}\right)^2 MW C$$

« LS » désigne l'intensité de lumière diffusée, «  $\left(\frac{I_{\theta}}{I_0}\right)$  » est le ratio de l'intensité de la lumière diffusée à l'angle  $\theta$  ( $I_{\theta}$ ) sur l'intensité de la lumière incidente ( $I_0$ ), « K » est une constante et «  $\left(\frac{dn}{dc}\right)$  » est l'incrément de l'indice de réfraction qui vaut 0.185 mL.g<sup>-1</sup> pour les protéines. Les deux inconnues de cette équation sont les variables « MW » et « C » qui désignent la masse molaire et la concentration de l'échantillon protéique.

La concentration protéique « C » est reliée à la différence d'indice de réfraction «  $\Delta RI$  » entre l'échantillon et le tampon par la relation :

$$\Delta RI = \left(\frac{dn}{dc}\right) C$$

La combinaison des deux équations précédentes permet de définir la relation ci-dessous.

$$MW = \frac{\Delta LS}{K \left(\frac{dn}{dc}\right)^2 \frac{\Delta RI}{\left(\frac{dn}{dc}\right)}} = \frac{\Delta LS}{K \left(\frac{dn}{dc}\right) \Delta RI}$$

Dans une expérience de SEC-MALLS, la différence d'intensité de lumière diffusée «  $\Delta LS$  » est mesurée à différents angles par le détecteur MALLS (Figure 33). La différence d'indice de réfraction «  $\Delta RI$  » entre une chambre de référence contenant du tampon et la chambre d'échantillon contenant l'élution de la colonne d'exclusion de taille est mesurée par le réfractomètre. Un dispositif expérimental comportant un détecteur MALLS et un réfractomètre permet donc de déduire la masse molaire d'un échantillon pur élué sur colonne de filtration sur gel (Figure 33). L'évolution de la masse molaire de la protéine en fonction du volume d'élution renseigne sur sa monodispersité. Il faut noter ici qu'à cause de la relation de proportionnalité entre l'intensité de la lumière diffusée et la masse molaire de l'échantillon, les agrégats, même en très faible quantité, émettent un signal de diffusion de lumière important.

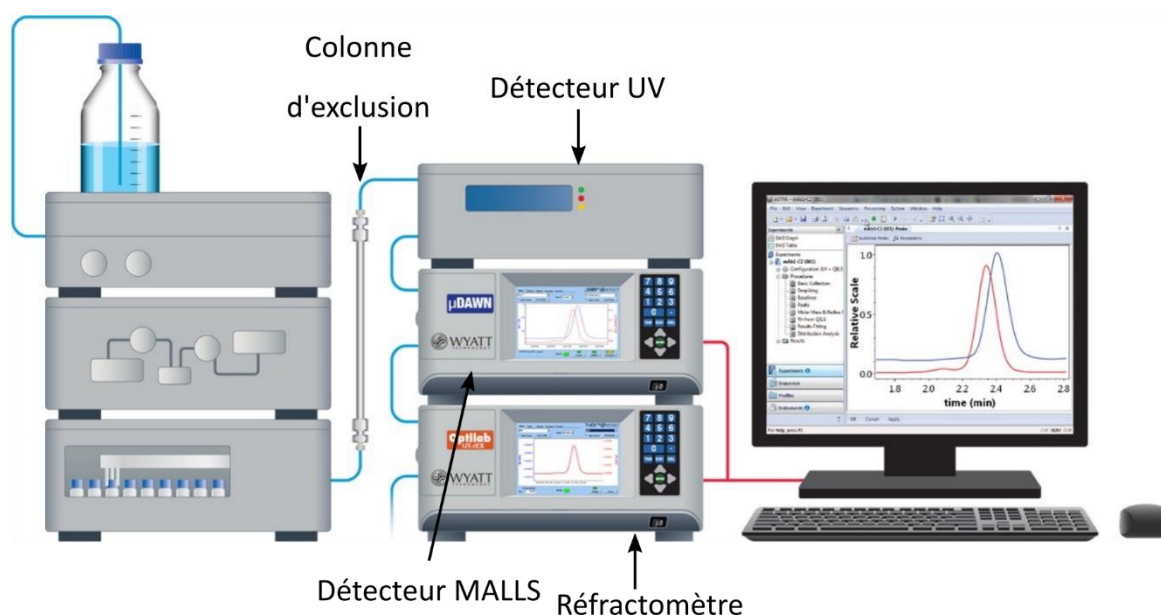


Figure 33. Dispositif expérimental des expériences de SEC-MALLS. Adapté du site du fournisseur Wyatt (<https://www.wyatt.com/>).

Les analyses de SEC-MALLS ont été effectuées à la plate-forme de biophysique de l'ISBG (« Integrated Structural Biology Grenoble »), avec l'aide de Caroline Mas (IBS), à l'exception de l'analyse de la protéine ClbK-PKS-Cy, qui a été effectuée par Aline Le Roy (IBS) à l'IBS. Les échantillons ont été concentrés à au moins  $2 \text{ mg.mL}^{-1}$  et injectés sur une colonne Superdex 200 Increase 10/300 GL (GE Healthcare). Le tampon de filtration sur gel propre à chaque protéine a été utilisé pour équilibrer la colonne et toutes les expériences ont été menées à température ambiante. L'instrumentation de la plate-forme de l'ISBG est constituée d'un détecteur MALLS Wyatt Dawn Heleos II, d'un réfractomètre Wyatt Optilab T-Rex, d'un détecteur UV Hitachi Elite LaChrom L-2400 et d'une pompe Hitachi Elite LaChrom L-2130. Les données ont été analysées avec le logiciel ASTRA 6.

### V.3. Caractérisation par photométrie de masse

La photométrie de masse est une autre méthode d'analyse de l'état oligomérique d'un échantillon protéique (Young *et al.*, 2018). Cette méthode permet d'analyser individuellement chaque particule et ainsi d'estimer directement la distribution de masse dans un échantillon, sans couplage avec une chromatographie d'exclusion. Le signal provenant de chaque particule individuelle étant mesuré, la quantité d'échantillon utilisée à chaque mesure est très faible, de l'ordre de 10-100 nM.

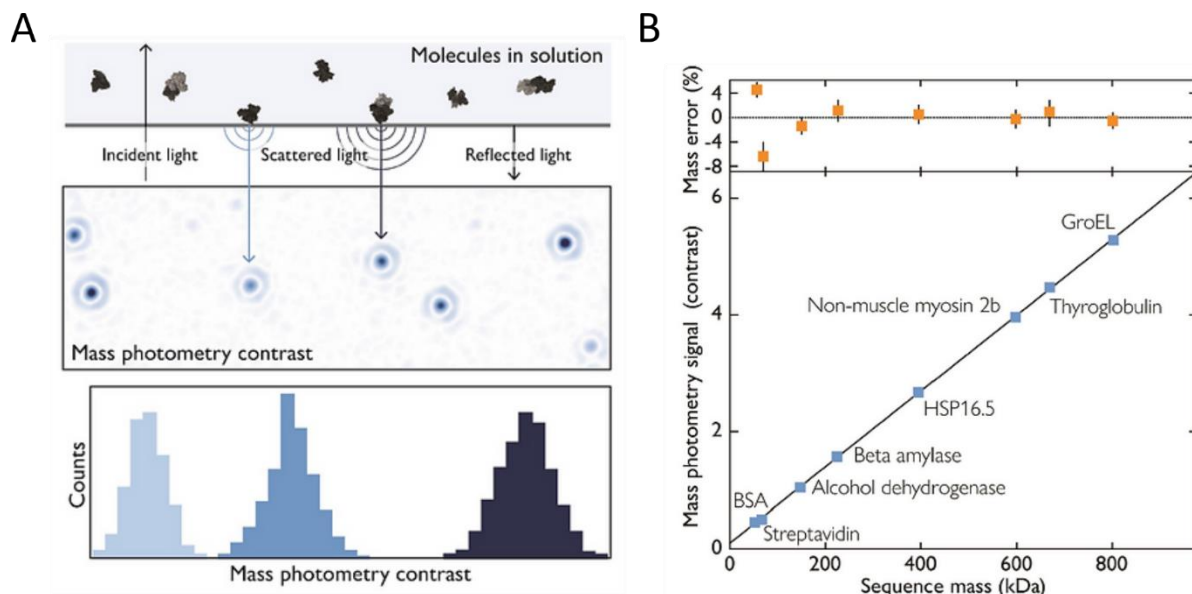


Figure 34. Principe de la photométrie de masse. A) Au cours d'une expérience de photométrie de masse, le signal de contraste des particules à l'interface eau-verre est enregistré. B) Le signal de contraste est directement proportionnel à la masse molaire de la protéine. D'après le site du fournisseur Refeyn (<https://www.refeyn.com/>).

Comme la technique du SEC-MALLS, cette technique est basée sur la diffusion de lumière mais plus précisément sur la diffusion à l'interface eau-verre (Figure 34A). La lumière diffusée par une particule à l'interface eau-verre interfère avec la lumière réfléchiée par la surface et le signal d'interférence est proportionnel à la masse molaire de la particule. L'échantillon est déposé sur une lamelle en verre et un film est enregistré à partir des images de la lamelle prises à différents temps. Un premier film est enregistré en présence de tampon sur la lamelle, il permet d'ajuster le focus. Lorsque l'échantillon protéique est mélangé au tampon, chaque événement de contact des particules avec la lamelle (en anglais « binding event ») est enregistré sous forme d'histogramme de contraste de diffusion. Ce signal de contraste est proportionnel à la masse molaire de la particule donc plus une particule est grosse, plus le contraste de diffusion est élevé. Une calibration avec un mélange de protéines de masse connue permet d'estimer des masses molaires à partir des signaux de contraste (Figure 34B).

Les expériences de photométrie de masse ont été effectuées à la plate-forme de biophysique de l'ISBG, sur l'instrument Refeyn One<sup>MP</sup>, avec l'aide de Caroline Mas. Les films ont été enregistrés avec le logiciel AcquireMP et analysés avec le logiciel DiscoverMP.

## VI. Caractérisation de l'activité enzymologique du domaine d'adénylation de ClbK

Le domaine d'adénylation (domaine A) des NRPS catalyse deux réactions (Figure 14A). Au cours de la réaction d'adénylation, le domaine A sélectionne un acide aminé et l'adényle grâce à l'hydrolyse de l'ATP, ce qui libère un groupement pyrophosphate inorganique (PP<sub>i</sub>). Au cours de la réaction de transfert, le domaine A charge l'acide aminé ainsi activé sur le groupement thiol du bras phosphopantéthéine (PPant) du domaine PCP.

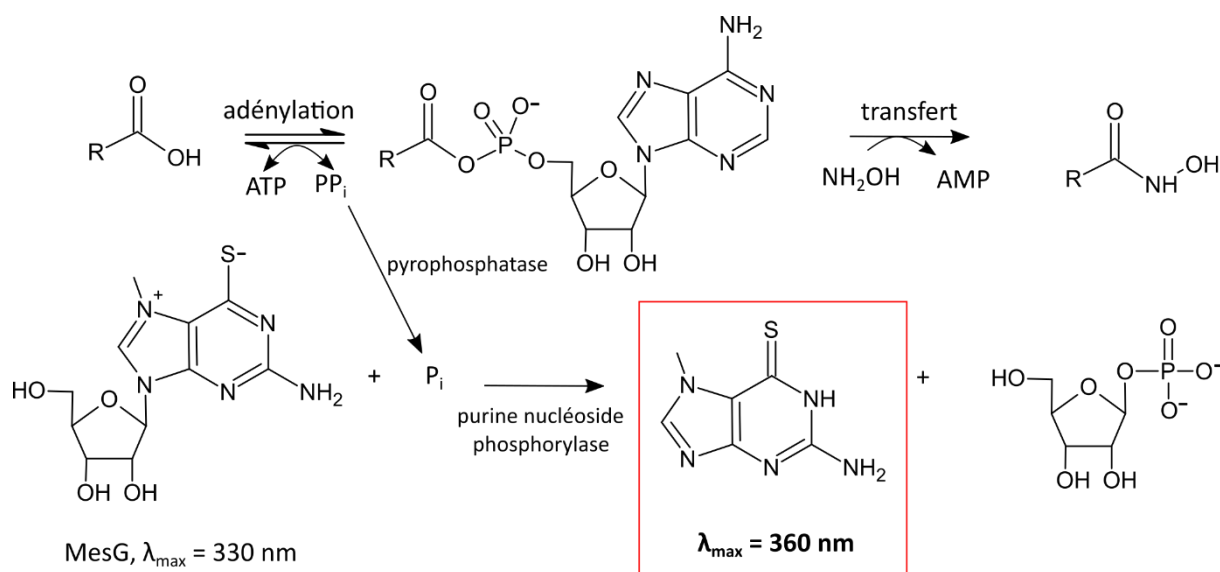


Figure 35. Mesure de l'activité d'adénylation du domaine d'adénylation (domaine A) par le dosage MesG-hydroxylamine. Le pyrophosphate inorganique (PP<sub>i</sub>) libéré au cours de la réaction d'adénylation est dosé par le 7-méthyl-6-thioguanosine (MesG). Adapté de Duckworth *et al.*, 2016.

Le dosage de l'activité d'adénylation établi par Duckworth *et al.* (Duckworth *et al.*, 2016) repose sur le dosage indirect du PP<sub>i</sub> libéré au cours de la réaction d'adénylation (Figure 35). Grâce à l'enzyme pyrophosphatase (IP), le PP<sub>i</sub> est hydrolysé en phosphate inorganique (P<sub>i</sub>) qui réagit avec la molécule 7-méthyl-6-thioguanosine (MesG) en présence de l'enzyme purine nucléoside phosphorylase (PNP) pour former notamment le composé 7-méthylthioguanine, absorbant la lumière à 360 nm. L'hydroxylamine se substitue au domaine PCP en acceptant le substrat activé.

Ce test d'activité du domaine d'adénylation a été appliqué à la protéine ClbK entière. Bien que ClbK possède un domaine PCP qui accepte théoriquement le substrat activé, l'hydroxylamine a été conservée dans l'essai avec ClbK afin qu'un des substrats de la réaction de transfert soit maintenu en excès. Le tampon de réaction contient du chlorure de

magnésium stabilisant l'ATP. Des solutions stock à 100 mM de chacun des vingt acides aminés ont été préparés par solubilisation dans du tampon de réaction, à l'exception des acides aminés acide aspartique, acide glutamique, asparagine, tryptophane et tyrosine, qui ont été solubilisés dans du tampon 1 M HCl dont le pH a ensuite été neutralisé par ajout de soude. Les solutions stocks 100 mM ATP (Sigma, solubilisé dans 1 M Hepes pH 7.5), 1 mM MesG (ChemCruz, solubilisé dans le DMSO), 100 U/mL de l'enzyme IP (ThermoFisher), 100 U/mL de l'enzyme PNP (Sigma, solubilisée dans H<sub>2</sub>O), 2M d'hydroxylamine (Sigma, solubilisée dans 2 M NaOH) ont été préparées. Les tests ont été effectués dans une plaque 96 puits à fond plat adaptée aux UV (Greiner) et le volume de réaction était de 100  $\mu$ L.

Plusieurs tests ont permis d'établir l'ordre optimal d'ajout des réactifs pour diminuer le signal parasite. L'acide aminé a d'abord été déposé dans le puits puis un mélange contenant le tampon, l'ATP, le MesG, l'hydroxylamine, les enzymes IP et PNP a été ajouté. L'ajout de ClbK a déclenché la réaction. La concentration de ClbK a été optimisée et la concentration finale de 40 nM a été retenue. Le mélange final contient donc 100 mM Hepes pH 7.5, 200 mM NaCl, 5 mM DTT, 5 mM MgCl<sub>2</sub>, 5 mM ATP, 0.1 mM MesG, 0.04 U IP, 0.1 U PNP, 150 mM hydroxylamine, 5 mM acide aminé et 40 nM ClbK. Chaque réaction a été réalisée en triplicat. L'absorbance à 360 nm a été enregistrée pendant une vingtaine de minutes.

Pour l'analyse, la pente de la zone linéaire de chaque courbe  $A_{360 \text{ nm}} = f(\text{temps})$  a été mesurée. Les valeurs de pentes négatives ont été considérées comme nulles. Les valeurs de chaque triplicat ont été moyennées et un contrôle négatif a indiqué que le signal en absence d'acide aminé était négligeable. Les valeurs de pente ( $\text{AU.s}^{-1}$ ) ont été converties en vitesse initiale de réaction ( $\mu\text{M.min}^{-1}$ ) selon le protocole décrit par Duckworth *et al.*. La pente en  $\text{AU.s}^{-1}$  a d'abord été multipliée par 60 pour être convertie en  $\text{ua.min}^{-1}$ . Les pentes ont ensuite été normalisées par le chemin optique estimé à 0.29 cm d'après les dimensions du puits (valeur de pente alors en  $\text{AU.cm}^{-1}.\text{min}^{-1}$ ). La valeur a ensuite été divisée par le coefficient d'extinction du 7-méthylthioguanine ( $\epsilon = 11000 \text{ M}^{-1}.\text{cm}^{-1}$ ) et divisée par deux car deux molécules de 7-méthylthioguanine sont formées par molécule d'acide aminé adénylé (valeur de vitesse initiale en  $\text{M.min}^{-1}$ ). Cette valeur multipliée par  $10^6$  est la valeur finale de la vitesse initiale en  $\mu\text{M.min}^{-1}$ .

## VII. Introduction à la diffusion des rayons X en vue d'études structurales

Deux techniques de biologie structurale reposent sur l'interaction entre les rayons X et les électrons des protéines : la cristallographie aux rayons X et la diffusion aux petits angles (SAXS). Pour visualiser des objets de taille  $d$ , un rayonnement de longueur d'onde inférieure à cette distance  $d$  est nécessaire. Les rayons X ayant une longueur d'onde de l'ordre l'ångström (Å), ils sont utilisés pour étudier la structure des protéines à l'échelle atomique puisque les distances interatomiques sont également de l'ordre de l'ångström (Putnam *et al.*, 2007). Les électrons de la protéine diffusent les rayons X de façon élastique, c'est-à-dire sans perte d'énergie entre l'onde incidente et l'onde diffusée. La différence principale entre ces deux techniques est l'état de la protéine, qui est sous forme cristalline dans un cas et en solution dans l'autre cas. Dans le contexte de la diffusion par un cristal, les ondes diffusées interfèrent entre elles de façon constructive ou destructive, on parle alors de diffraction, qui est un cas particulier de diffusion. Les ondes diffusées ou diffractées appartiennent à l'espace réciproque. Dans cet espace, les informations aux petits angles renseignent sur les grandes distances et les informations aux grands angles renseignent sur les petites distances. Le passage de l'espace réel à l'espace réciproque est une opération mathématique appelée transformée de Fourier. La déduction d'informations structurales à partir des données de diffusion/diffraction, c'est-à-dire le passage de l'espace réciproque à l'espace réel, est permis par la transformée de Fourier inverse.

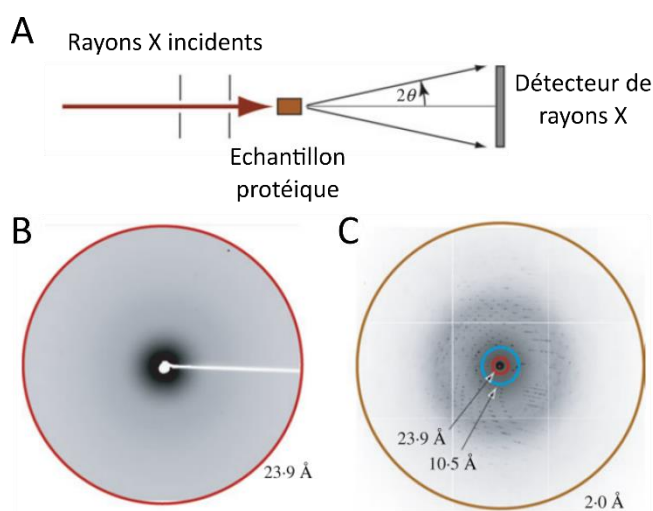


Figure 36. Interaction des rayons X avec l'échantillon protéique au cours d'une expérience de SAXS ou de cristallographie. A) Dispositif expérimental. B) Cliché de diffusion enregistré lors d'une expérience de SAXS. C) Cliché de diffraction enregistré lors d'une expérience de cristallographie aux rayons X. Le cercle rouge de la figure C correspond au cercle rouge de la figure B. Adapté de Putnam *et al.*, 2007.

Lors d'une expérience de SAXS ou de cristallographie aux rayons X, un échantillon est irradié par des rayons X et le signal de diffusion généré est enregistré sur un détecteur (Figure 36A). Lors d'une expérience de SAXS, l'échantillon est en solution et le signal provenant de chaque particule est moyenné, générant des signaux de faible intensité, qui ne diffusent qu'à des petits angles (Figure 36B). Le cliché de diffusion enregistré sur le détecteur est constitué de cercles concentriques d'intensité variable. A l'inverse, le cliché de diffraction d'un cristal protéique illuminé par des rayons X est un cliché en deux dimensions avec des tâches de diffraction correspondant aux zones d'interférences constructives (Figure 36C). Une rotation du cristal permet de collecter une série de clichés de diffraction en deux dimensions et ainsi de reconstituer un signal en trois dimensions. Le cristal, composé d'un très grand nombre de molécules protéiques, joue un rôle d'amplificateur du signal. Le signal de diffraction est donc enregistré à grands angles, ce qui correspond à des signaux de haute résolution.

## VIII. Caractérisation structurale par cristallographie aux rayons X

VIII.1. Tests de stabilité de la protéine dans différents tampons par essai Thermofluor

La température  $T_m$  est un indicateur de la stabilité des protéines et de leur probabilité de cristallisation. Au cours d'une expérience de dénaturation par augmentation de température, la température caractéristique  $T_m$  est la température à laquelle la moitié de la protéine est sous forme dépliée alors que l'autre moitié est repliée. Cette température peut être déterminée grâce à la technique du Thermofluor (Pantoliano *et al.*, 2001), utilisant un colorant qui fluoresce uniquement au contact des régions hydrophobes des protéines, généralement le SYPRO Orange (Figure 37). Dans une protéine repliée, le colorant n'a pas accès aux régions hydrophobes car elles sont enfouies au cœur de la protéine. Au cours d'une expérience de dénaturation thermique, l'augmentation de la température conduit d'abord à une augmentation de la fluorescence, traduisant la dénaturation de la protéine. Après avoir atteint un maximum, la fluorescence diminue, traduisant la précipitation de la protéine. Le  $T_m$  est calculé sur la phase de dénaturation, en lisant la température à laquelle la fluorescence est intermédiaire entre la fluorescence initiale et la fluorescence maximale.

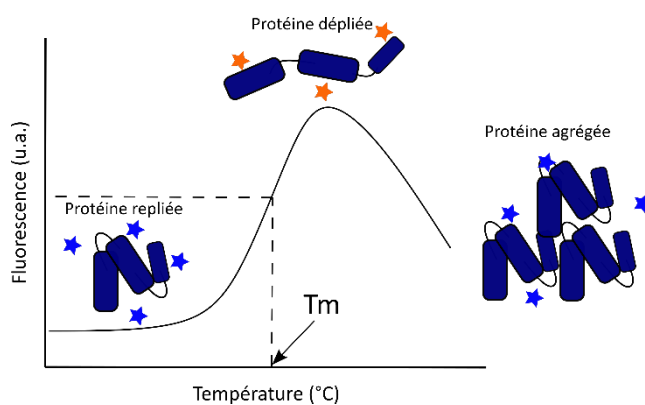


Figure 37. Schéma d'une expérience de Thermofluor. La température caractéristique  $T_m$  est calculée sur la phase d'augmentation de la fluorescence, correspondant à la phase de dénaturation de la protéine.

La plate-forme HTX (« High Throughput Crystallization », ISBG Grenoble) a développé une méthode basée sur la technique du Thermofluor afin de tester l'effet de différents tampons sur le  $T_m$  de la protéine (Dupeux *et al.*, 2011) et donc sur sa stabilité. Le kit RUBIC buffer screen (Molecular Dimensions) permet de tester l'effet de la nature du tampon, de sa concentration, du pH et de la concentration en sel sur le  $T_m$  de la protéine. Une fois le tampon optimal déterminé, le kit RUBIC additives (Molecular Dimensions) teste l'effet d'additifs tels que le sel, les ions, les détergents sur la température  $T_m$  de la protéine.



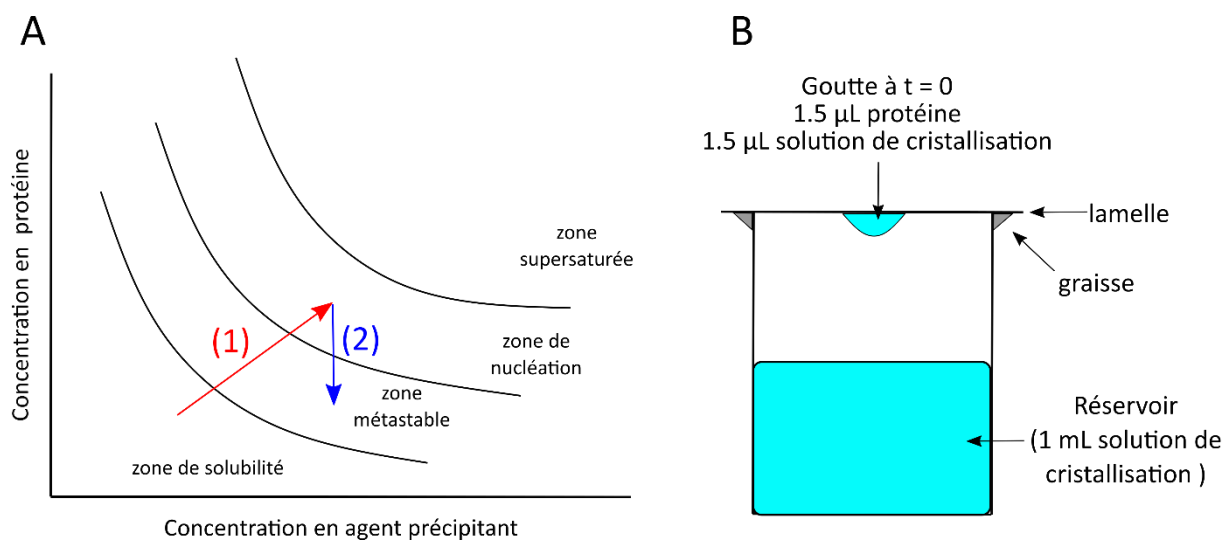
VIII.2. Cristallogénèse

Figure 38. Cristallogénèse des protéines. A) Diagramme de phase en deux dimensions. La cristallogénèse se divise en deux phases : la nucléation (1) et la croissance (2). B) Schéma du dispositif expérimental employé pour la méthode de diffusion de vapeur en goutte suspendue.

L'étude structurale par cristallographie aux rayons X nécessite l'obtention de la protéine d'intérêt sous forme de cristal, c'est-à-dire d'un réseau tri-dimensionnel ordonné. La cristallisation nécessite le passage de la protéine en solution sous forme solide et se divise en deux étapes : la nucléation puis la croissance (Figure 38A). La méthode de diffusion de vapeur en goutte suspendue a été utilisée pour les tests de cristallisation (McPherson *et al.*, 1995) (Figure 38B). Une goutte contenant la protéine et la solution de cristallisation est déposée sur une lamelle qui est ensuite retournée et scellée hermétiquement au-dessus d'un réservoir contenant un grand volume de la même solution de cristallisation. Des échanges de vapeur s'établissent alors entre la goutte et le réservoir. La solution de cristallisation étant moins concentrée dans la goutte que dans le réservoir, l'eau s'évapore préférentiellement de la goutte vers le réservoir, conduisant à une augmentation de la concentration en protéine et en agent précipitant dans la goutte. Ce phénomène diminue la solubilité de la protéine jusqu'à atteindre des conditions où la nucléation est possible (Figure 38A). La cristallisation conduit à une diminution de la concentration protéique dans la solution de la goutte. L'apparition de nouveaux cristaux n'est plus possible mais les cristaux existants peuvent croître dans la zone métastable. Les premiers tests de cristallisation ont été menés par la plate-forme HTX (ISBG Grenoble). Les meilleures conditions ont été reproduites manuellement selon la méthode décrite précédemment en plaques 24 puits.

### VIII.3. Congélation des cristaux et enregistrement des données de diffraction

Les cristaux obtenus ont été pêchés individuellement à l'aide de boucles puis trempés dans une solution cryoprotectrice avant d'être vitrifiés dans l'azote liquide. La congélation très rapide par vitrification est nécessaire pour limiter la formation de glace, qui détériore le réseau cristallin. Maintenir l'échantillon à très basse température permet de limiter le phénomène de radiolyse, c'est-à-dire la génération de radicaux libres à partir des molécules de solvant présentes dans le cristal, pouvant l'endommager.

Les données de diffraction sur les cristaux ont été collectées grâce au diffractomètre de l'IBS, aux lignes de lumières ID23-1, ID23-2, ID30A-3 et ID30B de l'ESRF (European Synchrotron Radiation Facility, Grenoble). La diffraction d'une centaine de cristaux a également été testée grâce à la ligne de lumière entièrement automatique ID30A-1 de l'ESRF (Bowler *et al.*, 2015).

Les données de diffraction à plus haute résolution sur la protéine ClbK-PKS\* ont été collectées sur la ligne ID30A-1. Le cristal avait été obtenu à température ambiante en mélangeant 1.5  $\mu\text{L}$  de la solution de cristallisation (0.1 M MES pH 6.5, 19 % PEG 3350) avec 1.5  $\mu\text{L}$  de la protéine ClbK-PKS\* concentrée à 15  $\text{mg}\cdot\text{mL}^{-1}$  et conservée à 4°C pendant deux semaines. Ce cristal avait été congelé après trempage dans une solution 0.1 M MES pH 6.5, 18 % PEG 3350, 25 % glycérol.

### VIII.4. Traitement des données de diffraction

Le traitement des données de diffraction et la résolution de la structure de ClbK-PKS\* ont été effectués par Pauline Macheboeuf. Le traitement des données se divise en plusieurs étapes et a été réalisé avec la suite de programmes XDS (Kabsch, 2010). A la première étape d'indexation, la répartition des tâches de diffractions est analysée et les paramètres de maille du cristal sont déterminés. La maille est l'unité élémentaire du cristal qui se répète par translation. A la deuxième étape d'intégration, les intensités des tâches de diffraction sont déterminées. Ces intensités sont ensuite mises à l'échelle puis moyennées afin de comparer des réflexions équivalentes par symétrie. A cette étape, la qualité des données est estimée.

### VIII.5. Phasage et affinement de la structure

Les ondes diffusées par le cristal sont décrites par une amplitude et une phase. L'information d'amplitude est déduite de l'intensité des tâches de diffraction. En revanche, l'information de phase est perdue au cours de l'expérience de diffraction : c'est le problème de la phase. L'information de phase est nécessaire afin de calculer la carte de densité électronique par transformée de Fourier inverse. Le problème de la phase peut être résolu par phasage expérimental ou par remplacement moléculaire.

Une méthode de phasage expérimental est le phasage par diffusion anormale. La position des diffuseurs anormaux dans le cristal peut être déterminée grâce à l'excitation du cristal à une longueur d'onde particulière, proche du seuil d'absorption du diffuseur anormal. Cette technique nécessite l'obtention de cristaux de la protéine d'intérêt en présence d'un diffuseur anormal. Une première stratégie de cristallisation de la protéine ClbK-PKS sous forme sélénométhionée a été testée. Une deuxième stratégie de co-cristallisation avec deux cristallophores, fournis par Eric Girard (IBS), a été investiguée (Engilberge *et al.*, 2017).

Le problème de la phase peut également être résolu par remplacement moléculaire (Dodson, 2021). Cette méthode nécessite une structure d'une protéine homologue ayant une identité de séquence supérieure à 30 %. Le programme Phaser (McCoy *et al.*, 2007) a permis la résolution du problème de la phase en utilisant le modèle de ClbK-PKS sans l'ACP généré par AlphaFold2 (Jumper *et al.*, 2021). La structure a ensuite été complétée grâce aux programmes de la suite CCP4 (Winn *et al.*, 2011) par des cycles de reconstruction manuelle par Coot (Emsley & Cowtan, 2004), d'optimisation de phase par PARROT (Cowtan, 2010) et de reconstruction automatique par BUCCANEER (Cowtan, 2006). Plusieurs cycles de reconstruction manuelle et d'affinement par REFMAC (Murshudov *et al.*, 2011) ont permis l'obtention de la structure cristallographique finale. La structure finale a été déposée sur le site de la PDB sous le code d'accèsion 7ZLB.

## IX. Caractérisation structurale par diffusion des rayons X aux petits angles (SAXS)

### IX.1. Dispositif expérimental et acquisition des données

Toutes les expériences de SAXS ont été effectuées sur la ligne de lumière BM29 de l'ESRF (European Synchrotron Radiation Facility, Grenoble), équipée d'un détecteur Pilatus2M, avec l'aide de Petra Pernot (ESRF) et Mark Tully (ESRF) (Pernot *et al.*, 2013). Les échantillons ont été illuminés par des rayons X à une longueur d'onde  $\lambda = 0.992 \text{ \AA}$  et le détecteur a été positionné à 2.83 m de l'échantillon. Cette ligne de lumière permet d'effectuer des mesures directes (en *batch*) ou couplées avec une chromatographie d'exclusion (SEC-SAXS). Pour une expérience de mesure directe, le signal de diffusion est mesuré sur deux échantillons : un échantillon protéique et l'échantillon de tampon correspondant. La protéine ClbK pure a été concentrée jusqu'à 50  $\mu\text{L}$  à 4  $\text{mg.mL}^{-1}$  avec un concentrateur Vivaspin 500 100 kDa (GE Healthcare) et le tampon élué à travers le concentrateur a été conservé comme échantillon tampon. Pour une expérience de SEC-SAXS, une colonne d'exclusion de taille est connectée à l'instrumentation du SAXS, ce qui permet d'éliminer de potentiels agrégats. Le signal de diffusion est mesuré pour toutes les fractions éluées de la colonne, à un taux d'acquisition d'une image par seconde. Pour l'expérience de SEC-SAXS, 70  $\mu\text{L}$  de la protéine ClbK pure, concentrée à 9  $\text{mg.mL}^{-1}$ , ont été injectés sur la colonne d'exclusion Superdex 200 Increase 10/300 GL, équilibrée dans le dernier tampon de purification, et la protéine a été éluée à un débit de 0.75  $\text{mL.min}^{-1}$ .

### IX.2. Soustraction du signal de diffusion du tampon

Toutes les analyses ont été effectuées avec les logiciels ATSAS 3.0.3 développés par le groupe de Dimitri Svergun (EMBL Hambourg) (Manalastas-Cantos *et al.*, 2021). Contrairement à la cristallographie aux rayons X, la technique de SAXS est une méthode de contraste : l'intensité de diffusion mesurée sur un échantillon protéique lors d'une expérience de SAXS résulte de la somme de la densité électronique de la protéine en elle-même et de celle du tampon. L'intensité du tampon doit être mesurée séparément à celle de la protéine dans son tampon et doit être soustraite à celle-ci afin d'en déduire l'intensité diffusée de la protéine en elle-même. Pour une expérience de SEC-SAXS, les images correspondant à l'échantillon et au tampon sont sélectionnées grâce au logiciel CHROMIXS (Panjkovich & Svergun, 2018). Le

logiciel PRIMUS est utilisé pour établir une moyenne des images du tampon et de l'échantillon et pour soustraire l'intensité du tampon à celle de l'échantillon (Konarev *et al.*, 2003).

Cette intensité après soustraction,  $I(s)$ , dépend du vecteur de diffusion  $s$ , en  $\text{nm}^{-1}$ , qui est défini par  $s = 4\pi \sin(\theta)/\lambda$ , avec  $2\theta$  l'angle incident décrit à la Figure 36A et  $\lambda$  la longueur d'onde (Mertens & Svergun, 2010). La courbe  $\log(I(s))$  en fonction de  $s$  après soustraction de l'effet du tampon est donc le signal caractéristique de diffusion de la protéine d'intérêt.

### IX.3. Analyse de Guinier

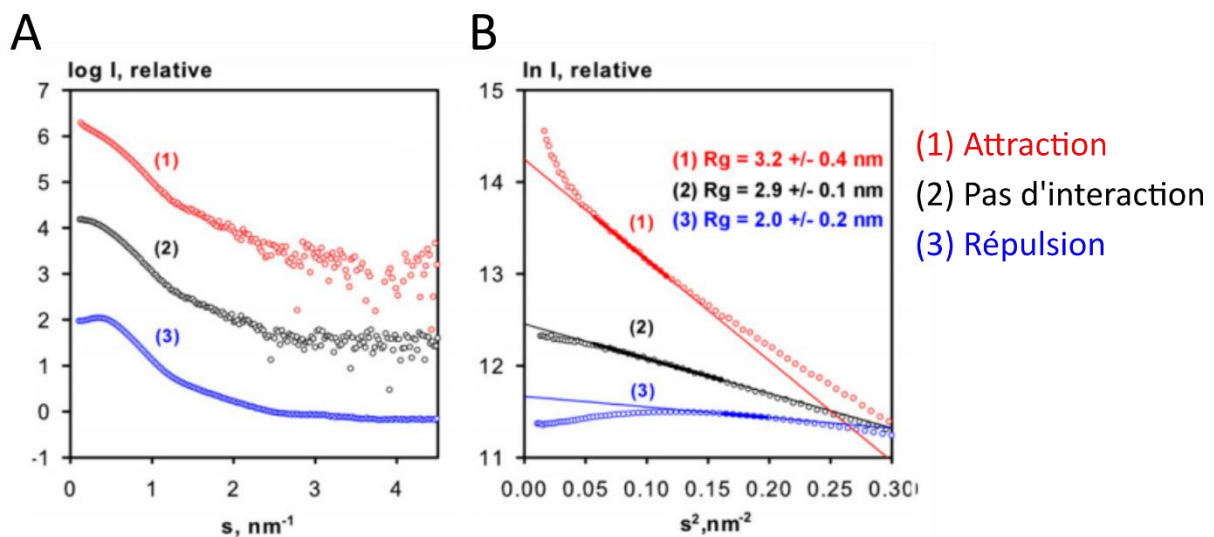


Figure 39. Courbes de diffusion (A) et représentations de Guinier correspondantes (B) d'échantillons agrégé (1), ne présentant pas d'interaction entre les particules (2) ou présentant de la répulsion entre particules (3). Adapté de Mertens & Svergun, 2010.  $R_g$  : rayon de giration.

A très petits angles et pour une protéine monodisperse, l'équation de Guinier est supposée valide :

$$I(s) = I(0) \exp\left(-\frac{1}{3} R_g^2 s^2\right)$$

Avec  $I(0)$  l'intensité diffusée à angle nul et  $R_g$  le rayon de giration.

En appliquant la fonction logarithme népérien à cette équation, on obtient :

$$\ln(I(s)) \simeq \ln(I(0)) - \frac{1}{3} R_g^2 s^2$$

La représentation de Guinier des données de diffusion ( $\ln(I(s))$  en fonction de  $s^2$ ), permet ainsi d'estimer la qualité de l'échantillon. Une courbe non linéaire est indicatrice de la répulsion ou l'attraction entre les particules (Figure 39). Une courbe linéaire pour des valeurs de  $sR_g < 1.3$

témoigne de l'absence d'interactions entre particules. La pente de cette droite est  $-1/3 Rg^2$  et l'ordonnée à l'origine est  $\ln(I(0))$  : les paramètres  $Rg$  et  $I(0)$  peuvent donc être extraits de cette courbe par régression linéaire. Le paramètre  $I(0)$  peut servir à l'estimation de la masse moléculaire de l'échantillon protéique mais nécessite une connaissance préalable de la concentration de l'échantillon, ce qui est difficile pour un échantillon étudié en SEC-SAXS.

#### IX.4. Fonction de distribution des distances

L'application de la transformée de Fourier inverse à la courbe de diffusion permet d'obtenir une fonction dans l'espace réel, la fonction de distribution des distances  $P(r)$  :

$$P(r) = \frac{r^2}{2\pi^2} \int_0^\infty s^2 I(s) \frac{\sin sr}{sr} ds$$

Cette fonction est une représentation des distances intramoléculaires de la protéine en solution et est un indicateur de la forme de la protéine (Figure 40). La dimension maximale de la protéine ( $D_{max}$ ) est lue à l'intersection de la courbe avec l'axe des abscisses. Par exemple, la fonction de distribution des distances d'une protéine globulaire est une gaussienne présentant un maximum à  $D_{max}/2$ . La fonction de distribution des distances est évaluée grâce au logiciel GNOM (Svergun, 1992).

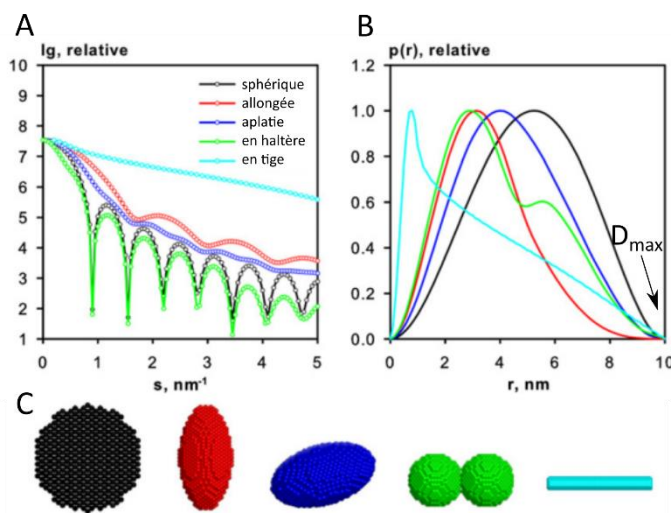


Figure 40. Courbes de diffusion (A) et fonctions de distribution des distances correspondantes (B) calculées pour des formes géométriques distinctes (C). Adapté de Mertens & Svergun, 2010.

Le rayon de giration  $Rg$  et l'intensité diffusée à angle nul  $I(0)$  peuvent être calculés grâce à la fonction de distribution des distances.

$$Rg^2 = \frac{\int_0^{D_{max}} r^2 P(r) dr}{2 \int_0^{D_{max}} P(r) dr}$$

$$I(0) = 4\pi \int_0^{D_{max}} P(r) dr$$

Ces calculs sont complémentaires à ceux effectués par l'approximation de Guinier et plus précis car ils utilisent l'ensemble de la courbe de diffusion et pas seulement la portion aux très petits angles.

#### IX.5. Graphique de Kratky

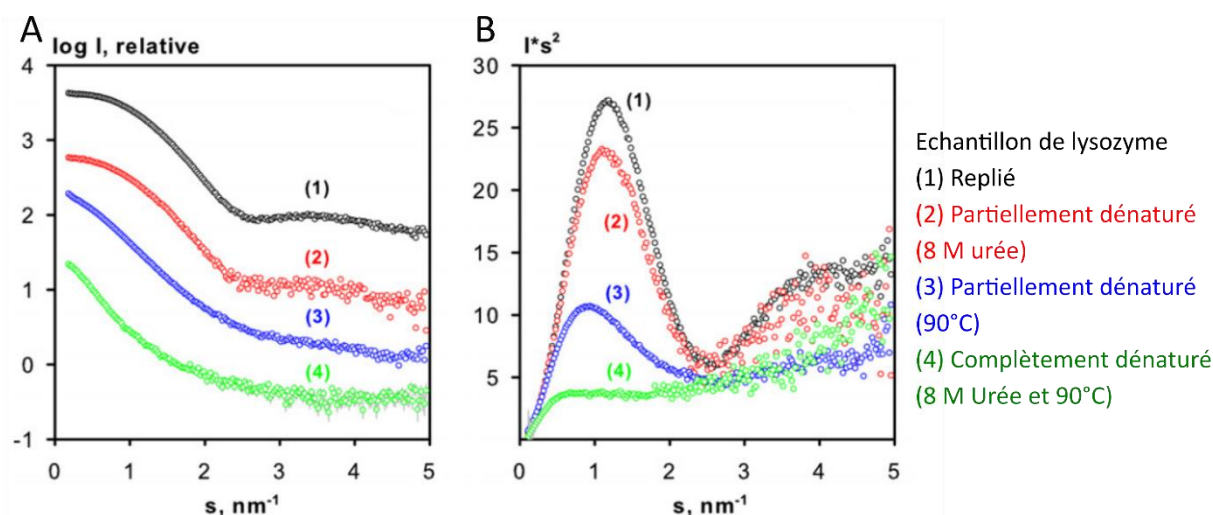


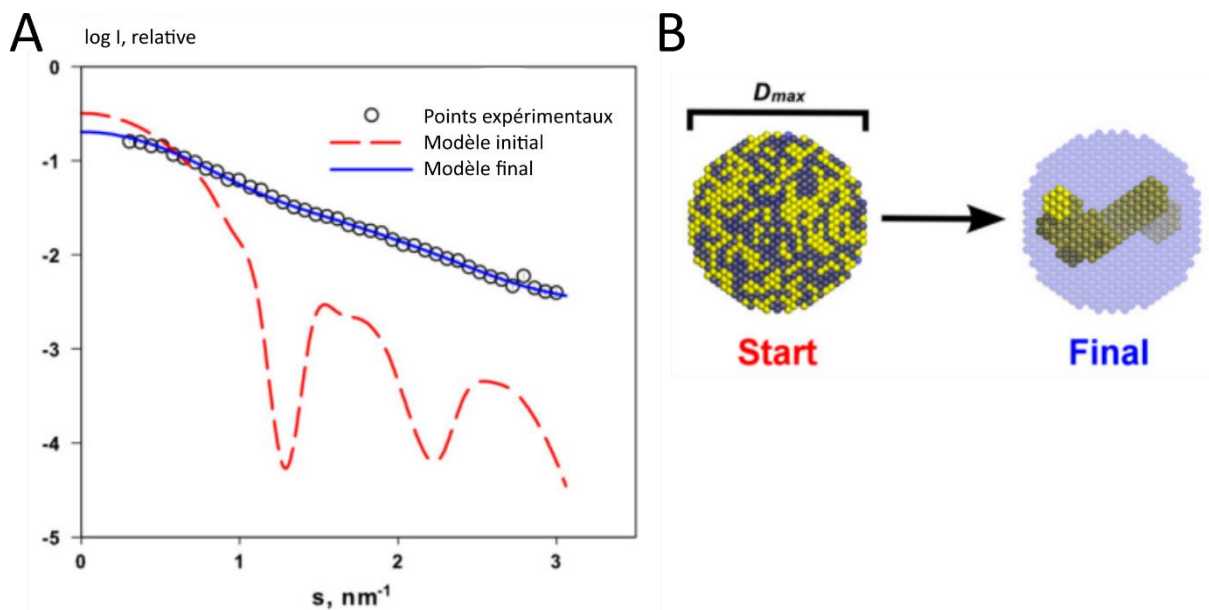
Figure 41. Représentation de Kratky des données de diffusion. Adapté de Mertens & Svergun, 2010. A) Courbes de diffusion d'échantillons de lysozyme à différents degrés de dénaturation. B) Représentations de Kratky de ces mêmes échantillons.

La représentation de Kratky ( $s^2I(s)$  en fonction de  $s$ ) est un indicateur de l'état de repliement de la protéine. Le graphique de Kratky des protéines repliées présente un pic bien défini à des faibles valeurs de  $s$  alors que le graphique des protéines dépliées ne présente pas ce pic caractéristique mais plutôt une augmentation continue des valeurs de  $I(s)s^2$  (Figure 41). La représentation de Kratky est effectuée grâce au logiciel PRIMUS.

#### IX.6. Modélisation *ab initio*

Les données SAXS sont généralement ambiguës, ce qui signifie que plusieurs modèles peuvent expliquer la courbe de diffusion expérimentale. L'ambiguïté des données pour un échantillon monodisperse peut être évaluée *a priori* grâce au programme AMBIMETER (Petoukhov & Svergun, 2015). Brièvement, une bibliothèque de profils de diffusion normalisés a été générée à partir de plus de 14000 formes topologiques. Le nombre de topologies générant des profils de diffusion similaires à la courbe expérimentale est un indicateur de l'ambiguïté des données. Un score inférieur à 1.5 indique une forte probabilité d'obtenir un

modèle unique expliquant les données alors qu'un score supérieur à 2.5 suggère qu'il est peu probable d'obtenir un modèle unique expliquant les données.



Des modèles en trois dimensions peuvent être générés à partir de la courbe de diffusion à une dimension, notamment grâce aux logiciels DAMMIN/DAMMIF (Svergun, 1999 ; Franke & Svergun, 2009). La génération d'un modèle par le programme DAMMIN repose sur un processus d'essai-erreur utilisant la méthode d'optimisation du recuit simulé (Figure 42). Ce processus débute à partir d'un volume initial, généralement sphérique, contraint par la dimension maximale  $D_{max}$  déterminée expérimentalement. Ce volume est occupé par des pseudo-atomes, assignés arbitrairement comme appartenant à la protéine ou au solvant. A chaque cycle, un nouveau modèle est généré par ré-assignement d'un pseudo-atome. Une contrainte de pénalité impose la continuité entre les pseudo-atomes de la protéine, afin de garantir la compaction des modèles. L'adéquation entre la courbe de diffusion théorique du modèle final et la courbe de diffusion des données expérimentales peut être estimée grâce au coefficient  $\chi^2$ . Un autre indicateur de la qualité du modèle est l'estimation de la masse molaire du modèle, qui peut être comparée à la masse molaire attendue. Le programme



DAMMIF, quarante fois plus rapide que DAMMIN, repose sur le même principe que DAMMIN, à l'exception du volume initial, qui n'est pas contraint.

Pour ClbK, vingt modèles ont été générés par le programme DAMMIF. Lorsque l'ambiguïté des données est élevée, il est conseillé de regrouper les modèles *ab initio* en différents *clusters* grâce au programme DAMCLUST (Petoukhov *et al.*, 2012). Les modèles sont d'abord tous comparés deux à deux et la similarité entre chaque couple de modèle est quantifiée grâce au paramètre NSD (« normalized spatial discrepancy »). Les modèles ayant un faible NSD sont ensuite regroupés en *clusters*. Pour les *clusters* non isolés, un modèle représentatif de chaque *cluster* est obtenu par moyennation grâce au programme DAMAVER (Volkov *et al.*, 2003).

Indépendamment du regroupement en *clusters*, un modèle *ab initio* affiné peut être généré à partir des vingt modèles préalablement générés (Volkov *et al.*, 2003). Les modèles sont d'abord comparés entre eux grâce à DAMSEL et les modèles présentant un NSD éloigné du NSD moyen sont écartés pour la suite de l'analyse. Les modèles conservés sont ensuite alignés par DAMSUP et moyennés par DAMAVER. A partir du modèle moyenné, le programme DAMSTART génère un modèle qui sert de volume initial pour une dernière étape d'affinement par DAMMIN.

## X. Caractérisation structurale par microscopie électronique à transmission

### X.1. Principe de la méthode

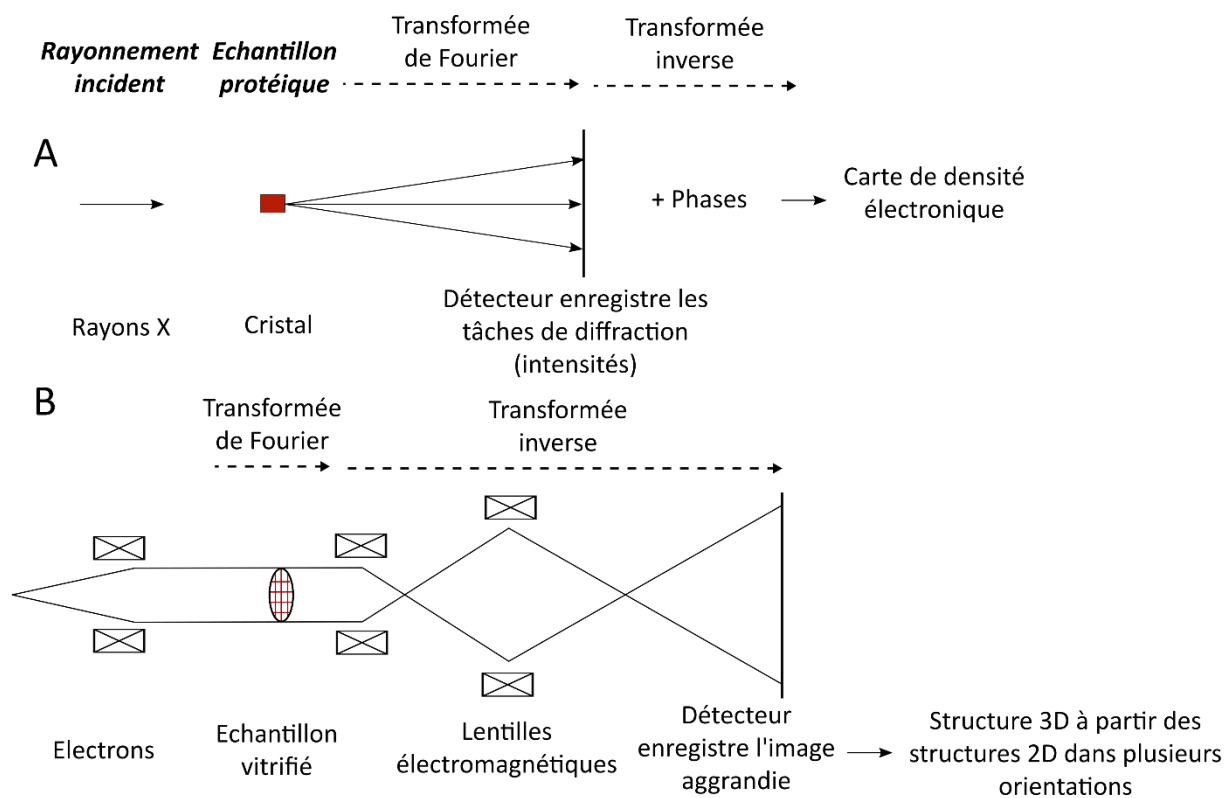


Figure 43. Comparaison schématique des techniques de cristallographie aux rayons X (A) et de cryo-microscopie électronique (B).

La cryo-microscopie électronique (cryo-ME) est une méthode structurale qui repose sur l'interaction entre les électrons du microscope et les atomes de la protéine. Les électrons ayant une longueur d'onde bien plus courte que les rayons X, de l'ordre de  $10^{-12}$  m, ils ne limitent pas la résolution des données mais ils sont très énergétiques et doivent donc être utilisés à faible dose afin de ne pas endommager l'échantillon. La technique de cryo-ME repose sur l'utilisation de lentilles électromagnétiques, selon le même principe que la microscopie optique utilisant des lentilles en verre (Figure 43). Contrairement à la cristallographie aux rayons X, l'information de phase n'est pas perdue (Saibil, 2000). Des images des particules en deux dimensions dans différentes orientations permettent la reconstitution de la structure en trois dimensions de la protéine. La technique de coloration négative permet la reconstruction de la surface de la particule, à basse résolution, alors que la technique de cryo-microscopie

électronique permet la reconstitution de la structure interne de la protéine, à moyenne ou haute résolution.

Le vide est maintenu dans la colonne du microscope afin d'empêcher l'interaction parasite entre les électrons et les molécules de l'air : les échantillons doivent donc être solidifiés avant d'être insérés dans le microscope. En coloration négative, les échantillons sont plongés dans une solution de métaux lourds puis séchés alors qu'en cryo-ME, ils sont vitrifiés (Figure 44).

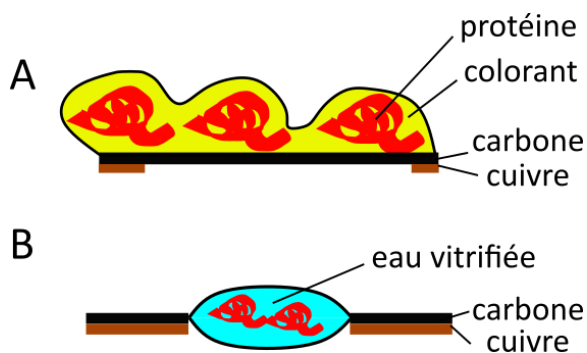


Figure 44. Environnement de l'échantillon dans une expérience de coloration négative (A) ou de cryo-ME (B).

### X.2. Coloration négative

Les atomes constitutifs des protéines sont des atomes légers ayant un faible pouvoir diffuseur. L'utilisation de colorants riches en atomes lourds permet une augmentation du contraste. Les zones contenant seulement du colorant apparaissent noires. A l'inverse, les zones contenant l'échantillon protéique sont moins riches en colorant et apparaissent blanches. La protéine est donc détectée par l'absence de colorant, d'où le nom de coloration négative.

Les grilles de coloration négative ont été préparées grâce à la technique de flottaison sur carbone (Valentine *et al.*, 1968). 4  $\mu\text{L}$  de protéine concentrée à 0.01-0.1  $\text{mg.mL}^{-1}$  sont adsorbés sur un support carbone-mica. Ce support est plongé dans un bain de colorant sodium silico tungstate (SST) pendant quelques secondes. Le mica se détache alors du carbone et le colorant se dépose sur le carbone. Le carbone est récupéré à l'aide d'une grille en cuivre et l'excès de colorant est séché avant visualisation au microscope. Le colorant SST a été choisi pour son pH neutre et son grain fin, adapté aux particules non symétriques. Les expériences de coloration négative ont été réalisées à la plate-forme de microscopie électronique de l'IBS. Les observations sur la protéine ClbK ont été effectuées avec le microscope Tecnai 12 (FEI) par

Alexandre Dos Santos (IBS) et Guy Schoehn (IBS) et les observations sur la protéine ClbK-PKS-Cy ont été effectuées par Mayara Miyachiro (IBS) avec le microscope Tecnai F20 (FEI).

L'information d'un grand nombre de particules peut être moyennée afin d'augmenter le ratio signal sur bruit des images. Les classifications des particules à partir des images de coloration négative ont été effectuées par Daphna Fenel à l'aide du programme RELION 2.1. Les particules ont d'abord été sélectionnées manuellement, ce qui a permis de générer des classes qui ont été utilisées comme références pour une phase de sélection automatique.

### X.3. Cryo-microscopie électronique

Un inconvénient de la technique de coloration négative est l'état déshydraté de l'échantillon. La technique de cryo-ME permet l'observation d'échantillons hydratés (Carroni & Saibil, 2016). Les échantillons de cryo-ME sont vitrifiés, c'est-à-dire refroidis suffisamment rapidement pour que l'eau entourant la protéine se solidifie sans former de cristaux de glace. L'échantillon est déposé sur la grille, qui est ensuite plongée très rapidement dans l'éthane liquide à -180°C. Des images des grilles sont ensuite collectées grâce au microscope puis les particules sont sélectionnées, alignées et classifiées. Un modèle en trois dimensions peut alors être reconstruit à partir des différentes classes deux dimensions. La congélation des échantillons et l'analyse au microscope ont été effectuées par Guy Schoehn.



# Résultats



## I. Etudes biochimiques et structurales de ClbK entière

### I.1. Objectifs de l'étude de ClbK entière

Le premier objectif de cette partie était l'obtention d'un protocole de purification le plus simple possible, reproductible et permettant d'obtenir des quantités de protéine pure suffisantes pour débiter des études structurales. L'étude de l'activité d'un domaine enzymatique de ClbK était un contrôle permettant de valider le protocole de purification. Un deuxième objectif était la détermination de l'état oligomérique de la protéine ClbK. Enfin, l'objectif majeur de cette partie était l'étude structurale de ClbK afin d'obtenir la première structure d'une protéine hybride PKS-NRPS. Le but était de tester plusieurs techniques de biologie structurale, complémentaires entre elles, afin de déterminer une ou plusieurs structures de ClbK et éventuellement de proposer un modèle de son cycle catalytique.

### I.2. Mise en place d'un protocole d'expression et de purification de ClbK

#### I.2.a. Expression de ClbK

Variable	Conditions testées	
Vecteur d'expression	Promoteur lactose : pET-15b, pET-21a	Promoteur arabinose : pBADM-10, pBADM-30
Quantité d'inducteur	15 g.L <sup>-1</sup> lactose ; 0.2 mM, 1 mM IPTG	0.001, 0.005, 0.01, 0.05, 0.2, 2 % arabinose
Souche d'expression	BL21(DE3), BL21(DE3) Bap1, BL21(DE3) pLysS, BL21(DE3) STAR, C41(DE3)	
Milieu d'expression	LB, TB	
Température et durée d'expression	37°C-3h, 18°C sur la nuit	

Tableau 4. Conditions d'expression de ClbK testées. Pour chaque paramètre testé, la meilleure condition est indiquée en rouge.

La mise en place d'un protocole d'expression et de purification reproductible est un prérequis aux études structurales. Ce protocole a nécessité plusieurs mois de tests afin d'obtenir des rendements de purification acceptables pour des études structurales. La comparaison de l'expression de ClbK dans plusieurs conditions indique que des conditions d'induction modérées, notamment à basse température et en présence de faibles quantités



d'inducteurs, permettent d'induire l'expression d'une quantité importante de ClbK (Tableau 4).

### I.2.b. Lyse et chromatographie d'affinité

Le choix du tampon de lyse a fait l'objet de nombreux tests de pH, de concentration en chlorure de sodium et d'additifs (Tableau 5). Malgré cette optimisation, une partie significative de ClbK est perdue dans la fraction non soluble après la lyse (Figure 45).

Variable	Conditions testées	
Tampon de lyse	50 mM MES pH 6, 50 mM Hepes pH 7.5	
Chlorure de sodium	50 mM, 150 mM, 300 mM, 500 mM, 1 M	
Agent réducteur	5 mM DTT, 10 mM $\beta$ -mercaptoéthanol, 5 mM TCEP	
Additif	5 % DMSO, 5 % Ethanol, 1 % Triton, 100 mM Urée, 10 % Glycérol, 0.05-0.5 M L-arginine, 0.05 M L-glutamate, 1 % Chaps, Benzonase, tablettes d'inhibiteurs de protéases	
Méthode de chromatographie d'affinité	Colonne HisTrap, résine Ni-IDA	Colonne GSTrap
Imidazole dans le tampon de lyse	0, 2, 5, 10, 20, 30 mM	

Tableau 5. Conditions testées pour la lyse et la chromatographie d'affinité visant à purifier ClbK. Pour chaque paramètre testé, la meilleure condition est indiquée en rouge.

De nombreuses conditions ont également été testées pour la première étape de purification de ClbK, la chromatographie d'affinité. La protéine ClbK a été fusionnée soit à une étiquette poly-histidines, en N-terminal ou en C-terminal, soit à une étiquette GST en N-terminal (Tableau 5). Dans les trois conditions et quelle que soit la durée d'incubation de la protéine avec la résine correspondante, la protéine se fixe très peu à la résine d'affinité (Tableau 5). Ainsi, la majorité de ClbK est retrouvée dans la fraction non retenue après l'incubation avec la résine d'affinité (Figure 45). L'accroche de la protéine est meilleure sur résine de nickel en *batch* plutôt que sur colonne et l'augmentation de la quantité de résine permet d'augmenter la quantité de ClbK retenue. L'ajout d'imidazole dans le tampon de lyse diminue encore la quantité de ClbK retenue sur la résine donc le tampon de lyse, utilisé pour la lyse et les lavages, n'en contient pas (Tableau 5), ce qui conduit à un grand nombre de contaminants dans les fractions d'élution de la chromatographie (Figure 45). Certains de ces contaminants émettent un signal positif en Western Blot révélé avec des anticorps anti-histidines, suggérant que ce sont des dégradations de la partie C-terminale de la protéine.

1 mM d'EDTA, un chélateur de cations divalents indispensables à certaines protéases, est ainsi rajouté directement dans les fractions d'éluion de la chromatographie d'affinité afin de limiter la dégradation de ClbK.

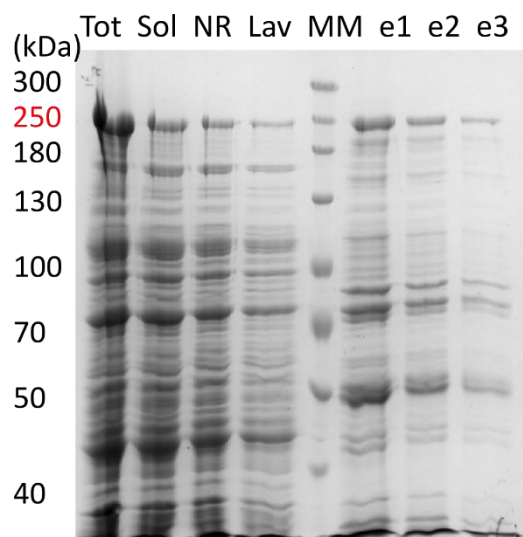


Figure 45. Gel SDS 7.5 % visualisant la pureté de ClbK (240 kDa) aux étapes de lyse et de chromatographie d'affinité sur résine de nickel. Tot : fraction totale, Sol : fraction soluble, NR : fraction non retenue, Lav : fraction de lavage, MM : marqueur de masse molaire, e1-e2-e3 : fractions d'éluions 1, 2 et 3.

### I.2.c. Chromatographie échangeuse d'anions

Le point isoélectrique de ClbK (5.6) est bien inférieur au pH du tampon de purification (7.5), ce qui permet de la purifier de ses nombreux contaminants par une chromatographie échangeuse d'anions (Figure 46). ClbK étant la dernière protéine éluée de la colonne, cette étape élimine la plupart des contaminants de ClbK.

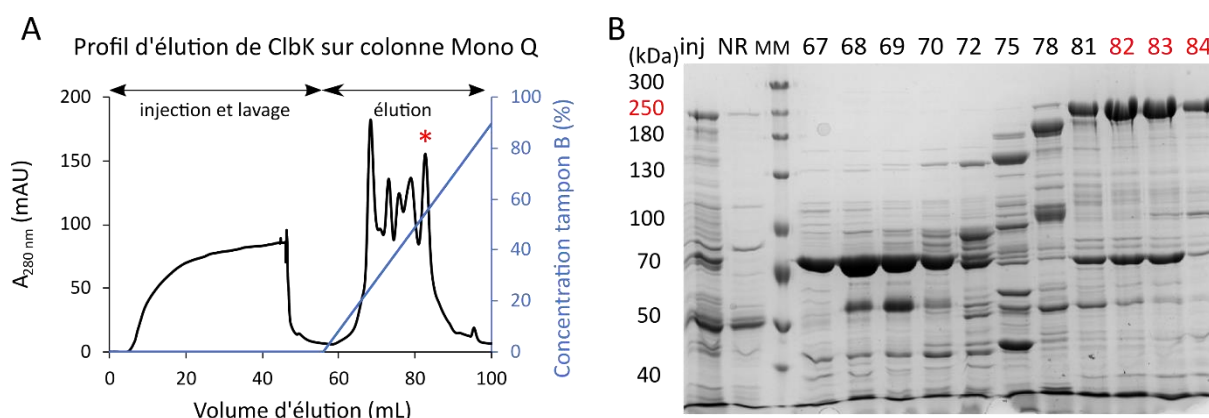


Figure 46. Deuxième étape de purification de ClbK (240 kDa) par chromatographie échangeuse d'anions. A) Profil d'éluion. Le pic contenant ClbK est indiqué par un astérisque rouge. B) Gel SDS 7.5 %. Inj : fraction injectée, NR : fraction non retenue, MM : marqueur de masse molaire. Les numéros au-dessus du gel correspondent aux volumes des fractions d'éluion du profil en A. Les fractions conservées pour la dernière étape de purification sont indiquées en rouge.

### I.2.d. Chromatographie d'exclusion de taille

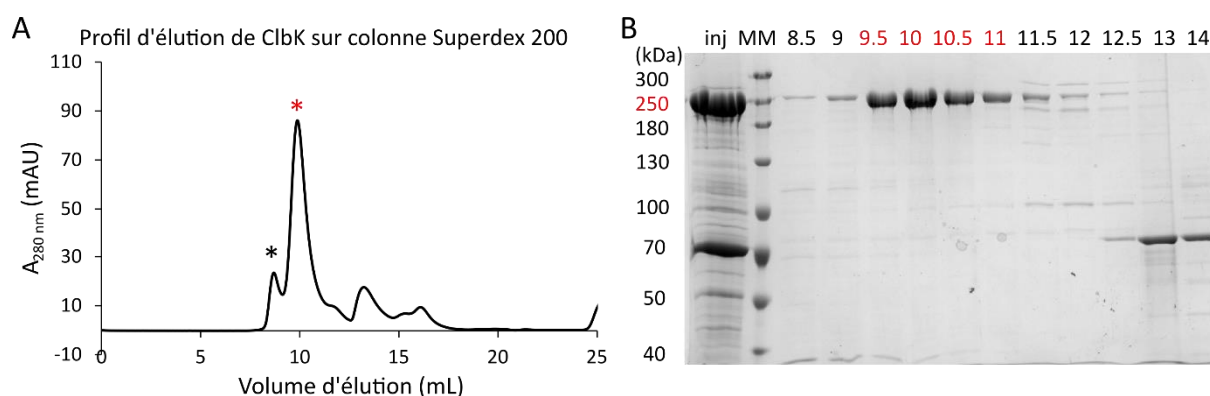


Figure 47. Troisième étape de purification de ClbK par chromatographie d'exclusion de taille. A) Profil d'éluion. Les pics contenant ClbK sont indiqués par des astérisques. B) Gel SDS 7.5 %. MM : marqueur de masse molaire, inj : fraction injectée. Les numéros au-dessus du gel correspondent aux volumes des fractions. Les fractions en rouge ont été conservées pour analyse ultérieure.

Une dernière étape de chromatographie d'exclusion de taille permet l'obtention d'un échantillon pur (Figure 47). ClbK est purifiée dans un tampon 50 mM Hepes pH 7.5, 200 mM NaCl, 5 mM DTT sur une colonne Superdex 200 Increase 10/300 GL (volume mort  $V_0$  : 8 mL). ClbK est retrouvée dans deux pics, un pic minoritaire éluant à 8.5 mL, juste après le volume mort et un pic majoritaire éluant à 10 mL. La hauteur du premier pic relativement au deuxième varie de 10 % jusqu'à 100 % selon la purification. La réinjection sur la même colonne d'exclusion du deuxième pic, après concentration, conduit à la reformation des deux pics, ce qui indique que les deux formes sont en équilibre.

Le rendement de purification est d'environ 0.1 mg de protéine pure par litre de culture. Ce rendement étant faible et la taille de l'étiquette poly-histidines étant négligeable par rapport à la taille totale de ClbK, il a été choisi de ne pas ajouter d'étape supplémentaire de clivage de l'étiquette. Les tentatives de caractérisation de ClbK par spectrométrie de masse n'ont pas abouti, probablement à cause de la masse molaire élevée de ClbK.

### I.3. Détermination de l'état oligomérique de ClbK

#### I.3.a. Diffusion de lumière laser multi-angles couplée à une chromatographie d'exclusion de taille (SEC-MALLS)

La polydispersité et l'état oligomérique de ClbK ont d'abord été étudiés grâce à la technique de SEC-MALLS. Un échantillon de ClbK pur après chromatographie échangeuse d'anions (50  $\mu\text{L}$ , 2  $\text{mg}\cdot\text{mL}^{-1}$ ) a été injecté sur une colonne d'exclusion Superdex 200 (Figure 48). Le premier pic est polydisperse, c'est-à-dire que la distribution des masses à l'intérieur du pic est hétérogène. Les masses estimées des protéines de ce pic sont de l'ordre de plusieurs mégadaltons, il s'agit donc d'agrégats solubles. Les agrégats, même en très petite quantité, diffusent fortement la lumière, ce qui explique la hauteur élevée du premier pic dans cette expérience.

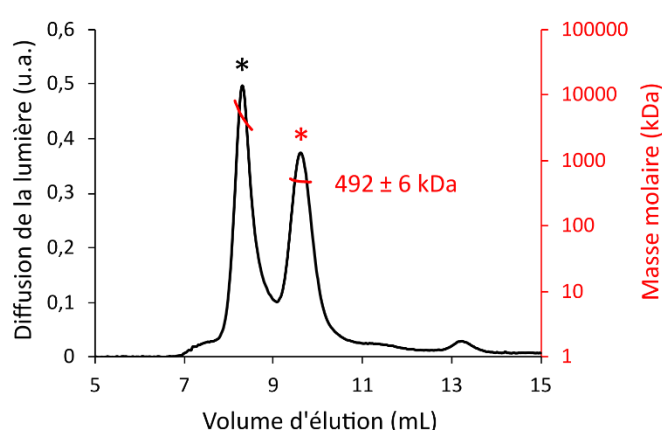


Figure 48. Analyse de ClbK par SEC-MALLS sur colonne Superdex 200 Increase 10/300 GL (25 mL). La masse molaire théorique de ClbK monomérique est 240 kDa. Les deux premiers pics correspondent aux deux pics de la Figure 47. Le premier pic contient ClbK sous forme agrégée et le deuxième pic contient ClbK sous forme dimérique. Les pics minoritaires après 11 mL sont des contaminants.

Le deuxième pic est monodisperse : la distribution des masses à l'intérieur du pic est homogène. La masse molaire estimée ( $\approx 492$  kDa) est deux fois supérieure à la masse molaire théorique du monomère de ClbK ( $\approx 240$  kDa), ce qui indique que ClbK forme un dimère en solution. L'analyse de plusieurs échantillons de ClbK à différentes concentrations en SEC-MALLS a confirmé ces résultats. L'état dimérique de ClbK a également été confirmé en SEC-SAXS (voir chapitre I.7).

Il a été montré que la protéine hybride PKS-NRPS HMWP1 forme à la fois des dimères et des monomères, qui se séparent sur colonne d'exclusion de taille (Sieber *et al.*, 2002). Aucun pic de ClbK sous forme monomérique n'a jamais été détecté par chromatographie d'exclusion de taille, ce qui suggère que la forme dimérique de ClbK est stable.

### I.3.b. Photométrie de masse

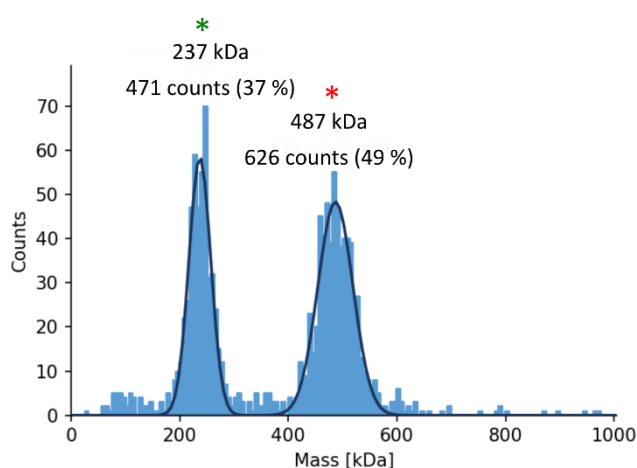


Figure 49. Analyse de ClbK par photométrie de masse. L'échantillon utilisé provient du deuxième pic de filtration sur gel ( $0.005 \text{ mg.mL}^{-1}$ ). La masse molaire estimée et le nombre d'événements enregistrés (« counts ») sont indiqués au-dessus de chaque pic. Le pic indiqué par un astérisque rouge correspond au deuxième pic de la Figure 48 (ClbK dimérique) alors que le pic indiqué par un astérisque vert (ClbK monomérique) n'est pas détecté par les autres méthodes.

L'état oligomérique de ClbK a été étudié avec une autre technique de biophysique, la photométrie de masse. Cette technique est complémentaire au SEC-MALLS car elle ne nécessite pas de colonne d'exclusion de taille et utilise très peu d'échantillon, qui est donc très dilué. L'analyse d'un échantillon du pic dimérique (pic 2 de la Figure 47, astérisque rouge), dilué dans le tampon d'exclusion de taille, montre que deux populations de ClbK coexistent : une population de monomères et une population de dimères (Figure 49). Ainsi, à faible concentration, le dimère de ClbK se dissocie en monomères. La dissociation du dimère en monomères a été décrite plusieurs fois pour d'autres protéines dont la forme physiologique est dimérique, notamment pour certaines PKS (Herbst, Huitt-Roehl, *et al.*, 2018 ; Wang *et al.*, 2021 ; Bagde *et al.*, 2021).

En résumé, l'analyse de l'état oligomérique de ClbK indique que ClbK est dimérique mais se dissocie en monomères à très faibles concentrations. Les PKS sont dimériques et la plupart des NRPS sont monomériques donc, à ce stade de l'analyse, on peut supposer que ClbK dimérise grâce à son module PKS.

#### I.4. Détermination de la préférence de substrat du domaine d'adénylation de ClbK

Le domaine d'adénylation de ClbK (domaine ClbK-A) catalyse deux réactions : il active sélectivement un acide aminé (adénylation) puis il le charge sur le bras PPant du domaine PCP (transfert). Des tests d'activité réalisés par Brachmann *et al.* sur le domaine ClbK-A purifié ont montré que le substrat du domaine ClbK-A est la L-cystéine (Brachmann *et al.*, 2015). Nous avons testé l'activité du domaine ClbK-A en utilisant la protéine ClbK entière comme catalyseur. Le premier objectif était de vérifier si notre protocole de purification de ClbK

maintenait son activité enzymatique intacte. Le deuxième objectif était de comparer l'activité enzymatique du domaine ClbK-A excisé et du domaine ClbK-A dans le contexte de la protéine entière.

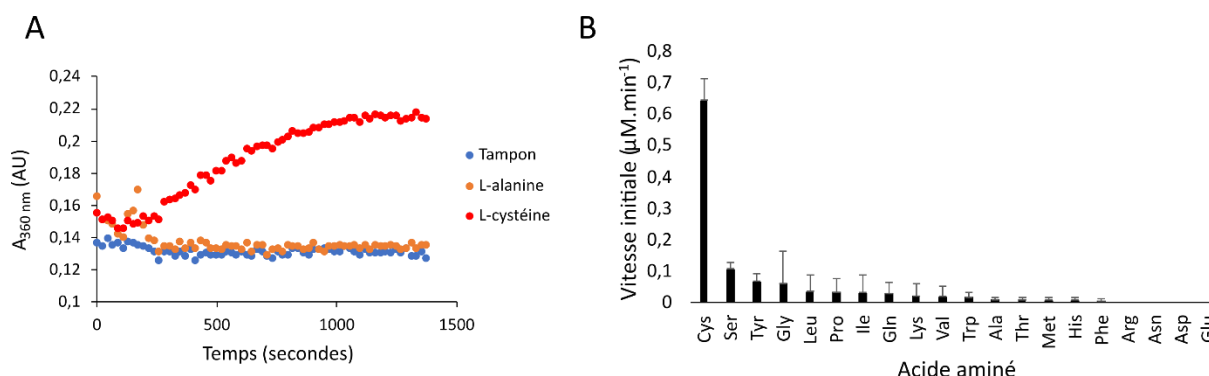


Figure 50. Test de l'activité du domaine A de ClbK par dosage du  $PP_i$  libéré lors de la réaction d'adénylation en utilisant la protéine ClbK entière comme catalyseur. A) Evolution de l'absorbance à 360 nm en fonction du temps en présence des réactifs nécessaires au dosage et de tampon, de L-alanine ou de L-cystéine. La pente du domaine linéaire de la courbe a servi à calculer les vitesses initiales  $V_0$  de la figure B. B) Vitesses initiales de réaction d'adénylation mesurées pour les vingt acides aminés. Les mesures ont été effectuées en triplicat.

Selon le dosage établi par Duckworth *et al.* (Duckworth *et al.*, 2016), en présence des composés nécessaires aux réactions, la libération de pyrophosphate inorganique ( $PP_i$ ) lors de la réaction d'adénylation conduit à la disparition du 7-méthylthioguanosine (MesG) et à l'apparition du 7-méthylthioguanine, molécule qui absorbe à 360 nm (Figure 35). Ainsi, en présence du substrat accepté par ClbK-A, une augmentation de l'absorbance à 360 nm traduit l'apparition de  $PP_i$  et donc la consommation du substrat par ClbK-A. En l'absence d'acide aminé, l'absorbance à 360 nm reste stable (Figure 50A). En revanche, en présence de L-cystéine, l'absorbance augmente de façon linéaire pendant une dizaine de minutes, jusqu'à atteindre un plateau. Cette augmentation de l'absorbance n'est visible qu'en présence de L-cystéine, et dans une moindre mesure de L-sérine et de L-tyrosine (Figure 50B) : le substrat de ClbK-A est bien la L-cystéine. Ce résultat confirme l'intégrité catalytique du domaine ClbK-A dans le contexte de la protéine entière. Il indique également que, dans le cas de ClbK, il n'y a pas de différence majeure entre l'activité du domaine A excisé et l'activité du domaine A dans le contexte de la protéine entière, même si l'acceptation des substrats L-sérine et L-tyrosine n'avait pas été rapportée dans le contexte de domaine ClbK-A excisé (Brachmann *et al.*, 2015).

Les résidus du domaine A impliqués dans la reconnaissance de la L-cystéine ont été mis en évidence grâce à la structure de la NRPS PchE en présence de cystéinyl-AMP (Wang *et*

*al.*, 2022). Les neuf résidus du domaine A<sub>core</sub> identifiés chez PchE sont conservés chez ClbK (Figure 51), suggérant un mode de sélection du substrat conservé. De façon surprenante, l'analyse de séquence du domaine A<sub>core</sub> du deuxième module de ClbJ, dont le substrat prédit est aussi la L-cystéine, a révélé que celui-ci présente une identité de séquence parfaite avec le sous-domaine A<sub>core</sub> de ClbK (Figure 51). Les autres domaines du module NRPS de ClbK et du deuxième module NRPS de ClbJ ne présentent pas une identité de séquence de 100 %. Une hypothèse pour expliquer cette observation serait un évènement de duplication récent du sous-domaine A<sub>core</sub>.

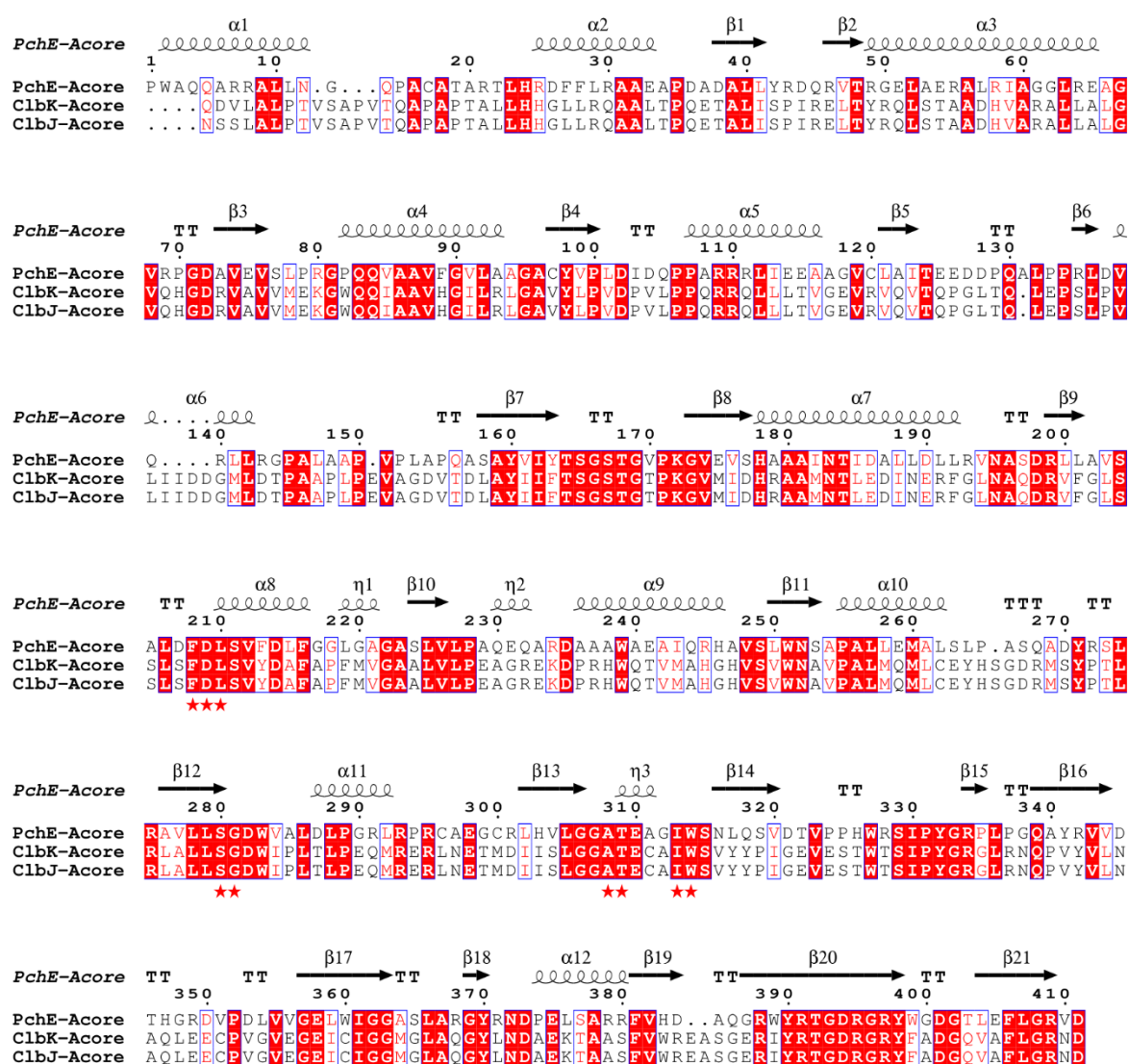


Figure 51. Analyse d'alignements de séquences des sous-domaines A<sub>core</sub> de PchE, ClbK et ClbJ, effectuée par le serveur ESPript. Les résidus conférant la spécificité du sous-domaine A<sub>core</sub> pour la cystéine, identifiés chez PchE, sont indiqués par une étoile rouge (Wang *et al.*, 2022). Les résidus en blanc sur un fond rouge sont strictement conservés et les résidus dans une colonne entourée en bleu ont des propriétés physico-chimiques similaires. Les sous-domaines A<sub>core</sub> de ClbK et de ClbJ ont 100 % d'identité de séquence.

La confirmation de l'utilisation de la L-cystéine par le domaine ClbK-A était un prérequis pour la synthèse d'une molécule stabilisant les domaines A et PCP de ClbK. En effet, les inhibiteurs AVS ont été utilisés à plusieurs reprises pour stabiliser les domaines A et PCP dans la conformation de thiolation (Gulick & Aldrich, 2018). Récemment, un inhibiteur cystéine-AVS a même été utilisé pour l'étude du complexe BmdB<sub>M2</sub>-BmdC (Fortinez *et al.*, 2022). Une collaboration avec des chimistes est en cours afin de synthétiser cet inhibiteur cystéine-AVS, qui pourrait permettre de stabiliser le module NRPS de ClbK lors d'études structurales.

### 1.5. Éléments suggérant la présence d'un cofacteur FMN associé à ClbK

Les échantillons concentrés de ClbK présentent une couleur jaune caractéristique (Figure 52A). Le spectre UV-Visible d'un échantillon de ClbK pur et concentré révèle deux pics à 360 et 450 nm (Figure 52B). Ces pics sont la signature d'un cofacteur flavine (Macheroux, 1999) (Figure 52C). Il est démontré que les domaines oxydases (Ox) de NRPS utilisent un cofacteur flavine mononucléotide (FMN) (Pang *et al.*, 2020 ; Fortinez *et al.*, 2022). Il est donc probable qu'un cofacteur FMN d'*E. coli* s'associe au domaine ClbK-Ox durant l'expression de ClbK et ne se dissocie pas durant sa purification. L'identité du cofacteur pourrait être confirmée par une expérience de chromatographie en phase liquide à haute performance (Fortinez *et al.*, 2022). Dans l'optique d'une étude ciblée du domaine ClbK-Ox, l'ajout de riboflavine, le précurseur du cofacteur FMN, durant l'expression pourrait augmenter la quantité de cofacteur FMN associée à ClbK (Skiba *et al.*, 2018).

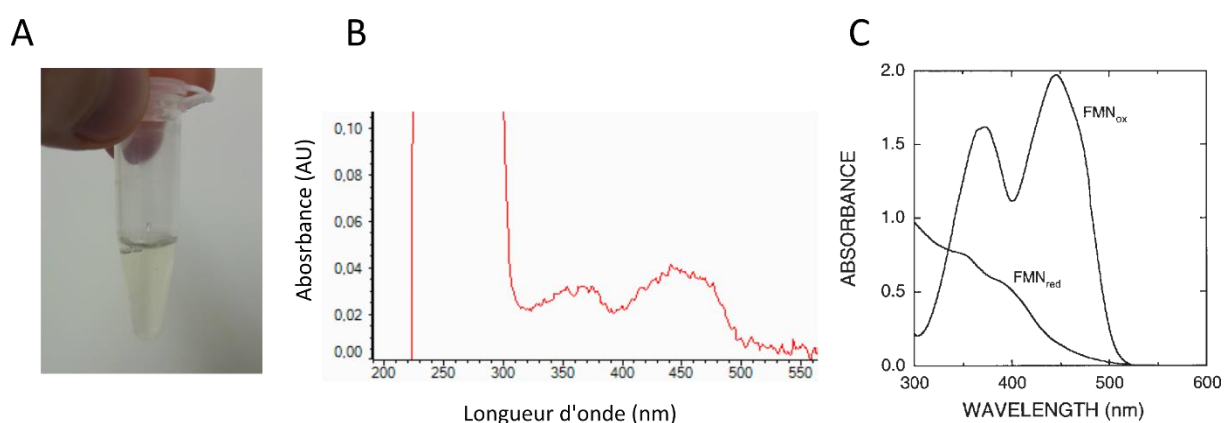


Figure 52. Co-purification de ClbK avec un cofacteur flavine. A) Echantillon de ClbK concentré à environ 2 mg.mL<sup>-1</sup>, avant injection sur colonne d'exclusion, présentant une couleur jaune caractéristique. B) Spectre UV-visible d'un échantillon pur de ClbK. C) Spectres UV-Visible du cofacteur FMN sous forme oxydée et réduite. D'après Macheroux, 1999.



## I.6. Essais de cristallisation de ClbK

### I.6.a. Premiers tests de cristallisation

Les premiers tests de cristallisation sur ClbK ont été menés par la plate-forme de cristallisation HTX (« High Throughput Crystallization ») de l'ISBG Grenoble. Ces essais de cristallisation ont été effectués à 20°C et trois concentrations de ClbK ont été testées : 2.8, 5.6 et 13 mg.mL<sup>-1</sup>. La concentration intermédiaire semble la plus propice à la cristallisation car environ la moitié des gouttes réalisées avec ClbK à cette concentration étaient claires et l'autre moitié étaient précipitées. Les seuls cristaux obtenus sont apparus après plus de deux mois et ne présentaient pas de signal UV, ce qui suggère qu'il s'agissait de cristaux de sels (Figure 53). Des tests de diffraction au diffractomètre de l'IBS sur certaines conditions HTX reproduites manuellement ont révélé que les cristaux reproduits étaient des cristaux de sel.

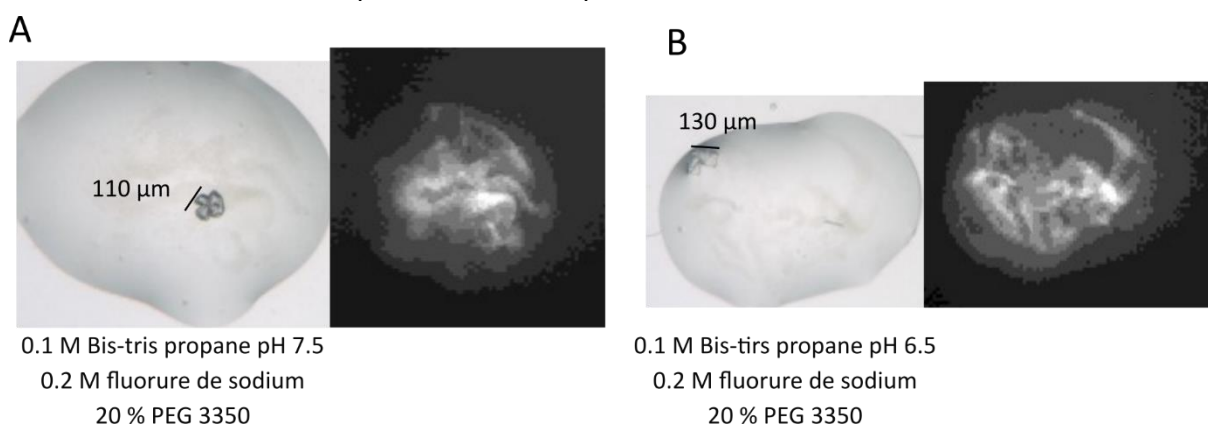


Figure 53. Exemples de cristaux obtenus par mélange de ClbK (2.8 mg.mL<sup>-1</sup>) avec différentes solutions de cristallographie (images HTX). Pour chaque condition, l'image de gauche est en lumière visible et celle de droite est en lumière UV. Les cristaux sont apparus après 60 jours et les photographies ont été prises après 90 jours.

### I.6.b. Comparaison de tampons pouvant stabiliser ClbK

La température  $T_m$  est un indicateur de la stabilité des protéines et peut être mesurée par la méthode du Thermofluor (Dupeux *et al.*, 2011) (Figure 37). Il a notamment été montré que 49 % des protéines ayant une température  $T_m$  supérieure ou égale à 45° cristallisent alors que seulement 23 % des protéines ayant une température  $T_m$  inférieure ou égale à 43°C cristallisent. La composition du tampon impacte la  $T_m$ .

La  $T_m$  de ClbK, concentrée à 2.8 mg.mL<sup>-1</sup> dans son tampon de filtration sur gel, est de 43°C, ce qui suggère que la probabilité de ClbK de cristalliser dans ce tampon est faible. Une comparaison de la température  $T_m$  de ClbK dans 96 conditions a été menée par la plate-forme

HTX. Bien qu'aucun tampon testé n'augmente grandement le  $T_m$  de ClbK, d'autres effets du tampon sur la stabilité de ClbK ont pu être mis en évidence (Figure 54).

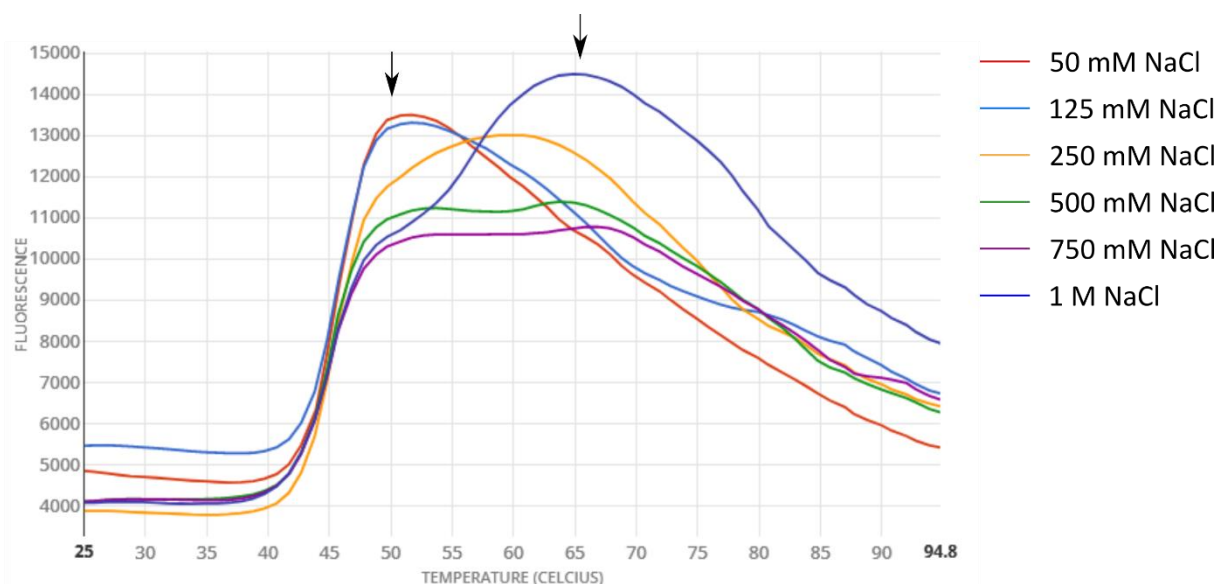


Figure 54. Courbes de dénaturation de ClbK ( $1 \text{ mg.mL}^{-1}$ ) en fonction de la concentration en chlorure de sodium (NaCl). Les six tampons testés sont composés de 50 mM Hepes pH 7.5 et d'une concentration variable en chlorure de sodium. L'augmentation de la fluorescence traduit la dénaturation de ClbK alors que la diminution de la fluorescence s'explique par la précipitation de la protéine. Selon les conditions, un ou deux pics de fluorescence, indiqués par des flèches, sont observés.

Le profil de la courbe de dénaturation varie selon la concentration en chlorure de sodium dans le tampon testé (Figure 54). A faible concentration en sel (50 et 125 mM), un pic de fluorescence unique est observé alors qu'à partir de 500 mM de chlorure de sodium, un deuxième pic apparaît. Plusieurs hypothèses pourraient expliquer l'existence de ce deuxième pic. Par exemple, le premier pic pourrait correspondre au dépliement d'une partie des domaines de ClbK et le deuxième pic au dépliement du reste des domaines. A faible concentration en sel, la précipitation de la protéine surviendrait avant le dépliement du deuxième groupe de domaines alors qu'à haute concentration en sel, les deux groupes de domaines se dissocieraient successivement avant la précipitation. Les deux pics pourraient également correspondre à la dénaturation de deux formes oligomériques de ClbK distinctes. Les expériences de SEC-MALLS ayant montré une forme dimérique de ClbK ont été effectuées à faible concentration en chlorure de sodium (1.3.a ci-dessus). Le premier pic correspondrait donc à la dénaturation de la forme dimérique de ClbK et le second pic à la dénaturation de la forme monomérique. Selon cette hypothèse, une faible concentration en sel stabiliserait la forme dimérique de ClbK alors qu'une forte concentration en sel favoriserait la dissociation

du dimère. Le même type de profil à deux pics est observable dans un tampon 100 mM citrate pH 6.

### I.6.c. Tests de cristallisation du mutant ClbK-SER

L'absence de condition de cristallisation prometteuse pour ClbK nous a conduit à construire un mutant de ClbK, ayant une probabilité accrue de cristalliser. Nous avons utilisé la technique de réduction d'entropie de surface (SER) (Longenecker *et al.*, 2001), déjà employée avec succès dans l'équipe Pathogénie Bactérienne (Wong & Dessen, 2014). Cette technique repose sur la mutation en alanines de certains résidus chargés de surface, ce qui permet de réduire l'entropie de surface localement et, en théorie, de favoriser la cristallisation. Les résidus Lys1718, Glu1719 et Glu 1721 ont été identifiés par le serveur SERp (Goldschmidt *et al.*, 2007) comme ayant une entropie de surface élevée. La construction ClbK-SER a été générée par mutation de ces trois résidus localisés dans le *linker* précédant le domaine Ox.

Le mutant ClbK-SER a été exprimé et purifié selon le même protocole que la protéine non mutée. Des essais de cristallisation ont d'abord été conduits à 20°C par la plate-forme HTX sur un échantillon de ClbK-SER concentré à 1.8 mg.mL<sup>-1</sup>. La plupart des gouttes étaient encore claires après trois mois, indiquant que la concentration en protéine était trop faible pour favoriser la cristallisation. Un deuxième échantillon, concentré à 3.2 mg.mL<sup>-1</sup>, a donc été testé pour la cristallisation. Cet échantillon a été incubé avec un excès de L-cystéine, le substrat du domaine ClbK-A. Entre la moitié et deux tiers des gouttes étaient encore claires après trois mois, indiquant que la concentration était encore trop faible pour favoriser la cristallisation.

En conclusion, aucune condition de cristallisation n'a pu être déterminée ni pour ClbK ni pour le mutant ClbK-SER dans les conditions testées. L'incubation avec des inhibiteurs stabilisant un ou plusieurs domaines de ClbK, par exemple la molécule cystéine-AVS, pourrait permettre de stabiliser une conformation de ClbK et ainsi favoriser sa cristallisation.

### I.7. Analyse structurale de ClbK en solution par diffusion des rayons X aux petits angles (SAXS)

La technique de diffusion des rayons X aux petits angles (SAXS), comme la cristallographie aux rayons X, repose sur l'interaction des rayons X avec la matière mais l'échantillon protéique est en solution dans une expérience de SAXS et non sous forme cristalline. La technique de SAXS permet donc de s'affranchir de l'étape limitante de cristallisation de la protéine. En revanche, la protéine étant en solution, les différentes conformations adoptées sont moyennées donc les informations structurales obtenues sont à basse résolution.

#### I.7.a. Comparaison des expériences de SAXS en mesure directe ou en SEC-SAXS

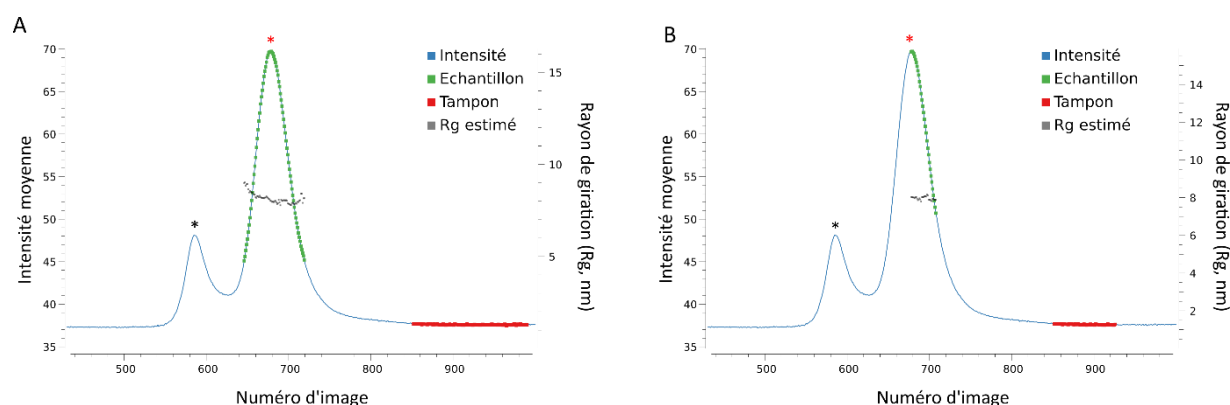


Figure 55. Profil de diffusion de la protéine ClbK analysée en SEC-SAXS, visualisé grâce au logiciel CHROMIXS. Les points en vert correspondent aux images choisies pour l'échantillon et les points en rouge correspondent aux images choisies pour le tampon. Les points en gris correspondent au rayon de giration (Rg) estimé pour chaque image de l'échantillon. Les deux pics sont équivalents aux deux pics de la Figure 48 (pic 1 : agrégat, pic 2 : dimère). A) Estimation du Rg pour l'ensemble du deuxième pic. B) Estimation du Rg pour la deuxième moitié du deuxième pic. Ces images d'échantillon et de tampon ont été utilisées pour le reste de l'analyse.

Le signal de diffusion de ClbK a été mesuré à la fois en SEC-SAXS et en *batch*. Pour l'expérience de SEC-SAXS, un échantillon pur de ClbK correspondant au deuxième pic de filtration sur gel (70  $\mu$ L à 9 mg.mL<sup>-1</sup>) a été réinjecté sur une colonne de filtration sur gel Superdex 200. Le profil de l'intensité des fractions d'élution indique la présence de deux espèces, qui correspondent aux formes agrégée et dimérique de ClbK (Figure 55). Le logiciel CHROMIXS permet le choix des images correspondant à l'échantillon protéique et au tampon. Ce logiciel permet également l'estimation du rayon de giration (Rg) pour différentes portions du pic d'échantillon. Le Rg estimé pour le deuxième pic n'est pas homogène car il est plus élevé au début du pic qu'à la fin du pic (Figure 55A), ce qui suggère que la fin du premier pic,

contenant des agrégats, se superpose avec le début du deuxième pic contenant des dimères. Afin d'éliminer la contribution de la forme agrégée et de se focaliser sur une population homogène et monodisperse, les images de l'échantillon ont été choisies parmi la deuxième moitié du deuxième pic uniquement (Figure 55B). La courbe de diffusion rouge présentée Figure 56A résulte de la soustraction des intensités moyennées pour la zone d'échantillon par les intensités moyennées pour la zone de tampon.

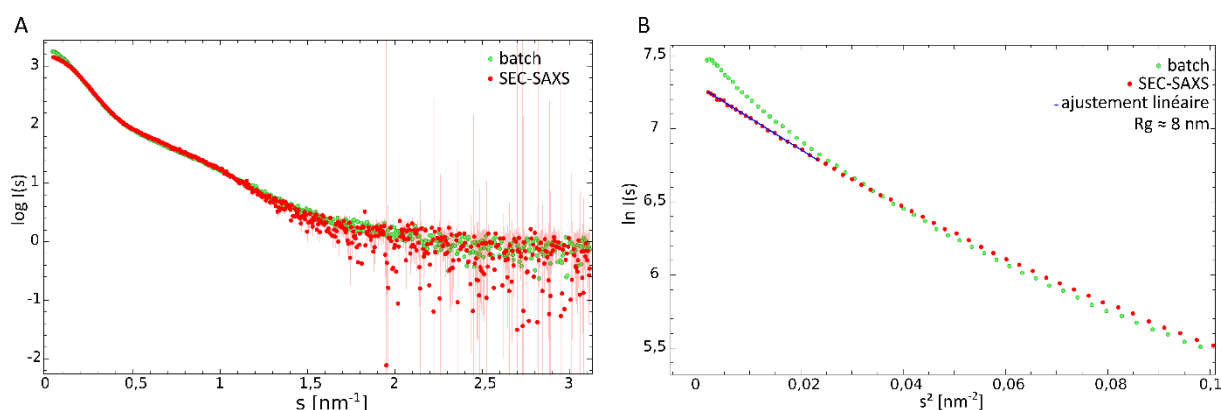


Figure 56. Comparaison des expériences de SAXS en mesure directe (*batch*, en vert) et en SEC-SAXS (rouge) pour la protéine ClbK. A) Courbes de diffusion de ClbK, mesurée en *batch* ou en SEC-SAXS. B) Représentation de Guinier des courbes de diffusion de ClbK en *batch* ou en SEC-SAXS.

L'analyse de ClbK en SEC-SAXS associée au choix adéquat des zones d'échantillon et de tampon permettent donc de distinguer la contribution du dimère de celle de l'agrégat dans le signal d'intensité. En revanche, la technique de SEC-SAXS dilue l'échantillon, elle diminue donc la résolution des données. Afin d'obtenir des données à plus haute résolution, des expériences de SAXS sur ClbK ont aussi été menées en *batch*. Bien que les courbes de diffusion de ClbK en *batch* et en SEC-SAXS aient globalement la même allure, le signal aux plus petits angles varie beaucoup entre les deux types d'expériences (Figure 56A). La représentation de Guinier se focalise sur ces très petits angles et la linéarité de la courbe est un bon indicateur de la qualité de l'échantillon. La courbe de Guinier de l'échantillon de ClbK mesuré en SEC-SAXS est linéaire, ce qui indique l'absence d'interactions entre particules dans cet échantillon (Figure 56B, courbe rouge). En revanche, la courbe de Guinier n'est pas linéaire pour l'échantillon mesuré en *batch*, ce qui suggère la présence d'agrégats (Figure 56B, courbe verte). La technique de SAXS en mesure directe n'est donc pas adaptée pour l'étude de ClbK concentrée, les résultats qui suivent porteront donc sur l'analyse de ClbK en SEC-SAXS. Les données ont été coupées après  $1.7 \text{ nm}^{-1}$  car le signal est très bruité au-delà de cette valeur (Figure 56A).

### I.7.b. Détermination des paramètres structuraux

La pente de la portion linéaire de la courbe de Guinier donne accès au rayon de giration de la protéine, qui est de  $7.95 \pm 0.02$  nm pour ClbK (Figure 56B). Cette distance correspond au rayon qu'aurait la protéine si elle avait la forme d'une sphère. Le rayon de giration extrait de la fonction de distribution des distances est de 8.2 nm, ce qui est légèrement plus élevé que celui extrait de la courbe de Guinier. A titre de comparaison, le  $R_g$  estimé pour la PKS dimodulaire DEBS3 (MM = 2 x 334 kDa) est de  $8.3 \pm 0.2$  nm.

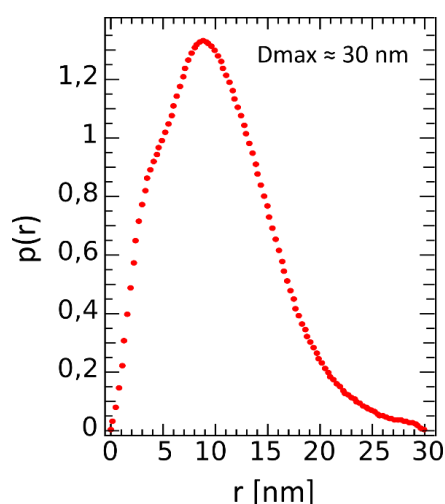


Figure 57. Fonction de distribution des distances de ClbK.  $D_{max}$  est la dimension maximale de la protéine.

La fonction de distribution des distances est la fonction de représentation des distances intramoléculaires. La forme de la courbe est caractéristique de la forme de la protéine et cette courbe donne accès à la dimension maximale ( $D_{max}$ ) de la protéine (Figure 40). Pour ClbK, la courbe présente un maximum à environ 10 nm et un épaulement à 5 nm, l'épaulement étant caractéristique des protéines multi-domaines (Figure 57). La forme de la courbe suggère que ClbK a une forme plutôt allongée, en ballon de rugby. La dimension maximale de la protéine est d'environ 30 nm.

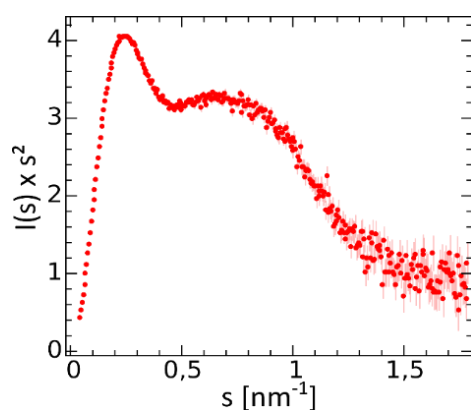


Figure 58. Graphique de Kratky de ClbK.

Le graphique de Kratky est un indicateur de la flexibilité de l'échantillon. Le graphique de Kratky pour ClbK montre deux pics distincts puis décroît et ne augmente pas ensuite (Figure 58). Pour les protéines désordonnées, la courbe ne redescend pas jusqu'à l'axe des abscisses et remonte ensuite. A  $1.5-2 \text{ nm}^{-1}$ , le signal de diffusion de ClbK est redescendu jusqu'à l'axe des abscisses mais le signal est très bruité après  $2 \text{ nm}^{-1}$ , il est donc incertain d'affirmer que le signal ne remonte pas ensuite. La présence des deux pics suggère toutefois que ClbK n'a pas un comportement de « beads on a string » (collier de perles) mais adopte un nombre limité de conformations.

### I.7.c. Modélisation de ClbK

A partir des courbes de diffusion 1D, des modèles 3D de la protéine responsable du signal de diffusion peuvent être générés, par exemple avec le logiciel DAMMIF. DAMMIF utilise un processus d'essai-erreur jusqu'à obtenir un modèle compact dont la courbe de diffusion théorique est la plus proche possible de la courbe de diffusion expérimentale (Figure 42). Plusieurs paramètres peuvent contraindre la géométrie des modèles générés, telle que la forme de la particule ou la symétrie. En l'absence de données structurales complémentaires sur ClbK, nous n'avons pas imposé de forme particulière. Bien que ClbK soit dimérique, nous n'avons pas non plus imposé de symétrie P2 car une asymétrie a été démontrée chez plusieurs PKS dimériques (Bagde *et al.*, 2021 ; Cogan *et al.*, 2021). Les 20 modèles ainsi générés sont représentés à la Figure 59.

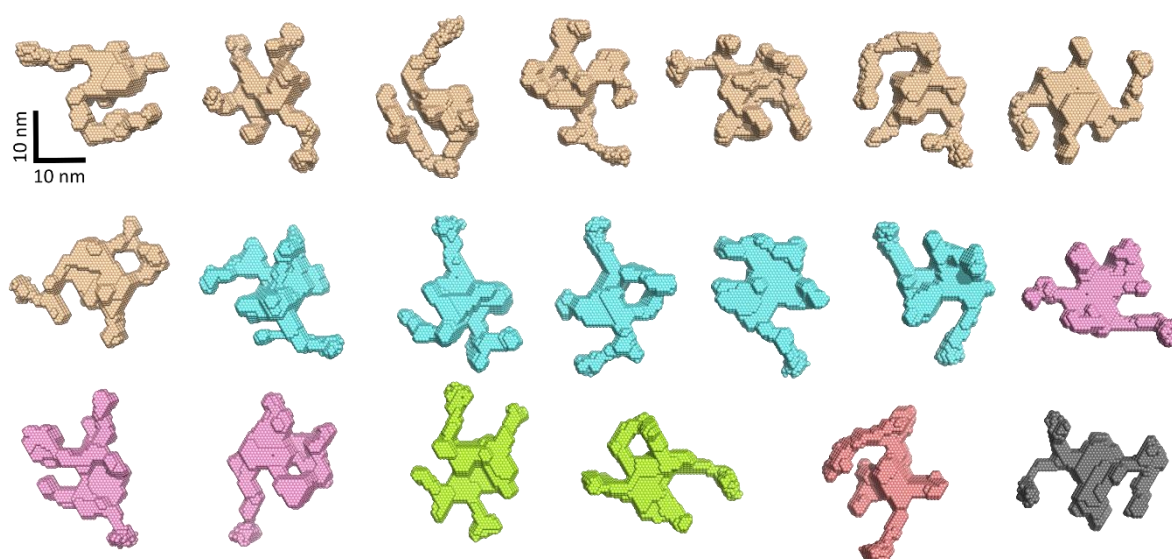


Figure 59. Modèles *ab initio* générés à partir des données de diffusion de ClbK. 20 modèles ont été générés par DAMMIF sans forme ni symétrie imposées. Modèles représentés avec Pymol. Voir la Figure 61 pour la signification des couleurs des modèles.

A partir de ces 20 modèles, un modèle moyen a été calculé qui a servi de forme initiale pour une étape d'affinement effectuée par le logiciel DAMMIN (Figure 60A). Ce modèle, comme les 20 modèles *ab initio* calculés précédemment, présente une valeur de  $\chi^2$  basse, inférieure à 1, indiquant que l'écart entre la courbe de diffusion théorique du modèle et la courbe de diffusion expérimentale est faible (Figure 60B).

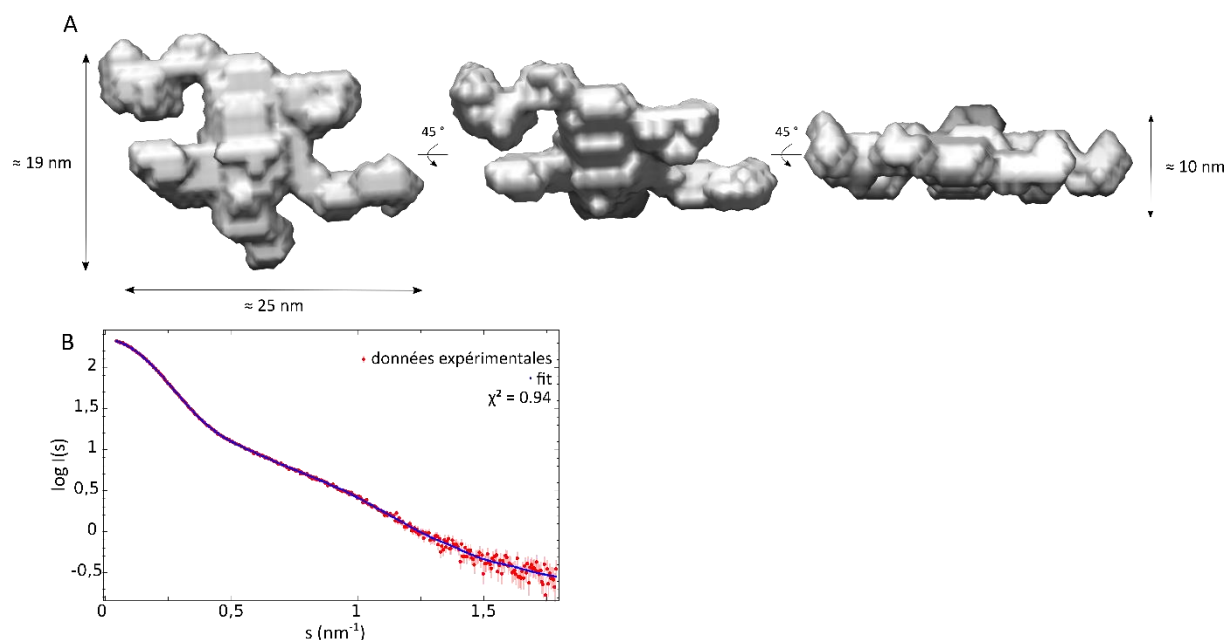


Figure 60. Modèle *ab initio* affiné de ClbK en solution généré par DAMMIN. A) Le modèle selon trois orientations, représenté avec Chimera. B) Ce modèle explique très bien les données de diffusion de ClbK ( $\chi^2 = 0.94$ ).

L'indice du  $\chi^2$  n'est pas un gage absolu de la fiabilité du modèle. En effet, les données SAXS présentent une certaine ambiguïté : plusieurs modèles peuvent expliquer la courbe de diffusion des données (Figure 59). Le score d'ambiguïté des données de diffusion de ClbK, calculé grâce au programme AMBIMETER, est de 2.6, ce qui indique que les données sont ambiguës. Pour des données ayant un score d'ambiguïté supérieur à 1.5, un regroupement des modèles en *clusters* est conseillé. Le programme DAMCLUST a permis de regrouper les modèles les plus proches entre eux en différents *clusters* et de générer un modèle moyen représentatif de chaque *cluster*. Deux des vingt modèles n'ont été inclus dans aucun *cluster* et les 18 autres modèles ont été regroupés en quatre *clusters* (Figure 61).



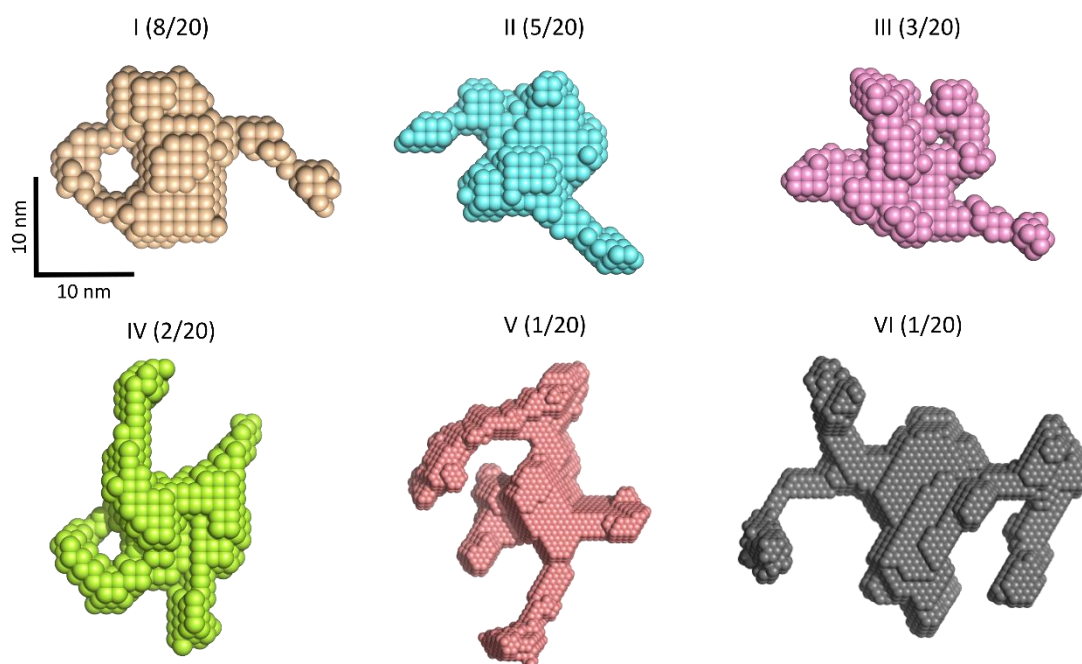


Figure 61. Modèles *ab initio* représentatifs de chaque *cluster* de ClbK. Les 20 modèles de la Figure 59 ont été comparés et regroupés en *clusters* par DAMCLUST. Le nombre de modèles contenus dans chaque *cluster* est indiqué et les modèles contenus dans chaque *cluster* sont représentés de la même couleur dans la Figure 59 et la Figure 61. Modèles représentés avec Pymol.

Malgré la grande diversité des modèles, des caractéristiques communes peuvent être déduites de la comparaison des modèles représentatifs de chaque *cluster* et du modèle affiné par DAMMIF. Premièrement, la forme et le volume des modèles confirment l'état dimérique de ClbK. En effet, les masses molaires estimées pour les modèles *ab initio* sont d'environ 430 kDa, ce qui est proche de la masse molaire de la forme dimérique de ClbK, 480 kDa. Bien que la symétrie P2 n'ait pas été imposée, cette symétrie est visible dans les modèles. Tous les modèles présentent un cœur compact et selon les *clusters*, entre deux et quatre extrémités allongées rattachées à ce cœur. Ces extrémités sont recourbées et isolent des cavités ressemblant aux chambres catalytiques décrites pour les PKS, fermées dans certains modèles et ouvertes dans d'autres. Le modèle de ClbK présenté en discussion prendra en compte ces éléments.

## 1.8. Observation de ClbK en microscopie électronique

La biologie structurale des NRPS/PKS a grandement bénéficié de la révolution de la microscopie électronique. En effet, une grande partie des structures de NRPS/PKS multi-domaines publiées récemment ont été obtenues par cryo-microscopie électronique (cryo-ME) (Dutta *et al.*, 2014 ; Whicher *et al.*, 2014 ; Katsuyama *et al.*, 2021 ; Bagde *et al.*, 2021 ; Cogan *et al.*, 2021 ; Fortinez *et al.*, 2022 ; Wang *et al.*, 2022). La cryo-ME permet l'obtention de structures à moyenne ou haute résolution à partir d'une quantité d'échantillon en solution limitée. Elle est adaptée aux protéines adoptant plusieurs conformations, comme ClbK. Toutefois, le nombre de conformations adoptées doit être faible pour pouvoir les distinguer en ME. L'étape préliminaire à l'analyse en cryo-ME est l'analyse en coloration négative, qui fournit une première estimation de l'homogénéité de l'échantillon sur la grille de ME.

### **1.8.a. Premières observations de ClbK en coloration négative : classification et problèmes rencontrés**

Des premiers tests de ME sur ClbK (deuxième pic de filtration sur gel, Figure 48) ont montré que l'observation de ClbK en coloration négative est optimale à une concentration de 0.02-0.05 mg.mL<sup>-1</sup> par le colorant sodium silico tungstate (SST). Les images indiquent que l'échantillon est apparemment homogène en taille et contient peu d'agrégats (Figure 62A).

Une classification a été effectuée sur plus de 3000 particules grâce au logiciel RELION (Figure 62B). Deux populations peuvent être distinguées : les deux-tiers des particules forment un assemblage mesurant 10-13 nm (les trois premières classes de la Figure 62B) et le tiers restant est plus petit et mesure 8-10 nm (les deux dernières classes de la Figure 62B). Deux hypothèses pourraient justifier la présence de ces deux populations. La première serait que la population de plus grande taille serait composée des particules dimériques de ClbK alors que la population de plus petite taille contiendrait les particules monomériques de ClbK. Cependant, les particules de grande taille mesurent seulement 10-13 nm, ce qui correspond à moins de la moitié de la dimension maximale ( $D_{max}$ ) de ClbK dimérique déterminée en SAXS. Même en tenant compte de la possibilité d'extrémités flexibles, ces tailles semblent trop faibles pour des particules de quasiment 500 kDa. La deuxième hypothèse serait que les deux populations reflètent deux orientations d'une même particule, ClbK sous forme monomérique. Cette hypothèse suggère que ClbK se dissocie en monomères sur la grille de ME en coloration négative.

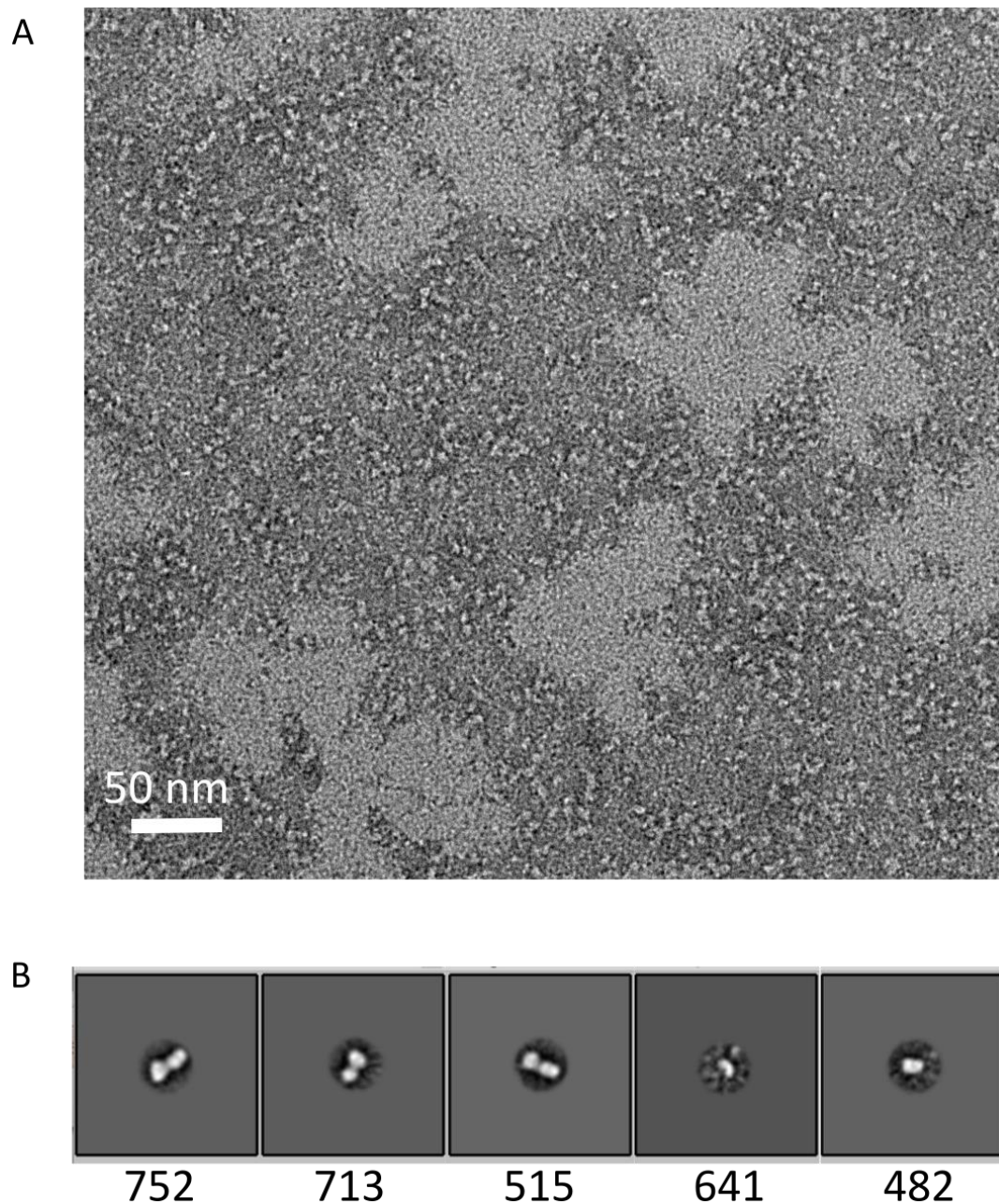


Figure 62. Observation de ClbK ( $0.05 \text{ mg.mL}^{-1}$ ) dans son tampon de filtration sur gel en coloration négative, colorée au sodium silico tungstate. A) Image représentative des observations de ClbK en coloration négative. Les larges zones blanches sont des zones sans colorant. B) Classes obtenues après classification de plus de 3100 particules par RELION. Le nombre de particules retrouvé dans chaque classe est indiqué sous chaque image.

L'observation de particules de ClbK par cryo-ME a montré des particules de taille équivalente à celles observées en coloration négative, ce qui indique que la dissociation de ClbK en monomères n'est pas due à l'interaction avec le colorant utilisé en coloration négative. La dissociation du dimère pourrait être causée par la dilution de l'échantillon en ME mais l'échantillon utilisé en photométrie de masse était dix fois plus dilué que celui utilisé en coloration négative et la moitié des particules étaient dimériques (Figure 49). Il semblerait donc que l'interaction de ClbK avec la grille de ME soit responsable de la dissociation du dimère. Ce problème de dissociation d'un dimère en ME a déjà été décrit pour certaines PKS (Herbst, Huitt-Roehl, *et al.*, 2018 ; Bagde *et al.*, 2021 ; Cogan *et al.*, 2021) mais l'optimisation du tampon utilisé pour la ME a permis de limiter la dissociation de certains de ces dimères.

#### **I.8.b. Observation de ClbK en absence de chlorure de sodium**

La dissociation du dimère de CTB1, une PKS itérative, a été minimisée grâce à une faible concentration en chlorure de sodium (50 mM NaCl) dans l'échantillon observé en cryo-ME (Herbst, Huitt-Roehl, *et al.*, 2018). Les données de Thermofluor collectées sur ClbK suggéraient également qu'une faible concentration en chlorure de sodium stabiliserait le dimère de ClbK (Figure 54). Il a donc été décidé d'observer ClbK dans un tampon faiblement concentré en sel. Nous avons d'abord modifié la composition du tampon de filtration sur gel utilisé pour la dernière étape de purification de ClbK (50 mM NaCl à la place de 200 mM NaCl précédemment). Ce protocole a fait chuter le rendement de la purification car ClbK est restée associée à ses contaminants en filtration sur gel. Une autre stratégie a donc été adoptée : une étape supplémentaire de dialyse dans un tampon sans sel a été ajoutée après la purification de ClbK selon les trois étapes décrites précédemment. L'observation de cet échantillon en coloration négative et en cryo-ME n'a pas permis de mettre en évidence des particules plus grosses pouvant correspondre à des dimères de ClbK (Figure 63A). L'analyse de cet échantillon en photométrie de masse n'a pas non plus permis de mettre en évidence une proportion accrue de dimères en absence de chlorure de sodium (Figure 63B et C).

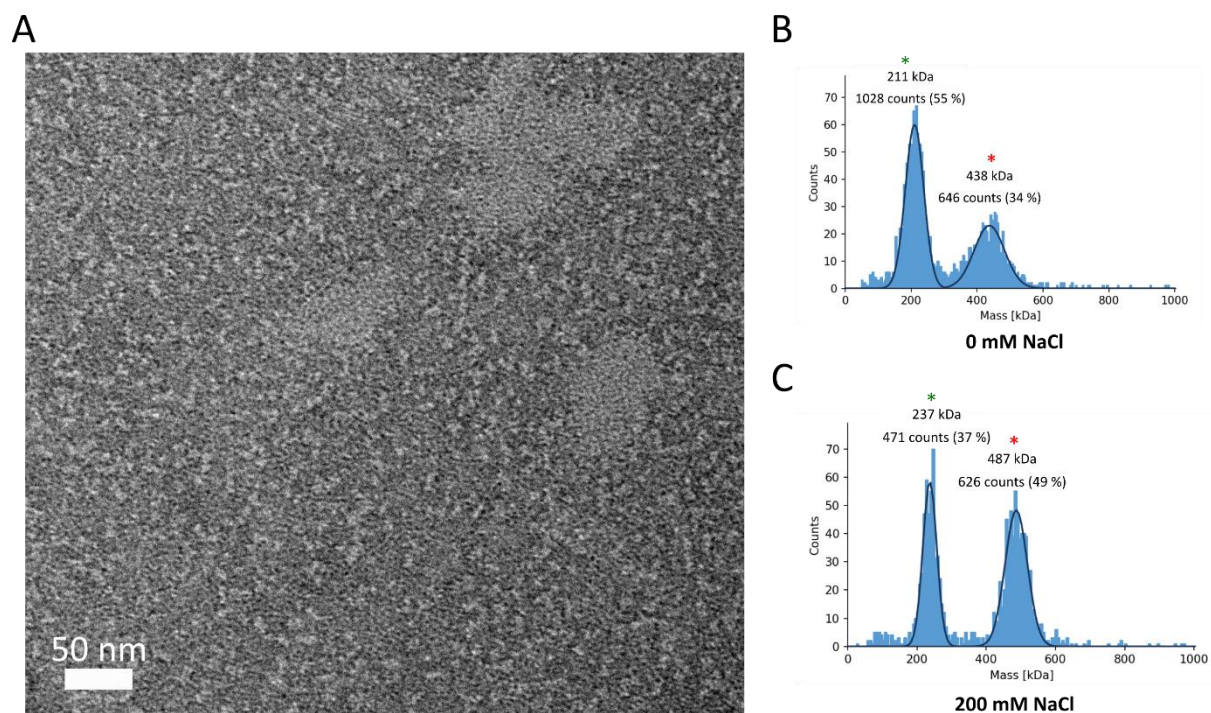


Figure 63. Observation de ClbK dans un tampon sans chlorure de sodium. A) Image de coloration négative de ClbK ( $0.03 \text{ mg.mL}^{-1}$ ) représentative de l'observation de la protéine ClbK dans un tampon sans chlorure de sodium, colorée au SST. B et C) Analyse par photométrie de masse d'un échantillon de ClbK ( $0.007 \text{ mg.mL}^{-1}$ ) sans sel (B) et d'un échantillon ( $0.005 \text{ mg.mL}^{-1}$ ) contenant 200 mM NaCl (C). La masse molaire estimée et le nombre d'événements (« counts ») enregistrés sont indiqués au-dessus de chaque pic.

### I.8.c. Observation de ClbK en présence d'agents de réticulation

Les agents de réticulation sont des molécules établissant des liaisons covalentes entre des molécules biologiques, par exemple les protéines. Ils permettent la stabilisation de complexes ou de dimères : le dimère de la PKS itérative LovB a par exemple été stabilisé avec un agent de réticulation avant observation en cryo-ME afin de limiter sa dissociation en monomères (Wang *et al.*, 2021).

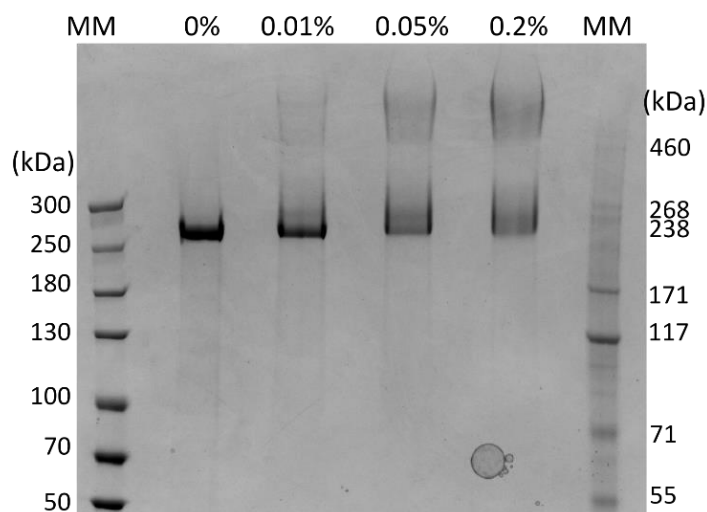


Figure 64. Gel en conditions dénaturantes visualisant l'effet du glutaraldéhyde (GA) sur ClbK. Un gel NuPage 3-8 % Tris-Acétate a été utilisé. Un échantillon de ClbK pur après filtration sur gel ( $0.15 \text{ mg.mL}^{-1}$ ) a été incubé avec trois concentrations de GA pendant 15 minutes.

Un premier test de réticulation de ClbK (dans un tampon contenant 200 mM NaCl) avec les agents de réticulation BS<sup>3</sup>, EGS et glutaraldéhyde (GA) a désigné le GA comme agent de réticulation le plus prometteur. En effet, l'augmentation de la concentration en GA conduit à la disparition de la bande de ClbK monomérique sur gel en conditions dénaturantes et à l'apparition d'une bande à la taille attendue pour la forme dimérique de ClbK (Figure 64). Aux concentrations testées, l'EGS est inefficace et le BS<sup>3</sup> n'agit que faiblement sur ClbK. La réinjection sur colonne de filtration sur gel d'un échantillon de ClbK ponté avec 0.2 % GA n'a pas conduit à la formation d'un pic d'agrégat. Par conséquent, l'étape de filtration sur gel a été omise pour l'observation en ME afin de ne pas diluer l'échantillon.

L'observation en coloration négative de l'échantillon de ClbK en sortie de filtration sur gel ponté par 0.2 % GA a confirmé l'absence d'agrégat (Figure 65A). Les images montrent que l'échantillon est plutôt homogène mais, à part quelques exceptions, les particules ne sont pas significativement plus grosses que dans l'échantillon non soumis au glutaraldéhyde (Figure 62A). L'analyse de cet échantillon en cryo-ME a confirmé l'absence de particule ayant une taille compatible avec une protéine de 500 kDa.

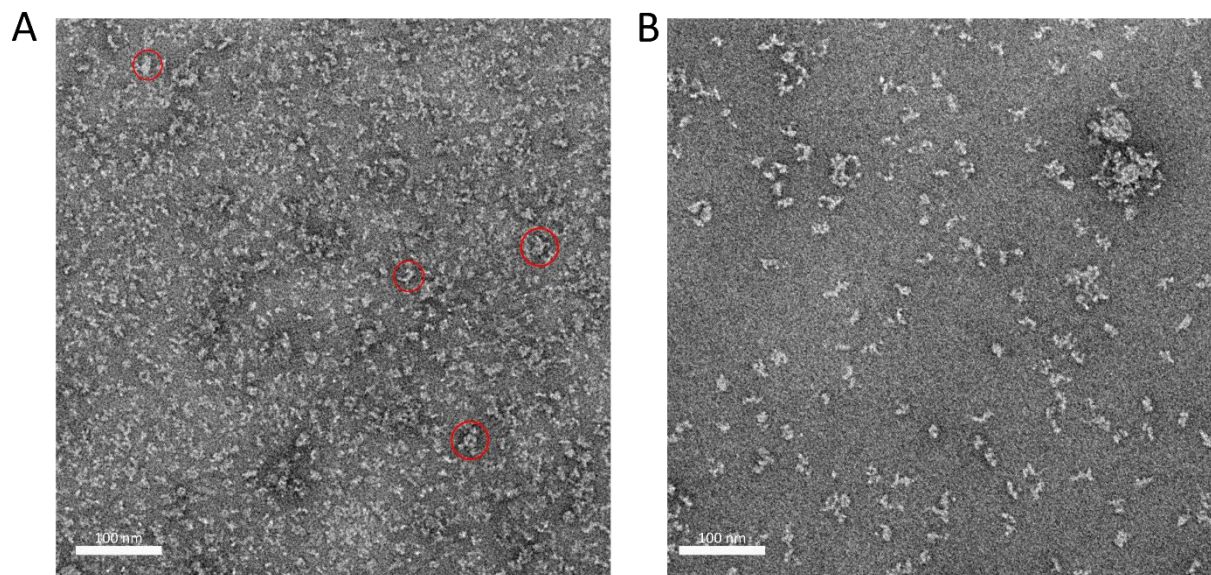


Figure 65. Observation en coloration négative de la protéine ClbK pontée par du glutaraldéhyde (GA). A) ClbK dans son tampon de filtration sur gel (contenant 200 mM NaCl) a été pontée par 0.2 % GA pendant 15 minutes. Les particules entourées en rouge ont un diamètre supérieur à 15 nm. B) ClbK dialysée dans un tampon sans sel a été pontée par 0.5 % GA pendant 15 minutes.

La stratégie de stabilisation par réticulation a aussi été employée sur un échantillon de ClbK dans un tampon sans chlorure de sodium. Le test de réticulation de ClbK en présence de quantités croissantes de GA a indiqué que des quantités plus importantes de GA sont nécessaires pour ponter le dimère de ClbK en l'absence de sel. L'échantillon de ClbK sans sel a donc été ponté avec 0.5 % GA. L'observation en coloration négative de cet échantillon montre des particules de taille importante mais les particules sont très hétérogènes et parfois même agrégées (Figure 65B). On peut supposer que la quantité plus importante d'agent de réticulation a causé l'agrégation d'une partie des protéines. L'hétérogénéité de l'échantillon pourrait être due à la plus grande quantité de GA ou à l'absence de sel.

La réticulation en elle-même ou couplée à l'absence de chlorure de sodium n'a donc pas permis de stabiliser la forme dimérique de ClbK pour une étude en cryo-ME. La technique du GraFix repose aussi sur l'utilisation d'agent de réticulation couplée à une ultracentrifugation (Stark, 2010). L'échantillon à stabiliser est déposé au sommet d'un tube contenant à la fois un gradient de glycérol et un gradient de GA, les plus fortes concentrations étant au fond du tube. Ce protocole permet de séparer différentes populations d'un échantillon après ultracentrifugation. Ce protocole a été testé plusieurs fois sur ClbK mais n'a pas permis de préparer un échantillon de qualité suffisante pour la ME. En effet, lors de la plupart des tests, ClbK est retrouvée dans la fraction au fond du tube, suggérant qu'elle est précipitée. De plus, les fractions contenant ClbK sont très concentrées en glycérol, ce qui est problématique pour la cryo-ME donc cette stratégie n'a pas été poursuivie.

#### 1.9. Bilan de l'étude de ClbK entière et perspectives

Durant cette thèse, un protocole de purification a été établi pour ClbK et ce protocole a été validé par le test de substrat du domaine d'adénylation de ClbK. Les expériences de biophysique ont montré que la protéine ClbK est dimérique, ce qui suggère que le module PKS de ClbK est responsable de la dimérisation de la protéine entière. Les premières études structurales ont été menées sur ClbK, notamment une étude de SEC-SAXS qui a permis de proposer des modèles de cette protéine. Les premières tentatives de cristallisation de ClbK n'ont quant à elles pas abouti, probablement à cause de la flexibilité de la protéine. La dissociation du dimère de ClbK empêche pour l'instant son analyse par cryo-ME.

De nombreuses options sont envisageables pour tenter d'obtenir la structure de ClbK entière. L'une d'elle est l'utilisation d'anticorps qui pourraient à la fois permettre de stabiliser une ou plusieurs conformations de ClbK et d'empêcher la dissociation du dimère. Cette stratégie a permis l'obtention de plusieurs structures à haute résolution des PKS Lsd14 et du premier module de DEBS (Bagde *et al.*, 2021 ; Cogan *et al.*, 2021 ; Guzman & Khosla, 2022) (Figure 25). Toutefois, ces anticorps reconnaissent principalement les hélices N-terminales de certaines PKS, les « docking domains », absentes chez ClbK. Une collaboration est en cours afin de synthétiser des nanocorps, des fragments d'anticorps de camélidés, dirigés spécifiquement contre ClbK (Uchański *et al.*, 2020). En effet, la petite taille de ces nanocorps et leur capacité à reconnaître spécifiquement une région variable ont fait de ces outils un véritable atout des études de biologie structurale. Ces nanocorps pourraient stabiliser une conformation particulière de ClbK et/ou stabiliser sa forme dimérique.

#### I.10. Choix d'une stratégie de « dissection-reconstruction » de ClbK

L'obtention de structures de NRPS/PKS contenant plus d'un module a souvent nécessité l'étude de fragments multi-domaines de ces méga-enzymes (Reimer, Aloise, Harrison, *et al.*, 2016 ; Reimer *et al.*, 2019 ; Davison *et al.*, 2014 ; Dorival *et al.*, 2016). Nous avons donc décidé de découper ClbK en fragments et de tester leur stabilité afin de découvrir les fragments les plus prometteurs pour l'obtention d'une structure à haute résolution. Ces données structurales combinées aux données à basse résolution sur la protéine entière obtenues par SAXS, pourraient permettre de construire un modèle de la protéine entière. Un autre objectif de cette approche était l'identification des éléments de ClbK nécessaires à sa dimérisation.

Une expérience de protéolyse limitée à la trypsine a d'abord été menée afin d'identifier des fragments stables de ClbK mais aucun fragment stable n'a pu être mis en évidence. Il a donc été décidé d'utiliser les structures de domaines homologues afin de délimiter des fragments de ClbK. Huit fragments ont été construits (Figure 30) et chacun de ces fragments a été cloné dans deux vecteurs, les plasmides pETM-11 (N-His) et pETM-30 (N-His, N-GST). Ainsi, l'expression, la solubilité et la fixation sur résine d'affinité ont été testées pour ces huit fragments. Quatre de ces fragments ont pu être exprimés et purifiés en quantité suffisante pour des analyses biochimiques ou structurales (Tableau 6).



Fragment	Rendement purification	Etat oligomérique	Cristallographie	ME	SEC-SAXS
ClbK	+	Dimérique	Plusieurs tests, cristaux de sels	Dissociation du dimère	Oui
ClbK-PKS	+++	Monomérique en solution, dimérique en cristallographie	Structure cristallographique	Homogène mais trop petit	Oui
ClbK-PKS-LK	+++	Monomérique	Ne cristallise pas dans les mêmes conditions que ClbK-PKS	Non	Non
ClbK-PKS-Cy	++	Monomérique	Plusieurs tests, pas de cristaux	Hétérogène	Oui
ClbK-PKS-A <sub>core</sub>	++	Monomérique	Non	Non	Non
ClbK-PKS-A <sub>sub</sub>	0	Non	Non	Non	Non
ClbK-NRPS <sub>court</sub>	0	Non	Non	Non	Non
ClbK-NRPS <sub>long</sub>	0	Non	Non	Non	Non
ClbK-A-Ox-A	0	Non	Non	Non	Non

Tableau 6. Récapitulatif des expériences menées sur les différentes constructions de ClbK. Pour les rendements de purification, +++, ++ et + signifient que plusieurs dizaines de milligrammes, plusieurs milligrammes ou moins d'un milligramme de protéine pure par litre de culture ont été obtenus. 0 indique que la protéine n'a pas pu être purifiée. Pour les colonnes 2 à 5, non signifie que l'expérience n'a pas été réalisée.

## II. Etudes biochimiques et structurales de ClbK-PKS

### II.1. Objectifs de l'étude de ClbK-PKS

La construction ClbK-PKS (85 kDa) contient les domaines KS, AT\* et ACP de ClbK mais n'inclut pas le *linker* connectant les modules PKS et NRPS (Figure 30). Un des objectifs de l'étude de la construction ClbK-PKS était de vérifier si, comme attendu, le module PKS était responsable de la dimérisation de la protéine hybride PKS-NRPS ClbK. L'autre objectif était la détermination de la structure à haute résolution de ce module PKS. Cette structure permettrait l'investigation des spécificités d'un module PKS à l'interface entre deux modules NRPS (Figure 29).

### II.2. Purification de ClbK-PKS

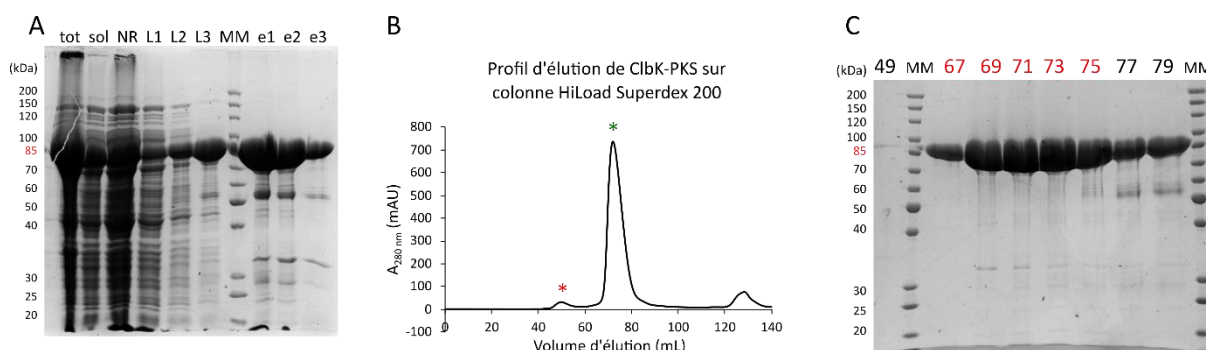


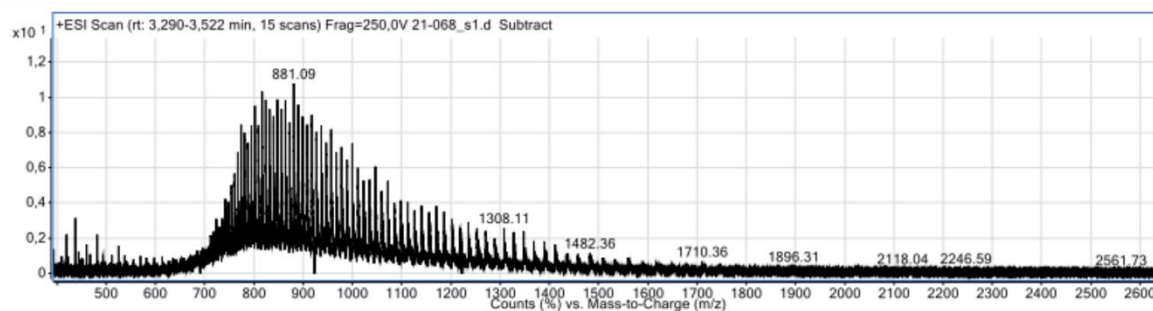
Figure 66. Purification de ClbK-PKS en deux étapes. A) Gel SDS 10 % visualisant la pureté de ClbK-PKS aux étapes de lyse et de chromatographie d'affinité sur résine de nickel. tot : fraction totale, sol : fraction soluble, NR : fraction non retenue, L1-L2-L3 : fractions de lavage 1, 2 et 3 avec du tampon de lyse contenant 0, 10 et 20 mM d'imidazole, MM : marqueur de masse molaire, e1-e2-e3 : fractions d'élution 1, 2 et 3. B) Profil d'élution de ClbK-PKS en filtration sur gel. Les pics indiqués par des astérisques contient ClbK-PKS. C) Gel SDS 10 % attestant de la pureté de ClbK-PKS après filtration sur gel. Les numéros au-dessus du gel correspondent aux volumes des fractions du profil en B).

Le fragment ClbK-PKS s'exprime en grande quantité, en particulier dans le milieu d'auto-induction. Cette protéine est soluble et s'accroche bien à la résine d'affinité quelle que soit l'étiquette présente en N-terminal. L'échantillon est suffisamment pur après chromatographie d'affinité pour être directement injecté sur colonne de filtration sur gel (Figure 66A). 5 mL de la fraction d'élution la plus concentrée ( $5 \text{ mg.mL}^{-1}$ ) sont injectés sur une colonne Superdex 200 de 120 mL. Le profil d'élution présente un pic minoritaire à 50 mL et un pic largement majoritaire à 72 mL (Figure 66B). Bien qu'un contaminant de 50 kDa soit retrouvé dans les dernières fractions du pic majoritaire, ce protocole permet d'obtenir environ

40 milligrammes de protéine pure par litre de culture (Figure 66C). L'étiquette poly-histidines n'a pas été coupée pour les premières caractérisations de ClbK-PKS.

### II.3. Détermination de l'état de modification post-traductionnelle de ClbK-PKS

A



B

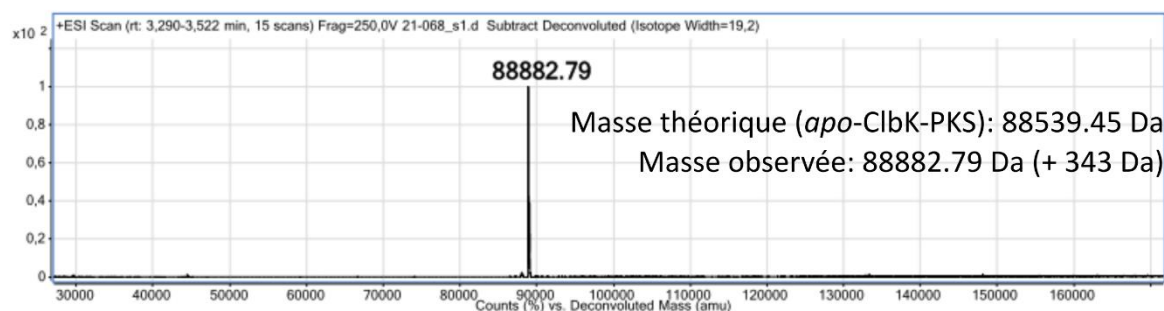


Figure 67. Analyse de ClbK-PKS par spectrométrie de masse (méthode LC-ESI TOF). A) Spectre de masse de ClbK-PKS. B) Spectre déconvolué.

Une analyse de ClbK-PKS par spectrométrie de masse a été menée afin de confirmer l'identité de la protéine purifiée et son état de modification post-traductionnelle. En effet, la protéine ClbK-PKS, comme ClbK, est exprimée dans la souche BL21(DE3) Bap1 d'*E. coli* (Pfeifer *et al.*, 2001) qui exprime aussi une PPTase ajoutant un bras PPant sur une sérine conservée des domaines CP (Figure 13). La masse observée pour ClbK-PKS présente un excédent de 343 Da par rapport à la masse attendue pour la forme *apo* de ClbK-PKS (Figure 67). Cet excès de masse correspond à la masse attendue pour le bras PPant, ce qui indique que ClbK-PKS est purifiée sous sa forme *holo*.

## II.4. Détermination de l'état oligomérique de ClbK-PKS

L'état oligomérique de ClbK-PKS, comme celui de ClbK entière, a été étudié par SEC-MALLS (Figure 68). La protéine a été concentrée à  $12 \text{ mg.mL}^{-1}$  avant injection de  $50 \text{ }\mu\text{L}$  sur une colonne Superdex 200 (25 mL). Les deux pics observés sont monodisperses. La masse molaire estimée pour le pic majoritaire, à 13.5 mL, est d'environ 81 kDa, ce qui correspond à la forme monomérique de ClbK-PKS (masse théorique : 89 kDa). La masse molaire estimée pour le pic minoritaire, à 12 mL, est d'environ 171 kDa, indiquant que ce pic contient la forme dimérique de ClbK-PKS. ClbK-PKS est donc majoritairement monomérique en solution.

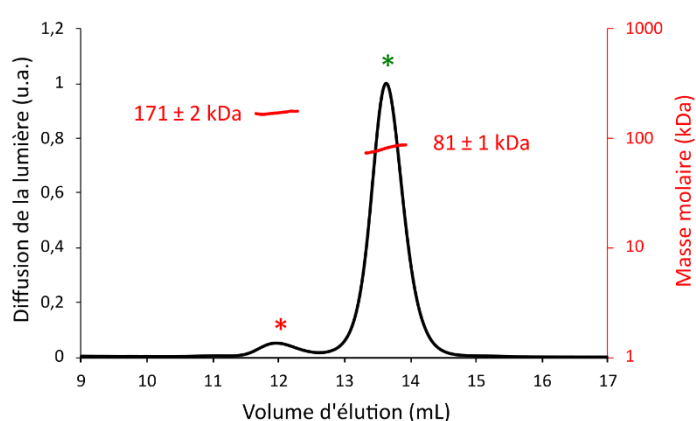


Figure 68. Analyse de ClbK-PKS par SEC-MALLS sur colonne Superdex 200 Increase 10/300 GL (25 mL). La masse molaire théorique du monomère de ClbK-PKS est 89 kDa donc le premier pic (astérisque rouge) contient la forme dimérique de la protéine et le deuxième pic (astérisque vert) contient sa forme monomérique.

L'état monomérique de ClbK-PKS est inattendu. En effet, les PKS sont généralement dimériques : elles dimérisent notamment par le domaine KS, que ClbK-PKS possède (A. T. Keatinge-Clay, 2012). Toutefois, la plupart des PKS contiennent d'autres éléments de dimérisation, tels que les « docking domains » ou le domaine déshydratase par exemple (Smith *et al.*, 2021 ; D. Gay *et al.*, 2013) et la dissociation de PKS dimériques a déjà été observée (Herbst *et al.*, 2016 ; Herbst, Huitt-Roehl, *et al.*, 2018). L'état oligomérique de ClbK-PKS dans le contexte de la protéine entière sera discutée dans les sections suivantes.

## II.5. Cristallisation de ClbK-PKS

### II.5.a. Premiers cristaux et premiers tests de diffraction de la protéine ClbK-PKS

Les très bons rendements de purification de ClbK-PKS ont permis de tester de très nombreuses conditions de cristallisation pour cette protéine. Au total, plus de quarante boîtes de cristallisation 24 puits ont été utilisées afin de tester une large gamme de conditions de cristallisation pour ClbK-PKS. Les premiers tests de cristallisation ont été menés par la plateforme HTX sur la protéine ClbK-PKS concentrée à 5, 15 et  $25 \text{ mg.mL}^{-1}$ . Les cristaux les plus

prometteurs sont apparus à 15 et 25 mg.mL<sup>-1</sup>, en quelques jours, dans la condition 0.1 M MMT pH 7, 25 % PEG 1500 (Figure 69). Ces cristaux émettaient un signal en lumière UV, indiquant qu'il ne s'agissait pas de cristaux de sel mais bien de cristaux de protéines (Figure 69).

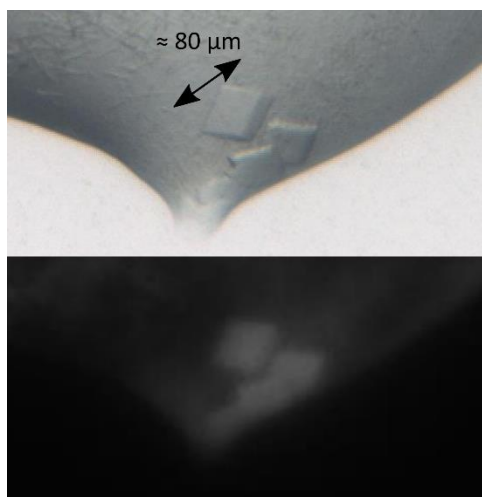


Figure 69. Premiers cristaux de la protéine ClbK-PKS (15 mg.mL<sup>-1</sup>). La condition de cristallisation est 0.1 M MMT pH 7, 25 % PEG 1500. Les cristaux sont apparus en moins de deux jours et les photographies en lumière visible (en haut) et UV (en bas) ont été prises après 10 jours.

Les conditions de cristallisation décrites dans le paragraphe précédent ont été reproduites et optimisées en plaques 24 puits en utilisant la technique de la goutte suspendue. Les cristaux ont pu être reproduits manuellement en diminuant la concentration de l'agent précipitant par rapport à la concentration initiale (0.1 M MMT pH 7, 15 % PEG 1500). Les premiers tests de diffraction de ces cristaux ont été menés sur la ligne ID30A-3 de l'ESRF. Malgré une apparente homogénéité, les cristaux de ClbK-PKS diffractent les rayons X de façon très hétérogène. En effet, dans la même goutte de cristallisation, certains cristaux n'ont pas diffracté les rayons X alors que d'autres les ont diffractés à une résolution variant de 4 à 10 Å.

De très nombreux paramètres de cristallisation ont été testés afin d'augmenter la résolution des données de diffraction. L'effet de la température, de la nature du tampon, du pH du tampon, de la nature du PEG, de la concentration en PEG, de la concentration en chlorure de sodium et de l'ajout d'additifs ont été testés. Au total, plus d'une cinquantaine de cristaux ont été testés pour la diffraction à l'ESRF mais aucun cristal de ClbK-PKS n'a diffracté à une résolution meilleure que 4 Å.

### II.5.b. Tests de cristallisation de la protéine mutée ClbK-PKS-S747A

Dans le but d'obtenir des cristaux de ClbK-PKS diffractant à plus haute résolution, un mutant ClbK-PKS-S747A a été généré. Bien que la protéine ClbK-PKS soit exprimée dans la souche BL21(DE3) Bap1 (Pfeifer *et al.*, 2001) et que la spectrométrie de masse ait confirmé la

présence du bras PPant sur ClbK-PKS (Figure 67), il est possible qu'une minorité de ClbK-PKS ne soit pas modifiée post-traductionnellement. Le mélange de formes *apo* et *holo* de la protéine ClbK-PKS pourrait générer des imperfections dans le cristal et ainsi diminuer la qualité des données de diffraction. Par exemple, les cristaux de la PKS Lsd14 purifiée sous forme *holo* diffractent peu les rayons X alors que les cristaux de la même protéine purifiée sous forme *apo* les diffractent à 2.4 Å (Bagde *et al.*, 2021). Une façon de ne travailler qu'avec la forme *apo* d'une protéine NRPS/PKS est de muter en alanine la sérine modifiable par le bras PPant (Tanovic *et al.*, 2008 ; Katsuyama *et al.*, 2021).

Le mutant ClbK-PKS-S747A a été exprimé et purifié dans des conditions quasi-identiques à la protéine non mutée. Le mutant a été exprimé dans la souche BL21(DE3) plutôt que la souche BL21(DE3) Bap1 employée pour la protéine non mutée et la concentration de chlorure de sodium dans le tampon de filtration sur gel a été diminuée de 150 mM à 50 mM. Le profil de filtration sur gel et le rendement de purification sont les mêmes que pour la protéine non mutée (Figure 66). Des tests de cristallisation ont été menés par la plate-forme HTX sur ce mutant, concentré à 5, 15 et 25 mg.mL<sup>-1</sup>. Ces tests, ainsi que des tests dans des boîtes 24 puits, ont montré que le mutant ClbK-PKS-S747A ne cristallise pas dans les conditions de cristallisation de la protéine non mutée. Des conditions de cristallisation différentes n'ont pas non plus pu être mises en évidence à partir de ces premiers tests de cristallisation. Le T<sub>m</sub> de la protéine mutée concentrée à 15 mg.mL<sup>-1</sup> est seulement de 39°C alors que la protéine ClbK-PKS non mutée présente un T<sub>m</sub> de 44°C à la même concentration. Ceci suggère que la mutation S747A diminue la stabilité de la protéine ClbK-PKS. L'étude de la stabilité de ClbK-PKS-S747A dans différents tampons par Thermofluor n'ayant pas permis de mettre en évidence un tampon stabilisant ce mutant, les tests de cristallisation sur ce mutant n'ont pas été poursuivis.

### **II.5.c. Cristallisation et diffraction des cristaux de ClbK-PKS après modification du protocole de purification (ClbK-PKS\*)**

L'étiquette poly-histidines peut favoriser ou empêcher la cristallisation selon la protéine. Nous avons décidé de cliver l'étiquette poly-histidines de la protéine ClbK-PKS sous forme *holo* et de tester la cristallisation de cette protéine, que nous appellerons ici ClbK-PKS\*. Le même protocole d'expression et de chromatographie d'affinité que pour ClbK-PKS a été utilisé. L'étiquette poly-histidines a ensuite été clivée par la protéase TEV et la protéine ClbK-

PKS\* générée a été purifiée par une deuxième chromatographie d'affinité. Les fractions non retenues sur la résine d'affinité contiennent la protéine ClbK-PKS\*, qui a été purifiée par une dernière étape de filtration sur gel (Figure 70A). Un profil d'élution similaire à celui de la protéine ClbK-PKS est observé, indiquant que la protéine ClbK-PKS\* est également majoritairement monomérique en solution (Figure 70B). Environ 20 milligrammes de protéine ClbK-PKS\* pure par litre de culture sont obtenus avec ce protocole. La caractérisation par spectrométrie de masse de la protéine ClbK-PKS\* a indiqué que l'étiquette poly-histidines est bien clivée et que la protéine est purifiée sous forme *holo* (Figure 70C).

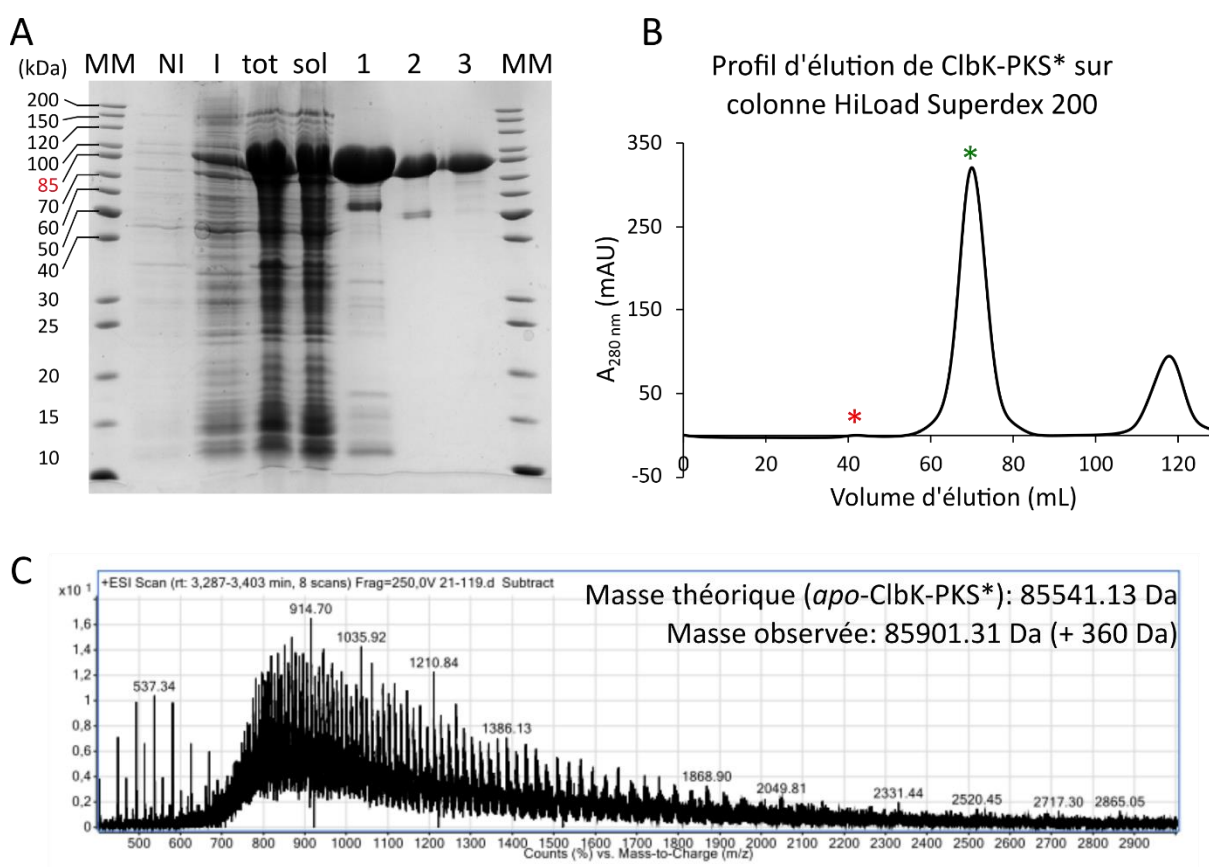
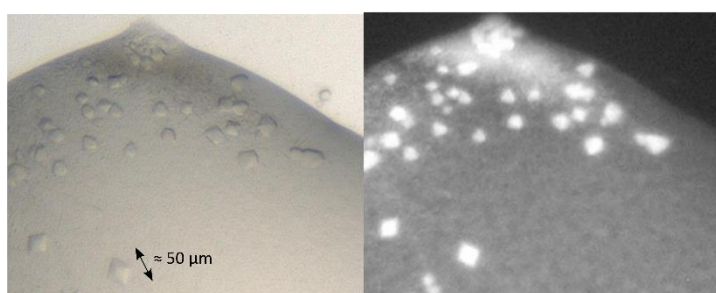


Figure 70. Purification de la protéine ClbK-PKS\* et caractérisation par spectrométrie de masse. A) Evaluation de la pureté de ClbK-PKS\* par gel SDS 12.5 %. MM : marqueur de masse molaire, NI : fraction avant induction, I : fraction induite, tot : fraction totale, sol : fraction soluble, 1 : fraction d'élution de chromatographie d'affinité, 2 : fraction après clivage de l'étiquette poly-histidines et chromatographie d'affinité inverse, 3 : fraction après chromatographie d'exclusion de taille. B) Profil d'élution de ClbK-PKS\* sur colonne d'exclusion de taille. C) Spectre de la protéine ClbK-PKS\* caractérisée par spectrométrie de masse. La différence d'environ 3 kDa entre les masses observées de ClbK-PKS (Figure 67) et ClbK-PKS\* confirme le clivage de l'étiquette poly-histidines pour la protéine ClbK-PKS\*. L'écart de 360 Da entre la masse observée et la masse théorique indique que la protéine ClbK-PKS\* est produite sous forme *holo*.

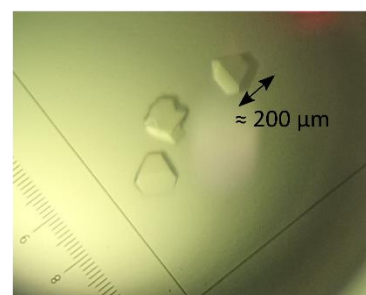
Les premiers tests de cristallisation menés par la plate-forme HTX sur la protéine ClbK-*PKS\** ont indiqué que cette protéine cristallise à la fois dans certaines des conditions de cristallisation de la protéine ClbK-*PKS* et dans de nouvelles conditions. Les nouvelles conditions (Figure 71A) n'ont pas pu être reproduites manuellement. En revanche, les conditions similaires aux conditions de cristallisation de ClbK-*PKS* ont pu être reproduites manuellement (Figure 71B). On peut noter la stabilité de la protéine ClbK-*PKS\**, qui cristallise après plusieurs semaines à 4 degrés et après congélation-décongélation à -80°C. La diffraction d'une centaine de cristaux a été testée sur la ligne automatique de l'ESRF, ID30-A-1 (Bowler *et al.*, 2015). Un de ces cristaux, obtenus à température ambiante dans la condition 0.1 M MES pH 6.5, 19 % PEG 3350 a diffracté les rayons X à 3 Å.

A



ClbK-*PKS\**, 15 mg.mL<sup>-1</sup>  
 0.1 M Bicine pH 9  
 10 % PEG 20000  
 2 % 1,4-Dioxane

B



ClbK-*PKS\**, 15 mg.mL<sup>-1</sup>  
 0.1 M MES pH 6.5  
 15 % PEG 3350

Figure 71. Cristaux de la protéine ClbK-*PKS\**. A) Cristaux obtenus par la plate-forme HTX. Les cristaux sont apparus en moins de deux jours et la photographie a été prise après six jours en lumière visible à gauche et lumière UV à droite. B) Cristaux obtenus après optimisation en plaque 24 puits. Les cristaux sont apparus en moins de deux jours et la photographie a été prise après cinq jours.

## II.6. Résolution de la structure de ClbK-*PKS\**

Afin de résoudre le problème de la phase, des méthodes de phasage expérimental par diffusion anormale ont d'abord été testées. La protéine ClbK-*PKS* a été produite sous forme sélénométhionée et la caractérisation de la protéine par spectrométrie de masse a montré que les 19 méthionines de la protéine native sont des sélénométhionines dans la protéine sélénométhionée. Les essais de cristallisation menés par la plate-forme HTX n'ont pas mis en évidence de condition de cristallisation pour cette protéine. En parallèle, des tentatives de co-cristallisation de ClbK-*PKS* avec deux molécules de cristallophores ont été menées (Engilberge *et al.*, 2017). Les tests de cristallisation par la plate-forme HTX ont montré que les cristallophores induisent la précipitation de la protéine ClbK-*PKS* dans la plupart des gouttes.



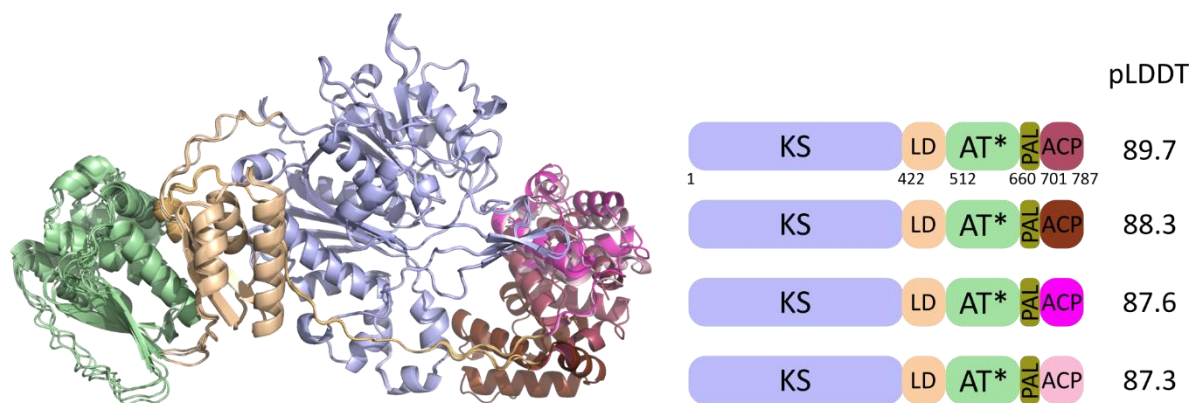


Figure 72. Modèles du module PKS de ClbK générés par AlphaFold2, alignés sur le domaine KS. Le pLDDT est un indicateur de la confiance que l'on peut accorder au modèle.

Une stratégie de remplacement moléculaire a ensuite été testée pour résoudre la structure de ClbK-PKS\*. Les paramètres de mailles suggéraient que l'unité asymétrique contenait six molécules de ClbK-PKS\*. L'absence d'homologues pour le domaine ClbK-AT\* et le nombre élevé de molécules dans la maille rendaient le remplacement moléculaire très compliqué. Des modèles de ClbK-PKS\* ont donc été générés avec le logiciel AlphaFold2 afin d'obtenir un bon modèle de remplacement moléculaire. Parmi les cinq modèles générés, quatre présentent un bon score associé à la prédiction, avec un pLDDT avoisinant les 0.90 (Figure 72). Dans ces quatre modèles, le di-domaine KS-AT\* se superpose bien mais le domaine ACP occupe des positions légèrement différentes entre les modèles. L'utilisation d'un modèle tronqué sans le domaine ACP comme modèle de remplacement moléculaire a permis la résolution de la structure de ClbK-PKS\*. Les statistiques des données de diffraction et d'affinement pour le meilleur jeu de données de ClbK-PKS\* sont résumées dans le Tableau 7.

Données de diffraction	
Résolution (Å)	48.98-3.0 (3.107-3.0)
Groupe d'espace	C2221
Paramètres de maille (Å)	188.52 234.51 257.79 90 90 90
Réflexions totales	223372 (22191)
Réflexions uniques	113539 (11185)
Redondance	2.0 (2.0)
Complétude (%)	99.65 (99.19)
I/σ	7.32 (0.39)
Wilson facteur B (Å <sup>2</sup> )	121.67
R-merge	0.09687 (2.017)
CC1/2	0.995 (0.157)

Affinement	
Réflexions utilisées	113466 (11185)
R-work	0.2189 (0.4148)
R-free	0.2576 (0.4495)
Résidus	2891
Rmsd liaisons (Å)	0.014
Rmsd angles (°)	2.20
Ramachandran favorisé (%)	83.94
Ramachandran autorisé (%)	11.41
Facteur B total (Å <sup>2</sup> )	114.51

Tableau 7. Statistiques des données de diffraction et d'affinement du meilleur jeu de données collecté sur ClbK-PKS\*. Les statistiques pour la tranche à plus haute résolution sont indiquées entre parenthèses.

## II.7. Analyse de la structure cristallographique de ClbK-PKS\*

### II.7.a. ClbK-PKS\* dimérise par le domaine KS

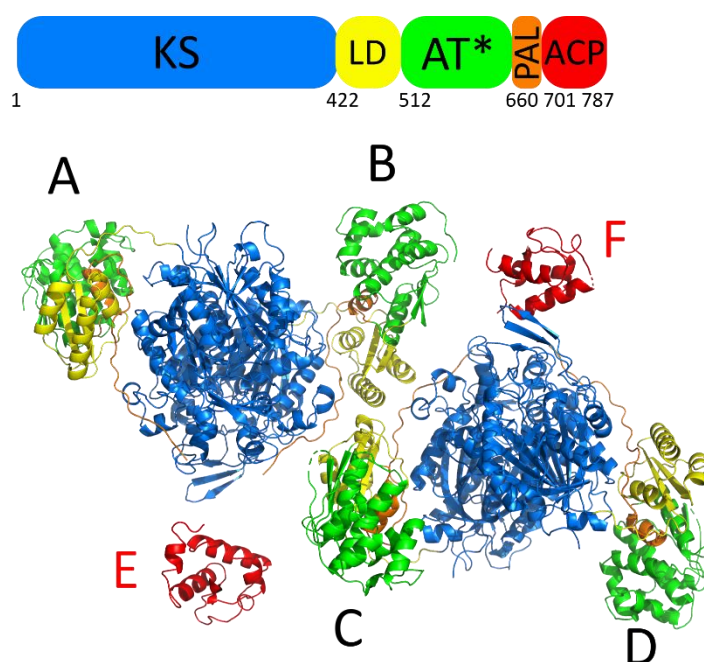


Figure 73. Structure de ClbK-PKS\*. Quatre molécules sont présentes dans l'unité asymétrique, formant deux homodimères. Un seul domaine ACP par dimère a été reconstruit mais cet ACP n'a pas pu être assigné à un monomère en particulier. KS : cétosynthase, LD : domaine de liaison KS-AT\*, AT\* : acyltransférase tronqué, PAL : « post-AT linker », ACP : « acyl carrier protein ».

La structure obtenue de ClbK-PKS\* comprend quatre molécules dans l'unité asymétrique, qui s'arrangent en deux homodimères : le dimère AB, formé par les chaînes A et B, et le dimère CD, formé par les chaînes C et D (Figure 73). Le RMSD de la superposition du dimère AB sur le dimère CD est de 0.210 Å, indiquant que les deux homodimères sont quasiment équivalents. Le seul élément de ClbK-PKS\* participant à la dimérisation est le domaine KS. L'alignement du dimère de KS de ClbK avec le dimère de KS du module 5 de DEBS (PDB : 2HG4) (Y. Tang *et al.*, 2006) et d'OzmQ-KS1 (PDB : 4OQJ) (Lohman *et al.*, 2015) indique un RMSD de 0.947 Å et 0.747 Å respectivement. L'interface de dimérisation de ClbK-PKS\* est

donc similaire à l'interface de dimérisation des autres PKS, qu'elles soient de type *cis*-AT ou *trans*-AT.

Les analyses de l'état oligomérique de ClbK-PKS ont montré que la protéine est majoritairement monomérique en solution (Figure 68) alors que ClbK-PKS\* est dimérique dans le cristal. On peut supposer que ClbK-PKS\* est en équilibre monomère-dimère en solution et que, dans les conditions de cristallisation décrites précédemment, le dimère cristallise préférentiellement au monomère, favorisant ainsi l'équilibre monomère-dimère dans le sens de la dimérisation. L'interface de dimérisation conservée avec les autres PKS suggère que le dimère observé en cristallographie n'est pas un artefact de cristallisation. Bien que la surface de dimérisation du domaine KS de ClbK-PKS\* (2070 Å<sup>2</sup>) soit comparable aux surfaces de dimérisation d'autres domaines KS, l'architecture dimérique des autres PKS repose souvent sur d'autres éléments de dimérisation que le domaine KS, tels que les « docking domains » N-terminaux et C-terminaux ou le domaine déshydratase (Smith *et al.*, 2021 ; A. Keatinge-Clay, 2008). L'absence d'éléments de dimérisation autres que le domaine KS chez ClbK-PKS\* pourrait expliquer sa dissociation en monomères en solution.

La dissociation de PKS dimériques en monomères a été observée à plusieurs reprises. Il a été montré que la dissociation du dimère d'une des PKS synthétisant l'érythromycine est limitée dans un tampon concentré en phosphate ou en citrate (Cogan *et al.*, 2021). Afin d'évaluer l'effet du citrate et du phosphate sur la dimérisation de ClbK-PKS\*, plusieurs chromatographies d'exclusion de taille de ClbK-PKS\* ont été réalisées dans différents tampons. Le même profil que celui observé en SEC-MALLS pour ClbK-PKS (Figure 68) est observé pour tous les tampons sauf celui contenant 500 mM de citrate. Étonnamment, le pic d'élution est décalé vers la fin de la colonne dans ce tampon, suggérant que la protéine est plus petite (Figure 74A). Bien que la dimérisation de ClbK-PKS\* dans le tampon citrate n'ait pu être observée en gel natif (Figure 74B), cette dimérisation dans le tampon citrate est suggérée par les analyses en photométrie de masse (Figure 74C et D). En effet, dans le tampon contrôle (25 mM Hepes pH 7.5, 50 mM NaCl), ClbK-PKS\* forme majoritairement des monomères alors que dans le tampon citrate (25 mM Hepes pH 7.5, 500 mM citrate), une quantité importante de ClbK-PKS\* dimérique est observée.

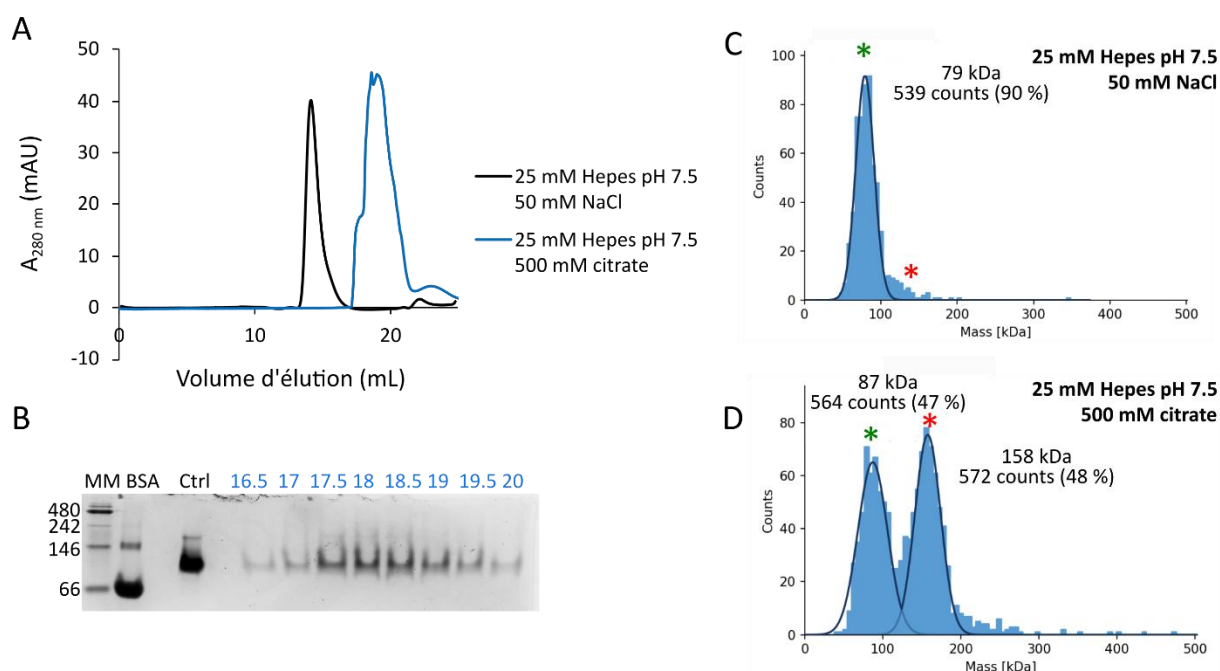


Figure 74. Effet du tampon sur l'état oligomérique de ClbK-PKS\*. A) Profils d'élution superposés de ClbK-PKS\* sur colonne Superdex 200 dans deux tampons différents. B) Gel natif des fractions de ClbK-PKS\* dans ces deux tampons. MM : marqueur de masse molaire, BSA : albumine de sérum bovin, contrôle de migration (66 et 132 kDa), ctrl : fraction la plus concentrée dans le tampon 25 mM Hepes pH 7.5, 50 mM NaCl. Les numéros des fractions en bleu correspondent aux volumes des fractions du profil d'élution dans le tampon citrate en A. C) Analyse d'une fraction de ClbK-PKS\* dans le tampon 25 mM Hepes pH 7.5, 50 mM NaCl par photométrie de masse. D) Analyse de la fraction à 17.5 mL de ClbK-PKS\* dans le tampon 25 mM Hepes pH 7.5, 500 mM citrate par photométrie de masse. La masse molaire estimée et le nombre d'événements enregistrés sont indiqués au-dessus de chaque pic. Masse molaire théorique du monomère de ClbK-PKS\* : 85.5 kDa. Pour les figures C et D, le pic marqué par un astérisque vert correspond à la forme monomérique de ClbK-PKS\* alors que le pic marqué par un astérisque rouge correspond à la forme dimérique de ClbK-PKS\*.

L'effet du tampon citrate, s'il est confirmé, pourrait être utilisé pour stabiliser la forme dimérique de ClbK-PKS\* pour des études en microscopie électronique. En effet, l'observation de ClbK-PKS\* en coloration négative a montré que la protéine est homogène mais trop petite pour l'étude par cryo-ME. L'effet du citrate pourrait aussi être évalué sur la protéine ClbK entière afin de stabiliser la forme dimérique de ClbK en ME.

## II.7.b. Description globale de la structure de ClbK-PKS\*

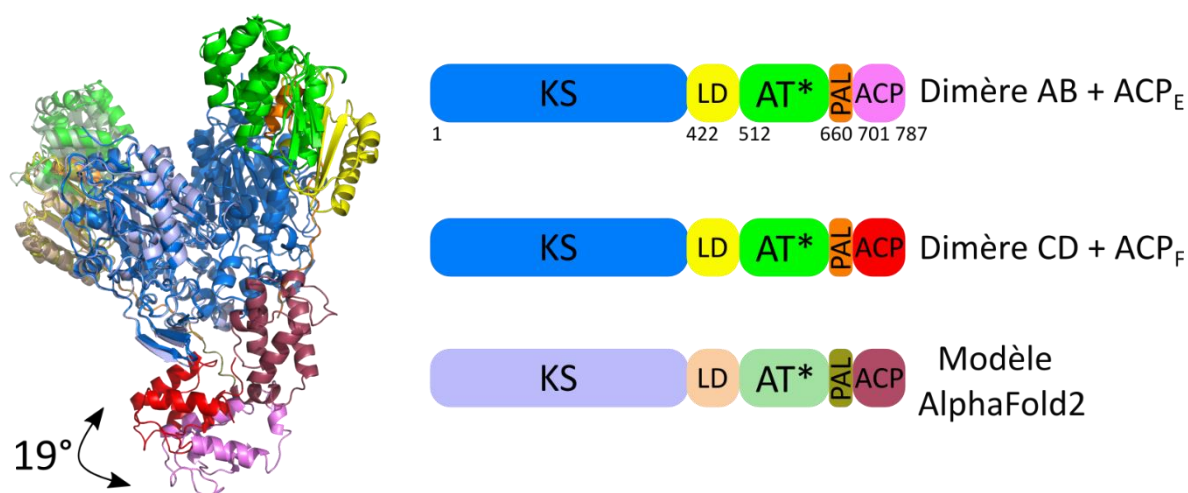


Figure 75. Superposition des deux dimères de ClbK-PKS\* avec le meilleur modèle généré par AlphaFold2.

Dans la structure de ClbK-PKS\*, les domaines cétyosynthase (KS, résidus 1 à 421), le domaine de liaison KS-AT (LD, résidus 422 à 511), le domaine acyltransférase tronqué (AT\*, résidus 512 à 659) et le *linker* post-AT (PAL, résidus 660 à 700) sont présents en quatre exemplaires alors que seulement deux domaines « acyl carrier protein » (ACP, résidus 701 à 787) ont pu être reconstruits. L'absence de densité électronique pour les derniers résidus du *linker* post-AT (PAL, résidus 695-700 manquants) ne permet pas d'assigner les domaines ACP à un di-domaine KS-AT. Les deux domaines ACP sont donc désignés par les chaînes E et F mais ne sont pas rattachés à une des quatre chaînes KS-LD-AT\*-PAL (Figure 73). La superposition du dimère CD sur le dimère AB montre que les deux ACP ont un positionnement légèrement différent dans les deux dimères. En effet, l'ACP<sub>F</sub> est plus proche du dimère CD que l'ACP<sub>E</sub> ne l'est du dimère AB et il faudrait une rotation de 19 degrés pour que l'ACP<sub>E</sub> se superpose à l'ACP<sub>F</sub> (Figure 75). Ces deux positions sont nettement différentes de la position de l'ACP du modèle AlphaFold2 : il faudrait une rotation de 145 degrés pour que l'ACP<sub>F</sub> se superpose à l'ACP du modèle AlphaFold2. La pertinence de la position de l'ACP de ClbK dans le contexte de son cycle catalytique sera discutée dans les paragraphes II.7.e et II.7.f ci-dessous.

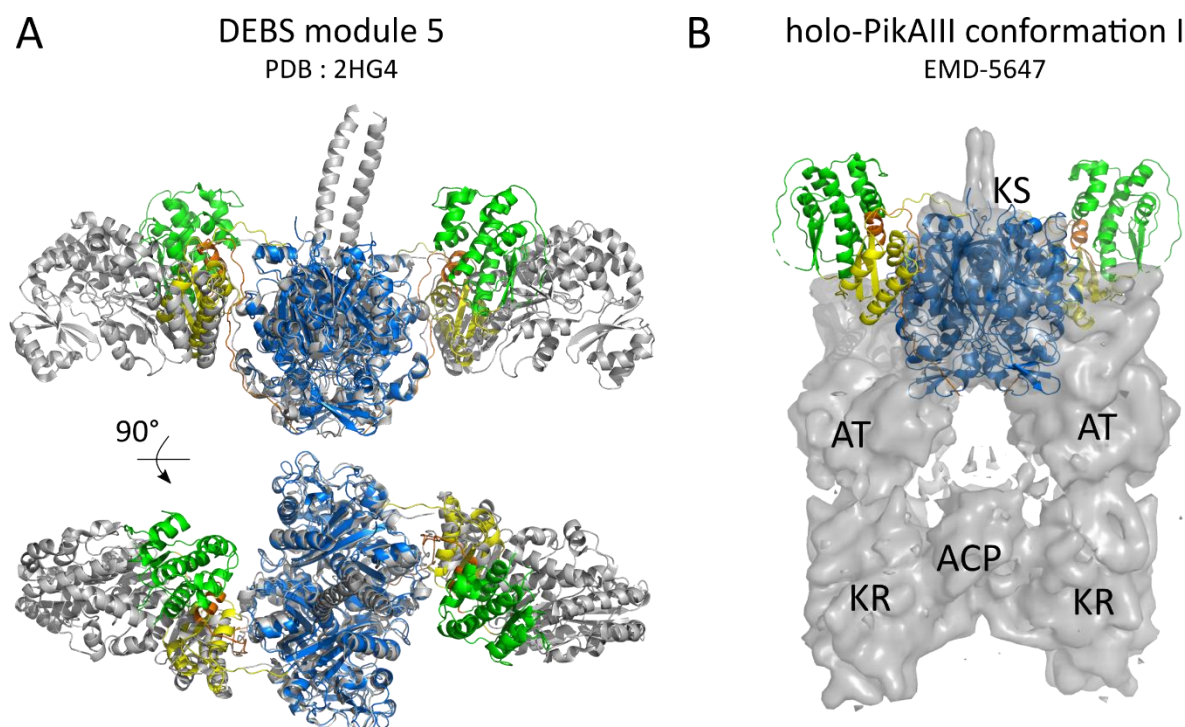
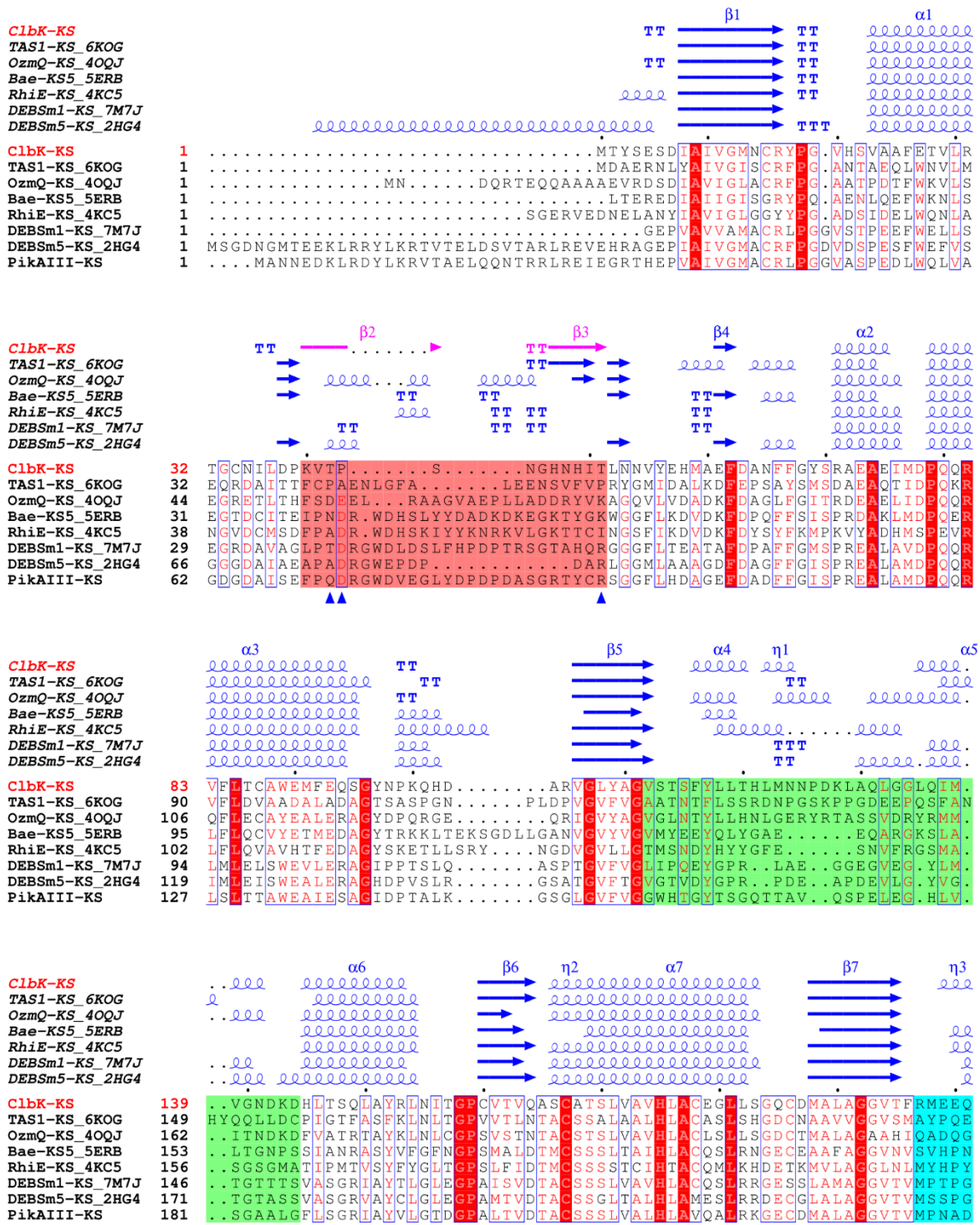


Figure 76. ClbK-PKS\* adopte une conformation étendue. ClbK est colorée avec les mêmes couleurs que dans la Figure 73 alors que les autres protéines sont colorées en gris. A) Superposition de la structure du di-domaine KS-AT\* de ClbK sur la structure cristallographique du di-domaine KS-AT appartenant au module 5 de DEBS (Y. Tang *et al.*, 2006). B) Positionnement du di-domaine KS-AT\* de ClbK dans la carte de densité électronique de PikAIII (Dutta *et al.*, 2014).

Deux conformations distinctes ont été décrites pour les di-domaines KS-AT des PKS *cis*-AT : la conformation étendue, majoritaire, notamment observée chez les protéines DEBS et la conformation en arche, décrite chez PikAIII uniquement. ClbK-PKS\* n'adopte pas cette conformation en arche mais présente une conformation globalement étendue avec un domaine AT\* légèrement rabattu vers le domaine KS par rapport aux structures canoniques des modules de DEBS (Figure 76). Cette position différente du domaine AT\* n'est pas due à un positionnement différent du LD, qui est quasiment superposable avec les structures canoniques de PKS *cis*-AT. Le positionnement resserré du domaine AT\* pourrait être caractéristique de ClbK ou être une conformation adoptée par ClbK à une étape de son cycle. En effet, il a été montré que des positionnements différents du domaine AT au cours du cycle du premier module de DEBS contrôlent l'accessibilité au site actif du domaine KS (Cogan *et al.*, 2021) (Figure 25C et D).

### II.7.c. Structure et spécificités du domaine KS de ClbK



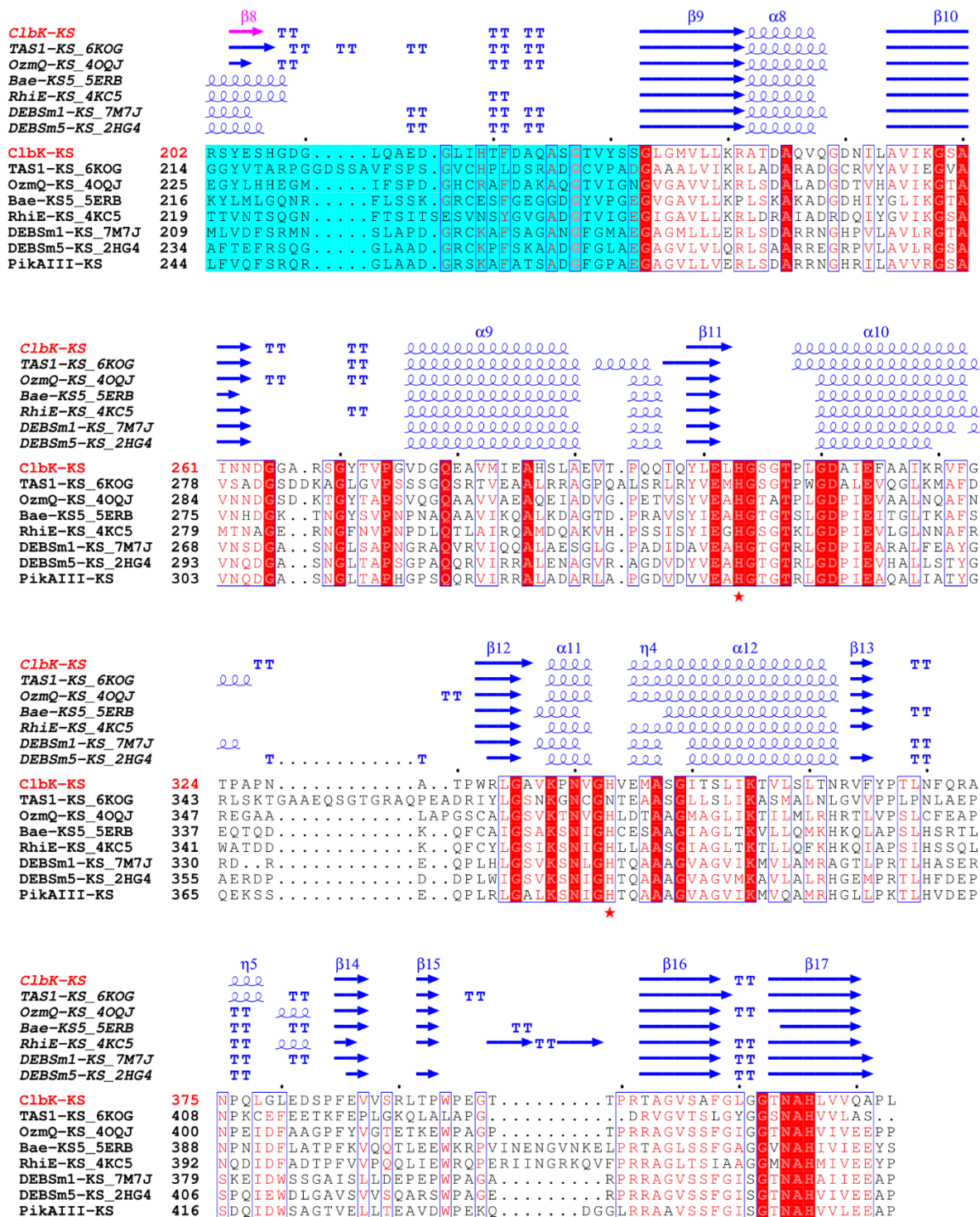


Figure 77. Analyse d'alignements de séquences de domaines KS, effectuée par le serveur ESPrpt. TAS1-KS et OzmQ-KS acceptent un substrat en provenance de NRPS, Bae-KS5 et RhiE-KS appartiennent à des PKS *trans*-AT et les KS de DEBS et PikAIII appartiennent à des PKS *cis*-AT. La *clasping loop*, la *dimer interface loop* et l'*active site cap* sont surlignées en saumon, vert pâle et cyan respectivement. Les trois résidus catalytiques sont indiqués par une étoile rouge. Les résidus de ClbK interagissant avec l'ACP sont indiqués par des triangles bleus. Les résidus en blanc sur un fond rouge sont strictement conservés. Dans une colonne entourée en bleu, au moins 70 % des résidus ont des propriétés physico-chimiques similaires et ces résidus similaires sont en rouge.



Le domaine KS de ClbK catalyse deux réactions (Figure 29). La première est l'aminacylation, c'est-à-dire le chargement du substrat donneur en provenance de ClbJ sur la cystéine 167 du domaine KS de ClbK. La deuxième est la condensation décarboxylative entre le substrat donneur et le substrat accepteur, l'unité aminomalonyl chargée sur le bras PPant de l'ACP. Le domaine KS de ClbK présente le repliement thiolase caractéristique des domaines KS et les résidus catalytiques cystéine 167, histidine 303 et histidine 343 sont conservés (A. T. Keatinge-Clay, 2012) (Figure 77).

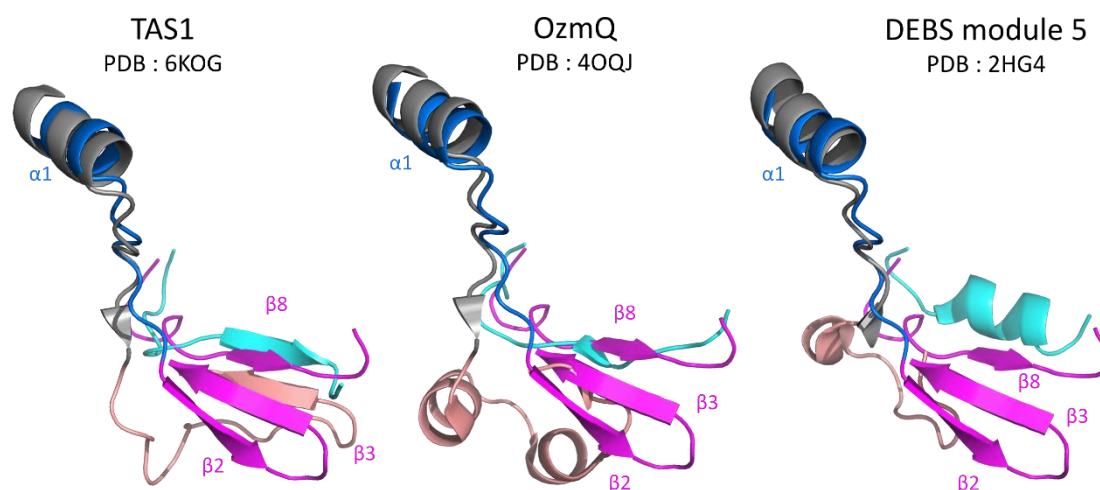


Figure 78. Alignement structural de la principale région variable du domaine KS de ClbK avec la même région chez d'autres KS. Les structures secondaires de ClbK sont numérotées. La *clasping loop* (β2 et β3) et l'*active site cap* (β8) de ClbK sont colorés en magenta alors que les régions homologues sont colorées en saumon et cyan pour les autres domaines KS.

Une étude comparative de plusieurs domaines KS de PKS *trans*-AT a permis de mettre en évidence plusieurs régions variables en séquence et en structure des domaines KS (Lohman *et al.*, 2015) (Figure 77 et Figure 78). L'une d'elles est la *clasping loop* (résidus 40 à 51 chez ClbK) qui se replie chez ClbK en deux brins bêta, les brins β2 et β3. Chez les domaines KS de TAS1 (Yun *et al.*, 2020) et d'OzmQ (Lohman *et al.*, 2015), qui sont deux domaines KS suivant des modules NRPS, la *clasping loop* se replie respectivement en un brin bêta unique et deux hélices alpha (Figure 78). Chez la plupart des autres KS, la *clasping loop* est non structurée. Le repliement observé uniquement chez les domaines KS en aval d'un module NRPS pourrait être une adaptation spécifique aux hybrides NRPS/PKS.

Plusieurs structures de PKS ont mis en évidence le rôle de la *clasping loop* pour l'interaction du domaine KS avec un autre domaine du même polypeptide (Figure 79). Chez la *trans*-AT RhiE, la *clasping loop* interagit avec le domaine de branchement non fonctionnel

(Bretschneider *et al.*, 2013). La *clasping loop* du premier module de DEBS permet l'interaction avec le domaine cétoréductase (KR) du même module (Cogan *et al.*, 2021). Nous pouvons faire l'hypothèse que les brins  $\beta 2$  et  $\beta 3$  de la *clasping loop* de ClbK ont un rôle équivalent et interagissent avec un domaine en aval du domaine KS de ClbK.

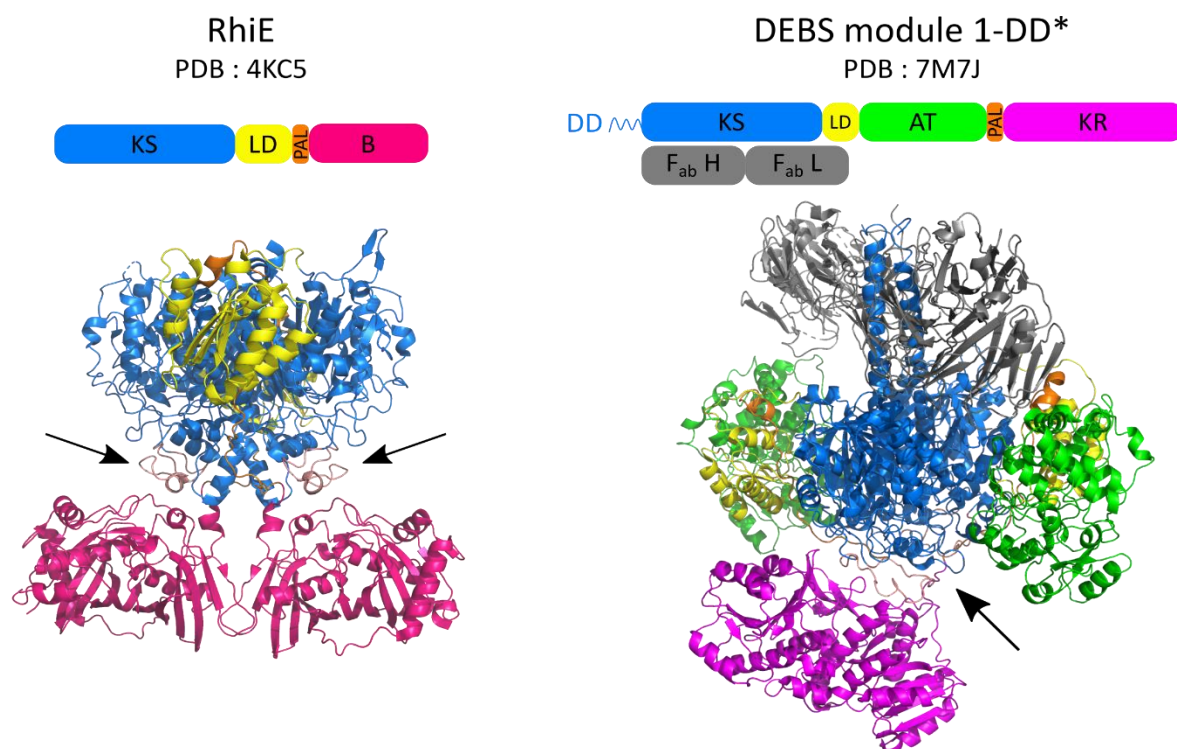


Figure 79. Rôle de la *clasping loop* du domaine KS dans l'interaction avec d'autres domaines de PKS chez RhiE et DEBS module 1. La *clasping loop* est colorée en saumon et indiquée par une flèche. B : domaine de branchement non fonctionnel, KR : cétoréductase, F<sub>ab</sub> H et F<sub>ab</sub> L : chaîne lourde et légère d'un fragment antigen binding.

L'*active site cap* de ClbK (résidus 197 à 231) se replie également en un brin bêta,  $\beta 8$ , qui se positionne à côté du brin  $\beta 3$  (Figure 78). Ce brin bêta est retrouvé uniquement chez les KS qui acceptent des substrats peptidiques, dont TAS1-KS (Yun *et al.*, 2020) et OzmQ-KS1 (Lohman *et al.*, 2015) (Figure 77 et Figure 78). Yun *et al.* ont fait l'hypothèse qu'un brin bêta à cette position permettrait au domaine KS d'accepter des substrats peptidiques, plus larges que les substrats traditionnels des PKS.

L'*active site cap* est localisé au niveau de l'entrée interne du KS, c'est-à-dire l'entrée du domaine KS utilisée par l'ACP accepteur, identifiée chez PikAIII (Dutta *et al.*, 2014). Chez les KS n'acceptant pas de substrat peptidique, l'*active site cap* se replie en une hélice qui obstrue l'entrée interne du domaine KS (Figure 77, Figure 78, DEBS module 5, Figure 80). La présence du brin  $\beta 8$  laisse l'entrée interne du domaine KS de ClbK béante (Figure 80) et la position de

l'ACP à proximité de cette entrée suggère qu'elle est fonctionnelle chez ClbK. La résolution intermédiaire des structures de PikAIII ne permet pas de déterminer si l'ouverture de l'entrée interne de PikAIII est également permise par la présence d'un brin bêta à la place d'une hélice alpha. L'entrée de côté traditionnelle est également ouverte chez ClbK (Figure 80).

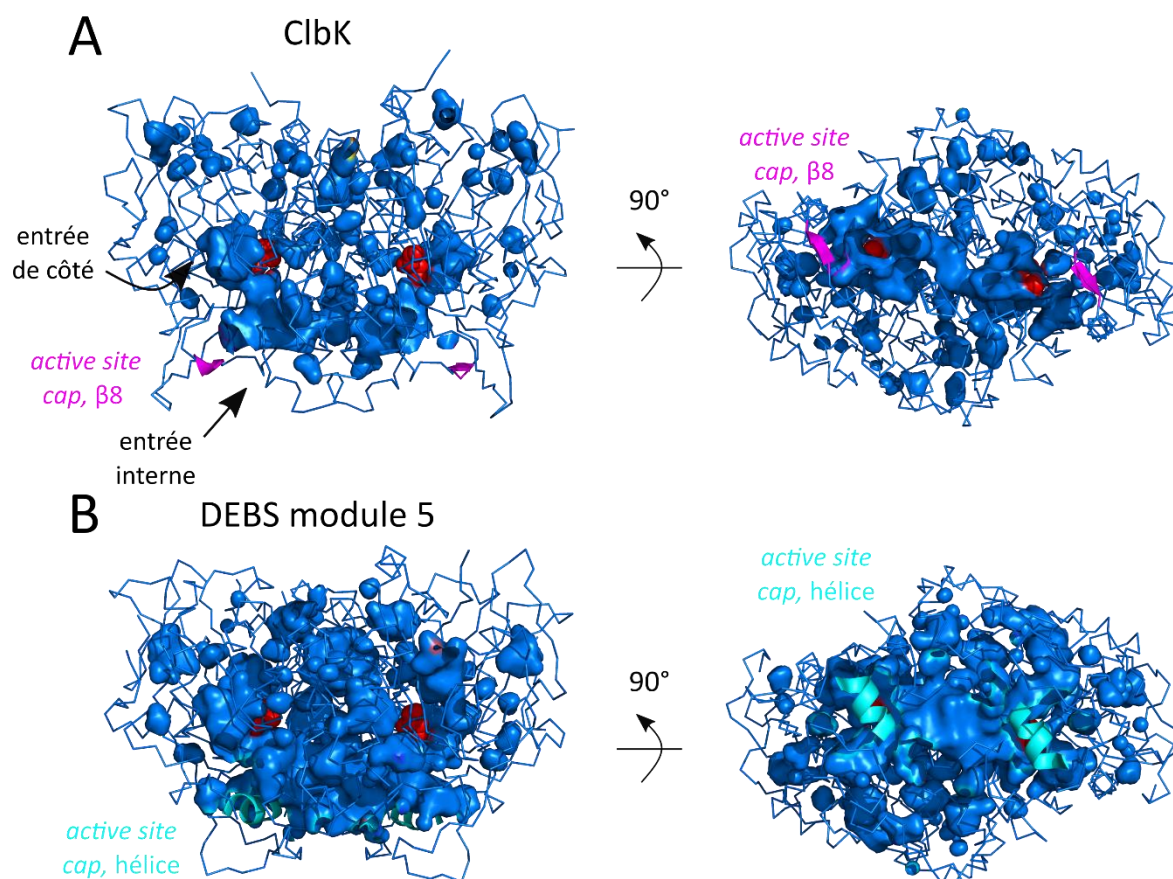


Figure 80. L'entrée interne du domaine KS de ClbK n'est pas obstruée. La cystéine catalytique du domaine KS est colorée par des sphères rouges. A) La visualisation des cavités du domaine de ClbK permet d'identifier deux entrées potentielles donnant accès à la cystéine catalytique. L'entrée interne du domaine KS de ClbK n'est pas obstruée par l'*active site cap* replié en un brin  $\beta 8$  et coloré en magenta. B) L'entrée interne du KS du cinquième module de DEBS (PDB : 2HG4, Y. Tang *et al.*, 2006) est fermée par l'hélice de l'*active site cap*, colorée en cyan.

Chez les domaines KS de PKS *trans*-AT, la nature du résidu précédant la cystéine catalytique (X-Cys) est corrélée à la chimie du substrat donneur. Par exemple, un résidu méthionine à cette position empêche l'acylation du KS par un substrat branché alors qu'un résidu alanine est compatible avec un substrat donneur branché (Jenner *et al.*, 2013). Les KS de PKS *trans*-AT acceptant un substrat peptidique ont généralement un résidu asparagine à la position X-Cys et ceux acceptant un cycle thiazole ou oxazole ont un résidu glycine ou sérine, supposé plus petit afin d'accommoder le cycle (Kohlhaas *et al.*, 2013). ClbK, qui accepte un

substrat contenant un cycle thiazole en provenance de ClbJ (Figure 29), possède effectivement un résidu sérine à la position X-Cys. A ma connaissance, les seuls domaines KS ayant un résidu sérine en position X-Cys sont des KS<sup>0</sup>, c'est-à-dire des domaines KS non compétents pour la condensation car ils ne possèdent pas la première histidine catalytique (Huang *et al.*, 2016) (Figure 81). Dans le cas de LnmI, il a été montré que le di-domaine KS<sup>0</sup>-ACP sert d'interface de transition entre un module NRPS et un module PKS. Le domaine KS de ClbK possédant un résidu sérine en position X-Cys et acceptant directement un substrat contenant un cycle thiazole est donc un nouveau type d'interface NRPS/PKS de PKS *trans*-AT.

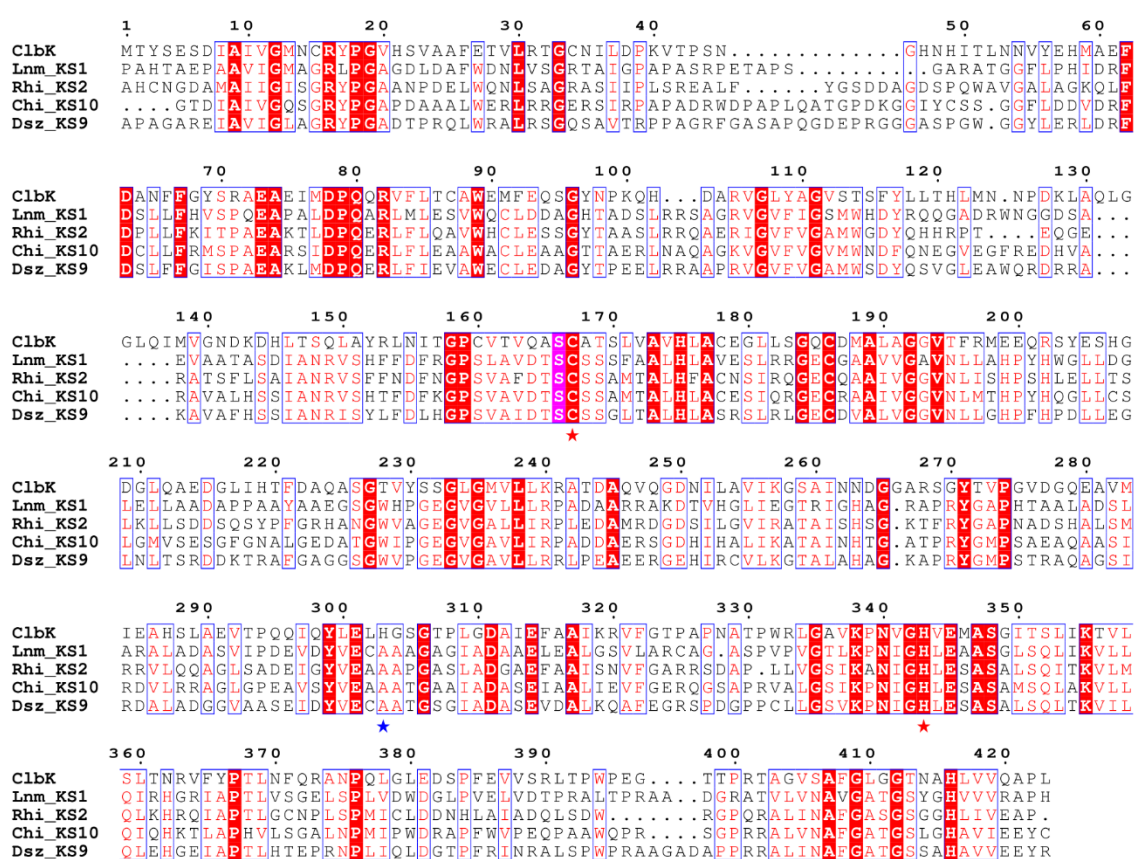


Figure 81. Analyse d'alignements de séquences de domaines KS de PKS *trans*-AT ayant une sérine en position X-Cys, effectuée par le serveur ESPript. Lnm, Rhi, Chi et Dsz désignent les métabolites leinamycine, rhizoxine, chivosazole et disorazole. Le substrat donneur de ClbK et Lnm\_KS1 possède un cycle thiazole alors que le substrat donneur des trois autres KS contient un cycle oxazole. Les trois résidus catalytiques sont marqués par des étoiles. Le résidu sérine précédant la cystéine catalytique est surligné en rose. A part ClbK, les autres KS sont des KS<sup>0</sup> incapables de catalyser la condensation car elles ne possèdent pas la première histidine catalytique (étoile bleue). Les résidus en blanc sur un fond rouge ou rose sont strictement conservés. Dans une colonne entourée en bleu, au moins 70 % des résidus ont des propriétés physico-chimiques similaires et ces résidus similaires sont en rouge.

### II.7.d. Le domaine AT de ClbK est partiellement dégénéré

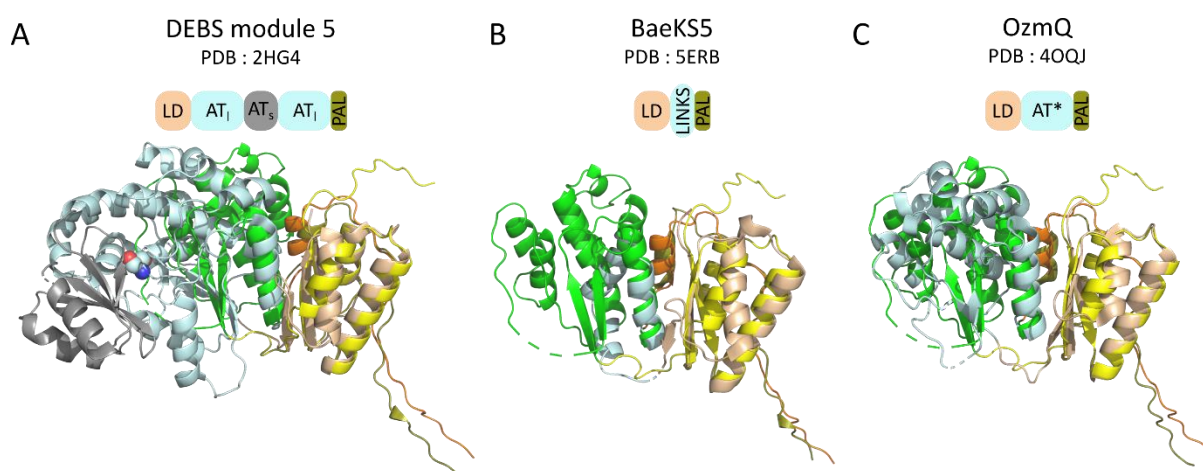


Figure 82. Alignement structural de la région post-KS de ClbK avec les régions post-KS de DEBS module 5 (A), BaeKS5 (B) et OzmQ (C). Les domaines de ClbK sont colorés avec les mêmes couleurs qu'à la Figure 73 (LD en jaune, AT\* en vert et PAL en orange). Les structures ont été alignées sur le domaine de liaison. AT<sub>I</sub> : grand sous-domaine AT, AT<sub>S</sub> : petit sous-domaine AT. LINKS : « Laterally-INTERacting Ketosynthase Sequence ». La sérine catalytique du domaine AT de DEBS module 5 est représentée par des sphères.

Le domaine acyltransférase (AT) sélectionne un substrat, futur substrat accepteur de la condensation, et le transfère au domaine ACP en aval du domaine KS. Le domaine AT est subdivisé en un grand sous-domaine, AT<sub>I</sub>, et un petit sous-domaine, AT<sub>S</sub>, et la sérine catalytique se situe à l'interface entre les deux sous-domaines (Figure 82A). Chez les PKS *cis*-AT, le domaine AT est localisé en aval du domaine KS et séparé de ce dernier par un domaine de liaison (LD). Chez les PKS *trans*-AT, bien que le domaine AT soit codé par une protéine indépendante, la plupart des KS possèdent également le domaine de liaison et des insertions de taille plus ou moins conséquente à la place du domaine AT. Cette insertion peut être réduite à quelques hélices, comme chez BaeKS5, qui participe à la synthèse de l'antibiotique bacillaene (D. C. Gay *et al.*, 2016) (Figure 82B et Figure 83). Chez BaeKS5, ces hélices constituent le motif LINKS (« Laterally-INTERacting Ketosynthase Sequence »), qui s'associe de façon homotypique et permet la construction de larges assemblages de PKS. ClbK possède une insertion bien plus large que ces hélices LINKS, similaire à l'insertion observée chez OzmQ (Figure 82C et Figure 83). Le domaine AT tronqué de ClbK est composé d'une partie du sous-domaine AT<sub>I</sub> et ne contient pas le sous-domaine AT<sub>S</sub>, ce qui explique sa non-fonctionnalité. Le domaine AT tronqué de ClbK, comme celui d'OzmQ, est donc probablement un ancien domaine AT fonctionnel, en cours de dégénérescence. Il est intéressant de noter qu'une certaine similarité de structure existe entre les domaines AT tronqués de ClbK et d'OzmQ, ce qui pourrait suggérer une fonction commune à ce reliquat d'AT.

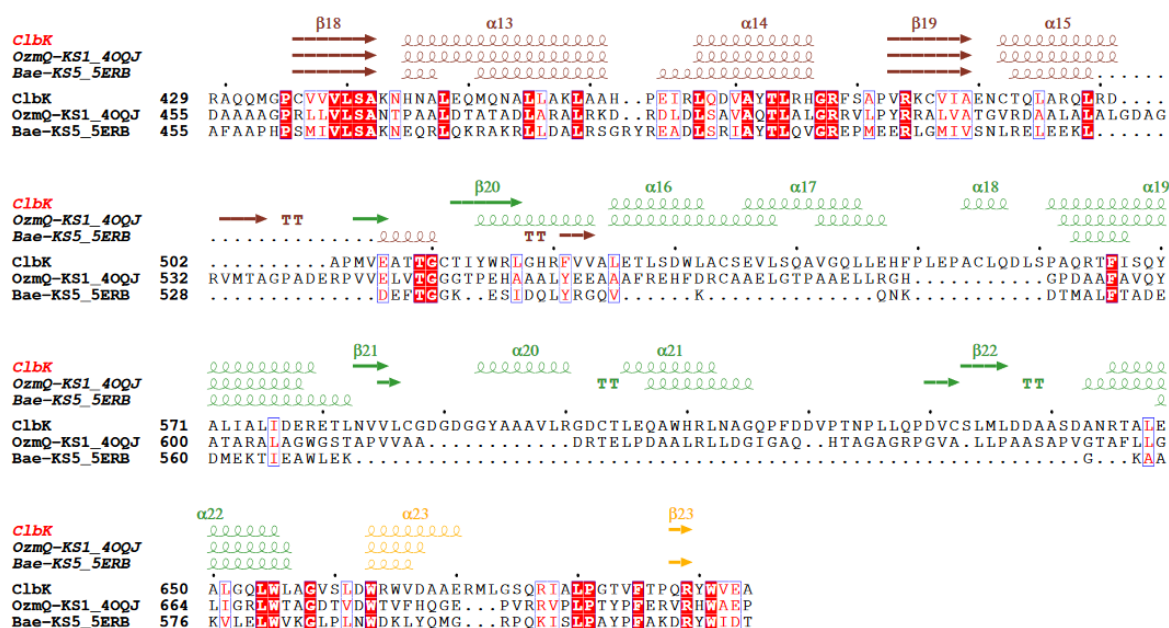


Figure 83. Analyse d'alignements de séquences post-KS de trois PKS de type *trans*-AT, effectuée par ESPript. Le domaine de liaison, le reliquat de domaine AT et le *linker* post-AT sont colorés en marron, vert et orange. Les résidus en blanc sur un fond rouge sont strictement conservés et les résidus dans une colonne entourée en bleu ont des propriétés physico-chimiques similaires.

Alors que l'insertion à la place du AT chez les PKS de type *trans*-AT est très variable, les *linkers* entourant cette insertion sont bien conservés, en séquence et en structure, entre eux, et avec les *linkers* entourant l'AT des PKS *cis*-AT (Figure 82 et Figure 83). Le domaine de liaison de ClbK possède le même repliement que les autres domaines de liaison, à l'exception du dernier brin bêta, qui est manquant chez ClbK (Figure 82 et Figure 83). Le *linker* post-AT de ClbK se replie contre le domaine KS et est lui aussi superposable aux *linkers* post-AT des PKS *cis*-AT et *trans*-AT (Figure 76 et Figure 82). Ceci suggère que l'absence d'un domaine AT fonctionnel impacte seulement la région du AT mais pas tout le module PKS. Les six derniers résidus du *linker* post-AT n'ont pas pu être modélisés faute de densité électronique, suggérant que la fin de ce *linker* est flexible et donc que le domaine ACP au bout du *linker* adopte plusieurs positions.

### II.7.e. Le domaine ACP est positionné à proximité de l'entrée interne du KS

ClbK- $PKS^*$  ne possède pas de domaine de modification optionnel donc le domaine ACP est directement connecté au domaine  $AT^*$  par le *linker* post- $AT$ . Deux domaines ACP sur quatre ont pu être partiellement reconstruits dans la structure de ClbK- $PKS^*$  : les sept derniers résidus de l'ACP<sub>E</sub> et les treize premiers résidus de l'ACP<sub>F</sub> sont manquants. La structure complète du domaine ACP de ClbK peut donc être reconstruite en superposant les deux chaînes (Figure 84). Le domaine ACP présente une architecture classique pour les domaines « carrier protein » puisqu'il comprend les quatre hélices traditionnelles, nommées H1, H2, H3, H4, et la sérine modifiable est localisée au début de l'hélice H2 (A. T. Keatinge-Clay, 2012) (Figure 84 et Figure 85).

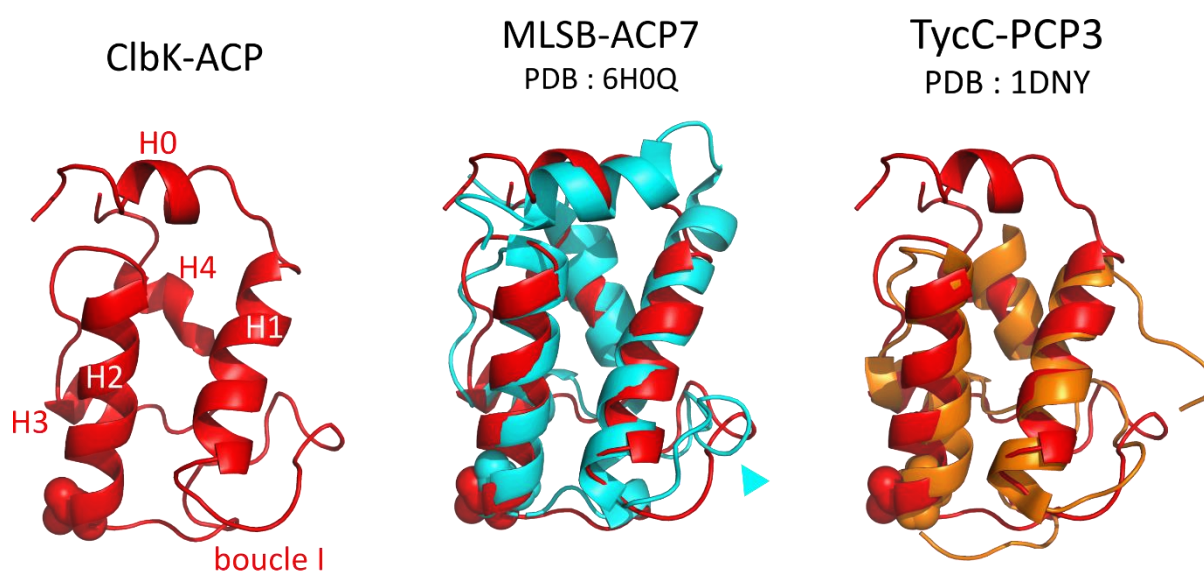


Figure 84. Structure du domaine ACP de ClbK et alignement structural avec l'ACP7 de MSLB et le PCP3 de TycC. La structure de l'ACP de ClbK a été obtenue en superposant les structures des deux domaines ACP, ACP<sub>E</sub> et ACP<sub>F</sub>. La sérine modifiée par le bras PPant est représentée par des sphères pour les trois domaines ACP. La flèche en cyan pointe la boucle I, plus longue chez MSLB-ACP7 que chez ClbK-ACP et TycC-PCP3.

Le domaine ACP de ClbK interagit à la fois avec des domaines PKS, la *trans*-AT ClbG et le domaine KS de ClbK, et avec un domaine NRPS, le domaine de cyclisation de ClbK. La structure de l'ACP de ClbK reflète cette versatilité NRPS/PKS puisque le domaine ACP présente à la fois des caractéristiques de domaines ACP et PCP. En effet, ClbK possède une hélice H0, identifiée chez MSLB-ACP7 (Figure 84 et Figure 85), plutôt typique des ACP de PKS *cis*-AT et rarement retrouvée chez les domaines PCP, ACP de PKS *trans*-AT ou d'hybride NRPS/PKS (Moretto *et al.*, 2019). Aucun rôle n'a pour l'instant été démontré pour cette hélice H0.

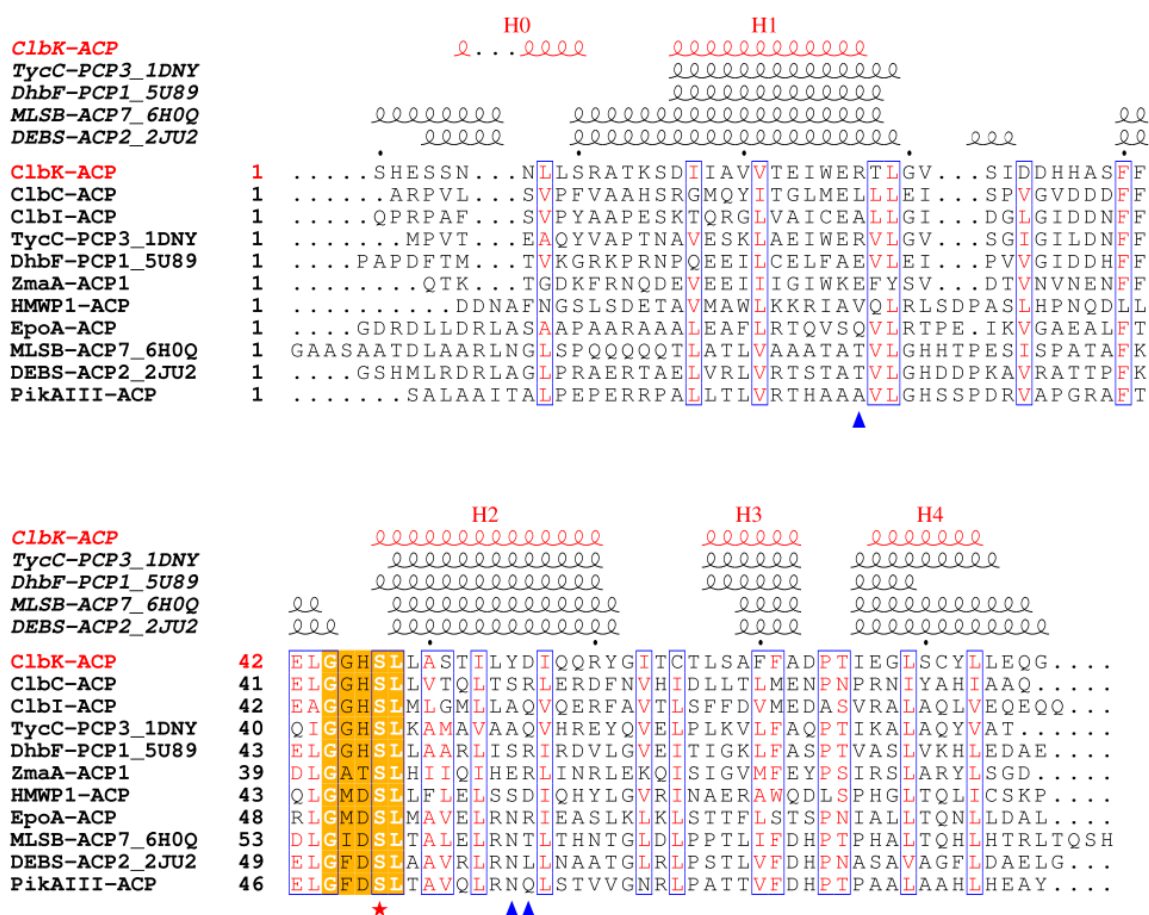


Figure 85. Analyse d’alignements de séquences de domaines « carrier protein » appartenant à des assemblages NRPS (TycC-PCP3 et Dhbf-PCP1), hybrides PKS/NRPS (ClbK-ACP, ClbC-ACP, ClbI-ACP, HMWP1-ACP et EpoA-ACP), PKS *cis*-AT (DEBS-ACP2 et MLSB-ACP7) et PKS *trans*-AT (ZmaA-ACP1). Analyse effectuée par ESPrpt. La sérine modifiée post-traductionnellement est marquée par une étoile rouge et le motif GxxSL auquel elle appartient est surligné en orange. Les résidus de ClbK-ACP interagissant avec le domaine KS sont marqués par un triangle bleu. Les résidus en blanc sur un fond orange sont strictement conservés. Dans une colonne entourée en bleu, au moins 70 % des résidus ont des propriétés physico-chimiques similaires et ces résidus similaires sont en rouge.

Des analyses bio-informatiques de centaines de séquences de domaines ACP de PKS *trans*-AT ont révélé que ces ACP co-évoluent avec le domaine en aval plutôt que le domaine en amont (Vander Wood & Keatinge-Clay, 2018). Ceci semble se confirmer pour le domaine ACP de ClbK puisque la sérine modifiée est localisée dans le motif **GGHSL**, typiquement retrouvé chez les domaines PCP qui délivrent leur substrat à un domaine de condensation, alors que les domaines ACP possèdent plutôt un motif **GxD<sub>1</sub>SL** (Figure 85). On peut noter que les domaines ACP de HMWP1 et d’EpoA, qui délivrent également leur substrat polycétide à un domaine de cyclisation ne présentent pas un motif **GGHSL** mais un motif **GMD<sub>1</sub>SL**. La boucle I de l’ACP de ClbK, entre les hélices 1 et 2 est plus courte que la boucle I des ACP des PKS *cis*-AT (Figure 84), ce qui semble aussi être une caractéristique des PCP et n’est pas retrouvée chez



les domaines ACP d'HMWP1 et EpoA (Figure 85). Les deux caractéristiques typiques des PCP, le motif GGHS<sub>L</sub> et la boucle I plus courte, sont également retrouvés chez ClbC et ClbI, les deux autres domaines ACP de l'îlot *pks* qui délivrent leur substrat à un domaine NRPS (Figure 9 et Figure 85). La comparaison de la séquence de ClbK-ACP avec celle de Zma-ACP, qui prend également en charge un substrat aminomalonyl n'a pas permis de mettre en évidence de motifs conservés entre les deux ACP (Figure 85).

Parmi les trois partenaires enzymatiques du domaine ACP de ClbK, qui sont dans l'ordre ClbG, ClbK-KS et ClbK-Cy (Figure 29), seul le domaine ClbK-KS est présent dans la structure de ClbK-PKS\*. Afin de déterminer si le domaine ACP de ClbK pouvait se trouver dans une conformation compatible avec la réaction de condensation catalysée par ClbK-KS, les interfaces du domaine ACP de ClbK avec le reste du module PKS ont été déterminées grâce au logiciel PISA. Sept interfaces ont été décrites dont deux couples d'interfaces très similaires (les couples 1-2 et 3-5), qui diffèrent seulement par l'identité de l'ACP (Figure 86 et Tableau 8). Selon l'interface, l'ACP contacte le domaine KS, le domaine de liaison, ces deux domaines à la fois ou le domaine AT\* (Figure 86). Bien que la densité électronique de la fin du *linker* post-AT ne soit pas continue, seulement six résidus du *linker* ne sont pas modélisés dans la structure de ClbK-PKS\*, ce qui contraint la position du premier résidu de l'ACP, la sérine 701, à moins de 30 Å par rapport au dernier résidu reconstruit du *linker* post-AT, l'alanine 694. Parmi les sept interfaces, seule l'interface 4 remplit ce critère et les distances sont compatibles avec l'appartenance de l'ACP<sub>F</sub> à la chaîne C ou D (Tableau 8). Les autres interfaces résultent certainement de l'arrangement cristallin de ClbK-PKS\*, elles ne seront donc pas décrites plus précisément ici.

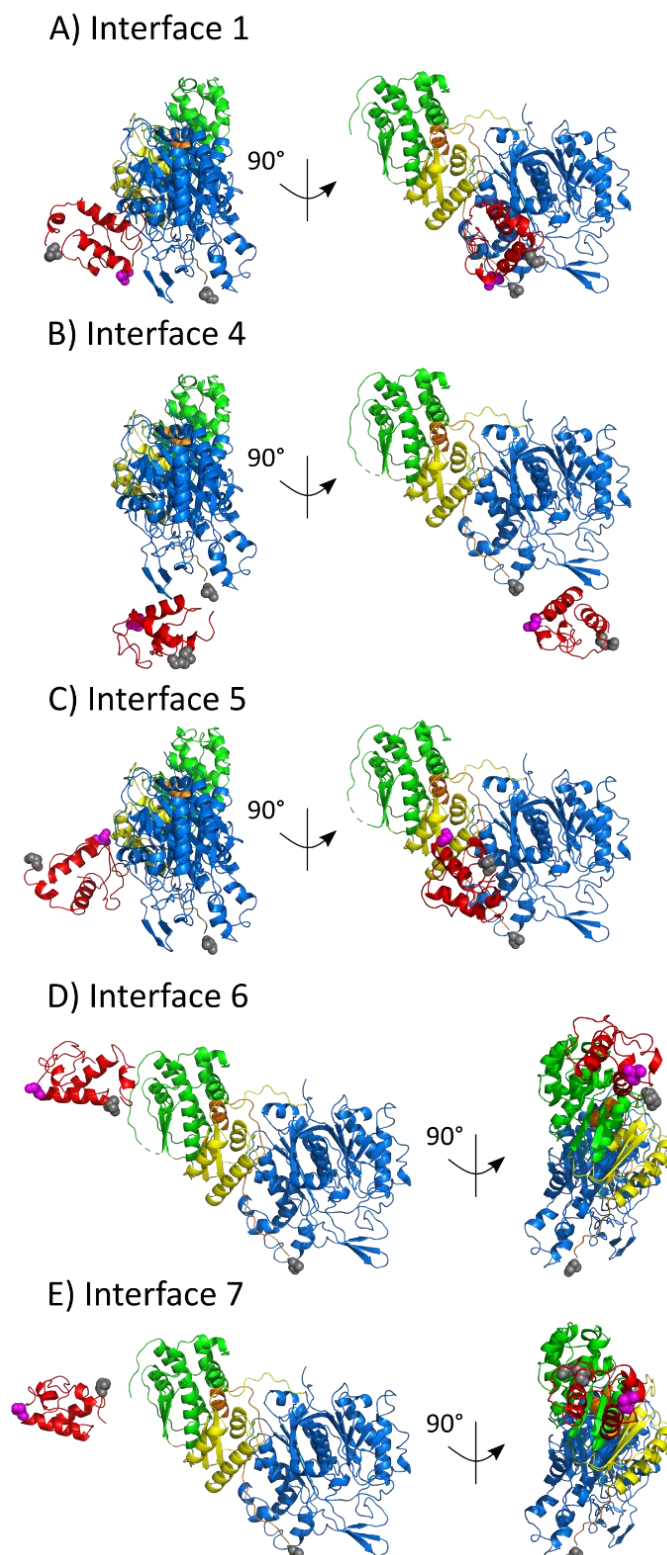


Figure 86. Structures des interfaces entre le domaine ACP et le reste du module PKS de ClbK, établies par PISA. Les interfaces 2 et 3 n'ont pas été représentées car elles sont très similaires aux interfaces 1 et 5 respectivement. La sérine 747 de l'ACP est représentée par des sphères violettes. Le dernier résidu du *linker* post-AT reconstruit (Ala694) ainsi que le premier résidu de l'ACP reconstruit (Ser701 pour les interfaces 1, 5 et 6 et Lys714 pour les interfaces 4 et 7) sont représentés par des sphères grises. Par souci de clarté, l'autre monomère de ClbK-PKS\* non contacté par l'ACP n'a pas été représenté.

Numéro	Identité ACP	Domaine(s) contacté(s) par l'ACP	Interface (Å <sup>2</sup> )	Distance A694-S701-chaîne contactée (Å)	Distance A694-S701-autre chaîne (Å)
1	E	KS et LD	354	50 (B)	67 (A)
2	F	KS et LD	327	50 (D)	67 (C)
3	F	LD	302	60 (A)	77 (B)
4	F	KS	280	<b>14 (D)</b>	<b>17 (C)</b>
5	E	LD	276	62 (C)	79 (D)
6	E	AT*	103	81 (D)	96 (C)
7	F	AT*	10	76 (B)	92 (A)

Tableau 8. Caractéristiques des interfaces entre le domaine ACP (chaîne E ou F) et le reste du module PKS de ClbK, déterminées par PISA. Les interfaces 1 et 2 sont très similaires, de même que les interfaces 3 et 5. La chaîne E a été superposée sur la chaîne F pour déterminer la distance A694-S701 des interfaces contenant l'ACP<sub>F</sub>, qui n'a pu être reconstruit qu'à partir du résidu 714. Pour chaque interface, deux distances ont été mesurées : la première est celle entre le premier résidu de l'ACP (Ser701) et le dernier résidu reconstruit du PAL (Ala694) du monomère contacté par l'ACP ; la seconde est la distance entre la Ser701 de l'ACP et l'Ala694 du PAL du monomère non contacté par l'ACP.

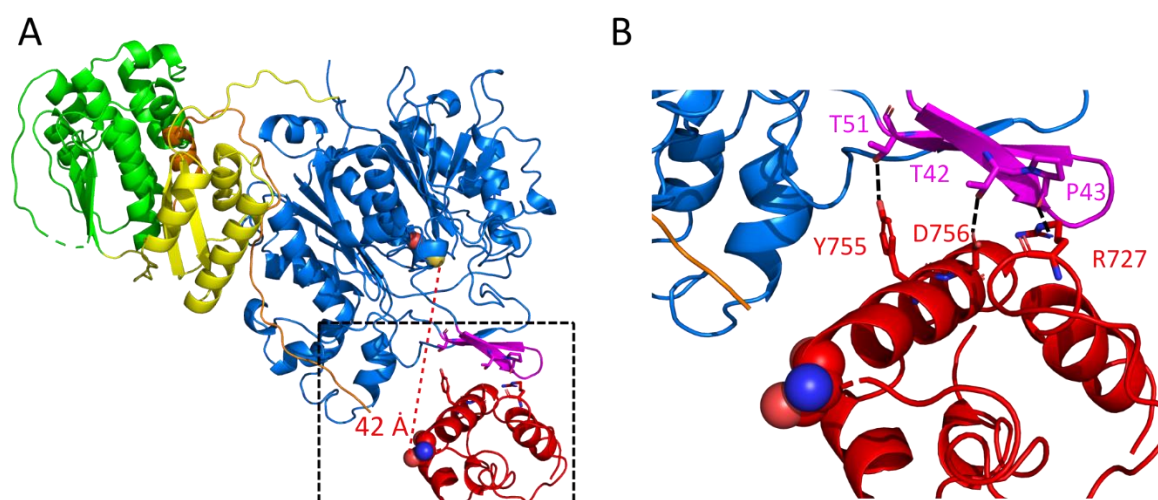


Figure 87. Détail de l'interface entre les domaines KS et ACP de ClbK (interface 4). A). La sérine 747 de l'ACP et la cystéine 167 du KS sont représentées par des sphères. B) Détail de la zone d'interaction entourée en A, déterminée par PISA. Les brins  $\beta 2$  et  $\beta 3$  de la *clasping loop* du domaine KS sont représentés en magenta.

Dans l'interface 4, le domaine ACP est positionné à proximité de l'entrée interne de ClbK et interagit avec les brins  $\beta 2$  et  $\beta 3$  de la *clasping loop* (Figure 87). Les interactions impliquent l'hélice H1 de l'ACP (R727 avec P43) et l'hélice H2 (Y755-T51 ; D756-T42). La position de l'ACP à proximité de l'entrée interne du domaine KS n'a été observé à ma

connaissance qu'une seule fois, chez PikAIII (Dutta *et al.*, 2014). Le positionnement de la structure de ClbK-PKS\* dans la carte de densité électronique de PikAIII chargé avec un substrat méthylmalonyl montre que le domaine ACP de ClbK occupe une position grossièrement équivalente à celle occupée par le domaine ACP de PikAIII (Figure 24, état 3 et Figure 88). Dutta *et al.* ont généré des mutants des résidus de la *dimer interface loop* (S164W) et de l'*active site cap* (N241W et D243W) bordant l'entrée interne de PikAIII afin de l'obstruer. L'activité catalytique de PikAIII étant très réduite chez chacun des mutants, la fonctionnalité de l'entrée interne a été démontrée. La résolution moyenne de la carte de densité électronique de PikAIII chargé avec du méthylmalonyl n'a pas permis d'identifier précisément les résidus à l'interface KS-ACP. On peut toutefois noter que les résidus des domaines KS et ACP de ClbK participant à l'interaction ne sont conservés ni avec PikAIII, ni avec les autres PKS (Figure 77 et Figure 85).

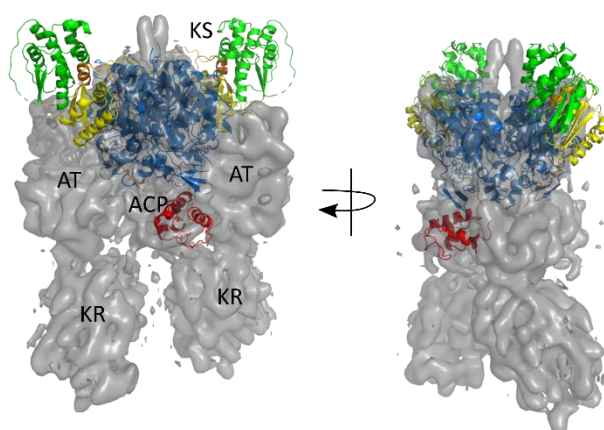


Figure 88. Superposition de ClbK-PKS\* dans la carte de densité électronique de PikAIII chargé avec du méthylmalonyl (EMD-5653).

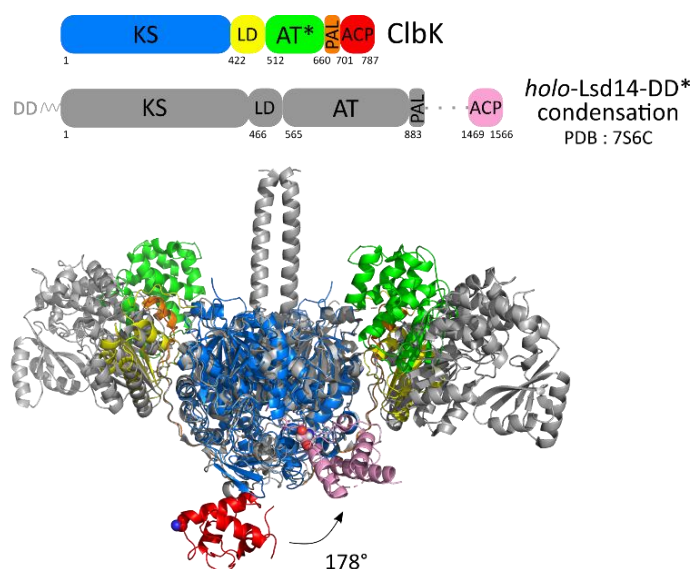
Bien que la présence du bras PPant de ClbK-PKS\* ait été confirmée par spectrométrie de masse (Figure 70C), aucune densité électronique n'est visible pour ce bras PPant dans la structure cristallographique, ce qui suggère que ce bras est flexible. 42 Å séparent la cystéine catalytique du domaine KS et la sérine 747 modifiée par le bras PPant, ce qui surpasse la longueur du bras PPant, d'environ 20 Å (Figure 87A). Le domaine ACP n'est donc pas dans une conformation compatible avec la réaction de condensation dans l'interface 4. Au vu de la petite distance à parcourir par l'ACP pour que la sérine 747 se situe à moins de 20 Å de la cystéine 167 du domaine KS, on peut faire l'hypothèse que la conformation observée en cristallographie est représentative d'un état pré ou post-condensation par le domaine KS. Le domaine ACP<sub>E</sub> est également positionné à proximité de l'entrée interne du domaine KS mais

est trop éloigné pour établir des liaisons avec ce domaine. Il pourrait également être représentatif d'un état pré ou post-condensation.

#### **II.7.f. Discussion sur la position de l'ACP dans la structure de ClbK-PKS\* et perspectives**

L'absence de densité électronique pour un domaine ACP par dimère de ClbK-PKS\* suggère que les deux domaines ACP n'ont pas un mouvement concerté et qu'ils adoptent différentes positions. La position de l'ACP est contrainte par le *linker* post-AT, qui se positionne contre le domaine KS. Le domaine ACP est positionné à proximité de l'entrée interne du domaine KS et contacte une région  $\beta 2$ -  $\beta 3$  du domaine KS spécifique à ClbK. La structure de ClbK-PKS\* révèle donc un nouveau type d'interface KS-ACP, qui pourrait être représentatif d'un état pré ou post-condensation. L'absence de densité électronique pour la fin du *linker* post-AT ne permet pas de définir à quelle chaîne appartient l'ACP et les distances sont compatibles avec l'appartenance de l'ACP<sub>F</sub> à la chaîne C ou D (Tableau 8, interface 4). Les travaux du Pr Khosla ont montré que le domaine ACP interagit avec le domaine KS de l'autre chaîne pour la condensation (Kao *et al.*, 1996 ; Cogan *et al.*, 2021). Si ce modèle est aussi valable pour le domaine ACP de ClbK, cela signifierait que le domaine ACP<sub>F</sub>, qui contacte le domaine KS de la chaîne D, appartiendrait à la chaîne C.

La fonctionnalité de l'entrée interne de ClbK doit être confirmée par des études structurales en présence de substrats. En effet, plusieurs études structurales ont démontré l'importance de la nature du substrat chargé sur le domaine ACP pour son positionnement (Dutta *et al.*, 2014 ; Whicher *et al.*, 2014 ; Bagde *et al.*, 2021). De plus, Rittner et Grininger ont souligné que le positionnement de l'ACP à l'entrée interne ne plaçait pas le substrat dans une conformation favorable pour la décarboxylation (Rittner & Grininger, 2014). Une structure de ClbK-PKS avec le substrat aminomalonyl chargé sur le domaine ACP ou un analogue de substrat du domaine KS pourrait permettre de confirmer la fonctionnalité de l'entrée interne du domaine KS. L'utilisation d'agent de réticulation pour stabiliser l'interaction entre les domaines ACP et KS de ClbK pourrait également offrir une vision plus précise de l'interaction entre ces deux domaines.



L'alignement structural de ClbK-PKS\* avec la PKS Lsd14 dans sa conformation de condensation (Bagde *et al.*, 2021) montre que l'entrée de côté de ClbK-KS n'est pas obstruée et est accessible à un domaine « carrier protein » (Figure 89). L'interaction de l'ACP avec le domaine KS au niveau de l'entrée de côté a été décrite à plusieurs reprises (Herbst, Huitt-Roehl, *et al.*, 2018 ; Bagde *et al.*, 2021 ; Cogan *et al.*, 2021). L'ACP de ClbK devrait effectuer une rotation de 178° pour parcourir les 40 Å séparant sa position à l'entrée interne et la position de l'ACP de Lsd14 à l'entrée de côté, ce qui impliquerait un mouvement du *linker* post-AT pour déplacer le domaine ACP.

Une hypothèse élégante serait que, comme chez PikAIII, les deux entrées séparées serviraient à la discrimination entre les substrats donneur et accepteur (Dutta *et al.*, 2014). Dans cette hypothèse, le domaine PCP de ClbJ, qui doit délivrer son substrat donneur à la cystéine 167 de ClbK-KS (Figure 29), utiliserait l'entrée de côté du domaine KS et le domaine ACP de ClbK utiliserait l'entrée interne pour présenter le substrat accepteur au domaine KS. Une structure du complexe entre ClbK-PKS et le domaine PCP de ClbJ chargé avec le substrat donneur de la condensation permettrait de déterminer quelle entrée est utilisée par le domaine PCP donneur.

### III. Etudes biochimiques et structurales d'autres fragments de ClbK

#### III.1. Objectifs de l'étude d'autres fragments de ClbK

Les expériences de SEC-MALLS sur ClbK entière ont démontré l'état dimérique de cette protéine (Figure 48). La structure cristallographique de ClbK-PKS\* a révélé que le domaine KS de ClbK est dimérique (Figure 73) mais les études biophysiques de ClbK-PKS ont montré que cette protéine est majoritairement monomérique en solution (Figure 68). La dimérisation par le domaine KS n'est donc pas suffisante pour maintenir le module ClbK-PKS dimérique en solution, et *a fortiori* pour maintenir le di-module ClbK dimérique. L'hypothèse la plus probable pour concilier ces faits est l'existence d'un ou plusieurs élément(s) de dimérisation qui, associé(s) à la dimérisation par le domaine KS, maintiendrait(en)t l'état dimérique de ClbK. Un objectif de cette partie était donc de déterminer le/les autres éléments de dimérisation de ClbK, en testant l'état oligomérique de fragments contenant des portions du module NRPS de ClbK.

A ce stade de l'étude, plusieurs éléments de ClbK sont en lice pour être des potentiels éléments de dimérisation. Des éléments de dimérisation en aval de l'ACP ont été décrits chez PikAIII et Lsd14 (Dutta *et al.*, 2014 ; Bagde *et al.*, 2021). Bien que ces régions, en C-terminal des protéines PikAIII et Lsd14, ne soient pas homologues au *linker* séparant le module PKS du module NRPS de ClbK, ce *linker* pourrait jouer un rôle dans la dimérisation de ClbK, à l'image du *linker* MAT-DH impliqué dans la dimérisation de la mFAS (Maier *et al.*, 2008). Jusqu'à très récemment, le paradigme concernant les NRPS était qu'il s'agissait de protéines monomériques (Tanovic *et al.*, 2008 ; Drake *et al.*, 2016 ; Reimer, Aloise, Harrison, *et al.*, 2016 ; Kreitler *et al.*, 2019 ; Reimer *et al.*, 2019). Un des rares contre-exemples était la NRPS VibF qui dimérise par un domaine de condensation inactif (Hillson & Walsh, 2003). Ces deux dernières années, l'état dimérique de modules NRPS a été démontré pour trois protéines. Deux d'entre elles, FmoA3 et PchE, dimérisent par le di-domaine Cy-A<sub>core</sub>, même si elles ne partagent pas exactement la même interface de dimérisation (Katsuyama *et al.*, 2021 ; Wang *et al.*, 2022). Les constructions ClbK-PKS-LK, ClbK-PKS-Cy et ClbK-PKS-A<sub>core</sub> ont donc permis d'analyser les rôles respectifs du *linker* PKS-NRPS, du domaine de cyclisation et du di-domaine Cy-A<sub>core</sub> dans la dimérisation de ClbK (Figure 30).

Un autre but de cette partie était de poursuivre la stratégie de « dissection-reconstruction » structurale de ClbK. La structure d'un fragment de ClbK contenant le module PKS et un ou plusieurs domaine(s) NRPS serait la première caractérisation structurale d'une protéine hybride PKS-NRPS.

### III.2. Purification des fragments ClbK-PKS-LK, ClbK-PKS-Cy et ClbK-PKS-A<sub>core</sub>

Les trois protéines ClbK-PKS-LK, ClbK-PKS-Cy et ClbK-PKS-A<sub>core</sub> ont été exprimées selon le même protocole que ClbK-PKS. La protéine ClbK-PKS-LK (91 kDa) a été purifiée selon le même protocole que la protéine ClbK-PKS, avec une pureté et un rendement de purification comparables. Les deux protéines ClbK-PKS-Cy (141 kDa) et ClbK-PKS-A<sub>core</sub> (187 kDa) ont été purifiées en trois étapes, similairement à la protéine ClbK entière (Figure 90 et Figure 91). Les rendements de purification sont d'environ 2.5 mg par litre de culture pour ClbK-PKS-Cy et 1 mg par litre de culture pour ClbK-PKS-A<sub>core</sub>.

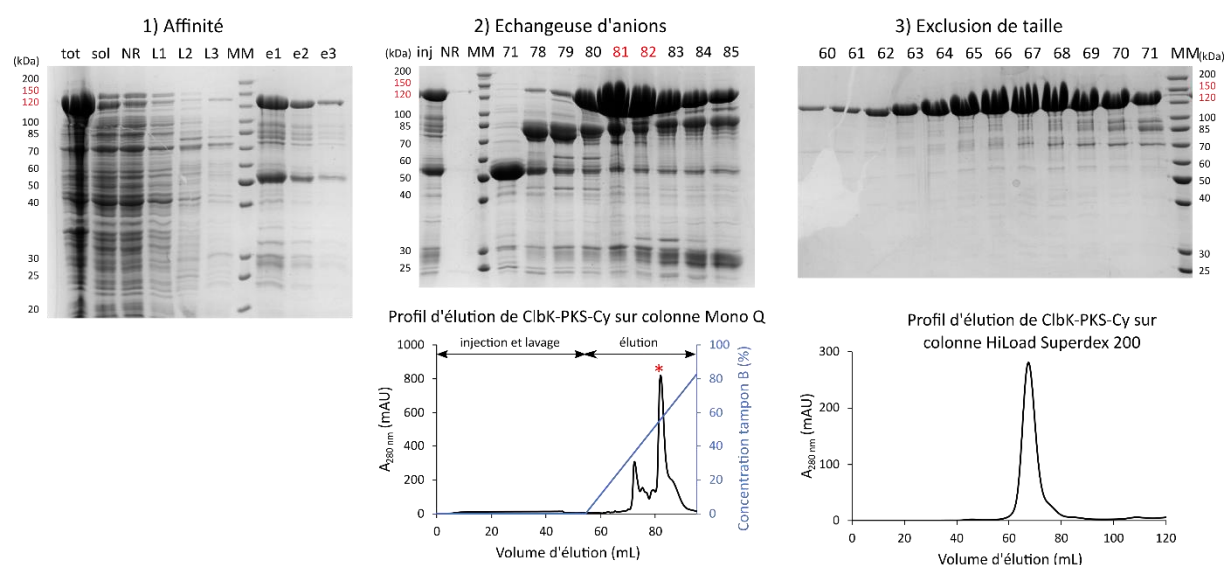


Figure 90. Purification de la protéine ClbK-PKS-Cy (141 kDa) en trois étapes. La pureté de la protéine ClbK-PKS-Cy a été estimée grâce à des gels SDS 10 %. tot : fraction totale, sol : fraction soluble, NR : fraction non retenue, L1-L2-L3 : fractions de lavage avec 0, 10 ou 20 mM d'imidazole, MM : marqueur de masse molaire, e1-e2-e3 : fractions d'éluion 1, 2 et 3, inj : fraction injectée. Les numéros au-dessus des gels correspondent aux volumes des fractions. Les fractions d'éluion 81 et 82 de l'échangeuse d'anions ont été injectées sur colonne d'exclusion de taille. Seules les huit premières fractions de la chromatographie d'exclusion de taille ont été conservées pour les analyses biochimiques et structurales.



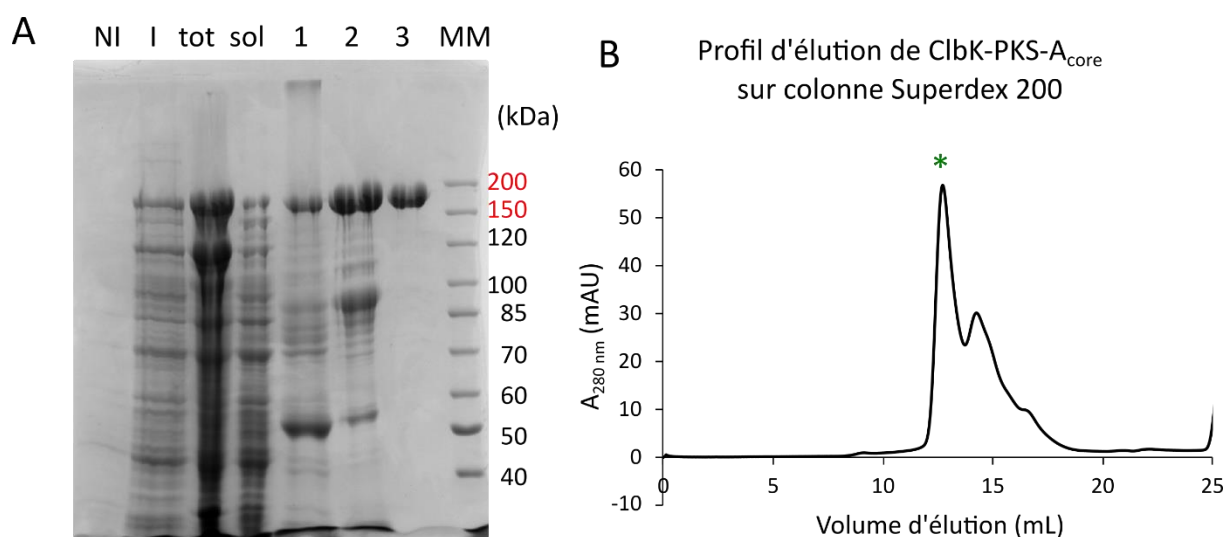


Figure 91. Purification de la protéine ClbK- $\text{PKS-A}_{\text{core}}$  (187 kDa). A) Evaluation de la pureté de ClbK- $\text{PKS-A}_{\text{core}}$  par gel SDS 7.5 %. NI : fraction avant induction, I : fraction induite, tot : fraction totale, sol : fraction soluble, 1 : fraction d'éluion de chromatographie d'affinité, 2 : fraction après chromatographie échangeuse d'anions, 3 : fraction après chromatographie d'exclusion de taille, MM : marqueur de masse molaire. B) Profil d'éluion de ClbK- $\text{PKS-A}_{\text{core}}$  sur colonne d'exclusion de taille. Le premier pic contient ClbK- $\text{PKS-A}_{\text{core}}$  et les pics suivants contiennent des contaminants.

### III.3. Détermination de l'état oligomérique des fragments ClbK- $\text{PKS-LK}$ , ClbK- $\text{PKS-Cy}$ et ClbK- $\text{PKS-A}_{\text{core}}$

L'analyse de ces fragments par SEC-MALLS a montré qu'ils sont tous les trois monomériques (Figure 92). L'addition du *linker* PKS-NRPS, du domaine Cy et du domaine  $\text{A}_{\text{core}}$  ne permettent donc pas de rétablir l'état dimérique de ClbK en solution. Ce résultat suggère que ClbK possède au moins un élément de dimérisation localisé en C-terminal du sous-domaine  $\text{A}_{\text{core}}$ . L'analyse par SEC-SAXS du fragment ClbK- $\text{PKS-Cy}$  a confirmé l'état monomérique de cette protéine (résultats non montrés). La forme monomérique de ClbK- $\text{PKS-Cy}$  ne prend pas en compte la dimérisation par le domaine KS démontrée grâce à la structure de ClbK- $\text{PKS}^*$ . Les données de SEC-SAXS sur ClbK- $\text{PKS-Cy}$  n'ont donc pas été jugées pertinentes dans le cadre de la protéine ClbK entière.

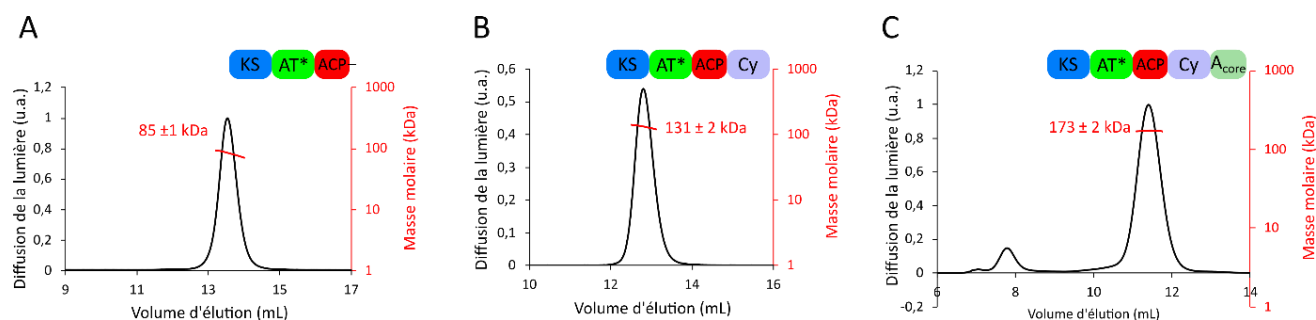


Figure 92. Analyse de fragments de ClbK par SEC-MALLS. 50  $\mu$ L de ClbK-PKS-LK (A, 12 mg.mL<sup>-1</sup>), ClbK-PKS-Cy (B, 3.5 mg.mL<sup>-1</sup>) et ClbK-PKS-A<sub>core</sub> (C, 2.3 mg.mL<sup>-1</sup>) ont été injectés sur une colonne Superdex 200 10/300 GL (25 mL). Les masses molaires théoriques monomériques de ClbK-PKS-LK, ClbK-PKS-Cy et ClbK-PKS-A<sub>core</sub> sont 91 kDa, 141 kDa et 187 kDa respectivement.

Des tests d'expression et de purification du fragment ClbK-PKS-A<sub>sub</sub> (Figure 30) ont été menés mais ils n'ont pas permis d'établir un protocole de purification pour ce fragment. Ce fragment, ainsi que les fragments ClbK-NRPS et ClbK-A-Ox-A ne s'accrochent pas à la résine de nickel (Tableau 6). Tous les fragments de ClbK se purifiant correctement par chromatographie d'affinité sont monomériques. Il semblerait donc que la dimérisation gêne la fixation sur les résines d'affinité des protéines ayant une étiquette poly-histidines. Il est également possible que les délimitations choisies pour certaines constructions ne soient pas optimales et empêchent la production de protéine sous forme stable.

#### III.4. Tentatives d'analyse structurale de ClbK-PKS-Cy

##### III.4.a. Essais de cristallogénèse

Les rendements de purification corrects pour ClbK-PKS-Cy ont permis d'envisager une étude cristallographique de ce fragment. Des tests de cristallisation à 8 et 11 mg.mL<sup>-1</sup> ont été menés par la plate-forme HTX. Aucune condition de cristallisation n'a pu être mise en évidence avec ces tests. Le T<sub>m</sub> de la protéine dans son tampon de filtration sur gel est de 43°C. Une expérience de Thermofluor a donc été menée afin de trouver un tampon stabilisant ClbK-PKS-Cy. Aucun tampon ayant un effet positif net sur le T<sub>m</sub> n'a pu être identifié.

### III.4.b. Observation en microscopie électronique à coloration négative

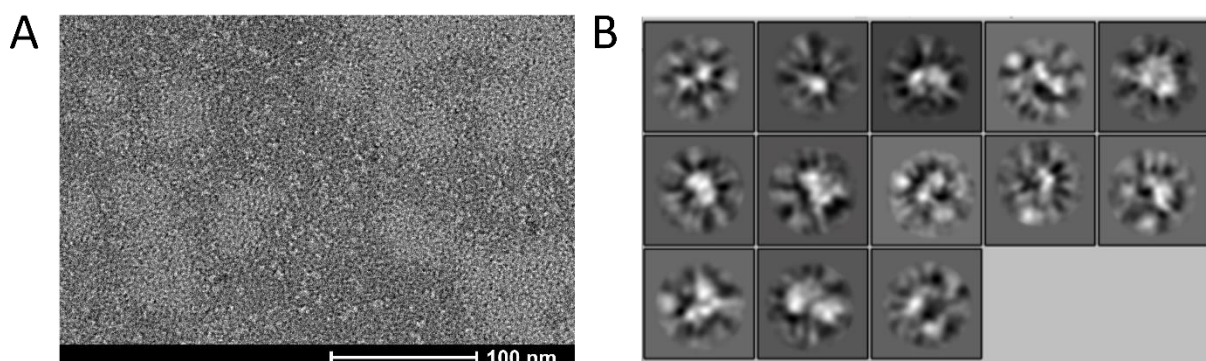


Figure 93. Observation de ClbK-PKS-Cy ( $0.02 \text{ mg.mL}^{-1}$ ) dans son tampon de filtration sur gel en coloration négative, colorée au sodium silico tungstate. A) Image représentative des observations de ClbK-PKS-Cy en coloration négative. Les zones blanches sont des zones sans colorant. B) Classes obtenues après classification de plus de 5800 particules par RELION.

Le fragment ClbK-PKS-Cy a été observé par microscopie électronique en coloration négative (Figure 93). La protéine ne forme pas d'agrégat mais la classification des particules révèle un très grand nombre de classes, qui pourraient être attribuées à un grand nombre de conformations adoptées par ClbK-PKS-Cy. La flexibilité de ce fragment pourrait expliquer la difficulté à déterminer des conditions de cristallisation. Plusieurs structures de domaines de cyclisation ou domaines équivalents ont été déterminées par cristallographie aux rayons X. Il est donc probable que ce ne soit pas le domaine ClbK-Cy en lui-même mais plutôt la flexibilité du *linker* connectant le module PKS au module NRPS qui soit responsable de la flexibilité de la protéine ClbK-PKS-Cy. Cette flexibilité inter-modulaire a déjà été décrite chez les NRPS grâce à plusieurs structures cristallographiques de fragments bi-modulaires de LgrA (Reimer *et al.*, 2019) et à l'observation en microscopie électronique de la NRPS DhbF (Tarry *et al.*, 2017).

### III.5. Bilan de l'études des fragments ClbK-PKS-LK, ClbK-PKS-Cy et ClbK-PKS-A<sub>core</sub> et perspectives

L'analyse par SEC-MALLS des fragments ClbK-PKS-LK, ClbK-PKS-Cy et ClbK-PKS-A<sub>core</sub> a montré que ces fragments sont monomériques en solution, comme le module PKS de ClbK. Il semblerait donc que le deuxième élément de dimérisation de ClbK soit situé en aval du domaine A<sub>core</sub>. Les structures récemment publiées par Fortinez *et al.* ont prouvé l'état dimérique de l'enzyme à activité oxydase BmdC et ont montré que la dimérisation par BmdC permet de maintenir l'état dimérique du module NRPS BmdB<sub>M2</sub> (Fortinez *et al.*, 2022). Le module NRPS de ClbK a la même composition en domaines que le complexe BmdB<sub>M2</sub>-BmdC.

En revanche, le domaine oxydase de ClbK ne possède pas le domaine N-terminal RREL de BmdC participant à sa dimérisation. Les auteurs de la structure du complexe Bmd<sub>M2</sub>-BmdC ont réussi à insérer le domaine Ox privé du domaine RREL dans BmdB<sub>M2</sub> et ont montré que cette protéine est dimérique et catalytiquement active (Fortinez *et al.*, 2022). Le domaine Ox de ClbK est donc le meilleur candidat pour être le deuxième élément de dimérisation de ClbK.

Une étude structurale du fragment ClbK-PKS-Cy a été initiée, sans succès pour l'instant. La construction ClbK-PKS-A<sub>core</sub> pourrait aussi devenir une cible structurale et les mêmes techniques employées sur ClbK-PKS pourraient s'appliquer à ce fragment. En particulier, le clivage de l'étiquette poly-histidines pourrait faciliter la cristallisation. Une approche structurale par cryo-ME est aussi envisageable mais, comme pour ClbK entière, des conditions d'observation permettant de visualiser le dimère de ClbK-PKS-A<sub>core</sub> doivent d'abord être déterminées. L'étude structurale de ce fragment pourrait bénéficier de la stabilisation par des nanocorps, qui sont en cours de synthèse.



# Discussion et perspectives



## Résumé des informations acquises sur ClbK

Durant cette thèse, une caractérisation biochimique, biophysique, enzymatique et structurale de la protéine hybride PKS-NRPS ClbK a été menée. Deux méthodes biophysiques ont montré que la protéine ClbK s'assemble en un dimère de quasiment 500 kDa mais se dissocie partiellement en monomères à faibles concentrations. L'analyse de la protéine par SEC-SAXS a confirmé l'état dimérique en solution et a permis de déterminer certains paramètres structuraux de ClbK, en particulier sa dimension maximale, autour de 30 nm. Des modèles *ab initio* de la méga-enzyme ont été générés. Malgré une diversité de formes, ces modèles suggèrent la présence d'au moins deux chambres catalytiques. La caractérisation structurale à haute résolution de cette méga-enzyme n'a pas été fructueuse, probablement à cause de sa flexibilité. Une stratégie de « dissection-reconstruction » a alors été choisie afin de se focaliser sur la caractérisation structurale de fragments de ClbK.

La structure du module PKS de ClbK (2 x 85 kDa), de type *trans*-AT, a été résolue par cristallographie aux rayons X à 3 Å. Cette structure a révélé que le module PKS de ClbK dimérise par le domaine KS et adopte la conformation étendue observée classiquement chez les PKS. La densité électronique d'un seul domaine ACP par dimère est détectée, suggérant que les deux ACP n'adoptent pas un mouvement concerté. Le domaine ACP est positionné à proximité de l'entrée interne du domaine KS, et interagit avec une région du domaine KS retrouvée spécifiquement chez ClbK.

La structure à haute résolution de ClbK-PKS\* pourra être utile à la caractérisation structurale de fragments de ClbK contenant ce module PKS et un ou des domaines NRPS, notamment ClbK-PKS-Cy ou ClbK-PKS-A<sub>core</sub>, dont l'étude a été initiée pendant cette thèse. Des perspectives pour obtenir la structure de ClbK-PKS\* dans une conformation catalytiquement active et d'élucider la structure haute résolution de ClbK entière ou d'un de ses fragments ont été proposées aux sections I.9, II.7.f, et III.5 ci-dessus.

En solution, le fragment PKS, ainsi que les fragments PKS-LK, PKS-Cy et PKS-A<sub>core</sub> sont tous monomériques. Ces résultats, combinés à l'état dimérique de la protéine entière, suggèrent qu'un ou plusieurs domaine(s) en C-terminal du sous-domaine A<sub>core</sub> participe(nt) à la dimérisation de la protéine ClbK entière. Le domaine Ox est le meilleur candidat pour être le deuxième élément de dimérisation de ClbK.



## Modèle possible de l'architecture de ClbK

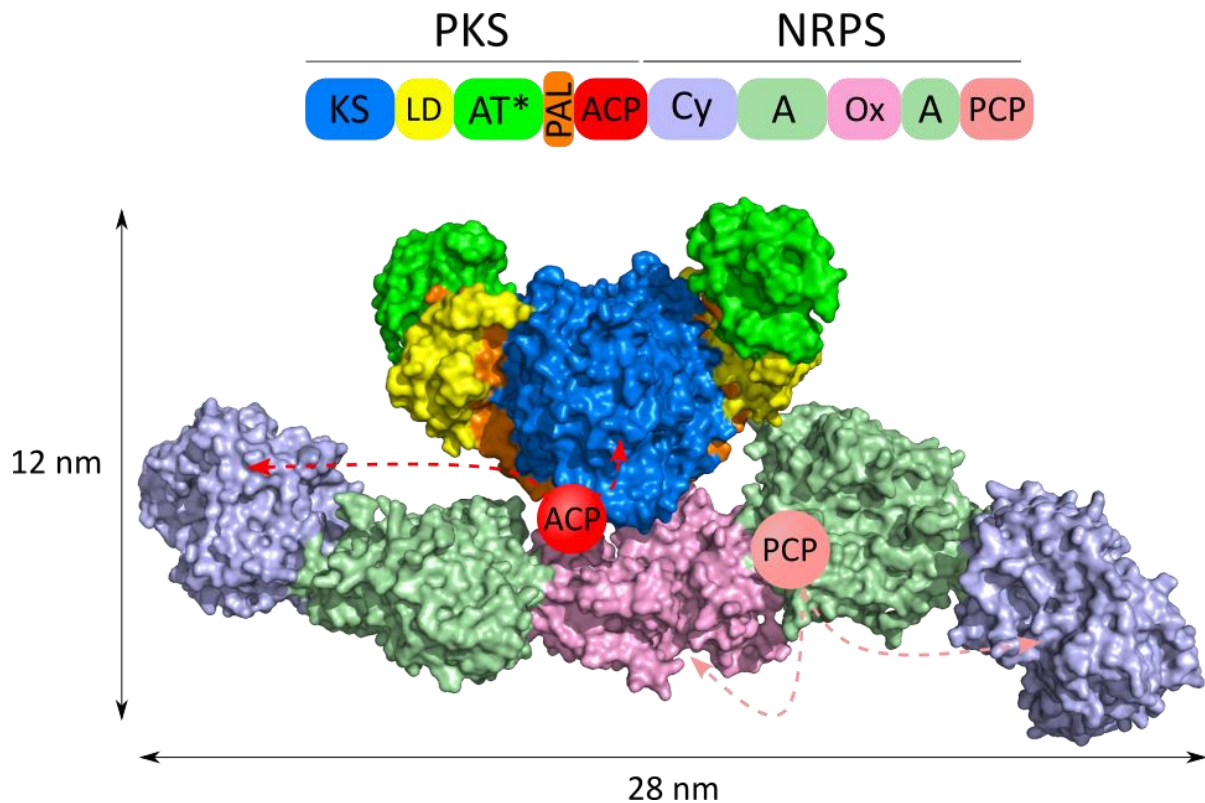


Figure 94. Modèle de l'architecture de l'hybride PKS-NRPS ClbK. Ce modèle a été construit en juxtaposant la structure de ClbK-PKS\* déterminée durant cette thèse avec la structure du dimère Bmd<sub>M2</sub>-BmdC en conformation de thiolation, déterminée par Fortinez *et al.* (PDB : 7LY7), possédant la même composition en domaines que le module NRPS de ClbK. Les résidus 1 à 141 du domaine oxydase BmdC ont été retirés de la structure du complexe BmdB<sub>M2</sub>-BmdC car le domaine Ox de ClbK ne possède pas cette région. Par simplicité, un seul domaine ACP et un seul domaine PCP ont été représentés. Les flèches indiquent les mouvements que doivent effectuer les domaines ACP et PCP pour interagir avec leurs partenaires catalytiques.

La structure du module PKS de ClbK a été résolue pendant cette thèse mais pas celle de son module NRPS. Toutefois, la structure du complexe entre Bmd<sub>M2</sub> (Cy-A-PCP) et BmdC (Ox), possédant la même composition en domaines que le module NRPS de ClbK, a été dévoilée cette année (Fortinez *et al.*, 2022) (Figure 18C). Le dimère adopte une conformation très étendue puisque 28 nm séparent les deux extrémités du complexe, ce qui est équivalent à la dimension maximale de ClbK, d'environ 30 nm, déterminée par SAXS. Un modèle de ClbK a donc été construit en superposant la structure cristallographique de ClbK-PKS\* avec la structure du complexe BmdB<sub>M2</sub>-BmdC (Figure 94). ClbK ne possède ni le domaine RREL de BmdC, ni le *linker* suivant ce domaine donc les résidus 1 à 141 de BmdC ont été retirés. Ce modèle est compatible avec les modèles *ab initio* déterminés à partir des données SAXS

(Figure 60 et Figure 61), qui suggèrent que ClbK possède une architecture allongée délimitant des chambres catalytiques, similairement aux PKS itératives. La principale différence entre ce modèle et les structures en X des PKS itératives est que les deux branches du X sont de longueur inégale, ce qui reflète la différence de taille entre les modules PKS et NRPS de ClbK.

Dans le modèle présenté à la Figure 94, le domaine ACP est positionné à proximité de l'entrée interne du domaine KS et le domaine PCP est positionné au niveau du domaine A, dans la conformation de thiolation. Ces deux domaines doivent délivrer leurs substrats à plusieurs domaines enzymatiques. Comme discuté précédemment, le domaine ACP n'est pas exactement positionné correctement pour une interaction productive avec le domaine KS mais nécessiterait un léger repositionnement pour l'être. En revanche, dans le modèle présenté, le domaine ACP est relativement loin du domaine de Cy. La position relative de ces deux domaines est contrainte par le *linker* ACP-Cy de 25 résidus séparant les deux domaines et la distance entre les deux domaines dépasse la distance que peut couvrir un *linker* de 25 résidus. On peut supposer que le *linker* post-AT, se terminant par 30 résidus non structurés, peut se décoller du domaine KS et ainsi rapprocher le domaine ACP du domaine Cy. Le domaine PCP doit quant à lui visiter les domaines A, Cy et Ox. Comme mentionné par Fortinez *et al.*, la disposition des trois sites catalytiques de ces domaines en arc limite la distance à parcourir pour le domaine PCP (Fortinez *et al.*, 2022).

## Hypothèses sur l'interaction de ClbK avec ses partenaires enzymatiques

Interaction avec l'enzyme ClbG

Le domaine AT de ClbK, non fonctionnel, est complété par l'enzyme discrète ClbG, qui charge le domaine ACP de ClbK avec un substrat aminomalonyl ; ClbG appartient donc au module PKS de ClbK au sens large. Bien que plusieurs structures de complexes *trans*-AT-ACP aient été résolues (Miyanaga *et al.*, 2016 ; Miyanaga, Ouchi, *et al.*, 2018), le mode de reconnaissance du domaine ACP par l'enzyme *trans*-AT dans le contexte du module n'est pas établi (Kosol *et al.*, 2018). En particulier, on ne sait pas si le domaine *trans*-AT interagit avec d'autres domaines que l'ACP au moment du chargement de celui-ci. La conservation du domaine de liaison KS-AT chez les PKS *trans*-AT suggère que celui-ci pourrait servir à ancrer le domaine *trans*-AT mais cela n'a pas été démontré. Le domaine AT tronqué de ClbK pourrait aussi servir de plate-forme d'ancrage pour ClbG.

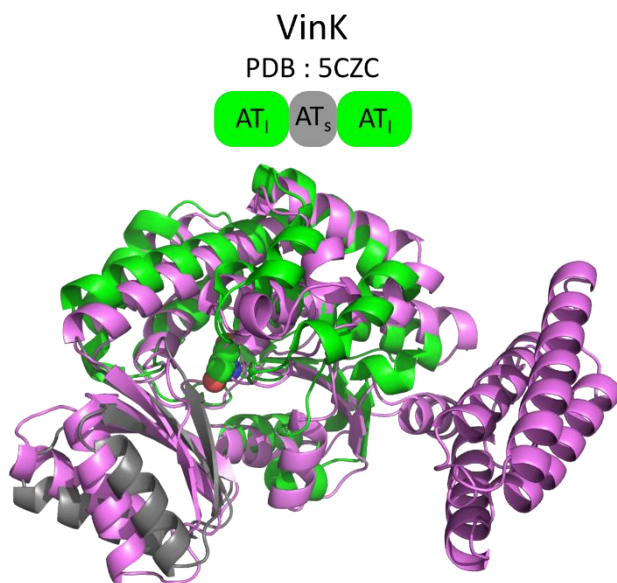


Figure 95. Superposition de la structure de la *trans*-AT VinK et du meilleur modèle AlphaFold2 de ClbG (en rose). La sérine catalytique de VinK est représentée par des sphères. En plus des sous-domaines AT<sub>1</sub> et AT<sub>s</sub>, ClbG possède un sous-domaine de 100 résidus en C-terminal.

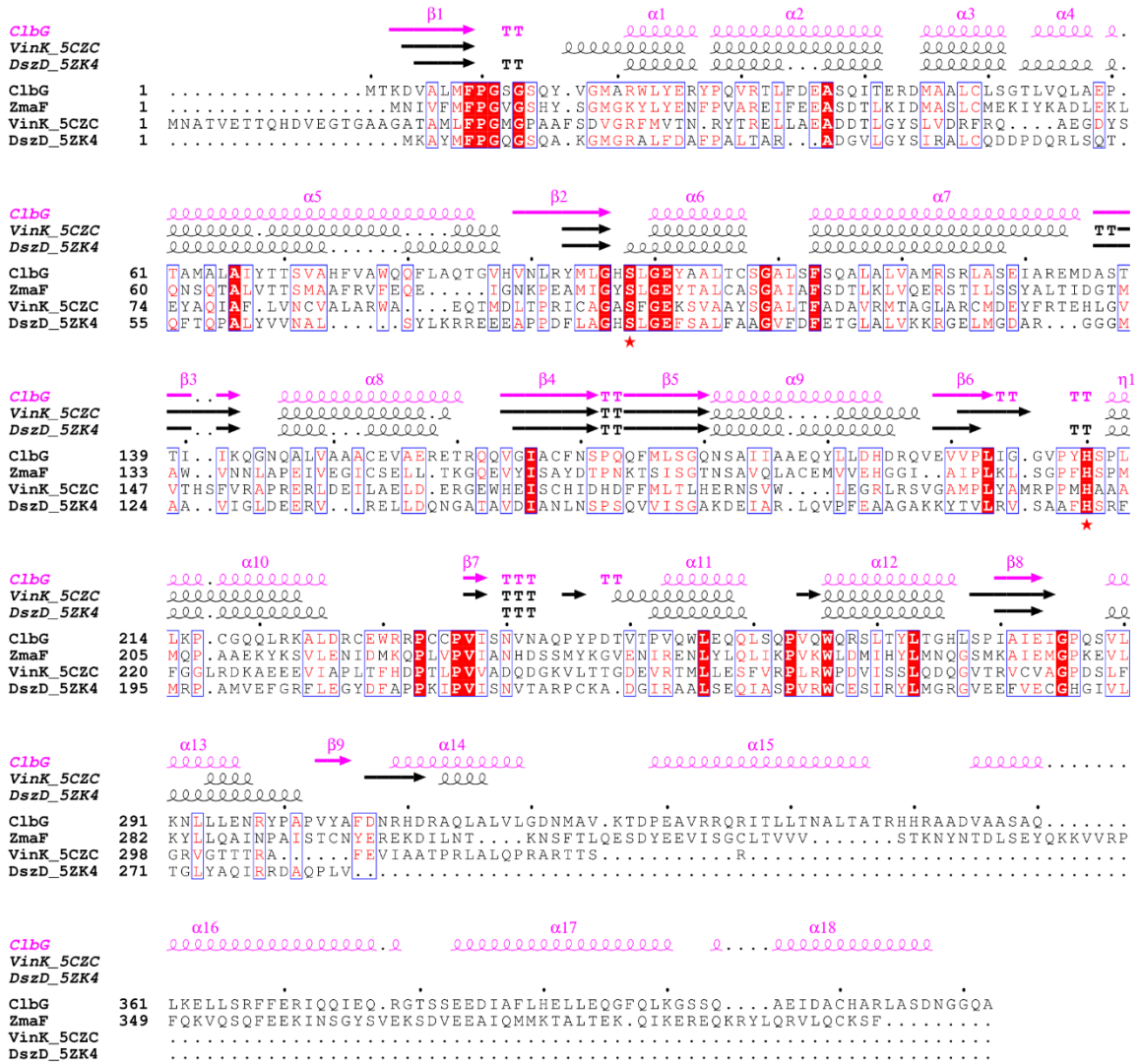


Figure 96. Analyse d’alignements de séquences de domaines *trans*-AT, effectuée par ESPrInt. La sérine et l’histidine catalytiques sont indiquées par des étoiles rouges. Les structures secondaires indiquées pour ClbG proviennent du modèle AlphaFold2. Les résidus en blanc sur un fond rouge sont strictement conservés. Dans une colonne entourée en bleu, au moins 70 % des résidus ont des propriétés physico-chimiques similaires et ces résidus similaires sont en rouge.

Un modèle de ClbG a été généré par AlphaFold2. Ce modèle indique qu’en plus des sous-domaines AT<sub>large</sub> et AT<sub>small</sub>, ClbG possède un sous-domaine de 100 résidus en C-terminal, constitué de quatre longues hélices (Figure 95). La comparaison de la séquence de ClbG avec celles d’autres enzymes *trans*-AT indique que ce sous-domaine est absent chez la plupart des domaines AT discrets, sauf peut-être chez ZmaF, qui possède également un sous-domaine d’une centaine de résidus en C-terminal (Figure 96). ZmaF intervient dans la biosynthèse de l’antibiotique zwittermicine A en transférant une unité aminomalonyl depuis le domaine ACP isolé ZmaH sur un domaine ACP inclus dans la PKS bi-modulaire ZmaA (Carpenter & Williams, 2018). ZmaF et ClbG jouent donc le même rôle dans la biosynthèse de l’unité aminomalonyl

puisque ClbG transfère également un substrat aminomalonyl depuis le domaine ACP isolé ClbE sur le domaine ACP de ClbK (Brachmann *et al.*, 2015). Le sous-domaine C-terminal identifié chez ZmaF et ClbG pourrait jouer un rôle dans la reconnaissance de l'ACP isolé et/ou dans l'ancrage au module PKS pour transférer l'unité aminomalonyle sur le domaine ACP. Afin de clarifier le rôle de ce sous-domaine, j'ai effectué le clonage, des tests d'expression et de purification de la protéine ClbG et des études sont en cours dans l'équipe en vue d'obtenir une structure du complexe entre ClbK-PKS et ClbG (résultats non montrés).

#### Interactions avec les autres méga-enzymes Clb

Les méga-enzymes localisées sur différents polypeptides interagissent grâce aux « docking domains », de courtes régions généralement localisées aux extrémités de ces polypeptides (Smith *et al.*, 2021). ClbK est impliquée dans deux interfaces hybrides NRPS/PKS séparées, une en amont avec le deuxième module NRPS de ClbJ, et une en aval avec la PKS ClbO (Figure 9 et Figure 29). Il a été proposé que des interactions électrostatiques entre « docking domains » jouent un rôle majeur dans les reconnaissances des partenaires aux interfaces séparées NRPS/PKS (Miyanaga, Kudo, *et al.*, 2018). L'importance des résidus chargés positivement à l'extrémité C-terminale de la NRPS EpoB pour la reconnaissance de la PKS EpoC en aval a été démontrée biochimiquement (Liu *et al.*, 2004).

L'extrémité C-terminale de ClbJ est très riche en résidus basiques puisqu'elle en possède 9 sur les 28 derniers résidus (**RAVNQTQQRDASRSRAQQRLVRRHQQR**). Ces résidus basiques pourraient interagir avec des résidus acides de ClbK, bien que seulement deux résidus acides soient retrouvés avant le premier brin bêta du domaine KS de ClbK (**MTYSESD**). ClbJ ne possédant pas de domaine Ox, le cycle thiazoline qu'elle incorpore dans la pré-colibactine est oxydé par le domaine Ox de ClbK et il semblerait que l'oxydation de ce cycle thiazoline soit catalysée avant la prise en charge du substrat contenant le cycle par le module PKS de ClbK (Trautman *et al.*, 2017). Cela signifierait que le domaine PCP<sub>2</sub> de ClbJ, avant de délivrer son substrat au domaine KS de ClbK, devrait aussi délivrer son substrat au domaine Ox de ClbK. Le *linker* à l'extrémité C-terminale de ClbJ pourrait aussi jouer un rôle dans l'interaction avec le domaine Ox de ClbK.

Similairement, l'extrémité C-terminale de ClbK possède une séquence post-PCP riche en résidus chargés mais elle contient presque autant de résidus acides (6) que basiques (7) sur

les 35 derniers résidus (**H**D**I****V****L****A****Q****A****E****V****S****Q****V****S****A****A****R****E****N****L****A****L****R****R****K****R****A****Q****Q****G****E****K****G****D****E**). Le domaine KS de ClbO est quant à lui précédé par une courte séquence contenant trois résidus acides et un résidu basique (**M****A****K****D****D****F****T****C****G****S****L****D**). Les constructions ClbK-NRPS<sub>court</sub> et ClbK-NRPS<sub>long</sub> visaient à étudier le rôle de cette séquence post-PCP mais aucune des deux constructions n'a pu être purifiée (Tableau 6).

L'état dimérique de ClbK, et en particulier de ClbK-PKS\* pose la question de l'état oligomérique de ses partenaires enzymatiques. En particulier, il a déjà été montré que dans certains systèmes hybrides NRPS/PKS, les DD excisés des NRPS sont dimériques, ce qui suggère que les NRPS de ces systèmes hybrides sont elles-mêmes dimériques (Richter *et al.*, 2008). Ainsi, la NRPS bi-modulaire ClbJ précédant ClbK pourrait être dimérique. Les premiers tests que j'ai effectués sur la protéine ClbJ indiquent que la protéine s'exprime mais est peu soluble et ne s'accroche pas aux résines d'affinité (résultats non montrés). Un travail d'optimisation de l'expression et de la purification de cette protéine est nécessaire afin de poursuivre la caractérisation structurale de la chaîne d'assemblage NRPS/PKS de la colibactine initiée durant cette thèse.

Le mode d'interaction de ClbK avec ses partenaires enzymatiques pose la question de la formation d'un méga-complexe synthétisant la colibactine chez *E. coli*. A ma connaissance, la seule étude portant sur l'organisation subcellulaire d'un assemblage NRPS/PKS est celle menée par Straight *et al.* sur l'architecture des méga-enzymes NRPS/PKS Pks participant à la biosynthèse de l'antibiotique bacillaene chez *Bacillus subtilis* (Straight *et al.*, 2007). L'observation par microscopie à fluorescence de protéines fluorescentes fusionnées aux protéines Pks a révélé que ces méga-enzymes colocalisent dans la cellule. Un méga-complexe contenant les protéines Pks a même été observé en microscopie électronique à proximité de la membrane cellulaire et la taille du méga-complexe a été estimée à 10-100 MDa. L'assemblage du complexe est supposé reposer sur l'association homotypique du motif LINKS, retrouvé chez certaines PKS de type *trans*-AT (D. C. Gay *et al.*, 2016). ClbK ne possède pas le motif LINKS mais une insertion bien plus grosse à cette position, qui pourrait éventuellement jouer le même rôle que le motif LINKS dans l'assemblage d'un méga-complexe synthétisant la colibactine, si ce dernier existe.

## Intérêt des travaux pour l'ingénierie des NRPS/PKS

Depuis leur découverte, l'incroyable complexité et diversité des métabolites synthétisés par les NRPS/PKS ont fait rêver à une reprogrammation de ces chaînes d'assemblage afin de générer de nouveaux métabolites d'intérêt biologique. Des méthodes très diverses ont été employées à cette fin, telles que la mutation ponctuelle des domaines d'adénylation, la substitution de ces domaines ou le remplacement de modules complets (Kalkreuter & Williams, 2018). Malgré des efforts considérables, la plupart des tentatives de reprogrammation ont échoué ou ont diminué drastiquement le rendement de biosynthèse du métabolite souhaité. Une nouvelle stratégie basée sur une redéfinition du module, que nous appellerons ici « unité d'échange », a montré des succès prometteurs.

Cette nouvelle définition du module PKS a été proposée après la découverte que les domaines ACP des PKS co-évoluent avec le domaine KS en aval plutôt que celui en amont (A. T. Keatinge-Clay, 2017 ; Vander Wood & Keatinge-Clay, 2018). Chez les PKS *cis*-AT, l'unité d'échange débute après le domaine KS, inclut le domaine AT, les domaines optionnels, le domaine ACP mais aussi le domaine KS en aval et maintient ainsi l'interface native entre ces deux derniers domaines. La conséquence de cette nouvelle définition est que la préservation de l'interface ACP<sub>n-1</sub>-KS<sub>n</sub> au sein de chaînes d'assemblage modifiées pourrait fournir de meilleurs rendements que l'échange de modules selon la définition classique. Cette approche a été employée avec succès récemment en supprimant 7 des 25 modules PKS *cis*-AT participant à la biosynthèse des molécules anticancéreuses stambomycines pour synthétiser de nouvelles molécules de mini-stambomycines (Su *et al.*, 2022). L'ingénierie de l'unité d'échange A-PCP-C selon la même redéfinition du module a aussi été payante chez les NRPS (Bozhüyük *et al.*, 2018).

Les résultats obtenus durant cette thèse sont en faveur de la redéfinition du module chez les hybrides PKS/NRPS également. En effet, la structure de ClbK-PKS\* a révélé à la fois des adaptations du domaine KS pour le module NRPS en amont et des adaptations du domaine ACP pour le module NRPS en aval. Par exemple, le domaine ACP de ClbK possède des caractéristiques de domaine PCP, en l'occurrence une boucle I courte et un motif GGHSL. Ces adaptations sont retrouvées chez les autres domaines ACP de l'îlot *pks* ayant une interface PKS/NRPS mais ne semblent pas être la règle générale des domaines ACP délivrant leur

substrat à un domaine NRPS. Ces caractéristiques spécifiques des domaines ACP interagissant avec des domaines NRPS seront à considérer dans l'ingénierie des chaînes d'assemblage hybrides NRPS/PKS.

Les résultats de cette thèse ont été soumis à publication et sont actuellement en révision.





# Références



- Alekseyev, V. Y., Liu, C. W., Cane, D. E., Puglisi, J. D., & Khosla, C. (2007). Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Science: A Publication of the Protein Society*, *16*(10), 2093-2107. <https://doi.org/10.1110/ps.073011407>
- Alonzo, D. A., Chiche-Lapierre, C., Tarry, M. J., Wang, J., & Schmeing, T. M. (2020). Structural basis of keto acid utilization in nonribosomal depsipeptide synthesis. *Nature Chemical Biology*, *16*(5), 493-496. <https://doi.org/10.1038/s41589-020-0481-5>
- Arthur, J. C., Gharaibeh, R. Z., Mühlbauer, M., Perez-Chanona, E., Uronis, J. M., McCafferty, J., Fodor, A. A., & Jobin, C. (2014). Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. *Nature Communications*, *5*, 4724. <https://doi.org/10.1038/ncomms5724>
- Arthur, J. C., Perez-Chanona, E., Mühlbauer, M., Tomkovich, S., Uronis, J. M., Fan, T.-J., Campbell, B. J., Abujamel, T., Dogan, B., Rogers, A. B., Rhodes, J. M., Stintzi, A., Simpson, K. W., Hansen, J. J., Keku, T. O., Fodor, A. A., & Jobin, C. (2012). Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science (New York, N.Y.)*, *338*(6103), 120-123. <https://doi.org/10.1126/science.1224820>
- Auvray, F., Perrat, A., Arimizu, Y., Chagneau, C. V., Bossuet-Greif, N., Massip, C., Brugère, H., Nougayrède, J.-P., Hayashi, T., Branchu, P., Ogura, Y., & Oswald, E. (2021). Insights into the acquisition of the pks island and production of colibactin in the Escherichia coli population. *Microbial Genomics*, *7*(5), 000579. <https://doi.org/10.1099/mgen.0.000579>
- Bagde, S. R., Mathews, I. I., Fromme, J. C., & Kim, C.-Y. (2021). Modular polyketide synthase contains two reaction chambers that operate asynchronously. *Science (New York, N.Y.)*, *374*(6568), 723-729. <https://doi.org/10.1126/science.abi8532>
- Bakkenist, C. J., & Kastan, M. B. (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature*, *421*(6922), 499-506. <https://doi.org/10.1038/nature01368>
- Berger, H., & Meyer, T. F. (2021). Mechanistic dissection unmasks colibactin as a prevalent mutagenic driver of cancer. *Cancer Cell*, *39*(11), 1439-1441. <https://doi.org/10.1016/j.ccell.2021.10.010>
- Bloudoff, K., Fage, C. D., Marahiel, M. A., & Schmeing, T. M. (2017). Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(1), 95-100. <https://doi.org/10.1073/pnas.1614191114>
- Bloudoff, K., & Schmeing, T. M. (2017). Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochimica Et Biophysica Acta. Proteins and Proteomics*, *1865*(11 Pt B), 1587-1604. <https://doi.org/10.1016/j.bbapap.2017.05.010>
- Bondarev, V., Richter, M., Romano, S., Piel, J., Schwedt, A., & Schulz-Vogt, H. N. (2013). The genus Pseudovibrio contains metabolically versatile bacteria adapted for symbiosis. *Environmental Microbiology*, *15*(7), 2095-2113. <https://doi.org/10.1111/1462-2920.12123>
- Bonhomme, S., Dessen, A., & Macheboeuf, P. (2021). The inherent flexibility of type I non-ribosomal peptide synthetase multienzymes drives their catalytic activities. *Open Biology*, *11*(5), 200386. <https://doi.org/10.1098/rsob.200386>
- Bonnet, M., Buc, E., Sauvanet, P., Darcha, C., Dubois, D., Pereira, B., Déchelotte, P., Bonnet, R., Pezet, D., & Darfeuille-Michaud, A. (2014). Colonization of the human gut by E. coli

- and colorectal cancer risk. *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 20(4), 859-867. <https://doi.org/10.1158/1078-0432.CCR-13-1343>
- Bossuet-Greif, N., Dubois, D., Petit, C., Tronnet, S., Martin, P., Bonnet, R., Oswald, E., & Nougayrède, J.-P. (2016). Escherichia coli C1bS is a colibactin resistance protein. *Molecular Microbiology*, 99(5), 897-908. <https://doi.org/10.1111/mmi.13272>
- Bossuet-Greif, N., Vignard, J., Taieb, F., Mirey, G., Dubois, D., Petit, C., Oswald, E., & Nougayrède, J.-P. (2018). The Colibactin Genotoxin Generates DNA Interstrand Cross-Links in Infected Cells. *MBio*, 9(2), e02393-17. <https://doi.org/10.1128/mBio.02393-17>
- Bowler, M. W., Nurizzo, D., Barrett, R., Beteva, A., Bodin, M., Caserotto, H., Delagenière, S., Dobias, F., Flot, D., Giraud, T., Guichard, N., Guijarro, M., Lentini, M., Leonard, G. A., McSweeney, S., Oskarsson, M., Schmidt, W., Snigirev, A., von Stetten, D., ... Mueller-Dieckmann, C. (2015). MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *Journal of Synchrotron Radiation*, 22(6), 1540-1547. <https://doi.org/10.1107/S1600577515016604>
- Bozhüyük, K. A. J., Fleischhacker, F., Linck, A., Wesche, F., Tietze, A., Niesert, C.-P., & Bode, H. B. (2018). De novo design and engineering of non-ribosomal peptide synthetases. *Nature Chemistry*, 10(3), 275-281. <https://doi.org/10.1038/nchem.2890>
- Brachmann, A. O., Garcie, C., Wu, V., Martin, P., Ueoka, R., Oswald, E., & Piel, J. (2015). Colibactin biosynthesis and biological activity depend on the rare aminomalonyl polyketide precursor. *Chemical Communications (Cambridge, England)*, 51(66), 13138-13141. <https://doi.org/10.1039/c5cc02718g>
- Braud, A., Hannauer, M., Mislin, G. L. A., & Schalk, I. J. (2009). The Pseudomonas aeruginosa pyochelin-iron uptake pathway and its metal specificity. *Journal of Bacteriology*, 191(11), 3517-3525. <https://doi.org/10.1128/JB.00010-09>
- Bretschneider, T., Heim, J. B., Heine, D., Winkler, R., Busch, B., Kusebauch, B., Stehle, T., Zocher, G., & Hertweck, C. (2013). Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature*, 502(7469), 124-128. <https://doi.org/10.1038/nature12588>
- Brignole, E. J., Smith, S., & Asturias, F. J. (2009). Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nature Structural & Molecular Biology*, 16(2), 190-197. <https://doi.org/10.1038/nsmb.1532>
- Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F., & Weissman, K. J. (2003). The structure of docking domains in modular polyketide synthases. *Chemistry & Biology*, 10(8), 723-731. [https://doi.org/10.1016/s1074-5521\(03\)00156-x](https://doi.org/10.1016/s1074-5521(03)00156-x)
- Brotherton, C. A., & Balskus, E. P. (2013). A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. *Journal of the American Chemical Society*, 135(9), 3359-3362. <https://doi.org/10.1021/ja312154m>
- Buc, E., Dubois, D., Sauvanet, P., Raisch, J., Delmas, J., Darfeuille-Michaud, A., Pezet, D., & Bonnet, R. (2013). High prevalence of mucosa-associated E. coli producing cyclomodulin and genotoxin in colon cancer. *PloS One*, 8(2), e56964. <https://doi.org/10.1371/journal.pone.0056964>
- Callaway, E. (2020). Revolutionary cryo-EM is taking over structural biology. *Nature*, 578(7794), 201-201. <https://doi.org/10.1038/d41586-020-00341-9>
- Campbell, N. A. (2012). *Biologie 9e édition*. <https://www.decitre.fr/livres/biologie-9782761350655.html>

- Cancer Genome Atlas Network. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407), 330-337. <https://doi.org/10.1038/nature11252>
- Carpenter, S. M., & Williams, G. J. (2018). Extender Unit Promiscuity and Orthogonal Protein Interactions of an Aminomalonyl-ACP Utilizing Trans-Acyltransferase from Zwittermicin Biosynthesis. *ACS chemical biology*, 13(12), 3361-3373. <https://doi.org/10.1021/acscchembio.8b00867>
- Carroni, M., & Saibil, H. R. (2016). Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods (San Diego, Calif.)*, 95, 78-85. <https://doi.org/10.1016/j.ymeth.2015.11.023>
- Chagneau, C. V., Garcie, C., Bossuet-Greif, N., Tronnet, S., Brachmann, A. O., Piel, J., Nougayrède, J.-P., Martin, P., & Oswald, E. (2019). The Polyamine Spermidine Modulates the Production of the Bacterial Genotoxin Colibactin. *MSphere*, 4(5), e00414-19. <https://doi.org/10.1128/mSphere.00414-19>
- Chagneau, C. V., Massip, C., Bossuet-Greif, N., Fremez, C., Motta, J.-P., Shima, A., Besson, C., Le Faouder, P., Cénac, N., Roth, M.-P., Coppin, H., Fontanié, M., Martin, P., Nougayrède, J.-P., & Oswald, E. (2021). Uropathogenic *E. coli* induces DNA damage in the bladder. *PLoS Pathogens*, 17(2), e1009310. <https://doi.org/10.1371/journal.ppat.1009310>
- Chen, J., Byun, H., Liu, R., Jung, I.-J., Pu, Q., Zhu, C. Y., Tanchoco, E., Alavi, S., Degnan, P. H., Ma, A. T., Roggiani, M., Beld, J., Goulian, M., Hsiao, A., & Zhu, J. (2022). A commensal-encoded genotoxin drives restriction of *Vibrio cholerae* colonization and host gut microbiome remodeling. *Proceedings of the National Academy of Sciences of the United States of America*, 119(11), e2121180119. <https://doi.org/10.1073/pnas.2121180119>
- Cogan, D. P., Zhang, K., Li, X., Li, S., Pintilie, G. D., Roh, S.-H., Craik, C. S., Chiu, W., & Khosla, C. (2021). Mapping the catalytic conformations of an assembly-line polyketide synthase module. *Science (New York, N.Y.)*, 374(6568), 729-734. <https://doi.org/10.1126/science.abi8358>
- Cougnoux, A., Gibold, L., Robin, F., Dubois, D., Pradel, N., Darfeuille-Michaud, A., Dalmaso, G., Delmas, J., & Bonnet, R. (2012). Analysis of structure-function relationships in the colibactin-maturing enzyme ClbP. *Journal of Molecular Biology*, 424(3-4), 203-214. <https://doi.org/10.1016/j.jmb.2012.09.017>
- Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica. Section D, Biological Crystallography*, 62(Pt 9), 1002-1011. <https://doi.org/10.1107/S0907444906022116>
- Cowtan, K. (2010). Recent developments in classical density modification. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 4), 470-478. <https://doi.org/10.1107/S090744490903947X>
- Crosby, J., & Crump, M. P. (2012). The structural role of the carrier protein--active controller or passive carrier. *Natural Product Reports*, 29(10), 1111-1137. <https://doi.org/10.1039/c2np20062g>
- Cuevas-Ramos, G., Petit, C. R., Marcq, I., Boury, M., Oswald, E., & Nougayrède, J.-P. (2010). *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, 107(25), 11537-11542. <https://doi.org/10.1073/pnas.1001261107>

- Davison, J., Dorival, J., Rabeharindranto, H., Mazon, H., Chagot, B., Gruez, A., & Weissman, K. J. (2014). Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module. *Chemical Science*, *5*(8), 3081-3095. <https://doi.org/10.1039/C3SC53511H>
- Deans, A. J., & West, S. C. (2011). DNA interstrand crosslink repair and cancer. *Nature Reviews. Cancer*, *11*(7), 467-480. <https://doi.org/10.1038/nrc3088>
- Dejea, C. M., Fathi, P., Craig, J. M., Boleij, A., Taddese, R., Geis, A. L., Wu, X., DeStefano Shields, C. E., Hechenbleikner, E. M., Huso, D. L., Anders, R. A., Giardiello, F. M., Wick, E. C., Wang, H., Wu, S., Pardoll, D. M., Housseau, F., & Sears, C. L. (2018). Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science (New York, N.Y.)*, *359*(6375), 592-597. <https://doi.org/10.1126/science.aah3648>
- Dodge, G. J., Maloney, F. P., & Smith, J. L. (2018). Protein-protein interactions in « cis-AT » polyketide synthases. *Natural Product Reports*, *35*(10), 1082-1096. <https://doi.org/10.1039/c8np00058a>
- Dodson, E. (2021). Introduction to molecular replacement: a time perspective. *Acta Crystallographica. Section D, Structural Biology*, *77*(Pt 7), 867-879. <https://doi.org/10.1107/S2059798321004368>
- Dorival, J., Annaval, T., Risser, F., Collin, S., Roblin, P., Jacob, C., Gruez, A., Chagot, B., & Weissman, K. J. (2016). Characterization of Intersubunit Communication in the Virginiamycin trans-Acyl Transferase Polyketide Synthase. *Journal of the American Chemical Society*, *138*(12), 4155-4167. <https://doi.org/10.1021/jacs.5b13372>
- Dougherty, M. W., & Jobin, C. (2021). Shining a Light on Colibactin Biology. *Toxins*, *13*(5), 346. <https://doi.org/10.3390/toxins13050346>
- Dowling, D. P., Kung, Y., Croft, A. K., Taghizadeh, K., Kelly, W. L., Walsh, C. T., & Drennan, C. L. (2016). Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(44), 12432-12437. <https://doi.org/10.1073/pnas.1608615113>
- Drake, E. J., Miller, B. R., Shi, C., Tarrasch, J. T., Sundlov, J. A., Allen, C. L., Skiniotis, G., Aldrich, C. C., & Gulick, A. M. (2016). Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature*, *529*(7585), 235-238. <https://doi.org/10.1038/nature16163>
- Du, L., Sánchez, C., & Shen, B. (2001). Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metabolic Engineering*, *3*(1), 78-95. <https://doi.org/10.1006/mben.2000.0171>
- Dubois, D., Baron, O., Cougnoux, A., Delmas, J., Pradel, N., Boury, M., Bouchon, B., Bringer, M.-A., Nougayrède, J.-P., Oswald, E., & Bonnet, R. (2011). ClbP is a prototype of a peptidase subgroup involved in biosynthesis of nonribosomal peptides. *The Journal of Biological Chemistry*, *286*(41), 35562-35570. <https://doi.org/10.1074/jbc.M111.221960>
- Duckworth, B. P., Wilson, D. J., & Aldrich, C. C. (2016). Measurement of Nonribosomal Peptide Synthetase Adenylation Domain Activity Using a Continuous Hydroxylamine Release Assay. *Methods in Molecular Biology (Clifton, N.J.)*, *1401*, 53-61. [https://doi.org/10.1007/978-1-4939-3375-4\\_3](https://doi.org/10.1007/978-1-4939-3375-4_3)
- Dupeux, F., Röwer, M., Seroul, G., Blot, D., & Márquez, J. A. (2011). A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta*

- Crystallographica. Section D, Biological Crystallography*, 67(Pt 11), 915-919. <https://doi.org/10.1107/S0907444911036225>
- Dutta, S., Whicher, J. R., Hansen, D. A., Hale, W. A., Chemler, J. A., Congdon, G. R., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L., & Skiniotis, G. (2014). Structure of a modular polyketide synthase. *Nature*, 510(7506), 512-517. <https://doi.org/10.1038/nature13423>
- Dziubańska-Kusibab, P. J., Berger, H., Battistini, F., Bouwman, B. A. M., Iftexhar, A., Katainen, R., Cajuso, T., Crosetto, N., Orozco, M., Aaltonen, L. A., & Meyer, T. F. (2020). Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. *Nature Medicine*, 26(7), 1063-1069. <https://doi.org/10.1038/s41591-020-0908-2>
- Eklöf, V., Löfgren-Burström, A., Zingmark, C., Edin, S., Larsson, P., Karling, P., Alexeyev, O., Rutegård, J., Wikberg, M. L., & Palmqvist, R. (2017). Cancer-associated fecal microbial markers in colorectal cancer detection. *International Journal of Cancer*, 141(12), 2528-2536. <https://doi.org/10.1002/ijc.31011>
- Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica. Section D, Biological Crystallography*, 60(Pt 12 Pt 1), 2126-2132. <https://doi.org/10.1107/S0907444904019158>
- Engel, P., Vizcaino, M. I., & Crawford, J. M. (2015). Gut symbionts from distinct hosts exhibit genotoxic activity via divergent colibactin biosynthesis pathways. *Applied and Environmental Microbiology*, 81(4), 1502-1512. <https://doi.org/10.1128/AEM.03283-14>
- Engilberge, S., Riobé, F., Di Pietro, S., Lassalle, L., Coquelle, N., Arnaud, C.-A., Pitrat, D., Mulatier, J.-C., Madern, D., Breyton, C., Maury, O., & Girard, E. (2017). Crystallophore: a versatile lanthanide complex for protein crystallography combining nucleating effects, phasing properties, and luminescence. *Chemical Science*, 8(9), 5909-5917. <https://doi.org/10.1039/c7sc00758b>
- Escobar-Páramo, P., Clermont, O., Blanc-Potard, A.-B., Bui, H., Le Bouguéneq, C., & Denamur, E. (2004). A Specific Genetic Background Is Required for Acquisition and Expression of Virulence Factors in *Escherichia coli*. *Molecular Biology and Evolution*, 21(6), 1085-1094. <https://doi.org/10.1093/molbev/msh118>
- Faïs, T., Cougnoux, A., Dalmasso, G., Laurent, F., Delmas, J., & Bonnet, R. (2016). Antibiotic Activity of *Escherichia coli* against Multiresistant *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 60(11), 6986-6988. <https://doi.org/10.1128/AAC.00130-16>
- Faïs, T., Delmas, J., Barnich, N., Bonnet, R., & Dalmasso, G. (2018). Colibactin: More Than a New Bacterial Toxin. *Toxins*, 10(4), E151. <https://doi.org/10.3390/toxins10040151>
- Feng, L., Gordon, M. T., Liu, Y., Basso, K. B., & Butcher, R. A. (2021). Mapping the biosynthetic pathway of a hybrid polyketide-nonribosomal peptide in a metazoan. *Nature Communications*, 12(1), 4912. <https://doi.org/10.1038/s41467-021-24682-9>
- Fortinez, C. M., Bloudoff, K., Harrigan, C., Sharon, I., Strauss, M., & Schmeing, T. M. (2022). Structures and function of a tailoring oxidase in complex with a nonribosomal peptide synthetase module. *Nature Communications*, 13(1), 548. <https://doi.org/10.1038/s41467-022-28221-y>
- Franke, D., & Svergun, D. I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography*, 42(2), 342-346. <https://doi.org/10.1107/S0021889809000338>



- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExpASY: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, *31*(13), 3784-3788. <https://doi.org/10.1093/nar/gkg563>
- Gatsios, A., Kim, C. S., & Crawford, J. M. (2021). Escherichia coli small molecule metabolism at the host-microorganism interface. *Nature Chemical Biology*, *17*(10), 1016-1026. <https://doi.org/10.1038/s41589-021-00807-5>
- Gay, D. C., Gay, G., Axelrod, A. J., Jenner, M., Kohlhaas, C., Kampa, A., Oldham, N. J., Piel, J., & Keatinge-Clay, A. T. (2014). A close look at a ketosynthase from a trans-acyltransferase modular polyketide synthase. *Structure (London, England: 1993)*, *22*(3), 444-451. <https://doi.org/10.1016/j.str.2013.12.016>
- Gay, D. C., Wagner, D. T., Meinke, J. L., Zogzas, C. E., Gay, G. R., & Keatinge-Clay, A. T. (2016). The LINKS motif zippers trans-acyltransferase polyketide synthase assembly lines into a biosynthetic megacomplex. *Journal of Structural Biology*, *193*(3), 196-205. <https://doi.org/10.1016/j.jsb.2015.12.011>
- Gay, D., You, Y.-O., Keatinge-Clay, A., & Cane, D. E. (2013). Structure and stereospecificity of the dehydratase domain from the terminal module of the rifamycin polyketide synthase. *Biochemistry*, *52*(49), 8916-8928. <https://doi.org/10.1021/bi400988t>
- Gérard, P. (2016). Gut microbiota and obesity. *Cellular and Molecular Life Sciences: CMLS*, *73*(1), 147-162. <https://doi.org/10.1007/s00018-015-2061-5>
- Goldschmidt, L., Cooper, D. R., Derewenda, Z. S., & Eisenberg, D. (2007). Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Science: A Publication of the Protein Society*, *16*(8), 1569-1576. <https://doi.org/10.1110/ps.072914007>
- Gulick, A. M. (2009). Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chemical Biology*, *4*(10), 811-827. <https://doi.org/10.1021/cb900156h>
- Gulick, A. M. (2017). Nonribosomal peptide synthetase biosynthetic clusters of ESKAPE pathogens. *Natural Product Reports*, *34*(8), 981-1009. <https://doi.org/10.1039/c7np00029d>
- Gulick, A. M., & Aldrich, C. C. (2018). Trapping interactions between catalytic domains and carrier proteins of modular biosynthetic enzymes with chemical probes. *Natural Product Reports*, *35*(11), 1156-1184. <https://doi.org/10.1039/c8np00044a>
- Guntaka, N. S., Healy, A. R., Crawford, J. M., Herzon, S. B., & Bruner, S. D. (2017). Structure and Functional Analysis of ClbQ, an Unusual Intermediate-Releasing Thioesterase from the Colibactin Biosynthetic Pathway. *ACS Chemical Biology*, *12*(10), 2598-2608. <https://doi.org/10.1021/acschembio.7b00479>
- Guzman, K. M., & Khosla, C. (2022). Fragment antigen binding domains (Fabs) as tools to study assembly-line polyketide synthases. *Synthetic and Systems Biotechnology*, *7*(1), 506-512. <https://doi.org/10.1016/j.synbio.2021.12.003>
- Healy, A. R., Wernke, K. M., Kim, C. S., Lees, N. R., Crawford, J. M., & Herzon, S. B. (2019). Synthesis and reactivity of precolibactin 886. *Nature Chemistry*, *11*(10), 890-898. <https://doi.org/10.1038/s41557-019-0338-2>
- Helfrich, E. J. N., Ueoka, R., Dolev, A., Rust, M., Meoded, R. A., Bhushan, A., Califano, G., Costa, R., Guggler, M., Steinbeck, C., Moreno, P., & Piel, J. (2019). Automated structure prediction of trans-acyltransferase polyketide synthase products. *Nature Chemical Biology*, *15*(8), 813-821. <https://doi.org/10.1038/s41589-019-0313-7>

- Herbst, D. A., Huitt-Roehl, C. R., Jakob, R. P., Kravetz, J. M., Storm, P. A., Alley, J. R., Townsend, C. A., & Maier, T. (2018). The structural organization of substrate loading in iterative polyketide synthases. *Nature Chemical Biology*, *14*(5), 474-479. <https://doi.org/10.1038/s41589-018-0026-3>
- Herbst, D. A., Jakob, R. P., Zähringer, F., & Maier, T. (2016). Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature*, *531*(7595), 533-537. <https://doi.org/10.1038/nature16993>
- Herbst, D. A., Townsend, C. A., & Maier, T. (2018). The architectures of iterative type I PKS and FAS. *Natural Product Reports*, *35*(10), 1046-1069. <https://doi.org/10.1039/c8np00039e>
- Hertweck, C. (2009). The biosynthetic logic of polyketide diversity. *Angewandte Chemie (International Ed. in English)*, *48*(26), 4688-4716. <https://doi.org/10.1002/anie.200806121>
- Hillson, N. J., & Walsh, C. T. (2003). Dimeric structure of the six-domain VibF subunit of vibriobactin synthetase: mutant domain activity regain and ultracentrifugation studies. *Biochemistry*, *42*(3), 766-775. <https://doi.org/10.1021/bi026903h>
- Homburg, S., Oswald, E., Hacker, J., & Dobrindt, U. (2007). Expression analysis of the colibactin gene cluster coding for a novel polyketide in Escherichia coli. *FEMS Microbiology Letters*, *275*(2), 255-262. <https://doi.org/10.1111/j.1574-6968.2007.00889.x>
- Huang, Y., Tang, G.-L., Pan, G., Chang, C.-Y., & Shen, B. (2016). Characterization of the Ketosynthase and Acyl Carrier Protein Domains at the LnmI Nonribosomal Peptide Synthetase-Polyketide Synthase Interface for Leinamycin Biosynthesis. *Organic Letters*, *18*(17), 4288-4291. <https://doi.org/10.1021/acs.orglett.6b02033>
- Iftekhar, A., Berger, H., Bouznad, N., Heuberger, J., Boccellato, F., Dobrindt, U., Hermeking, H., Sigal, M., & Meyer, T. F. (2021). Genomic aberrations after short-term exposure to colibactin-producing E. coli transform primary colon epithelial cells. *Nature Communications*, *12*(1), 1003. <https://doi.org/10.1038/s41467-021-21162-y>
- Iskandar, K., Murugaiyan, J., Hammoudi Halat, D., Hage, S. E., Chibabhai, V., Adukkadukkam, S., Roques, C., Molinier, L., Salameh, P., & Van Dongen, M. (2022). Antibiotic Discovery and Resistance: The Chase and the Race. *Antibiotics*, *11*(2), 182. <https://doi.org/10.3390/antibiotics11020182>
- Izoré, T., & Cryle, M. J. (2018). The many faces and important roles of protein-protein interactions during non-ribosomal peptide synthesis. *Natural Product Reports*, *35*(11), 1120-1139. <https://doi.org/10.1039/c8np00038g>
- Jenner, M., Frank, S., Kampa, A., Kohlhaas, C., Pöplau, P., Briggs, G. S., Piel, J., & Oldham, N. J. (2013). Substrate specificity in ketosynthase domains from trans-AT polyketide synthases. *Angewandte Chemie (International Ed. in English)*, *52*(4), 1143-1147. <https://doi.org/10.1002/anie.201207690>
- Jiang, Y., Stornetta, A., Villalta, P. W., Wilson, M. R., Boudreau, P. D., Zha, L., Balbo, S., & Balskus, E. P. (2019). Reactivity of an Unusual Amidase May Explain Colibactin's DNA Cross-Linking Activity. *Journal of the American Chemical Society*, *141*(29), 11489-11496. <https://doi.org/10.1021/jacs.9b02453>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589. <https://doi.org/10.1038/s41586-021-03819-2>

- Kabsch, W. (2010). XDS. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 2), 125-132. <https://doi.org/10.1107/S0907444909047337>
- Kalkreuter, E., & Williams, G. J. (2018). Engineering enzymatic assembly lines for the production of new antimicrobials. *Current Opinion in Microbiology*, 45, 140-148. <https://doi.org/10.1016/j.mib.2018.04.005>
- Kao, C. M., Pieper, R., Cane, D. E., & Khosla, C. (1996). Evidence for two catalytically independent clusters of active sites in a functional modular polyketide synthase. *Biochemistry*, 35(38), 12363-12368. <https://doi.org/10.1021/bi9616312>
- Kaper, J. B., Nataro, J. P., & Mobley, H. L. (2004). Pathogenic Escherichia coli. *Nature Reviews. Microbiology*, 2(2), 123-140. <https://doi.org/10.1038/nrmicro818>
- Kapur, S., Chen, A. Y., Cane, D. E., & Khosla, C. (2010). Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proceedings of the National Academy of Sciences of the United States of America*, 107(51), 22066-22071. <https://doi.org/10.1073/pnas.1014081107>
- Katsuyama, Y., Sone, K., Harada, A., Kawai, S., Urano, N., Adachi, N., Moriya, T., Kawasaki, M., Shin-Ya, K., Senda, T., & Ohnishi, Y. (2021). Structural and Functional Analyses of the Tridomain-Nonribosomal Peptide Synthetase FmoA3 for 4-Methyloxazoline Ring Formation. *Angewandte Chemie (International Ed. in English)*, 60(26), 14554-14562. <https://doi.org/10.1002/anie.202102760>
- Katz, L., & Baltz, R. H. (2016). Natural product discovery: past, present, and future. *Journal of Industrial Microbiology & Biotechnology*, 43(2-3), 155-176. <https://doi.org/10.1007/s10295-015-1723-5>
- Keatinge-Clay, A. (2008). Crystal structure of the erythromycin polyketide synthase dehydratase. *Journal of Molecular Biology*, 384(4), 941-953. <https://doi.org/10.1016/j.jmb.2008.09.084>
- Keatinge-Clay, A. T. (2012). The structures of type I polyketide synthases. *Natural Product Reports*, 29(10), 1050-1073. <https://doi.org/10.1039/c2np20019h>
- Keatinge-Clay, A. T. (2017). Polyketide Synthase Modules Redefined. *Angewandte Chemie (International Ed. in English)*, 56(17), 4658-4660. <https://doi.org/10.1002/anie.201701281>
- Kim, C. S., Turocy, T., Moon, G., Shine, E. E., & Crawford, J. M. (2021). Escherichia coli-Derived  $\gamma$ -Lactams and Structurally Related Metabolites Are Produced at the Intersection of Colibactin and Fatty Acid Biosynthesis. *Organic Letters*, 23(17), 6895-6899. <https://doi.org/10.1021/acs.orglett.1c02461>
- Kohlhaas, C., Jenner, M., Kampa, A., Briggs, G. S., Afonso, J. P., Piel, J., & Oldham, N. J. (2013). Amino acid-accepting ketosynthase domain from a trans-AT polyketide synthase exhibits high selectivity for predicted intermediate. *Chemical Science*, 4(8), 3212-3217. <https://doi.org/10.1039/C3SC50540E>
- Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., & Svergun, D. I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, 36(5), 1277-1282. <https://doi.org/10.1107/S0021889803012779>
- Kosol, S., Jenner, M., Lewandowski, J. R., & Challis, G. L. (2018). Protein-protein interactions in trans-AT polyketide synthases. *Natural Product Reports*, 35(10), 1097-1109. <https://doi.org/10.1039/c8np00066b>
- Kreitler, D. F., Gemmell, E. M., Schaffer, J. E., Wenczewicz, T. A., & Gulick, A. M. (2019). The structural basis of N-acyl- $\alpha$ -amino- $\beta$ -lactone formation catalyzed by a nonribosomal

- peptide synthetase. *Nature Communications*, *10*(1), 3432. <https://doi.org/10.1038/s41467-019-11383-7>
- Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology*, *372*(3), 774-797. <https://doi.org/10.1016/j.jmb.2007.05.022>
- Kudo, F., Miyanaga, A., & Eguchi, T. (2019). Structural basis of the nonribosomal codes for nonproteinogenic amino acid selective adenylation enzymes in the biosynthesis of natural products. *Journal of Industrial Microbiology & Biotechnology*, *46*(3-4), 515-536. <https://doi.org/10.1007/s10295-018-2084-7>
- Kwak, E. L., & Chung, D. C. (2007). Hereditary colorectal cancer syndromes: an overview. *Clinical Colorectal Cancer*, *6*(5), 340-344. <https://doi.org/10.3816/CCC.2007.n.002>
- La pénicilline I - Découverte d'un antibiotique.* (s. d.). CultureSciences-Chimie. Consulté 17 mars 2022, à l'adresse <https://culturesciences.chimie.ens.fr/thematiques/chimie-organique/chimie-pharmaceutique/la-penicilline-i-decouverte-d-un-antibiotique>
- Labby, K. J., Watsula, S. G., & Garneau-Tsodikova, S. (2015). Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis. *Natural Product Reports*, *32*(5), 641-653. <https://doi.org/10.1039/c4np00120f>
- Lavelle, A., & Sokol, H. (2020). Gut microbiota-derived metabolites as key actors in inflammatory bowel disease. *Nature Reviews. Gastroenterology & Hepatology*, *17*(4), 223-237. <https://doi.org/10.1038/s41575-019-0258-z>
- Levy, M., Kolodziejczyk, A. A., Thaiss, C. A., & Elinav, E. (2017). Dysbiosis and the immune system. *Nature Reviews. Immunology*, *17*(4), 219-232. <https://doi.org/10.1038/nri.2017.7>
- Li, V. S. W. (2020). Modelling intestinal inflammation and infection using 'mini-gut' organoids. *Nature Reviews. Gastroenterology & Hepatology*, 1-2. <https://doi.org/10.1038/s41575-020-00391-4>
- Li, X., Sevillano, N., La Greca, F., Deis, L., Liu, Y.-C., Deller, M. C., Mathews, I. I., Matsui, T., Cane, D. E., Craik, C. S., & Khosla, C. (2018). Structure-Function Analysis of the Extended Conformation of a Polyketide Synthase Module. *Journal of the American Chemical Society*, *140*(21), 6518-6521. <https://doi.org/10.1021/jacs.8b02100>
- Li, Z.-R., Li, J., Cai, W., Lai, J. Y. H., McKinnie, S. M. K., Zhang, W.-P., Moore, B. S., Zhang, W., & Qian, P.-Y. (2019). Macrocyclic colibactin induces DNA double-strand breaks via copper-mediated oxidative cleavage. *Nature Chemistry*, *11*(10), 880-889. <https://doi.org/10.1038/s41557-019-0317-7>
- Liu, F., Garneau, S., & Walsh, C. T. (2004). Hybrid nonribosomal peptide-polyketide interfaces in epothilone biosynthesis: minimal requirements at N and C termini of EpoB for elongation. *Chemistry & Biology*, *11*(11), 1533-1542. <https://doi.org/10.1016/j.chembiol.2004.08.017>
- Lohman, J. R., Ma, M., Osipiuk, J., Nocek, B., Kim, Y., Chang, C., Cuff, M., Mack, J., Bigelow, L., Li, H., Endres, M., Babnigg, G., Joachimiak, A., Phillips, G. N., & Shen, B. (2015). Structural and evolutionary relationships of « AT-less » type I polyketide synthase ketosynthases. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41), 12693-12698. <https://doi.org/10.1073/pnas.1515460112>
- Longenecker, K. L., Garrard, S. M., Sheffield, P. J., & Derewenda, Z. S. (2001). Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI. *Acta Crystallographica. Section D, Biological Crystallography*, *57*(Pt 5), 679-688. <https://doi.org/10.1107/s09074444901003122>

- Lowry, B., Li, X., Robbins, T., Cane, D. E., & Khosla, C. (2016). A Turnstile Mechanism for the Controlled Growth of Biosynthetic Intermediates on Assembly Line Polyketide Synthases. *ACS Central Science*, 2(1), 14-20. <https://doi.org/10.1021/acscentsci.5b00321>
- Macheroux, P. (1999). UV-visible spectroscopy as a tool to study flavoproteins. *Methods in Molecular Biology (Clifton, N.J.)*, 131, 1-7. <https://doi.org/10.1385/1-59259-266-X:1>
- Maier, T., Leibundgut, M., & Ban, N. (2008). The crystal structure of a mammalian fatty acid synthase. *Science (New York, N.Y.)*, 321(5894), 1315-1322. <https://doi.org/10.1126/science.1161269>
- Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I., & Franke, D. (2021). ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *Journal of Applied Crystallography*, 54(1), 343-355. <https://doi.org/10.1107/S1600576720013412>
- Martin, P., Marcq, I., Magistro, G., Penary, M., Garcie, C., Payros, D., Boury, M., Olier, M., Nougayrède, J.-P., Audebert, M., Chalut, C., Schubert, S., & Oswald, E. (2013). Interplay between siderophores and colibactin genotoxin biosynthetic pathways in *Escherichia coli*. *PLoS Pathogens*, 9(7), e1003437. <https://doi.org/10.1371/journal.ppat.1003437>
- Masschelein, J., Sydor, P. K., Hobson, C., Howe, R., Jones, C., Roberts, D. M., Ling Yap, Z., Parkhill, J., Mahenthalingam, E., & Challis, G. L. (2019). A dual transacylation mechanism for polyketide synthase chain release in enacyloxin antibiotic biosynthesis. *Nature chemistry*, 11(10), 906-912. <https://doi.org/10.1038/s41557-019-0309-7>
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, 40(Pt 4), 658-674. <https://doi.org/10.1107/S0021889807021206>
- McPherson, A., Malkin, A. J., & Kuznetsov, Y. G. (1995). The science of macromolecular crystallization. *Structure (London, England: 1993)*, 3(8), 759-768. [https://doi.org/10.1016/s0969-2126\(01\)00211-8](https://doi.org/10.1016/s0969-2126(01)00211-8)
- Mertens, H. D. T., & Svergun, D. I. (2010). Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of Structural Biology*, 172(1), 128-141. <https://doi.org/10.1016/j.jsb.2010.06.012>
- Miller, D. A., Luo, L., Hillson, N., Keating, T. A., & Walsh, C. T. (2002). Yersiniabactin synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*. *Chemistry & Biology*, 9(3), 333-344. [https://doi.org/10.1016/s1074-5521\(02\)00115-1](https://doi.org/10.1016/s1074-5521(02)00115-1)
- Miyanaga, A., Iwasawa, S., Shinohara, Y., Kudo, F., & Eguchi, T. (2016). Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 113(7), 1802-1807. <https://doi.org/10.1073/pnas.1520042113>
- Miyanaga, A., Kudo, F., & Eguchi, T. (2018). Protein-protein interactions in polyketide synthase-nonribosomal peptide synthetase hybrid assembly lines. *Natural Product Reports*, 35(11), 1185-1209. <https://doi.org/10.1039/c8np00022k>
- Miyanaga, A., Ouchi, R., Ishikawa, F., Goto, E., Tanabe, G., Kudo, F., & Eguchi, T. (2018). Structural Basis of Protein-Protein Interactions between a trans-Acting Acyltransferase and Acyl Carrier Protein in Polyketide Disorazole Biosynthesis. *Journal of the American Chemical Society*, 140(25), 7970-7978. <https://doi.org/10.1021/jacs.8b04162>

- Moretto, L., Heylen, R., Holroyd, N., Vance, S., & Broadhurst, R. W. (2019). Modular type I polyketide synthase acyl carrier protein domains share a common N-terminally extended fold. *Scientific Reports*, *9*(1), 2325. <https://doi.org/10.1038/s41598-019-38747-9>
- Mori, S., Pang, A. H., Lundy, T. A., Garzan, A., Tsodikov, O. V., & Garneau-Tsodikova, S. (2018). Structural basis for backbone N-methylation by an interrupted adenylation domain. *Nature Chemical Biology*, *14*(5), 428-430. <https://doi.org/10.1038/s41589-018-0014-7>
- Mousa, J. J., Yang, Y., Tomkovich, S., Shima, A., Newsome, R. C., Tripathi, P., Oswald, E., Bruner, S. D., & Jobin, C. (2016). MATE transport of the E. coli-derived genotoxin colibactin. *Nature Microbiology*, *1*, 15009. <https://doi.org/10.1038/nmicrobiol.2015.9>
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica. Section D, Biological Crystallography*, *67*(Pt 4), 355-367. <https://doi.org/10.1107/S0907444911001314>
- Nougayrède, J.-P., Chagneau, C. V., Motta, J.-P., Bossuet-Greif, N., Belloy, M., Taieb, F., Gratadoux, J.-J., Thomas, M., Langella, P., & Oswald, E. (2021). A Toxic Friend: Genotoxic and Mutagenic Activity of the Probiotic Strain Escherichia coli Nissle 1917. *MSphere*, *6*(4), e0062421. <https://doi.org/10.1128/mSphere.00624-21>
- Nougayrède, J.-P., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz, E., Gottschalk, G., Buchrieser, C., Hacker, J., Dobrindt, U., & Oswald, E. (2006). Escherichia coli induces DNA double-strand breaks in eukaryotic cells. *Science (New York, N.Y.)*, *313*(5788), 848-851. <https://doi.org/10.1126/science.1127059>
- Nougayrède, J.-P., Taieb, F., De Rycke, J., & Oswald, E. (2005). Cyclomodulins: bacterial effectors that modulate the eukaryotic cell cycle. *Trends in Microbiology*, *13*(3), 103-110. <https://doi.org/10.1016/j.tim.2005.01.002>
- Olier, M., Marcq, I., Salvador-Cartier, C., Secher, T., Dobrindt, U., Boury, M., Bacquié, V., Pénary, M., Gaultier, E., Nougayrède, J.-P., Fioramonti, J., & Oswald, E. (2012). Genotoxicity of Escherichia coli Nissle 1917 strain cannot be dissociated from its probiotic activity. *Gut Microbes*, *3*(6), 501-509. <https://doi.org/10.4161/gmic.21737>
- Pang, B., Chen, Y., Gan, F., Yan, C., Jin, L., Gin, J. W., Petzold, C. J., & Keasling, J. D. (2020). Investigation of Indigoidine Synthetase Reveals a Conserved Active-Site Base Residue of Nonribosomal Peptide Synthetase Oxidases. *Journal of the American Chemical Society*, *142*(25), 10931-10935. <https://doi.org/10.1021/jacs.0c04328>
- Panjikovich, A., & Svergun, D. I. (2018). CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics*, *34*(11), 1944-1946. <https://doi.org/10.1093/bioinformatics/btx846>
- Pantoliano, M. W., Petrella, E. C., Kwasnoski, J. D., Lobanov, V. S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B. A., Lane, P., & Salemme, F. R. (2001). High-Density Miniaturized Thermal Shift Assays as a General Strategy for Drug Discovery. *Journal of Biomolecular Screening*, *6*(6), 429-440. <https://doi.org/10.1177/108705710100600609>
- Pérez-Berezo, T., Pujó, J., Martín, P., Le Faouder, P., Galano, J.-M., Guy, A., Knauf, C., Tabet, J. C., Tronnet, S., Barreau, F., Heuillet, M., Dietrich, G., Bertrand-Michel, J., Durand, T., Oswald, E., & Cenac, N. (2017). Identification of an analgesic lipopeptide produced by the probiotic Escherichia coli strain Nissle 1917. *Nature Communications*, *8*(1), 1314. <https://doi.org/10.1038/s41467-017-01403-9>
- Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., Morawe, C., Mueller-Dieckmann, C., Ohlsson, S.,

- Schmid, W., Surr, J., Theveneau, P., Zerrad, L., & McSweeney, S. (2013). Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *Journal of Synchrotron Radiation*, 20(Pt 4), 660-664. <https://doi.org/10.1107/S0909049513010431>
- Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V., & Svergun, D. I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, 45(Pt 2), 342-350. <https://doi.org/10.1107/S0021889812007662>
- Petoukhov, M. V., & Svergun, D. I. (2015). Ambiguity assessment of small-angle scattering curves from monodisperse systems. *Acta Crystallographica. Section D, Biological Crystallography*, 71(Pt 5), 1051-1058. <https://doi.org/10.1107/S1399004715002576>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605-1612. <https://doi.org/10.1002/jcc.20084>
- Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., & Khosla, C. (2001). Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science (New York, N.Y.)*, 291(5509), 1790-1792. <https://doi.org/10.1126/science.1058092>
- Pleguezuelos-Manzano, C., Puschhof, J., Rosendahl Huber, A., van Hoeck, A., Wood, H. M., Nomburg, J., Gurjao, C., Manders, F., Dalmaso, G., Stege, P. B., Paganelli, F. L., Geurts, M. H., Beumer, J., Mizutani, T., Miao, Y., van der Linden, R., van der Elst, S., Genomics England Research Consortium, Garcia, K. C., ... Clevers, H. (2020). Mutational signature in colorectal cancer caused by genotoxic pks+ *E. coli*. *Nature*, 580(7802), 269-273. <https://doi.org/10.1038/s41586-020-2080-8>
- Putnam, C. D., Hammel, M., Hura, G. L., & Tainer, J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics*, 40(3), 191-285. <https://doi.org/10.1017/S0033583507004635>
- Putze, J., Hennequin, C., Nougayrède, J.-P., Zhang, W., Homburg, S., Karch, H., Bringer, M.-A., Fayolle, C., Carniel, E., Rabsch, W., Oelschlaeger, T. A., Oswald, E., Forestier, C., Hacker, J., & Dobrindt, U. (2009). Genetic Structure and Distribution of the Colibactin Genomic Island among Members of the Family Enterobacteriaceae. *Infection and Immunity*, 77(11), 4696-4703. <https://doi.org/10.1128/IAI.00522-09>
- Reimer, J. M., Aloise, M. N., Harrison, P. M., & Schmeing, T. M. (2016). Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature*, 529(7585), 239-242. <https://doi.org/10.1038/nature16503>
- Reimer, J. M., Aloise, M. N., Powell, H. R., & Schmeing, T. M. (2016). Manipulation of an existing crystal form unexpectedly results in interwoven packing networks with pseudo-translational symmetry. *Acta Crystallographica Section D: Structural Biology*, 72(10), 1130-1136. <https://doi.org/10.1107/S2059798316013504>
- Reimer, J. M., Eivaskhani, M., Harb, I., Guarné, A., Weigt, M., & Schmeing, T. M. (2019). Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science (New York, N.Y.)*, 366(6466), eaaw4388. <https://doi.org/10.1126/science.aaw4388>
- Richter, C. D., Nietlispach, D., Broadhurst, R. W., & Weissman, K. J. (2008). Multienzyme docking in hybrid megasynthetases. *Nature Chemical Biology*, 4(1), 75-81. <https://doi.org/10.1038/nchembio.2007.61>

- Rittner, A., & Grininger, M. (2014). Modular polyketide synthases (PKSs): a new model fits all? *ChemBiochem: A European Journal of Chemical Biology*, *15*(17), 2489-2493. <https://doi.org/10.1002/cbic.201402432>
- Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, *42*(W1), W320-W324. <https://doi.org/10.1093/nar/gku316>
- Roy, R. S., Gehring, A. M., Milne, J. C., Belshaw, P. J., & Walsh, C. T. (1999). Thiazole and oxazole peptides: biosynthesis and molecular machinery. *Natural Product Reports*, *16*(2), 249-263. <https://doi.org/10.1039/a806930a>
- Russo, T. A., & Johnson, J. R. (2000). Proposal for a New Inclusive Designation for Extraintestinal Pathogenic Isolates of Escherichia coli: ExPEC. *The Journal of Infectious Diseases*, *181*(5), 1753-1754. <https://doi.org/10.1086/315418>
- Saibil, H. R. (2000). Macromolecular structure determination by cryo-electron microscopy. *Acta Crystallographica Section D: Biological Crystallography*, *56*(10), 1215-1222. <https://doi.org/10.1107/S09074444900010787>
- Savage, D. C. (1977). Microbial ecology of the gastrointestinal tract. *Annual Review of Microbiology*, *31*, 107-133. <https://doi.org/10.1146/annurev.mi.31.100177.000543>
- Schneider, T. L., Shen, B., & Walsh, C. T. (2003). Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry*, *42*(32), 9722-9730. <https://doi.org/10.1021/bi034792w>
- Schultz, M. (2008). Clinical use of E. coli Nissle 1917 in inflammatory bowel disease. *Inflammatory Bowel Diseases*, *14*(7), 1012-1018. <https://doi.org/10.1002/ibd.20377>
- Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology*, *14*(8), e1002533. <https://doi.org/10.1371/journal.pbio.1002533>
- Shou, Q., Feng, L., Long, Y., Han, J., Nunnery, J. K., Powell, D. H., & Butcher, R. A. (2016). A hybrid polyketide-nonribosomal peptide in nematodes that promotes larval survival. *Nature Chemical Biology*, *12*(10), 770-772. <https://doi.org/10.1038/nchembio.2144>
- Shrestha, E., Coulter, J. B., Guzman, W., Ozbek, B., Hess, M. M., Mummert, L., Ernst, S. E., Maynard, J. P., Meeker, A. K., Heaphy, C. M., Haffner, M. C., De Marzo, A. M., & Sfanos, K. S. (2021). Oncogenic gene fusions in nonneoplastic precursors as evidence that bacterial infection can initiate prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(32), e2018976118. <https://doi.org/10.1073/pnas.2018976118>
- Sieber, S. A., Linne, U., Hillson, N. J., Roche, E., Walsh, C. T., & Marahiel, M. A. (2002). Evidence for a monomeric structure of nonribosomal Peptide synthetases. *Chemistry & Biology*, *9*(9), 997-1008. [https://doi.org/10.1016/s1074-5521\(02\)00214-4](https://doi.org/10.1016/s1074-5521(02)00214-4)
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*, 539. <https://doi.org/10.1038/msb.2011.75>
- Silpe, J. E., Wong, J. W. H., Owen, S. V., Baym, M., & Balskus, E. P. (2022). The bacterial toxin colibactin triggers prophage induction. *Nature*, *603*(7900), 315-320. <https://doi.org/10.1038/s41586-022-04444-3>
- Skiba, M. A., Maloney, F. P., Dan, Q., Fraley, A. E., Aldrich, C. C., Smith, J. L., & Brown, W. C. (2018). PKS-NRPS Enzymology and Structural Biology: Considerations in Protein



- Production. *Methods in Enzymology*, 604, 45-88. <https://doi.org/10.1016/bs.mie.2018.01.035>
- Smith, H. G., Beech, M. J., Lewandowski, J. R., Challis, G. L., & Jenner, M. (2021). Docking domain-mediated subunit interactions in natural product megasynth(et)ases. *Journal of Industrial Microbiology & Biotechnology*, 48(3-4), kuab018. <https://doi.org/10.1093/jimb/kuab018>
- Stachelhaus, T., Mootz, H. D., & Marahiel, M. A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology*, 6(8), 493-505. [https://doi.org/10.1016/S1074-5521\(99\)80082-9](https://doi.org/10.1016/S1074-5521(99)80082-9)
- Stark, H. (2010). GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Methods in Enzymology*, 481, 109-126. [https://doi.org/10.1016/S0076-6879\(10\)81005-5](https://doi.org/10.1016/S0076-6879(10)81005-5)
- Stidham, R. W., & Higgins, P. D. R. (2018). Colorectal Cancer in Inflammatory Bowel Disease. *Clinics in Colon and Rectal Surgery*, 31(3), 168-178. <https://doi.org/10.1055/s-0037-1602237>
- Straight, P. D., Fischbach, M. A., Walsh, C. T., Rudner, D. Z., & Kolter, R. (2007). A singular enzymatic megacomplex from *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America*, 104(1), 305-310. <https://doi.org/10.1073/pnas.0609073103>
- Strakova, N., Korena, K., & Karpiskova, R. (2021). Klebsiella pneumoniae producing bacterial toxin colibactin as a risk of colorectal cancer development - A systematic review. *Toxicon: Official Journal of the International Society on Toxinology*, 197, 126-135. <https://doi.org/10.1016/j.toxicon.2021.04.007>
- Strieker, M., Tanović, A., & Marahiel, M. A. (2010). Nonribosomal peptide synthetases: structures and dynamics. *Current Opinion in Structural Biology*, 20(2), 234-240. <https://doi.org/10.1016/j.sbi.2010.01.009>
- Su, L., Hôtel, L., Paris, C., Chepkirui, C., Brachmann, A. O., Piel, J., Jacob, C., Aigle, B., & Weissman, K. J. (2022). Engineering the stambomycin modular polyketide synthase yields 37-membered mini-stambomycins. *Nature Communications*, 13, 515. <https://doi.org/10.1038/s41467-022-27955-z>
- Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography*, 25(4), 495-503. <https://doi.org/10.1107/S0021889892001663>
- Svergun, D. I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical Journal*, 76(6), 2879-2886. [https://doi.org/10.1016/S0006-3495\(99\)77443-6](https://doi.org/10.1016/S0006-3495(99)77443-6)
- Tang, J.-W., Liu, X., Ye, W., Li, Z.-R., & Qian, P.-Y. (2022). Biosynthesis and bioactivities of microbial genotoxin colibactins. *Natural Product Reports*. <https://doi.org/10.1039/d1np00050k>
- Tang, Y., Chen, A. Y., Kim, C.-Y., Cane, D. E., & Khosla, C. (2007). Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chemistry & Biology*, 14(8), 931-943. <https://doi.org/10.1016/j.chembiol.2007.07.012>
- Tang, Y., Kim, C.-Y., Mathews, I. I., Cane, D. E., & Khosla, C. (2006). The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proceedings of the National Academy of Sciences of the United States of America*, 103(30), 11124-11129. <https://doi.org/10.1073/pnas.0601924103>

- Tanovic, A., Samel, S. A., Essen, L.-O., & Marahiel, M. A. (2008). Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science (New York, N.Y.)*, *321*(5889), 659-663. <https://doi.org/10.1126/science.1159850>
- Tarry, M. J., Haque, A. S., Bui, K. H., & Schmeing, T. M. (2017). X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure (London, England: 1993)*, *25*(5), 783-793.e4. <https://doi.org/10.1016/j.str.2017.03.014>
- Tilg, H., Adolph, T. E., Gerner, R. R., & Moschen, A. R. (2018). The Intestinal Microbiota in Colorectal Cancer. *Cancer Cell*, *33*(6), 954-964. <https://doi.org/10.1016/j.ccell.2018.03.004>
- Tomlins, S. A., Rhodes, D. R., Perner, S., Dhanasekaran, S. M., Mehra, R., Sun, X.-W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., Lee, C., Montie, J. E., Shah, R. B., Pienta, K. J., Rubin, M. A., & Chinnaiyan, A. M. (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science (New York, N.Y.)*, *310*(5748), 644-648. <https://doi.org/10.1126/science.1117679>
- Trautman, E. P., Healy, A. R., Shine, E. E., Herzon, S. B., & Crawford, J. M. (2017). Domain-Targeted Metabolomics Delineates the Heterocycle Assembly Steps of Colibactin Biosynthesis. *Journal of the American Chemical Society*, *139*(11), 4195-4201. <https://doi.org/10.1021/jacs.7b00659>
- Tripathi, P., Shine, E. E., Healy, A. R., Kim, C. S., Herzon, S. B., Bruner, S. D., & Crawford, J. M. (2017). ClbS Is a Cyclopropane Hydrolase That Confers Colibactin Resistance. *Journal of the American Chemical Society*, *139*(49), 17719-17722. <https://doi.org/10.1021/jacs.7b09971>
- Tronnet, S., Floch, P., Lucarelli, L., Gaillard, D., Martin, P., Serino, M., & Oswald, E. (2020). The Genotoxin Colibactin Shapes Gut Microbiota in Mice. *MSphere*, *5*(4), e00589-20. <https://doi.org/10.1128/mSphere.00589-20>
- Tronnet, S., Garcie, C., Brachmann, A. O., Piel, J., Oswald, E., & Martin, P. (2017). High iron supply inhibits the synthesis of the genotoxin colibactin by pathogenic *Escherichia coli* through a non-canonical Fur/RyhB-mediated pathway. *Pathogens and Disease*, *75*(5). <https://doi.org/10.1093/femspd/ftx066>
- Tsunematsu, Y., Hosomi, K., Kunisawa, J., Sato, M., Shibuya, N., Saito, E., Murakami, H., Yoshikawa, Y., Iwashita, Y., Miyoshi, N., Mutoh, M., Ishikawa, H., Sugimura, H., Miyachi, M., Wakabayashi, K., & Watanabe, K. (2021). Mother-to-infant transmission of the carcinogenic colibactin-producing bacteria. *BMC Microbiology*, *21*(1), 235. <https://doi.org/10.1186/s12866-021-02292-1>
- Tufar, P., Rahighi, S., Kraas, F. I., Kirchner, D. K., Löhr, F., Henrich, E., Köpke, J., Dikic, I., Güntert, P., Marahiel, M. A., & Dötsch, V. (2014). Crystal structure of a PCP/Sfp complex reveals the structural basis for carrier protein posttranslational modification. *Chemistry & Biology*, *21*(4), 552-562. <https://doi.org/10.1016/j.chembiol.2014.02.014>
- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature*, *449*(7164), 804-810. <https://doi.org/10.1038/nature06244>
- Uchański, T., Pardon, E., & Steyaert, J. (2020). Nanobodies to study protein conformational states. *Current Opinion in Structural Biology*, *60*, 117-123. <https://doi.org/10.1016/j.sbi.2020.01.003>

- Valentine, R. C., Shapiro, B. M., & Stadtman, E. R. (1968). Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from *Escherichia coli*. *Biochemistry*, 7(6), 2143-2152. <https://doi.org/10.1021/bi00846a017>
- Vander Wood, D. A., & Keatinge-Clay, A. T. (2018). The modules of trans-acyltransferase assembly lines redefined with a central acyl carrier protein. *Proteins*, 86(6), 664-675. <https://doi.org/10.1002/prot.25493>
- Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., & Kinzler, K. W. (2013). Cancer Genome Landscapes. *Science (New York, N.Y.)*, 339(6127), 1546-1558. <https://doi.org/10.1126/science.1235122>
- Volkov, V. V., Svergun, D. I., & IUCr. (2003, avril 16). *Uniqueness of ab initio shape determination in small-angle scattering* [Text]. *Journal of Applied Crystallography*. <https://doi.org/10.1107/S0021889803000268>
- Wallenstein, A., Rehm, N., Brinkmann, M., Selle, M., Bossuet-Greif, N., Sauer, D., Bunk, B., Spröer, C., Wami, H. T., Homburg, S., von Büнау, R., König, S., Nougayrède, J.-P., Overmann, J., Oswald, E., Müller, R., & Dobrindt, U. (2020). ClbR Is the Key Transcriptional Activator of Colibactin Gene Expression in *Escherichia coli*. *MSphere*, 5(4), e00591-20. <https://doi.org/10.1128/mSphere.00591-20>
- Wami, H., Wallenstein, A., Sauer, D., Stoll, M., von Büнау, R., Oswald, E., Müller, R., & Dobrindt, U. (2021). Insights into evolution and coexistence of the colibactin- and yersiniabactin secondary metabolite determinants in enterobacterial populations. *Microbial Genomics*, 7(6), 000577. <https://doi.org/10.1099/mgen.0.000577>
- Wang, J., Li, D., Chen, L., Cao, W., Kong, L., Zhang, W., Croll, T., Deng, Z., Liang, J., & Wang, Z. (2022). Catalytic trajectory of a dimeric nonribosomal peptide synthetase subunit with an inserted epimerase domain. *Nature Communications*, 13(1), 592. <https://doi.org/10.1038/s41467-022-28284-x>
- Wang, J., Liang, J., Chen, L., Zhang, W., Kong, L., Peng, C., Su, C., Tang, Y., Deng, Z., & Wang, Z. (2021). Structural basis for the biosynthesis of lovastatin. *Nature Communications*, 12(1), 867. <https://doi.org/10.1038/s41467-021-21174-8>
- Weissman, K. J. (2015a). Uncovering the structures of modular polyketide synthases. *Natural Product Reports*, 32(3), 436-453. <https://doi.org/10.1039/c4np00098f>
- Weissman, K. J. (2015b). The structural biology of biosynthetic megaenzymes. *Nature Chemical Biology*, 11(9), 660-670. <https://doi.org/10.1038/nchembio.1883>
- Weissman, K. J., Timoney, M., Bycroft, M., Grice, P., Hanefeld, U., Staunton, J., & Leadlay, P. F. (1997). The molecular basis of Celmer's rules: the stereochemistry of the condensation step in chain extension on the erythromycin polyketide synthase. *Biochemistry*, 36(45), 13849-13855. <https://doi.org/10.1021/bi971566b>
- Wernke, K. M., Xue, M., Tirla, A., Kim, C. S., Crawford, J. M., & Herzon, S. B. (2020). Structure and bioactivity of colibactin. *Bioorganic & Medicinal Chemistry Letters*, 30(15), 127280. <https://doi.org/10.1016/j.bmcl.2020.127280>
- Whicher, J. R., Dutta, S., Hansen, D. A., Hale, W. A., Chemler, J. A., Dosey, A. M., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L., & Skiniotis, G. (2014). Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature*, 510(7506), 560-564. <https://doi.org/10.1038/nature13409>
- Williams, P. C., Wernke, K. M., Tirla, A., & Herzon, S. B. (2020). Employing chemical synthesis to study the structure and function of colibactin, a « dark matter » metabolite. *Natural Product Reports*, 37(11), 1532-1548. <https://doi.org/10.1039/d0np00072h>

- Wilson, M. R., Jiang, Y., Villalta, P. W., Stornetta, A., Boudreau, P. D., Carrá, A., Brennan, C. A., Chun, E., Ngo, L., Samson, L. D., Engelward, B. P., Garrett, W. S., Balbo, S., & Balskus, E. P. (2019). The human gut bacterial genotoxin colibactin alkylates DNA. *Science (New York, N.Y.)*, *363*(6428), eaar7785. <https://doi.org/10.1126/science.aar7785>
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica. Section D, Biological Crystallography*, *67*(Pt 4), 235-242. <https://doi.org/10.1107/S0907444910045749>
- Wong, S. G., & Dessen, A. (2014). Structure of a bacterial  $\alpha$ 2-macroglobulin reveals mimicry of eukaryotic innate immunity. *Nature Communications*, *5*, 4917. <https://doi.org/10.1038/ncomms5917>
- Xue, M., Kim, C. S., Healy, A. R., Wernke, K. M., Wang, Z., Frischling, M. C., Shine, E. E., Wang, W., Herzon, S. B., & Crawford, J. M. (2019). Structure elucidation of colibactin and its DNA cross-links. *Science (New York, N.Y.)*, *365*(6457), eaax2685. <https://doi.org/10.1126/science.aax2685>
- Xue, M., Shine, E., Wang, W., Crawford, J. M., & Herzon, S. B. (2018). Characterization of Natural Colibactin-Nucleobase Adducts by Tandem Mass Spectrometry and Isotopic Labeling. Support for DNA Alkylation by Cyclopropane Ring Opening. *Biochemistry*, *57*(45), 6391-6394. <https://doi.org/10.1021/acs.biochem.8b01023>
- Young, G., Hundt, N., Cole, D., Fineberg, A., Andrecka, J., Tyler, A., Olerinyova, A., Ansari, A., Marklund, E. G., Collier, M. P., Chandler, S. A., Tkachenko, O., Allen, J., Crispin, M., Billington, N., Takagi, Y., Sellers, J. R., Eichmann, C., Selenko, P., ... Kukura, P. (2018). Quantitative mass imaging of single biological macromolecules. *Science (New York, N.Y.)*, *360*(6387), 423-427. <https://doi.org/10.1126/science.aar5839>
- Yun, C.-S., Nishimoto, K., Motoyama, T., Shimizu, T., Hino, T., Dohmae, N., Nagano, S., & Osada, H. (2020). Unique features of the ketosynthase domain in a nonribosomal peptide synthetase-polyketide synthase hybrid enzyme, tenuazonic acid synthetase 1. *The Journal of Biological Chemistry*, *295*(33), 11602-11612. <https://doi.org/10.1074/jbc.RA120.013105>
- Zha, L., Jiang, Y., Henke, M. T., Wilson, M. R., Wang, J. X., Kelleher, N. L., & Balskus, E. P. (2017). Colibactin assembly line enzymes use S-adenosylmethionine to build a cyclopropane ring. *Nature Chemical Biology*, *13*(10), 1063-1065. <https://doi.org/10.1038/nchembio.2448>
- Zhong, E. D., Bepler, T., Berger, B., & Davis, J. H. (2021). CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods*, *18*(2), 176-185. <https://doi.org/10.1038/s41592-020-01049-4>



# Annexe 1



Séquences de ClbK et de ses fragments

Les résidus soulignés sont des résidus issus du vecteur d'expression.

**ClbK (codée par le vecteur pET-15b)**

Dans le mutant ClbK-SER, les trois résidus en gras ont été mutés en alanine.

MGSSHHHHHSSGLVPRGSHMTYSESDIAIVGMNCRYPGVHSVAAFETVLRTGCNILDPKVTPSNGHN  
 HITLNNVYEHMAEFDANFFGYSRAEAEIMDPQQRVFLTCAWEMFEQSGYNPKQHDARVGLYAGVSTSF  
 YLLTHLMNPNPKLAQLGGLQIMVGNDDKHLTSQLAYRLNITGPCVTVQASCATSLVAVHLACEGLLSGQ  
 CDMALAGGVTFRMEEQRSYESHGDGLQAEDGLIHTFDAQASGTVYSSGLGMVLLKRATDAQVQGDNIL  
 AVIKGSAINNDGGARSGYTPGVGDGQEAVMIEAHSLAETVPPQIQYLELHSGTPLGDAIEFAAIKRVFG  
 TPAPNATPWRLGAVKPNVGHVEMASGITSLIKTVLSLTNRVFPYPTLNFQRANPQLGLEDSPFEVVSRLTP  
 WPEGTTPTAGVSAFGLGGTNAHLVVQAPLSTPQARAQQMGPCVVVLSAKNHNALEQMQRNALLAKL  
 AAHPEIRLQDVAYTLRHGRFSAPVRKCVIAENCTQLARQLRDAPMVEATTGCTIYWRGLHGRFVVALETLS  
 DWLACSEVLSQAVGQLEHFLEPAQLDLSPAQRTEFISQYALIALIDERETLNVVLCGDGDDGGYAAAVLR  
 GDCTLEQAWHRLNAGQPFDVPTNPLLQPDVCSLMLDDAASDANRTALEALGQLWLAGVSLDWRWV  
 DAAERMLGSQRIALPGTVFTPQRYWVEAVRPATFSHESSNNLLSRATKSDIIAVVTEI WERTLGVSIDDHH  
 ASFFELGGHSLLASTILYDIQQRYGITCTLSAFFADPTIEGLSCYLLEQGGSETAVSALPDTVFAPDQQHLPPF  
 LTDVQQAYWVGRRKSLGLGNISTHIYVEYELQGLDETA FNRALNAVIARHSMLRAIVNDDGMQQILPNV  
 PEYHVAFYTTQCEDAFQQRCRELRDRTLSHQMIDCSRWPLFQMEVVVDPQQKARLHVSIDLLIADAWSLE  
 LFIRELAYHYRHPQAALPTLTYSFRDYVLTLSYKTPQFERARDYWRARIETLPPGPRLPLRTDPTKLENPT  
 FVRRSYCLSRAIWQRLKQAGQMSITPTLLLTGFAQVLARFSSSPHFSNLTLFNRLPLHADINHLIGDFTA  
 LTLLEIDMSQGETLQARANVIHSQWLWRDLNRLFGGIQVSRLLVQTHRDPKSVIPIVFTSLLNQYEASWE  
 TDDTLFNQPQDDLYSISQTPQVWLDHQVMERNGELHFNWDVVEQLFEPALMDQMFQCYCQLLHALA  
 QRPQLWHETQDVLALPTVSAPVTQAPAPTALLHHGLLRQAALTPQETALISPIRELYRQLSTAADHVAR  
 ALLALGVQHGDRAVVMMEKGWQQIAAVHGIIRLGA VYLPVDPVLPQRRQLLLTVGEVVRVQVTQPGLT  
 QLEPSLPVLIIDDGMLDTPAAPLPEVAGDVTDLAYIIFTSGSTGTPKGMIDHRAAMNTLEDINERFGLNA  
 QDRVFLSSLSFDLSVYDAFAPFMVGAALVLEAGREKDRHWQTVMAHGHVSVWNAV PALMQMLC  
 EYHSGDRMSYPTLRLALLSGDWIPLTLPEQMRERLNETMDIISLGGATECAIWSVYYPIGEVESTWTSIPY  
 GRGLRNQPVVYVNAQLEECVPGVEGEICIGGMGLAQQGYLND AEKTAASFVWREASGERIYRTGDRGRYF  
 ADGQVAFLGRNDTQVKVNGYRIELGEIERCIARHPDVEQSVVAVGNSQHRRLVAFAKLHDRHQAQAL  
 QAKEAEAAALAQGIIVNPAQRLAFKLEPHIRALDGLGIALTAPADSTRYIKRRSYRHFS AQKTTLAQLGQL  
 LSGLGQMRLPGLPFAKYAYASAGGLYPVQTYVYLHPDKIEEGVSGIYFDPRQSCLMPVAPEVELNSGFH  
 AGPNQSIADRAAFTLFMVADMAVISPFYQEAAWHFSVMEAGTLCHLLEEDAPRYGLGLCQLGMADFS  
 AVASHFQLSPHHRYVHCTVGGAIQEAASAAALLRDFSTYEKPKETAAPLDMQSYKDAMLRLRQLPDP  
 YMVPSDLMLATDFPLTANGKLDKRLQLQGEQIAHQRDGVGPIQVDSALQQRLVALWQEV LGVSHVS  
 AEDDFSLGGSSIELVRIQQALEAIIQEIPIVDLFRLP TIADVARYLDEQLHNLPAAHDIVLAQAEVSQVSA  
 ARENLALRRKRAQQGEGKDE



**ClbK-PKS (codée par le vecteur pETM-11)**

MKHHHHHPMSDYDIPTTENLYFQGAMGTYSESDIAIVGMNCRYPGVHSVAAFETVLRGTCNILDPKVT  
 PSNGHNNHITLNNVYEHMAEFDANFFGYSRAEAEIMDPQQRVFLTCAWEMFEQSGYNPKQHDARVGLY  
 AGVSTSFYLLTHLMNNDKLAQLGGLQIMVGNKDHLTSQLAYRLNITGPCVTVQASCATSLVAVHLAC  
 EGLLSGQCDMALAGGVTFRMEEQRSYESHGDGLQAEDGLIHTFDAQASGTVYSSGLGMVLLKRATDAQ  
 VQGDNILAVIKGSAINNDGGARSGYTPGVGDGQEAVMIEAHS LAEVTPQQIQYLELHGSGTPLGDAIEFA  
 AIKRVFGTPAPNATPWRLGAVKPNVGHVEMASGITSLIKTVLSLTNRVYPTLNFQRANPQLGLEDSPFV  
 VSRLTPWPEGTTPTAGVSAFGLGGTNAHLVVQAPLSTPQARAQMGPCVVVLSAKNHNALEQM  
 QNALLAKLAAHPEIRLQDVAYTLRHGRFSAPVRKCVIAENCTQLARQLRDAPMVEATTGCTIYWR  
 LGHRFVVALETSDWLACSEVLSQAVGQLEHFPLEPAQLDLSPAQRTEFISQYALIALIDERETLNVV  
 LCGDGDGGYAAVLRGDCTLEQAWHRLNAGQPFDVPTNPLLQPDVCSMLDDAASDANRTALEALG  
 QLWLAGVSLDWRWVDAERMLGSQRIALPGTVFTPQRYWVEAVRPATFSHESNNLLSRATKSDI  
 IAVVTEIWERTLGVSIDDHASFELGGHSLLASTILYDIQRYGITCTLSAFFADPTIEGLSCYLLEQ

**ClbK-PKS\* (codée par le vecteur pETM-11, après clivage de l'étiquette poly-histidines)**

GAMGTYSESDIAIVGMNCRYPGVHSVAAFETVLRGTCNILDPKVTPSNNGHNNHITLNNVYEHMAEFDANF  
 FGYSRAEAEIMDPQQRVFLTCAWEMFEQSGYNPKQHDARVGLYAGVSTSFYLLTHLMNNDKLAQLG  
 GLQIMVGNKDHLTSQLAYRLNITGPCVTVQASCATSLVAVHLACEGLLSGQCDMALAGGVTFRMEEQ  
 RSYESHGDGLQAEDGLIHTFDAQASGTVYSSGLGMVLLKRATDAQVQGDNILAVIKGSAINNDGGARSG  
 YTPGVGDGQEAVMIEAHS LAEVTPQQIQYLELHGSGTPLGDAIEFAAIKRVFGTPAPNATPWRLGAVKPN  
 VGHVEMASGITSLIKTVLSLTNRVYPTLNFQRANPQLGLEDSPFVVSRLTPWPEGTTPTAGVSAFGLG  
 GTNAHLVVQAPLSTPQARAQMGPCVVVLSAKNHNALEQM  
 QNALLAKLAAHPEIRLQDVAYTLRHGRFSAPVRKCVIAENCTQLARQLRDAPMVEATTGCTIYWR  
 LGHRFVVALETSDWLACSEVLSQAVGQLEHFPLEPAQLDLSPAQRTEFISQYALIALIDERETLNVV  
 LCGDGDGGYAAVLRGDCTLEQAWHRLNAGQPFDVPTNPLLQPDVCSMLDDAASDANRTALEALG  
 QLWLAGVSLDWRWVDAERMLGSQRIALPGTVFTPQRYWVEAVRPATFSHESNNLLSRATKSDI  
 IAVVTEIWERTLGVSIDDHASFELGGHSLLASTILYDIQRYGITCTLSAFFADPTIEGLSCYLLEQ

**CibK-PKS-A<sub>core</sub> (codée par le vecteur pETM-11)**

MKHHHHHHPMSDYDIPTTENLYFQGAMGTYSESDIAIVGMNCRYPGVHSVAAFETVLRGTCNILDPKVT  
 PSNGHNHITLNNVYEHMAEFDANFFGYSRAEAEIMDPQQRVFLTCAWEMFEQSGYNPKQHDARVGLY  
 AGVSTSFYLLTHLMNNPDKLAQLGGLQIMVGNKDHLTSQLAYRLNITGPCVTVQASCATSLVAVHLAC  
 EGLLSGQCDMALAGGVTFRMEEQRSYESHGDGLQAEDGLIHTFDAQASGTVYSSGLGMVLLKRATDAQ  
 VQGDNILAVIKGSAINNDGGARSGYTVPGVDGQEAVMIEAHS LAE VTPQQIQYLELHGSGTPLGDAIEFA  
 AIKRVFGTPAPNATPWRLGAVKPNVGHVEMASGITSLIKTVLSLTNRVFPYPTLNFQRANPQLGLEDSPFEV  
 VSRLTPWPEGTTTPRTAGVSAFGLGGTNAHLVVQAPLSTPQARAQQMGPCVVVLSAKNHNALEQMQRN  
 ALLAKLAAHPEIRLQDVAYTLRHGRFSAPVRKCVIAENCTQLARQLRDAPMVEATTGCTIYWRLGHRFVV  
 ALETLSDWLACSEVLSQAVGQLEHFPLEPAQLDLSPAQRTFISQYALIALIDERETLNVVLCGDGDGGYA  
 AAVLRGDCTLEQAWHRLNAGQPFDDVPTNPLLQPDVCSLMLDDAASDANRTALEALGQLWLAGVSLD  
 WRWVDAERMLGSQRIALPGTVFTPQRYWVEAVRPATFSHESNNLLSRATKSDIIVVTEIWERTLGVS  
 IDDHASFFELGGHSLLASTILYDIQQRYGITCTLSAFFADPTIEGLSCYLLEQGGSETAVSALPDTVFAPDQ  
 QHLFPFLTDVQQAYWVGRRKSGLGNISTHIYVEYELQGLDETA FN RAL NAVI ARH SMLRAIVNDDGMQ  
 QILPNVPEYHVAFYTTQCEDAFQQRCRELRDLSHQMIDCSRWPLFQMEVVVDPQQKARLHVSIDLLIA  
 DAWSLELFIRELAYHYRHPQAALPTLTYSFRDYVLTLSYKTPQFERARDYWRARIETLPPGPRPLRTDP  
 TKLENPTFVRRSYCLSRAIWQRLKTQAGQMSITPTLLLTGFAQVLARFSSSPHFSNLTLFNRLPLHADIN  
 HLIGDFTALTLEIDMSQGETLQARANVIHSQLWRDLNRLFGGIQVSRLLVQTHRDPKSVIPIVFTSLLN  
 QYEASWETDDTLFNQPQDDLYSISQTPQVWLDHQVMERNGELHFNWDVVEQLFEPALMDQMFQCY  
 CQLLHALAQRPQLWHEIQDVLALPTVSAPVTQAPAPTALLHHGLLRQAALTPQETALISPIRELTQRQST  
 AADHVARALLALGVQHGDRAVVMKEGWQQIAAVHGILRLGAVYLPVDPVLPQRRLQLLLTVGEVRV  
 QVTQPGLTQLEPSLPVLIIDDGMLDTPAAPLPEVAGDVTDLAYIIFTSGSTGTPKGVMMIDHRAAMNTLEDI  
 NERFGLNAQDRVFGLSSLSFDLSVYDAFAPFMVGAALVLPEAGREKDPRHWTVMAGHVSVWNAV  
 ALMQMLCEYHSGDRMSYPTLRLALLSGDWIPLTLPEQMRERLNETMDIISLGGATECAIWSVYPIGEVE  
 STWTSIPYGRGLRNQPVYVLNAQLEECVGVGEVEICIGGMGLAQQGYLNDAEKTAASFVWREASGERIYRT  
 GDRGRYFADGGQVAFLGRND

La protéine CibK-PKS-LK se termine à la leucine surlignée en rouge et la protéine CibK-PKS-Cy se termine à la thréonine soulignée en bleu



# **Annexe 2 : publication 1**



## Review



**Cite this article:** Bonhomme S, Dessen A, Macheboeuf P. 2021 The inherent flexibility of type I non-ribosomal peptide synthetase multienzymes drives their catalytic activities. *Open Biol.* **11**: 200386. <https://doi.org/10.1098/rsob.200386>

Received: 26 November 2020

Accepted: 5 May 2021

### Subject Area:

biochemistry/structural biology

### Keywords:

non-ribosomal peptide synthetases, flexibility, supramodular architecture

### Author for correspondence:

Pauline Macheboeuf

e-mail: [pauline.macheboeuf@ibs.fr](mailto:pauline.macheboeuf@ibs.fr)

# The inherent flexibility of type I non-ribosomal peptide synthetase multienzymes drives their catalytic activities

Sarah Bonhomme<sup>1</sup>, Andréa Dessen<sup>1,2</sup> and Pauline Macheboeuf<sup>1</sup>

<sup>1</sup>Univ. Grenoble Alpes, CNRS, CEA, IBS, F-38000 Grenoble, France

<sup>2</sup>Brazilian Biosciences National Laboratory (LNBio), CNPEM, Campinas 13084-971, São Paulo, Brazil

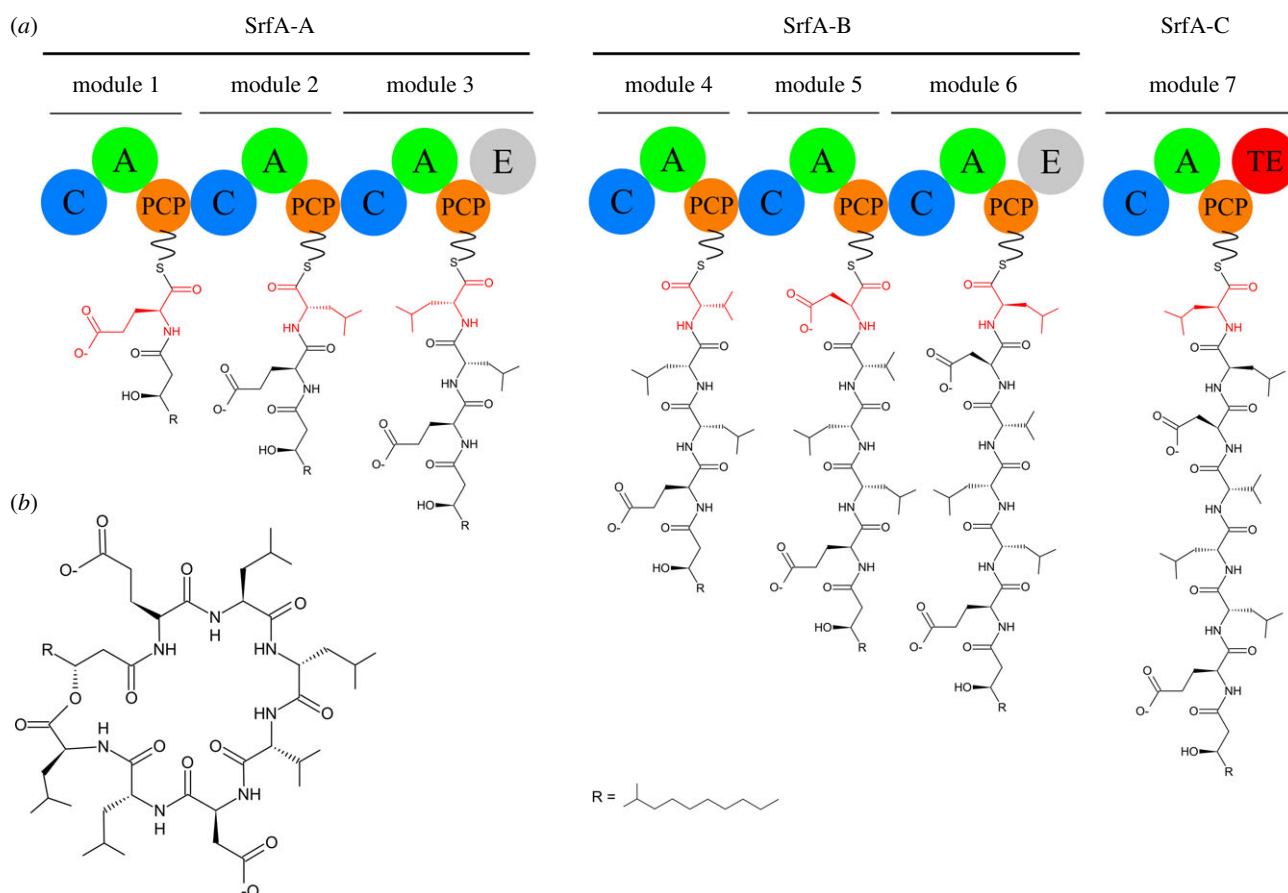
AD, 0000-0001-6487-4020; PM, 0000-0003-4523-9923

Non-ribosomal peptide synthetases (NRPSs) are multienzymes that produce complex natural metabolites with many applications in medicine and agriculture. They are composed of numerous catalytic domains that elongate and chemically modify amino acid substrates or derivatives and of non-catalytic carrier protein domains that can tether and shuttle the growing products to the different catalytic domains. The intrinsic flexibility of NRPSs permits conformational rearrangements that are required to allow interactions between catalytic and carrier protein domains. Their large size coupled to this flexibility renders these multi-domain proteins very challenging for structural characterization. Here, we summarize recent studies that offer structural views of multi-domain NRPSs in various catalytically relevant conformations, thus providing an increased comprehension of their catalytic cycle. A better structural understanding of these multienzymes provides novel perspectives for their re-engineering to synthesize new bioactive metabolites.

## 1. Introduction: the inherent flexibility of non-ribosomal peptide synthetases

Natural products are secondary metabolites synthesized by microorganisms in order to adapt to their environment [1]. Many of these natural products have been used for medical purposes, such as the antibiotics daptomycin and vancomycin [2] and the anti-cancer molecule bleomycin [3]. Among these natural products, non-ribosomal peptides (NRPs) make up a vast class of peptide-based metabolites, synthesized independently from the ribosome by large machineries named non-ribosomal peptide synthetases (NRPSs). For instance, the surfactin lipopeptide is synthesized by the 1 MDa surfactin NRPS from *Bacillus subtilis* (figure 1) [4]. Recent progress in whole genome sequencing has revealed the existence of numerous NRPS gene clusters among bacteria and fungi, mostly of unknown function [5–7]. Nevertheless, the structural understanding of the machineries that produce these metabolites has long remained limited to studies of isolated domains [8] until relatively recently, and has evolved dramatically in the last few years.

NRPSs are classified into two categories, type I and type II. In type II NRPSs, the incorporation of an amino acid into the metabolite necessitates the involvement of several domains carried by distinct proteins [9]. By contrast, the type I NRPS megaenzymes use an assembly line strategy (figure 1) with modules that act sequentially, each being responsible for the incorporation of an amino acid into the final metabolite. A NRPS assembly line is composed of an initiation module (module 1 in figure 1*a*), a termination module



**Figure 1.** Organization of the surfactin NRPS. (a) The surfactin NRPS is composed of three polypeptides: SrfA-A, SrfA-B and SrfA-C. Each polypeptide is composed of one or several modules, each one being responsible for the incorporation of an amino acid, highlighted in red in the growing metabolite. Module 1 is the initiation module, module 7 is the termination module and those in between are elongation modules. The addition of an amino acid into the metabolite requires the cooperation between domains, represented as coloured spheres. All surfactin NRPS modules possess a unique non-catalytic domain, the PCP (in orange) that tethers the growing metabolite. Each module also contains at least two catalytic domains: a condensation domain (C, in blue) and an adenylation domain (A, in green). Additionally, modules 3 and 6 possess an optional epimerization domain (E, in grey). Finally, the termination module ends with a thioesterase (TE, in red) domain that releases and cyclizes the surfactin molecule. (b) Chemical structure of the surfactin.

(module 7 in figure 1a) and one or several elongation modules (2 to 6 in figure 1a). In general, several modules are fused into a single polypeptide. Polyketide synthases (PKSs) that are also modular megaenzymes, adopt the same assembly line logic as NRPSs but they use small carbon chain substrates instead of amino acid substrates [10]. This similar strategy explains the existence of numerous hybrid NRPS/PKS assembly lines that produce hybrid peptide-polyketide metabolites [11].

Most modules are composed of two catalytic domains, the adenylation (A) and the condensation (C) domains (figure 1a); each NRPS module also incorporates the non-catalytic, but essential, peptidyl carrier protein (PCP) domain [12]. In this review, we will consider that a classical module starts with a C domain and ends with a PCP domain (figure 1). The PCP domain functions as an anchoring platform for shuttling substrates to the different catalytic domains within a module; it also allows the transport of the modified substrate from the upstream to the downstream module (figure 1a). PCP domains, whose masses are in the range of 10 kDa, have been mostly studied in isolated form by NMR [13–15]. They fold as a right-handed four helix bundle, with the four helices (I, II, III and IV) being connected by loops. At the N-terminus of helix II, all PCP domains possess a conserved serine residue that serves as an attachment

point for a phosphopantetheine arm (PPant arm). This post translational modification is catalysed by phosphopantetheinyl transferases (PPtases) that convert apo-PCPs into holo-PCPs [16]. The 20 Å long PPant arm displays a free thiol at its extremity that allows the loading of various substrates. In this review, the term ‘loaded-PCP’ will be used to refer to PCPs modified with a PPant arm and loaded with a substrate. Indeed, a number of NRPSs have been structurally characterized in various productive conformations by employing a promiscuous PPtase, such as Sfp from *Bacillus subtilis* [17], for the loading of substrates, mimics or dead-end inhibitors, onto the megaenzymes [18–22].

The PCP domain delivers substrates to the catalytic domains. First, the adenylation (A) domain activates the incoming acid monomer. A domains select very diverse monomers including  $\alpha$ -L- or  $\alpha$ -D-amino acids,  $\beta$ -amino acids or aryl acids [23]. Subsequently, the condensation (C) domain catalyses peptide bond formation between two PCP-tethered monomers. Many optional tailoring domains that further chemically modify the metabolite under construction can also be found in NRPSs [24]. For example, the epimerization (E) domain converts natural L-amino acids into D-amino acids (modules 3 and 6 in figure 1a). Finally, each assembly line ends with a domain, such as a thioesterase

(TE) or a reductase (Re) (module 7 in figure 1*a*), that releases the final product. This final stage can introduce further diversity in the peptide as the release can occur either by hydrolysis or cyclization. Interestingly, since surfactin release occurs via macrolactonization [25,26], surfactin metabolites contain both amide and ester bonds, thus deserving the designation of 'depsipeptide' (figure 1*b*) [27].

NRPS flexibility allows the conformational changes required for the interactions between the PCP and the catalytic domains. However, this flexibility is a drawback to structural characterizations of large NRPS fragments and, accordingly, successful structural studies have often required the employment of chemical tools to reduce conformational heterogeneity [19–22]. Nevertheless, characterizing the movements of NRPS multienzymes is a requirement for the detailed understanding of these fascinating machineries. In this review, we focus on recent aspects of NRPS flexibility that allow PCP movements during a catalytic cycle, by describing both the successive conformations adopted by these enzymes during a cycle as well as movements that engender passage from one conformation to the next.

## 2. Known non-ribosomal peptide synthetase structures

The field of NRPS structural biology achieved major breakthroughs in the last five years with the publication of several crystal structures of multi-domain NRPSs, largely due to the work of the Schmeing and Gulick groups (figure 2) [19–22,28,29]. These results add to a high number of structures of individual catalytic domains and PCP-containing didomains, solved by NMR, X-ray crystallography or a combination of both techniques [8]. Each domain adopts the same fold in different structures, independent of the number of domains present in the protein. Nevertheless, understanding the organization of full modules as well as module–module interactions is essential to provide a better insight into these assembly lines. For a long time, however, the intrinsic flexibility of NRPSs prevented the structural characterization of full modules at high resolution. The first structure of a full module, that of the termination module of surfactin, SrfA-C, was solved by X-ray crystallography in 2008 (figures 1*a* and 2*a,b*) [30]. Since then, crystal structures of three other termination modules have been solved: AB3403 from an uncharacterized pathway of *Acinetobacter baumannii*, the enterobactin termination module EntF and ObiF1 from the obafluorin assembly line (figure 2*c–e*) [20,29,30]. In 2017, a combined effort in X-ray crystallography and negative staining electron microscopy (EM) provided insights into the structures of the last two modules of DhbF involved in the synthesis of bacillibactin (figure 2*f,g*) [21]. Lastly, the structural elucidation of the dimodular protein LgrA, from the gramicidin synthetase complex that contains an initiation and an elongation module, was a breakthrough in the comprehension of supra-modular NRPS organization. Indeed, 12 crystal structures of LgrA fragments complexed with substrates, substrate analogues and dead-end inhibitors were solved and provided a full picture of the catalytic cycle of a dimodular NRPS (figures 2*h–j* and 3) [19,22].

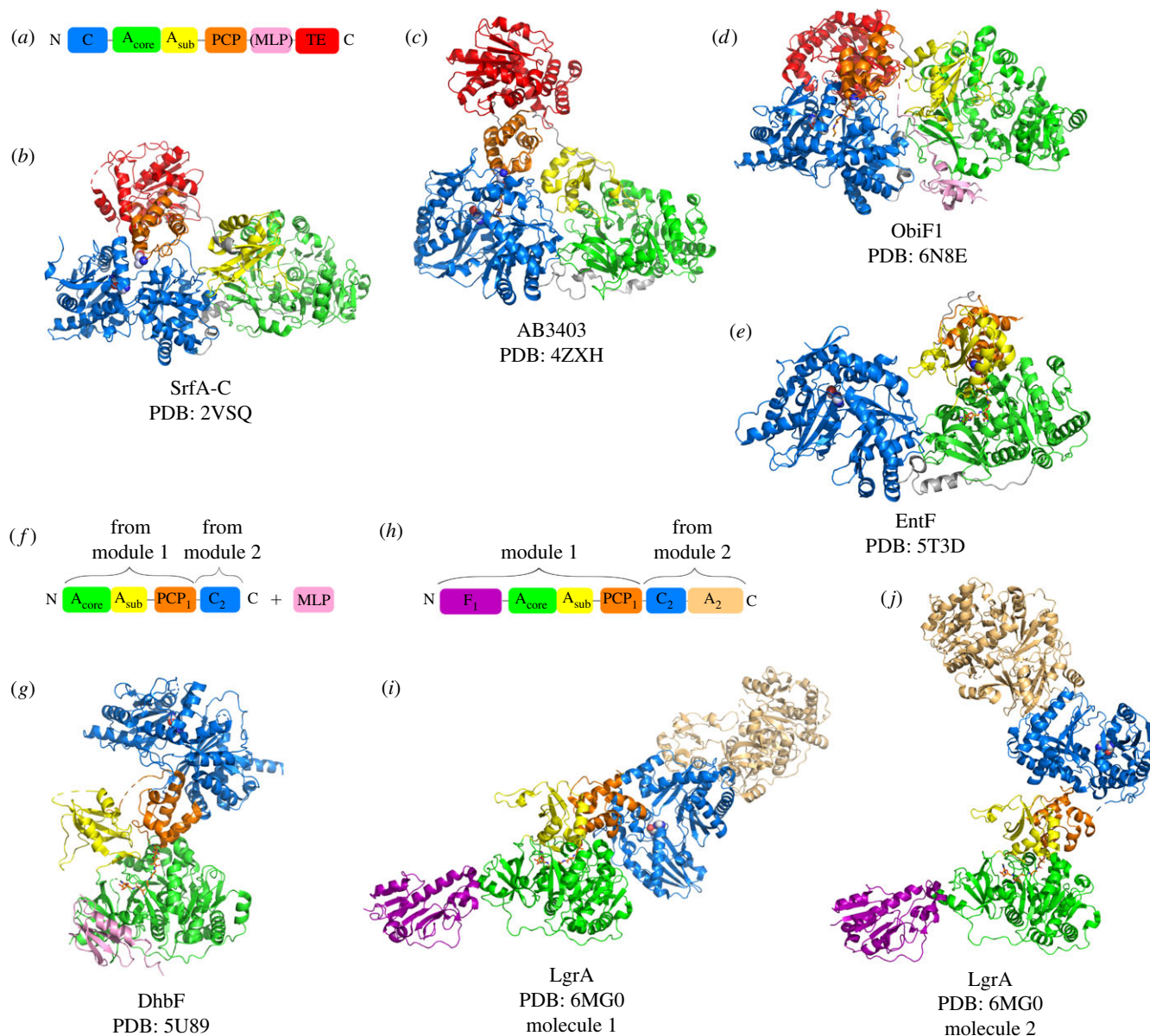
## 3. Loading of the amino acid onto the peptidyl carrier protein domain

The adenylation (A) domain is divided into two subdomains: the N-terminal subdomain,  $A_{\text{core}}$ , consists of around 400 amino acids, while the C-terminal subdomain,  $A_{\text{sub}}$ , comprises around 100 amino acids [31]. The A domain catalyses two reactions: the activation of the acid monomer using ATP (adenylation) and its subsequent transfer to the PCP domain (thiolation or thioesterification). The A domain is able to adopt several conformations that have been described as 'the domain alternation cycle' and which are supported by several structures of complete NRPS modules (figure 3) [19,20,22,31]. Remarkably, the thiolation state has been characterized multiple times using non-hydrolysable analogues (figure 3*b*, thiolation conformation) [18–20,22]. For example, the structure of PA1221, a natural A-PCP didomain NRPS from *Pseudomonas aeruginosa*, was obtained both in its apo form and in a loaded form locked in the thiolation conformation through the use of the inhibitor valyl-adenosine vinylsulfonamide (AVS) [32]. In the apo form, the electron density for the PCP domain was absent, suggesting the domain was flexible, whereas the whole didomain was visible in the presence of the AVS inhibitor, indicating that it stabilized the A-PCP interface. This suggests that the numerous crystal structures of NRPSs in the thiolation conformation do not reflect a preferentially adopted conformation *in vivo* but, more likely, a conformation that favours crystallization.

The cycle starts when the A domain, in an open conformation, is available for substrate binding (figure 3*b*, open conformation) [31]. The  $A_{\text{core}}$  subdomain contains the monomer binding pocket that accommodates ATP and the acid monomer. Upon substrate binding, a 30° rotation of the  $A_{\text{sub}}$  subdomain leads to the closed conformation, which is then suitable for adenylation (figure 3*b*, closed conformation) [19]. This conformation allows the entry of an  $A_{\text{sub}}$  loop into the  $A_{\text{core}}$  subdomain; this loop contains a conserved catalytic lysine that stabilizes the acid substrate and ATP [33]. Subsequently, after adenylation and pyrophosphate release, a 140° rotation of the  $A_{\text{sub}}$  subdomain allows the conversion between the closed and the thiolation conformations, the latter being able to catalyse thioesterification (figure 3*b*, thiolation conformation) [31].

The rotations of the  $A_{\text{sub}}$  subdomain are facilitated by a flexible hinge region containing a conserved aspartic acid or a lysine, located in the  $A_{\text{core}}-A_{\text{sub}}$  linker [34]. The structure of an adenylate-forming enzyme with the hinge residue mutated into proline revealed an enzyme blocked in the adenylation conformation. Consistent with the structure, the mutant enzyme was still capable of adenylation, but not of thiolation [34]. The importance of the hinge residue has also been demonstrated in the context of the multi-domain NRPS EntF, in which the same hinge residue mutation abolished enterobactin production [35]. Therefore, the flexibility of the hinge residue in the  $A_{\text{core}}-A_{\text{sub}}$  linker is essential to allow the movement of the  $A_{\text{sub}}$  subdomain relative to the  $A_{\text{core}}$ , which is necessary to allow conformational changes of the whole module. Indeed, rotations of the  $A_{\text{sub}}$  subdomain drive movement of the PCP owing to the linker connecting the two domains. Analysis of the A-PCP linker region reveals that it contains multiple prolines, absent in standalone A domains [35]. These prolines might rigidify





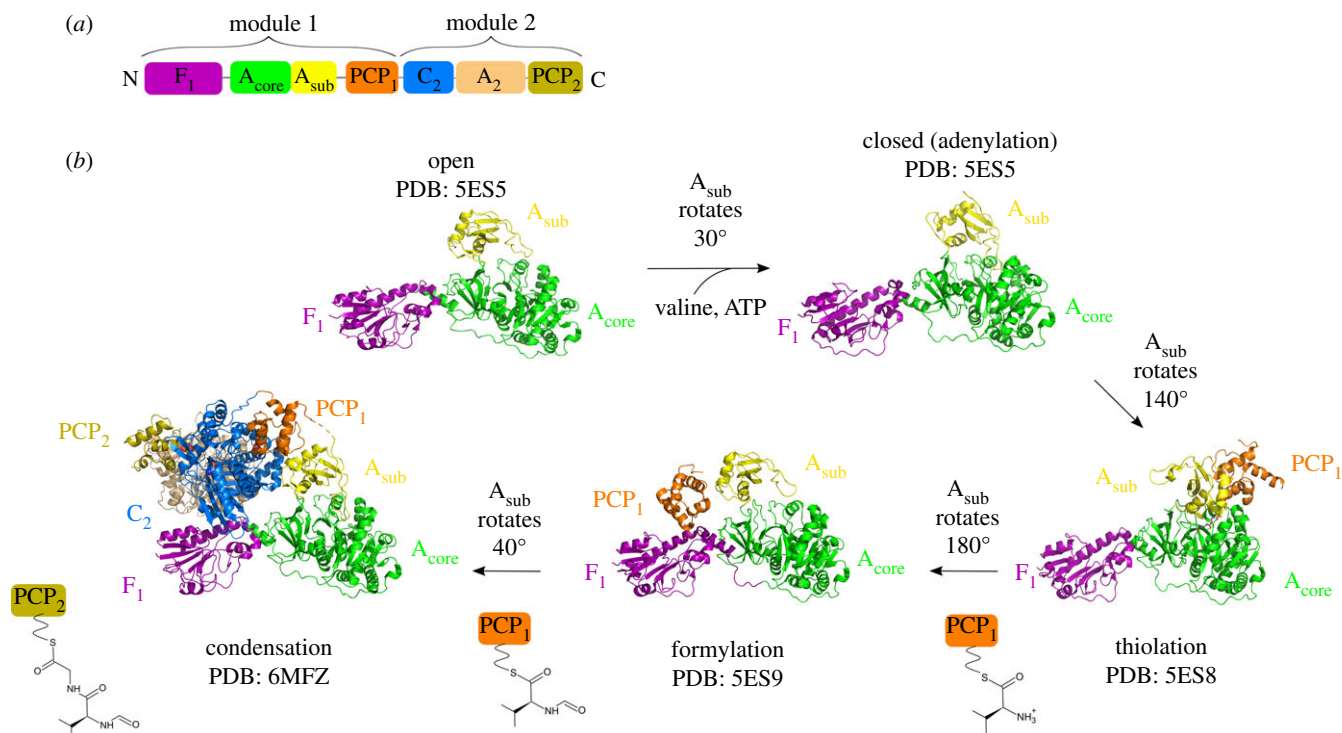
**Figure 2.** Structural gallery of NRPS modules. If present, the PPant arm attached to the PCP domain is represented as sticks. (a,b) Domain organization and crystal structures of the termination modules SrfA-C (PDB code: 2VSQ) [19], (c) AB-3403 (PDB code: 4ZXH) [21], (d) ObiF1 (PDB code: 6N8E) [22] and (e) EntF (PDB code: 5T3D) [21]. In addition to the PCP domain, these modules are composed of condensation (c), adenylation (a) and thioesterase (TE) domains. No electron density was detected for the TE domain of EntF. The catalytic His of the C domain is represented in spheres. (f,g) Domain organization and crystal structure of DhbF, a cross-module construct (PDB code: 5U89) [16]. (h–j) Domain organization and crystal structures of a five-domain construct of LgrA (PDB code: 6MG0) [24]. LgrA starts with a formylation (f) domain. The domain in pink in ObiF1 (d) and DhbF (g) represents MLP, an A domain activator.

the A-PCP linker, thus facilitating movements of the PCP domain in concert with the movements of the A<sub>sub</sub> subdomain [35].

#### 4. Modification of the peptidyl carrier protein-tethered amino acid

The vast diversity of NRPs arises in part from the action of tailoring domains, such as cyclization (Cy), epimerization (E), formylation (F), ketoreductase (KR), methyltransferase (Met) and oxidase (Ox) domains, that modify the peptide under construction [19,36–40]. It is not unusual for tailoring domains to be inserted within A domains, that are then called interrupted A domains [39,41]. The structure of LgrA in the formylation conformation provides insight into the mechanism of amino acid modification by a tailoring

domain, in this case the formylation of PCP-bound L-valine by an F domain using a formyltetrahydrofolate cofactor (figure 3b, formylation conformation) [19]. Since the F and PCP domains are separated by the large A domain (figure 3a), a substantial conformational change must occur to allow the PCP domain to position its PPant arm in the F active site. The A<sub>sub</sub> subdomain rotates 180° from its position in the thiolation conformation while the PCP domain rotates 75°, thus moving 60 Å away from its thiolation position (compare figure 3b, thiolation and formylation conformations). The interaction surface between the F and PCP domains is very limited. Unfortunately, there is no structure of the full dimodular LgrA in the formylation conformation. It is worth noting that the structure of dimodular LgrA (figure 3b, condensation conformation) reveals that the position occupied by the PCP<sub>1</sub> domain in the formylation conformation is occupied by the C domain at a later stage of the catalytic cycle [22].



**Figure 3.** Catalytic cycle of LgrA, a dimodular NRPS. (a) Domain organization of LgrA. For clarity, the inactive epimerization domain of LgrA, that follows the PCP<sub>2</sub> domain, is not shown. (b) Catalytic cycle of LgrA illustrated by five crystal structures (PDB codes: 5ES5, 5ES8, 5ES9 and 6MFZ) [23,24]. Four different structures of module 1 reveal details about the catalytic cycle of an initiation module. The PCP<sub>1</sub> domain is disordered in the open and closed states. First, the open state allows binding of valine and ATP. The closed state is the conformation which is relevant for the activation of valine by ATP (i.e. the adenylation state). The thiolation conformation captures the transfer of valine from the A<sub>1</sub> domain to the PPant arm of the PCP<sub>1</sub> domain. In the formylation conformation, the F<sub>1</sub> domain adds a formyl group to valine, still attached to the PCP<sub>1</sub> domain. The structure of the full dimodular NRPS allows the visualization of the condensation conformation. After condensation, Gly activated by module 2 is covalently bound to formyl-Val by a peptide bond. In all presented structures, the F<sub>1</sub>-A<sub>core</sub> bidomain adopts similar conformations whereas the A<sub>sub</sub> subdomain and the PCP<sub>1</sub> domain are positioned differently. Rotations of the A<sub>sub</sub> subdomain induce movements of the PCP<sub>1</sub> domain due to the linker between them.

## 5. Elongation of the donor peptide chain with an acceptor amino acid

Elongation is the only reaction that necessitates interactions between domains belonging to different modules. This reaction is catalysed by a condensation (C) domain or, more rarely, by a cyclization (Cy) also called heterocyclization (HC) domain, that catalyses cyclization after condensation [42]. If the two modules belong to the same polypeptide, PCP and C domains are directly connected by a linker (figure 3*b*, condensation conformation). If they belong to different polypeptides, docking domains facilitate the interaction between the upstream PCP and the downstream C domain [43]. The C domain of an elongation module  $n$  catalyses peptide bond formation between the growing chain carried by the upstream PCP domain (donor PCP), that belongs to module  $n-1$ , and the activated amino acid carried by the downstream PCP domain (acceptor PCP) located on the same module  $n$  (figure 3*b*, condensation conformation) [42]. The growing peptide chain is directly transferred from one PCP domain to the next, without attachment to the C domain. Thus, the two PCP domains must bind simultaneously to two different binding sites on the C domain; these are referred to as donor and acceptor binding sites. However, the C domain must discriminate between the two PCP domains to maintain the directionality of the assembly line. Therefore, the condensation reaction results from the interaction between three domains (the C, the donor PCP and the acceptor PCP domains).

The C domain adopts the V-shaped pseudo-dimeric fold seen in the chloramphenicol acyltransferase family [42]. It is divided into N-terminal and C-terminal lobes, and the active site is located inside the N-terminal lobe, at the centre of a tunnel formed by the interface between the two lobes (figure 2). The structures of three termination modules, SrfA-C, AB3403 and ObiF1, show the acceptor PCP domain docked onto the C domain acceptor site (figure 2*a-d*) [20,29,30]. In SrfA-C, although the PCP domain is in its apo form due to a Ser-Ala mutation, the Ala is located 16 Å away from the catalytic His of the C domain, suggesting that this structure is compatible with a productive condensation reaction (figure 2*b*) [30]. Both holo-AB3403 and holo-ObiF1 show the PCP domain docked onto the C acceptor site with the PPant arm inserted through a tunnel that allows the positioning of the final thiol in proximity to the catalytic His of the C domain (figure 2*c,d*) [20,30]. It is worth noting that, as opposed to what was observed for the PCP interaction with the A domain, no substrate or inhibitor was needed to favour the interaction between the acceptor PCP and C domains.

Several dimodular LgrA structures have revealed for the first time the productive interaction between a donor PCP domain and its corresponding C domain (figure 3*b*, condensation conformation) [22]. Indeed, in four structures of LgrA, the donor PCP<sub>1</sub> domain is docked in the C<sub>2</sub> donor site, presenting its conserved Ser towards the catalytic His of the C domain, the two residues being separated by less than 20 Å. The donor PCP<sub>1</sub> domain globally has the same

orientation in these structures and three of them show electron density for the PPant arm in the donor tunnel of the C domain. However, the structure of  $F_1$ - $A_1$ -PCP<sub>1</sub>-C<sub>2</sub> with f-Val loaded onto the PPant arm of PCP<sub>1</sub> reveals that the reactive thioester group must slightly modify its position in order to be properly positioned for attack by the acceptor amino acid. The authors hypothesized that the donor substrate could only be correctly positioned in the presence of the PCP-bound-acceptor substrate in the active site of the C domain [22]. The importance of the loading status of the carrier protein domain for megaenzyme conformation has already been demonstrated for the modular PKS megaenzymes that incorporate domains analogous to the C, A and PCP of the NRPS systems [44,45]. Indeed, structures of the PikAIII PKS module loaded with various substrates obtained by cryo-electron microscopy revealed that the acyl carrier protein (ACP) domain adopts dramatically different positions according to the nature of the substrate loaded onto the ACP.

A detailed comprehension of the condensation reaction requires structural information on a NRPS including at least a donor PCP, an acceptor PCP and a condensation domain. The structure of holo-LgrA  $F_1$ - $A_1$ -PCP<sub>1</sub>-C<sub>2</sub>-A<sub>2</sub>-PCP<sub>2</sub> reveals both donor and acceptor PCPs docked onto a single C domain (figure 3*b*, condensation conformation) [22]. 29 Å separate the two PCP Ser residues loaded with their 20 Å-long PPant arms. The Ser from the acceptor PCP<sub>2</sub> is located 15 Å away from the catalytic His of the C domain while the Ser from the donor PCP<sub>1</sub> is 18 Å away from it. These distances are compatible with the proximity between substrates required for nucleophilic attack. Unfortunately, there is no electron density for the PCP PPant arms so the detailed interaction of donor and acceptor substrates cannot be deduced from this structure.

The detailed view of PCP-bound-substrates in the condensation conformation can be obtained using a mechanism-based probe, recently designed by the Gulick and Aldrich groups [46]. Although the structure of LgrA  $F_1$ - $A_1$ -PCP<sub>1</sub>-C<sub>2</sub>-A<sub>2</sub>-PCP<sub>2</sub> shows that an inhibitor is not necessary to lock a dimodular NRPS in the condensation conformation, this new chemical probe stabilized the interaction between the donor PCP, the acceptor PCP and the C domains. The enterobactin assembly line served as a model to prove the functionality of this probe [46]. The authors were able to mimic the PPant arm loaded with the natural substrate of the donor PCP by replacing the whole acyl-thioester portion of the substrate by a non-hydrolysable analogue incorporating a ketone functionality, thus preventing the release of the loaded substrate. The resulting crypto-PCP was shown to bind to the donor site of the C domain and these results allowed the construction of a model where the crypto-PCP inserts its unnatural PPant arm into the donor tunnel. The authors assumed that the pantetheine probe would then react with the natural acceptor substrate loaded on the acceptor PCP, forming an imine bond instead of the natural peptide bond formed between substrates. In this configuration, both PCP domains should be docked onto the C domain and linked together via their PPant arms connected through an imine bond. Therefore, this probe should help stabilize the interaction between the C and PCP domains during condensation and could lead to more crystal structures of bimodular NRPS locked in the condensation conformation.

The structure of the LgrA PCP<sub>1</sub> domain has been solved in association with its three catalytic partners ( $A_1$ ,  $F_1$  and  $C_2$  domains, figure 3*b*); therefore, the comparison of the three crystal structures provides insights into the conformational changes that allow the PCP<sub>1</sub> domain to shuttle between its three partners [19,22]. As described above, the large movements required for the PCP domain to reach its different catalytic partners are mainly driven by conformational changes of the  $A_{\text{sub}}$  subdomain that are transferred to the PCP domain by the  $A_{\text{sub}}$ -PCP rigid linker. Shifting from the formylation to the condensation conformation, the PCP<sub>1</sub> domain must cross 30 Å, achieved through a rotation of 40° of the  $A_{\text{sub}}$  subdomain (compare figure 3*b*, formylation and condensation conformations). Similarly, after condensation, the PCP<sub>1</sub> domain must detach from the C domain and travel back 50 Å to return to the  $A_1$  active site (compare figure 3*b*, condensation and thiolation conformations), achieved by a rotation of 150° of the  $A_{\text{sub}}$  subdomain. Even in the absence of structures that show the second LgrA module in the thiolation conformation, we can easily extrapolate that the movements seen in module 1 could be similar in module 2.

Interestingly, a NRPS module can start a second catalytic cycle before the first one is complete [20]. For example, the structure of LgrA in the condensation conformation (figure 3*b*) shows the PCP<sub>1</sub> domain in the peptide donation conformation while the  $A_1$  domain is in the closed conformation and can thus catalyse adenylation [22]. After adenylation, the aminoacyl-AMP is tightly sequestered in the A domain active site in the absence of the available PCP [47,48]. Subsequently, the  $A_1$  domain can catalyse thiolation as soon as the  $C_2$  domain has catalysed condensation which will liberate the PCP<sub>1</sub> domain. This decoupling between different domain activities likely increases the synthesis rate of NRPSs.

## 6. Release of the peptidyl carrier protein-tethered peptide

The structures of four termination modules harbouring a TE domain (C-A-PCP-TE) are now available, i.e. SrfA-C, AB3403, EntF and ObiF1 (figure 2*a-e*) [20,28–30]. In all four crystal structures, the TE positions are dramatically different (figure 2*a-d*), suggesting that the TE domain is most probably a mobile element. Negative staining EM images of EntF showed the TE domain in various positions compared to the other domains and no density was observable for the TE domain in the corresponding crystal structure (figure 2*e*), confirming that the EntF TE domain can adopt multiple conformations [20]. The ObiF1 module has an unusual domain organization, since the TE domain is followed by a MbtH-like protein (MLP) [30]. Interestingly, in these conditions, the MLP domain anchors the TE domain to the module (figure 2*d*). These elements suggest that the high mobility of the TE domain is due to the flexibility of the short PCP-TE linker and to the fact that, in general, no successive domain imposes structural restraints on the final TE domain. As none of the structures of these four termination modules revealed the interaction between the PCP and the TE domains, it was characterized through the crystal structure of the EntF PCP-TE didomain [49]. The authors used a phosphopantetheinyl-based inhibitor loaded onto the PCP domain that stabilized the transient interaction between the

PCP and TE domains [49], thus providing details of a productive PCP–TE interaction.

## 7. Non-ribosomal peptide synthetases flexibility at the supramodular scale, unrelated to the catalytic cycle

As described in the previous sections, NRPS flexibility allows the conformational rearrangements that are required for the PCP domain to interact with its catalytic partners. It is then legitimate to wonder whether NRPS flexibility is only restricted to the movements that shuttle PCP-tethered substrates to the different NRPS active sites or if there is flexibility at the supramodular scale, unrelated to the catalytic cycle. In other words, are there architectural rules governing the relationships between successive modules, or is their relationship random? In addition to the crystal structure of bimodular LgrA, structural models of multimodular NRPSs derive from low resolution techniques or from the combination of crystal structures. Indeed, in 2016, Marahiel and co-workers proposed a helical model for a hypothetical 7-module NRPS assembly by combining the C-A-PCP structure of the SrfA-C termination module (figure 2*b*) with the PCP-C cross-module structure of TycC from the tyrocidine synthetase [30,50,51]. The helical axis was occupied by the PCP domains and each module was rotated by 120° relative to the previous one.

Several EM observations indicated that NRPSs probably adopt a more flexible architecture than the helical model mentioned above. An early negative staining EM observation of a fungal 11-module NRPS, responsible for the synthesis of cyclosporin, pictured this 1.7 MDa machinery as an assembly of globular moieties, most likely modules, that could adopt either very compact or elongated structures [52]. It led to the hypothesis that NRPS modules are arranged as ‘beads on a chain’, suggesting that an NRPS assembly line would not adopt any specific architecture.

More recently, the dimodular NRPS DhbF (C<sub>1</sub>-A<sub>1</sub>-PCP<sub>1</sub>-C<sub>2</sub>-A<sub>2</sub>-PCP<sub>2</sub>+MLP) was also observed by negative staining EM [21]. Despite the presence of AVS inhibitors that limited its conformational heterogeneity, DhbF adopted a continuum of conformations as diverse as an elongated shape, an L shape or a very compact shape. Most particles could be sorted into five classes that differed by the relative positions of the first module in relation to the second one. Therefore, although the flexibility inside a module is limited due to the stable conformation adopted by the C-A didomain, it seems that there are few limitations to the position one module can adopt relatively to the adjacent one. These data favour an irregular architecture for NRPSs; however, the fact that the number of classes is limited to five suggests that the supramodular architecture of NRPSs is not completely random. In the crystal structure of the A<sub>1</sub>-PCP<sub>1</sub>-C<sub>2</sub> cross-module (figure 2*f,g*), there is no density for the PCP<sub>1</sub>-C<sub>2</sub> linker, suggesting that it could be flexible [21]. Therefore, the movements of the PCP<sub>1</sub>-C<sub>2</sub> linker combined with the absence of strong intermodule interactions could explain the various conformations adopted by the dimodular DhbF.

The six recent dimodular LgrA crystal structures provided further evidence that NRPSs do not adopt a unique stable architecture but rather a few conformations among a myriad of possibilities [22]. One striking example confirmed

the flexibility of the PCP<sub>1</sub>-C<sub>2</sub> intermodule linker. Indeed, the F<sub>1</sub>-A<sub>1</sub>-PCP<sub>1</sub>-C<sub>2</sub>-A<sub>2</sub> variant was crystallized in the thiolation conformation for module 1 using a Val-AVS inhibitor and two molecules were found in the asymmetric unit (figure 2*h-j*). Within these two molecules, module 1 is identical but module 2 adopts two radically different positions. This behaviour results in two strikingly different LgrA shapes, reminiscent of the two DhbF structures observed by EM, one that is elongated and the other L-shaped. Therefore, it seems that locking one module in a specific conformation does not impose a unique conformation on the adjacent module. The most convincing evidence that NRPSs adopt a flexible architecture was obtained from the SAXS analyses of the LgrA F<sub>1</sub>-A<sub>1</sub>-PCP<sub>1</sub>-C<sub>2</sub>-A<sub>2</sub> construct [22]. They indicated that the conformations adopted in the crystal structures do not exactly reflect the conformations adopted in solution. To better estimate these, Reimer and co-workers used the ensemble optimization method to generate different models that took into account flexibility parameters [53]. The ensemble generated fit very well with the experimental data, thus confirming the flexibility of LgrA. However, it cannot be excluded that LgrA flexibility is only apparent and is an effect of the absence of the other components of the assembly line. Indeed, in addition to LgrA, the linear gramicidin NRPS is composed of three other proteins [54] that could restrain the conformations that LgrA can adopt.

## 8. Concluding remarks

The structural and functional studies of NRPS multienzymes are not limited to providing details regarding the production of complex metabolites but can also be applied to the discovery of new antibiotics. Indeed, the products of these machineries are often essential for bacterial virulence, hence targeting their biosynthesis is a promising strategy to fight microbial pathogens [6]. For instance, the multi-drug resistant *Klebsiella pneumoniae* uses several siderophores for iron acquisition, including the NRPs enterobactin and yersiniabactin [55–57]. Strains deficient for yersiniabactin production are much less virulent than the wild-type strains [58], suggesting that the yersiniabactin NRPS machinery could be potentially be explored as an antibacterial development target.

Moreover, the re-engineering of NRPS megaenzymes in order to produce new medically relevant molecules is of particular interest [59,60]. This prospect exists since the discovery of the modular organization of NRPSs [61]. To date, a straightforward strategy to re-engineer NRPS assembly lines to produce artificial peptides has been difficult to establish, although some successful reports of re-engineering were published [62–65]. Classical strategies using substitutions of A, C-A, PCP-C-A units or entire modules yielded only a small amount of synthesized peptide [59,60]. Recently, the Bode group successfully exploited a novel exchange strategy, using A-PCP-C exchange units (XUs) by fixing the borders of the XU within the flexible C and A domain linker [66]. They subsequently improved their strategy by dividing the C domain, placing the borders of the XU within the flexible linker that connects the N-terminal acceptor and C-terminal donor subdomains of C, yielding C<sub>Acc</sub>-A-PCP-C<sub>Don</sub> (XUs) [67]. This strategy allowed the authors to produce very high yields of novel NRPS peptides, paving the way for new biotechnological approaches that could optimize the production

of novel bioactive compounds through NRPS engineering. Therefore, an increased knowledge on the supramodular architecture of NRPSs, especially regarding the linker regions that allow enzyme flexibility, raises interesting perspectives for natural product re-engineering.

**Data accessibility.** This article has no additional data.

**Authors' contributions.** S.B., A.D. and P.M. wrote the paper together.

**Competing interests.** We declare we have no competing interests.

**Funding.** This study was funded by Fondation Innovations en Infectiologie (grant no. AO12-20), Ecole nationale Supérieure de Lyon and Ligue Contre le Cancer (grant no. M32917). Work on NRPS/PKS enzymes in the Bacterial Pathogenesis Laboratory is supported by grants from Le Comité de Savoie de la Ligue Contre le Cancer (to P.M.) and the FINOVI foundation (to A.D.). S.B. is the recipient of a PhD fellowship from the Ecole Nationale Supérieure (ENS) de Lyon.

**Acknowledgements.** The IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA).

## References

- Cragg GM, Newman DJ. 2013 Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta.* **1830**, 3670–3695. (doi:10.1016/j.bbagen.2013.02.008)
- Felnagle EA, Jackson EE, Chan YA, Podevels AM, Berti AD, McMahon MD, Thomas MG. 2008 Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharm.* **5**, 191–211. (doi:10.1021/mp700137g)
- Galm U, Wendt-Pienkowski E, Wang L, Huang S-X, Unsinn C, Tao M, Coughlin JM, Shen B. 2011 Comparative analysis of the biosynthetic gene clusters and pathways for three structurally related antitumor antibiotics: bleomycin, tallysomyin, and zorbamycin. *J. Nat. Prod.* **74**, 526–536. (doi:10.1021/np1008152)
- Menkhaus M, Ullrich C, Kluge B, Vater J, Vollenbroich D, Kamp RM. 1993 Structural and functional organization of the surfactin synthetase multienzyme system. *J. Biol. Chem.* **268**, 7678–7684. (doi:10.1016/S0021-9258(18)53010-6)
- Singh M, Chaudhary S, Sareen D. 2017 Non-ribosomal peptide synthetases: identifying the cryptic gene clusters and decoding the natural product. *J. Biosci.* **42**, 175–187. (doi:10.1007/s12038-017-9663-z)
- Gulick AM. 2017 Nonribosomal peptide synthetase biosynthetic clusters of ESKAPE pathogens. *Nat. Prod. Rep.* **34**, 981–1009. (doi:10.1039/C7NP00029D)
- Vassaux A, Meunier L, Vandenbol M, Baurain D, Fickers P, Jacques P, Leclère V. 2019 Nonribosomal peptides in fungal cell factories: from genome mining to optimized heterologous production. *Biotechnol. Adv.* **37**, 107449. (doi:10.1016/j.biotechadv.2019.107449)
- Izoré T, Cryle MJ. 2018 The many faces and important roles of protein-protein interactions during non-ribosomal peptide synthesis. *Nat. Prod. Rep.* **35**, 1120–1139. (doi:10.1039/C8NP00038G)
- Jaremko MJ, Davis TD, Corpuz JC, Burkart MD. 2020 Type II non-ribosomal peptide synthetase proteins: structure, mechanism, and protein-protein interactions. *Nat. Prod. Rep.* **37**, 355–379. (doi:10.1039/C9NP00047J)
- Hertweck C. 2009 The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed. Engl.* **48**, 4688–4716. (doi:10.1002/anie.200806121)
- Miyana A, Kudo F, Eguchi T. 2018 Protein-protein interactions in polyketide synthase-nonribosomal peptide synthetase hybrid assembly lines. *Nat. Prod. Rep.* **35**, 1185–1209. (doi:10.1039/C8NP00022K)
- Crosby J, Crump MP. 2012 The structural role of the carrier protein—active controller or passive carrier. *Nat. Prod. Rep.* **29**, 1111–1137. (doi:10.1039/C2np20062g)
- Koglin A *et al.* 2006 Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* **312**, 273–276. (doi:10.1126/science.1122928)
- Frueh DP, Arthanari H, Koglin A, Vosburg DA, Bennett AE, Walsh CT, Wagner G. 2008 Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* **454**, 903–906. (doi:10.1038/nature07162)
- Jaremko MJ, Lee DJ, Opella SJ, Burkart MD. 2015 Structure and substrate sequestration in the Pyoluteorin type II peptidyl carrier protein PtlL. *J. Am. Chem. Soc.* **137**, 11 546–11 549. (doi:10.1021/jacs.5b04525)
- Beld J, Sonnenschein EC, Vickery CR, Noel JP, Burkart MD. 2014 The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep.* **31**, 61–108. (doi:10.1039/C3NP70054B)
- Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C. 2001 Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790–1792. (doi:10.1126/science.1058092)
- Gulick AM, Aldrich CC. 2018 Trapping interactions between catalytic domains and carrier proteins of modular biosynthetic enzymes with chemical probes. *Nat. Prod. Rep.* **35**, 1156–1184. (doi:10.1039/C8NP00044A)
- Reimer JM, Aloise MN, Harrison PM, Schmeing TM. 2016 Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* **529**, 239–242. (doi:10.1038/nature16503)
- Drake EJ, Miller BR, Shi C, Tarrasch JT, Sundlov JA, Allen CL, Skiniotis G, Aldrich CC, Gulick AM. 2016 Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238. (doi:10.1038/nature16163)
- Tarry MJ, Haque AS, Bui KH, Schmeing TM. 2017 X-Ray crystallography and electron microscopy of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture. *Structure* **25**, 783–793.e4. (doi:10.1016/j.str.2017.03.014)
- Reimer JM, Eivaskhani M, Harb I, Guarné A, Weigt M, Schmeing TM. 2019 Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366**, eaaw4388. (doi:10.1126/science.aaw4388)
- Kudo F, Miyana A, Eguchi T. 2019 Structural basis of the nonribosomal codes for nonproteinogenic amino acid selective adenylation enzymes in the biosynthesis of natural products. *J. Ind. Microbiol. Biotechnol.* **46**, 515–536. (doi:10.1007/s10295-018-2084-7)
- Walsh CT, Chen H, Keating TA, Hubbard BK, Losey HC, Luo L, Marshall CG, Miller DA, Patel HM. 2001 Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr. Opin. Chem. Biol.* **5**, 525–534. (doi:10.1016/S1367-5931(00)00235-0)
- Tsang CC, Bruner SD, Kohli RM, Marahiel MA, Walsh CT, Sieber SA. 2002 Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase. *Biochemistry* **41**, 13 350–13 359. (doi:10.1021/bi026592a)
- Bruner SD, Weber T, Kohli RM, Schwarzer D, Marahiel MA, Walsh CT, Stubbs MT. 2002 Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrtE. *Structure* **10**, 301–310. (doi:10.1016/S0969-2126(02)00716-5)
- Alonzo DA, Schmeing TM. 2020 Biosynthesis of depsipeptides, or Depsi: the peptides with varied generations. *Protein Sci.* **29**, 2316–2347. (doi:10.1002/pro.3979)
- Miller BR, Drake EJ, Shi C, Aldrich CC, Gulick AM. 2016 Structures of a nonribosomal peptide synthetase module bound to Mbth-like proteins support a highly dynamic domain architecture. *J. Biol. Chem.* **291**, 22 559–22 571. (doi:10.1074/jbc.M116.746297)
- Kreitler DF, Gemmell EM, Schaffer JE, Wenciewicz TA, Gulick AM. 2019 The structural basis of N-acyl- $\alpha$ -amino- $\beta$ -lactone formation catalyzed by a nonribosomal peptide synthetase. *Nat. Commun.* **10**, 3432. (doi:10.1038/s41467-019-11383-7)
- Tanovic A, Samel SA, Essen L-O, Marahiel MA. 2008 Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659–663. (doi:10.1126/science.1159850)

31. Gulick AM. 2009 Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem. Biol.* **4**, 811–827. (doi:10.1021/cb900156h)
32. Mitchell CA, Shi C, Aldrich CC, Gulick AM. 2012 Structure of PA 1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry* **51**, 3252–3263. (doi:10.1021/bi300112e)
33. Branchini BR, Murtiashaw MH, Magyar RA, Anderson SM. 2000 The role of lysine 529, a conserved residue of the acyl-adenylate-forming enzyme superfamily, in firefly luciferase. *Biochemistry* **39**, 5433–5440. (doi:10.1021/bi9928804)
34. Wu R, Reger AS, Lu X, Gulick AM, Dunaway-Mariano D. 2009 The mechanism of domain alternation in the acyl-adenylate forming ligase superfamily member 4-chlorobenzoate: coenzyme A ligase. *Biochemistry* **48**, 4115–4125. (doi:10.1021/bi9002327)
35. Miller BR, Sundlov JA, Drake EJ, Makin TA, Gulick AM. 2014 Analysis of the linker region joining the adenylation and carrier protein domains of the modular nonribosomal peptide synthetases. *Proteins* **82**, 2691–2702. (doi:10.1002/prot.24635)
36. Dowling DP, Kung Y, Croft AK, Taghizadeh K, Kelly WL, Walsh CT, Drennan CL. 2016 Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl Acad. Sci. USA* **113**, 12 432–12 437. (doi:10.1073/pnas.1608615113)
37. Chen W-H, Li K, Guntaka NS, Bruner SD. 2016 Interdomain and intermodule organization in epimerization domain containing nonribosomal peptide synthetases. *ACS Chem. Biol.* **11**, 2293–2303. (doi:10.1021/acscchembio.6b00332)
38. Alonzo DA, Chiche-Lapierre C, Tarry MJ, Wang J, Schmeing TM. 2020 Structural basis of keto acid utilization in nonribosomal depsipeptide synthesis. *Nat. Chem. Biol.* **16**, 493–496. (doi:10.1038/s41589-020-0481-5)
39. Mori S, Pang AH, Lundy TA, Garzan A, Tsodikov OV, Garneau-Tsodikova S. 2018 Structural basis for backbone N-methylation by an interrupted adenylation domain. *Nat. Chem. Biol.* **14**, 428–430. (doi:10.1038/s41589-018-0014-7)
40. Pang B, Chen Y, Gan F, Yan C, Jin L, Gin JW, Petzold CJ, Keasling JD. 2020 Investigation of indigoidine synthetase reveals a conserved active-site base residue of nonribosomal peptide synthetase oxidases. *J. Am. Chem. Soc.* **142**, 10 931–10 935. (doi:10.1021/jacs.0c04328)
41. Labby KJ, Watsula SG, Garneau-Tsodikova S. 2015 Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis. *Nat. Prod. Rep.* **32**, 641–653. (doi:10.1039/C4NP00120F)
42. Bloudoff K, Schmeing TM. 2017 Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta Proteins Proteom.* **1865**, 1587–1604. (doi:10.1016/j.bbapap.2017.05.010)
43. Smith HG, Beech MJ, Lewandowski JR, Challis GL, Jenner M. In press. Docking domain-mediated subunit interactions in natural product megasynth(et)ases. *J. Ind. Microbiol. Biotechnol.*
44. Dutta S *et al.* 2014 Structure of a modular polyketide synthase. *Nature* **510**, 512–517. (doi:10.1038/nature13423)
45. Whicher JR *et al.* 2014 Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560–564. (doi:10.1038/nature13409)
46. Shi C, Miller BR, Alexander EM, Gulick AM, Aldrich CC. 2020 Design, synthesis, and biophysical evaluation of mechanism-based probes for condensation domains of nonribosomal peptide synthetases. *ACS Chem. Biol.* **15**, 1813–1819. (doi:10.1021/acscchembio.0c00411)
47. Keating TA, Suo Z, Ehmann DE, Walsh CT. 2000 Selectivity of the yersiniabactin synthetase adenylation domain in the two-step process of amino acid activation and transfer to a holo-carrier protein domain. *Biochemistry* **39**, 2297–2306. (doi:10.1021/bi992341z)
48. Ehmann DE, Shaw-Reid CA, Losey HC, Walsh CT. 2000 The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proc. Natl Acad. Sci. USA* **97**, 2509–2514. (doi:10.1073/pnas.040572897)
49. Liu Y, Zheng T, Bruner SD. 2011 Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chem. Biol.* **18**, 1482–1488. (doi:10.1016/j.chembiol.2011.09.018)
50. Samel SA, Schoenafinger G, Knappe TA, Marahiel MA, Essen L-O. 2007 Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* **15**, 781–792. (doi:10.1016/j.str.2007.05.008)
51. Marahiel MA. 2016 A structural model for multimodular NRPS assembly lines. *Nat. Prod. Rep.* **33**, 136–140. (doi:10.1039/C5NP00082C)
52. Hoppert M, Gentsch C, Schörgendorfer K. 2001 Structure and localization of cyclosporin synthetase, the key enzyme of cyclosporin biosynthesis in *Tolypocladium inflatum*. *Arch. Microbiol.* **176**, 285–293. (doi:10.1007/s002030100324)
53. Tria G, Mertens HDT, Kachala M, Svergun DI. 2015 Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCr* **2**, 207–217. (doi:10.1107/S205225251500202X)
54. Kessler N, Schuhmann H, Morneweg S, Linne U, Marahiel MA. 2004 The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. *J. Biol. Chem.* **279**, 7413–7419. (doi:10.1074/jbc.M309658200)
55. Shankar C, Veeraraghavan B, Nabarro LEB, Ravi R, Ragupathi NKD, Rupali P. 2018 Whole genome analysis of hypervirulent *Klebsiella pneumoniae* isolates from community and hospital acquired bloodstream infection. *BMC Microbiol.* **18**, 6. (doi:10.1186/s12866-017-1148-6)
56. Peltier F, Choquet M, Decroix V, Adjidi CC, Castelain S, Guiheneuf R, Pluquet E. 2019 Characterization of a multidrug-resistant *Klebsiella pneumoniae* ST607-K25 clone responsible for a nosocomial outbreak in a neonatal intensive care unit. *J. Med. Microbiol.* **68**, 67–76. (doi:10.1099/jmm.0.000884)
57. Remya PA, Shanthi M, Sekar U. 2019 Characterisation of virulence genes associated with pathogenicity in *Klebsiella pneumoniae*. *Indian J. Med. Microbiol.* **37**, 210–218. (doi:10.4103/ijmm.IJMM\_19\_157)
58. Lawlor MS, O'connor C, Miller VL. 2007 Yersiniabactin is a virulence factor for *Klebsiella pneumoniae* during pulmonary infection. *Infect. Immun.* **75**, 1463–1472. (doi:10.1128/IAI.00372-06)
59. Brown AS, Calcott MJ, Owen JG, Ackerley DF. 2018 Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. *Nat. Prod. Rep.* **35**, 1210–1228. (doi:10.1039/C8NP00036K)
60. Bozhüyük KA, Micklefield J, Wilkinson B. 2019 Engineering enzymatic assembly lines to produce new antibiotics. *Curr. Opin. Microbiol.* **51**, 88–96. (doi:10.1016/j.mib.2019.10.007)
61. Cane DE, Walsh CT, Khosla C. 1998 Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282**, 63–68. (doi:10.1126/science.282.5386.63)
62. Kegler C, Bode HB. 2020 Artificial splitting of a non-ribosomal peptide synthetase by inserting natural docking domains. *Angew. Chem. Int. Ed. Engl.* **59**, 13 463–13 467. (doi:10.1002/anie.201915989)
63. Calcott MJ, Owen JG, Ackerley DF. 2020 Efficient rational modification of non-ribosomal peptides by adenylation domain substitution. *Nat. Commun.* **11**, 4554. (doi:10.1038/s41467-020-18365-0)
64. Tietze A, Shi Y-N, Kronenwerth M, Bode HB. 2020 Nonribosomal peptides produced by minimal and engineered synthetases with terminal reductase domains. *Chembiochem* **21**, 2750–2754. (doi:10.1002/cbic.202000176)
65. Ruijne F, Kuipers OP. 2021 Combinatorial biosynthesis for the generation of new-to-nature peptide antimicrobials. *Biochem. Soc. Trans.* **49**, 203–215. (doi:10.1042/BST20200425)
66. Bozhüyük KAJ, Fleischhacker F, Linck A, Wesche F, Tietze A, Niesert C-P, Bode HB. 2018 De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281. (doi:10.1038/nchem.2890)
67. Bozhüyük KAJ *et al.* 2019 Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **11**, 653–661. (doi:10.1038/s41557-019-0276-z)







Certaines souches d'*Escherichia coli* résidant dans notre microbiote intestinal produisent la colibactine, un carcinogène dont le rôle a été démontré dans le cancer colorectal et soupçonné dans d'autres cancers. La biosynthèse de la colibactine, comme celle de nombreux métabolites secondaires microbiens d'intérêt médical, repose sur des méga-enzymes appartenant à la famille des synthétases de peptides non ribosomiaux (NRPS) et des synthèses de polycétides (PKS). Les NRPS/PKS sont constituées de domaines enzymatiques variés, regroupés en modules. Un module NRPS incorpore un acide aminé dans le métabolite en construction alors qu'un module PKS ajoute une chaîne acyle. En plus des domaines enzymatiques, chaque module intègre un domaine non enzymatique nommé « carrier protein », qui présente les substrats et le métabolite en construction aux autres domaines. Chez les NRPS/PKS de type I, plusieurs modules sont fusionnés au sein d'un même polypeptide. La taille et la flexibilité de ces protéines les rend très difficiles à caractériser, même si des avancées majeures ont eu lieu récemment, à la fois pour les NRPS et les PKS. En revanche, aucune protéine hybride NRPS/PKS fusionnée, composée d'un module NRPS associé à un module PKS, n'a été caractérisée structurellement.

Le but de cette thèse était la caractérisation structurale de l'hybride PKS-NRPS ClbK, responsable de l'incorporation dans la colibactine d'une unité aminomalonyl et d'un cycle thiazole potentiellement génotoxique. Ce travail a montré que ClbK forme un assemblage dimérique de 480 kDa. La protéine ClbK purifiée est capable d'activer la cystéine, le précurseur du cycle thiazole. La diffusion des rayons X aux petits angles a révélé que ClbK possède une forme allongée, de dimension maximale 30 nm et les modèles *ab initio* générés suggèrent la présence de plusieurs chambres catalytiques, déjà identifiées chez les PKS.

Parallèlement, la structure du module PKS de ClbK (85 kDa) a été résolue à 3 Å par cristallographie aux rayons X. Ce fragment forme un assemblage dimérique et possède plusieurs adaptations facilitant ses interactions avec les domaines NRPS en amont et en aval. Le domaine « carrier protein » est détecté à une position inhabituelle, observée une seule fois chez les PKS. La caractérisation de l'état oligomérique en solution de plusieurs fragments multi-domaines de ClbK suggère que ClbK dimérise grâce à deux domaines enzymatiques situés à ses extrémités N- et C-terminales. L'ensemble des données collectées au cours de cette thèse et les données structurales disponibles nous ont permis de proposer le premier modèle architectural d'une protéine hybride NRPS/PKS.

**Mots clés :** colibactine, microbiote, biologie structurale, synthase de polycétide (PKS), synthétase de peptides non ribosomiaux (NRPS), hybride NRPS/PKS, cancer colorectal, *E. coli pks+*.

Some *Escherichia coli* strains from the human gut microbiota synthesize colibactin, a causative agent of colorectal cancer and a potential agent involved in other cancers. As is the case for many valuable secondary metabolites produced by micro-organisms, colibactin is synthesized by non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs). NRPSs/PKSs use diverse enzymatic domains that are organized into modules. A NRPS module incorporates an amino acid during metabolite assembly, while a PKS module adds an acyl chain. In addition to enzymatic domains, each module integrates a non-enzymatic domain, named carrier protein, that delivers the substrates and the growing metabolite to the enzymatic domains. In type I NRPSs/PKSs, several modules are commonly fused into a single polypeptide. Even if the structural characterization of these megaenzymes is very challenging due to their size and flexibility, major breakthroughs occurred recently both in the NRPS and the PKS fields. However, no fused hybrid NRPS/PKS, composed of a NRPS module associated to a PKS module, has been structurally characterized to date.

The main goal of this thesis was the structural characterization of the hybrid PKS-NRPS ClbK, that incorporates into colibactin an aminomalonyl unit and a thiazole ring that could impart the genotoxic attributes of colibactin. This work demonstrated that ClbK is a 480 kDa dimer. The purified protein catalyzes the activation of cysteine, the precursor of the thiazole ring. Small angle X-ray scattering (SAXS) revealed that ClbK has an elongated shape with a maximal dimension of 30 nm. *Ab initio* models generated from SAXS data suggest that ClbK possesses catalytic chambers, a common feature of PKS proteins.

In parallel, the structure of the PKS module of ClbK (85 kDa) was solved at 3 Å by X-ray crystallography. The structure revealed that the protein is dimeric and harbors several adaptations allowing its interactions with NRPS domains. The carrier protein domain is tethered at an unusual position, observed only once in PKS proteins. Characterization of the oligomeric state of several multi-domain fragments of ClbK suggest that the megaenzyme dimerizes owing to two domains located at its N-terminal and C-terminal extremities. Data obtained during this work associated with other structural data allowed us to propose the first architectural model of a hybrid NRPS/PKS protein.

**Key words:** colibactin, microbiota, structural biology, polyketide synthase (PKS), non-ribosomal peptide synthetase (NRPS), hybrid NRPS/PKS, colorectal cancer, *pks+ E. coli*.