



**HAL**  
open science

# Monooxygénations catalysées par des polycétide synthases modulaires : caractérisation structurale et fonctionnelle

Romain Francois

► **To cite this version:**

Romain Francois. Monooxygénations catalysées par des polycétide synthases modulaires : caractérisation structurale et fonctionnelle. Biochimie, Biologie Moléculaire. Université de Lorraine, 2023. Français. NNT : 2023LORR0259 . tel-04585012

**HAL Id: tel-04585012**

**<https://theses.hal.science/tel-04585012>**

Submitted on 23 May 2024

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

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



**UNIVERSITÉ  
DE LORRAINE**

**BIBLIOTHÈQUES  
UNIVERSITAIRES**

## AVERTISSEMENT

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

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

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

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

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

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

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

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

## **Ecole Doctorale BioSE (Biologie-Santé-Environnement)**

### **Thèse**

Présentée et soutenue publiquement pour l'obtention du titre de

**DOCTEUR DE L'UNIVERSITE DE LORRAINE**

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

par **Romain FRANÇOIS**

## **Monooxygénations catalysées par des polycétide synthases modulaires : caractérisation structurale et fonctionnelle**

**Le 22 novembre 2023**

### **Membres du jury :**

<b>Rapporteurs :</b>	<b>Mme. Coralie BOMPARD</b>	<b>Chargée de recherche, CNRS, Lille</b>
	<b>M. Christophe CORRE</b>	<b>Professeur, University of Warwick, Coventry</b>
<b>Examineurs :</b>	<b>Mme. Sylvie LAUTRU</b>	<b>Chargée de recherche, CNRS, Gif-sur-Yvette</b>
	<b>M. Benoît GUILLOT</b>	<b>Professeur, Université de Lorraine, Nancy, président du jury</b>
	<b>Mme. Kira WEISSMAN</b>	<b>Professeure, Université de Lorraine, Nancy, directrice de thèse</b>
	<b>M. Arnaud GRUEZ</b>	<b>Maître de conférences, Université de Lorraine, Nancy, co-encadrant de thèse</b>

---

**UMR 7365 UL-CNRS IMoPA, Biopôle de l'Université de Lorraine, Campus Brabois  
Santé, 9 Avenue de la Forêt de Haye, BP 20199, 54505 Vandœuvre-lès-Nancy, France**



---

# *Remerciements*

---



*Je remercie profondément ma directrice de thèse, Mme. Kira J. Weissman, pour son soutien indéfectible durant ces trois années de thèse, pour ses conseils et les discussions enrichissantes que nous avons partagées. Merci pour m'avoir accordé ta confiance et m'avoir choisi pour rejoindre ton équipe. Ton optimisme et ta bienveillance m'ont sans aucun doute aidé à avancer et à persévérer pendant ces dernières années.*

*Je remercie mon co-encadrant de thèse, M. Arnaud Gruez, qui, le premier, m'a donné l'opportunité de travailler dans l'équipe du Pr. Weissman. Merci de m'avoir fait confiance lors d'un premier stage de Master 1, puis lors d'un second stage en Master 2, pour finalement entamer ensemble l'aventure du doctorat. Merci de m'avoir donné le goût de la biophysique et en particulier de la cristallographie et merci pour ces expériences hors du commun sur les lignes de lumière du synchrotron (bien qu'elles aient toujours été synonymes de nuits blanches).*

*Je remercie Mme. Sylvie Lautru et M. Christophe Corre pour l'honneur qu'ils m'ont fait d'avoir accepté de prendre part à mon jury de thèse.*

*Je remercie Mme. Coralie Bompard et M. Benoît Guillot pour avoir formé mon comité de suivi individuel pour trois ans et pour avoir accepté de devenir membres du mon jury de thèse. Merci pour les précieux conseils que vous m'avez fournis lors de nos entrevues et surtout pour vos encouragements malgré les difficultés rencontrées.*

*Je remercie Sabrina Collin pour avoir été, depuis cinq ans, une formatrice, une collègue et une amie. Merci d'avoir pris le temps de me former à l'expression et la purification de protéines recombinantes, merci pour tous tes conseils, pour ces longues discussions partagées au sujet du projet de thèse, au sujet du laboratoire, de nos vies, au sujet de tout et de rien.*

*Je remercie M. Jean-Malo Massicard de m'avoir pris sous son aile à mon arrivée en thèse. Merci de m'avoir formé aux techniques de biologie moléculaire de routine et merci pour toutes les discussions échangées. Il me semble que c'était hier que tu soutenais ta thèse et me passais le flambeau pour mener à bien la mienne.*

*Je remercie les autres maîtres de conférences de l'équipe PKS, M. Benjamin Chagot, M. Christophe Jacob et Mme. Lynn Gregory-Pauron pour les moments partagés et les conseils prodigués. Merci à vous d'avoir participé à ma formation, et ce depuis la licence.*

*Je remercie M. Bruno Charpentier, M. Stéphane Labialle, M. Arnaud Paul et Mme. Hélène Capelle pour m'avoir accueilli et encadré lors de mon tout premier stage de recherche en troisième année de licence.*

*Je remercie M. Christophe Muller et M. Aymeric Sanchez pour les innombrables allers-retours entre le réservoir d'azote liquide et la salle de cristallographie qui nous ont permis de congeler nos cristaux de protéines.*

*I would like to thank Jörn Piel for welcoming me in his team for six months in the Institute of Microbiology at ETH Zürich. Thank you for giving me the opportunity to work in your team, thank you for your kindness and joyfulness, and thank you for all your advice.*

*I would also like to thank Pr. Piel's team for the wonderful time I spent with them. Thank you for being fantastic colleagues, both brilliant scientists and good friends. Special thanks go to Franziska Hemmerling, Amy E. Fraley, Cora L. Dieterich, Mathijs F. J. Mabeesoone, Maria Dell and Hannah A. Minas, with whom I had the honour of working for six months that seemed far too short.*

*Je remercie les doctorants de l'équipe PKS, Mme. Magali Tirado, M. Arnaud Buzon, M. Serge Scat et M. Andrea Calderari pour ces deux ou trois années passées en votre compagnie. Merci pour*

*vosre bonne humeur et vosre soutien. Magali et Arnaud, je vous souhaite de la réussite pour le reste de vosre thèse. Serge et Andrea, courage pour la dernière ligne droite.*

*Je remercie tous ceux du laboratoire IMoPA avec qui j'ai passé ces trois dernières années, Roméo, Océane, Lauralie, Anne, Ghita, Kévin, Quentin, Allan, Julie, Jade, et ceux que je n'ai pas cités. Merci d'avoir été présents et d'avoir maintenu une bonne ambiance à toute épreuve.*

*Je remercie tous mes amis en dehors du laboratoire, présents à mes côtés depuis près de vingt ans pour certains. Merci à Anthonin, Janis, Lucas, Maxime, Axel, Clem, Thibault, Marine, Émeline, P-A, Benoît, Dylan, Quentin, Justine, Ophéline, Charline, Marc-Antoine, Hugo et tant d'autres. Merci d'être présents, que ce soit une fois par semaine ou une fois par an.*

*Je remercie du fond du cœur ma famille, qui a cru en moi pendant le long périple qu'ont constitué ces huit années d'études supérieures. Merci à mes parents de m'avoir soutenu, de m'avoir offert la chance de pouvoir en arriver là et d'avoir essayé de retenir le sujet de ma thèse. Merci à mon frère, et merci à ma sœur et son compagnon, leur soutien depuis Saint-Pierre-et-Miquelon me va droit au cœur.*

*Enfin, mes derniers remerciements vont à celle sans qui je n'aurais jamais pu en arriver là, Félicia Jeannelle. Merci d'avoir été là lors de ces trois années mouvementées, de m'avoir encouragé malgré les hauts et les bas et d'avoir continué à croire en moi dans les moments difficiles. Ton amour et ton soutien ont rendu tout cela possible.*







« *Schokolade ist keine Vorspeise* »

Duolingo



---

*Table des  
matières*

---



<b>État de l'art.....</b>	<b>1</b>
<b>I. Les polycétides, des métabolites spécialisés aux propriétés variées .....</b>	<b>1</b>
1. Métabolites spécialisés.....	1
2. Les polycétides .....	2
a. Une grande diversité fonctionnelle .....	2
i. Toxines.....	3
ii. Communication .....	4
iii. Régulation de processus biologiques .....	5
iv. Composants membranaires.....	7
v. Relations symbiotiques .....	7
b. Une grande diversité structurale.....	8
c. Propriétés biologiques et intérêt pharmacologique .....	9
i. Les métabolites spécialisés, source de composés bioactifs .....	10
ii. Les polycétides en médecine .....	13
<b>II. Stratégies de découverte/synthèse de nouveaux polycétides .....</b>	<b>16</b>
1. La biosphère .....	16
a. Exploiter la diversité naturelle .....	16
i. Une souche, beaucoup de composés.....	17
ii. La coculture .....	18
iii. Culture de microorganismes « non-cultivables » .....	20
b. Dévoiler le potentiel biosynthétique par <i>genome mining</i> .....	22
2. L'hémisynthèse.....	25
3. La synthèse chimique totale .....	27
4. L'ingénierie génétique.....	29
<b>III. Les polycétides synthèses.....</b>	<b>29</b>
1. Un principe de synthèse commun .....	29
2. Les PKS de type III .....	30

3. Les PKS de type II .....	32
4. Les PKS de type I .....	35
a. Les PKS de type I itératives.....	36
b. Les PKS de type I modulaires .....	39
c. Pertinence de la modularité pour l'ingénierie génétique .....	44
<b>IV. Déterminer la structure pour appréhender la fonction .....</b>	<b>48</b>
1. Les domaines canoniques de PKS .....	48
a. Les domaines acyltransférases .....	48
b. Les domaines porteurs d'acyles .....	50
c. Les domaines kétosynthases.....	51
d. Les domaines kétoréductases .....	54
e. Les domaines déshydratases .....	57
f. Les domaines énoyle réductases .....	60
g. Les domaines thioestérases .....	62
2. Organisation séquentielle et tridimensionnelle des PKS .....	66
a. Les modules et sous-unités .....	66
i. L'ACP au centre du module, perspective structurale.....	66
ii. Structure de modules et sous-unités.....	71
b. Les domaines de docking .....	76
<b>V. Caractérisation de domaines non-canoniques .....</b>	<b>80</b>
1. De nouvelles activités pour étendre la chimie des PKS .....	80
2. Activités monooxygénases .....	81
a. Les cytochromes P450 .....	81
b. Les monooxygénases flavine-dépendantes .....	82
i. Oxydation de type Baeyer-Villiger.....	83
ii. $\alpha$ -Hydroxylation.....	84
iii. Formation d'oxime .....	85
iv. Classes et structures de FMO .....	85



<b>Objectifs .....</b>	<b>89</b>
<b>Matériels et Méthodes .....</b>	<b>95</b>
<b>I. Matériel biologique et solutions d'usage courant.....</b>	<b>97</b>
1. Souches bactériennes .....	97
2. Vecteurs .....	97
3. Milieux de culture .....	98
4. Solutions d'usage courant .....	99
<b>II. Sous-clonage et mutagenèse des séquences codantes d'intérêt .....</b>	<b>100</b>
1. Amplification des séquences codantes .....	100
2. Restriction enzymatique .....	101
3. Clonage en vecteur d'expression .....	101
4. Génération des mutants de LbmA-Ox-ACP.....	102
<b>III. Transformation bactérienne par les vecteurs recombinants .....</b>	<b>103</b>
1. Préparation des bactéries compétentes.....	103
2. Transformation bactérienne.....	103
<b>IV. Production et purification des protéines recombinantes .....</b>	<b>104</b>
1. Production des protéines recombinantes .....	104
2. Extraction et purification des protéines recombinantes .....	104
a. Lyse cellulaire.....	104
b. Chromatographie d'affinité nickel .....	105
c. Protéolyse de l'étiquette His <sub>6</sub> -SUMO et seconde chromatographie d'affinité .....	105
d. Chromatographie d'exclusion stérique .....	105
e. Purification des mutants de LbmA-Ox-ACP .....	106
<b>V. Homogénéité des protéines en solution .....</b>	<b>106</b>
<b>VI. Criblage des conditions de solubilisation de la BVMO OocK.....</b>	<b>107</b>
<b>VII. Études structurales par cristallographie aux rayons X .....</b>	<b>109</b>
1. Cristallisation.....	110

a. Principe et techniques de cristallisation protéique .....	110
b. Criblage et affinement des conditions de cristallisation .....	112
2. Congélation des cristaux .....	113
3. Diffraction et collection des données.....	114
4. Traitement des données de diffraction et phasage.....	114
a. Traitement des données .....	114
b. Phasage par remplacement moléculaire .....	114
5. Construction et affinement de la structure .....	116
<b>VIII. Tests d'activité de LbmA-Ox-ACP et ses mutants.....</b>	<b>116</b>
<b>Résultats .....</b>	<b>119</b>
<b>I. Études structurales et fonctionnelles d'un domaine FMO générateur d'oxime au sein d'un module de PKS <i>trans</i>-AT .....</b>	<b>121</b>
<b>II. Etudes préliminaires d'enzymes discrètes et domaines BVMO : solubilisation et essais de cristallisation .....</b>	<b>134</b>
1. Les BVMO discrètes OocK et SbnP.....	134
a. Criblage des conditions de solubilisation de OocK .....	134
b. Purification des BVMO discrètes .....	140
c. Essais de cristallisation des BVMO discrètes.....	141
2. Le domaine LbmC-BVMO.....	142
a. Analyses bioinformatiques .....	142
b. Purification de LbmC-BVMO.....	144
c. Homogénéité en solution .....	145
d. Essais de cristallisation de LbmC-BVMO .....	146
<b>III. L'<math>\alpha</math>-hydroxylase discrète OocM : une autre classe de FMO au sein des modules de PKS <i>trans</i>-AT .....</b>	<b>146</b>
1. Analyses bioinformatiques .....	147
2. Purification de OocM .....	149
3. Caractérisation de OocM en solution : stabilité et oligomérisation .....	150

4. Essais de cristallisation.....	150
<b><i>Discussion et perspectives</i></b> .....	<b>152</b>
<b><i>Annexes</i></b> .....	<b>165</b>
<b><i>Bibliographie</i></b> .....	<b>271</b>



---

*Liste des  
abréviations*

---



2-PS	2-pyrone synthase
6-MSA	Acide 6-méthylsalicylique
A	Domaine d'adénylation
ACP	Protéine porteuse d'acyle
AH	Acylhydrolase
ARO	Aromatase
AT	Acyltransférase
$\beta$ hD	<i>Docking</i> en épingle $\beta$
BGC	Cluster de gènes biosynthétiques
BVMO	Monooxygénase de type Baeyer-Villiger
C	Domaine de condensation
$^{\circ}$ DD	Domaine de <i>docking</i> C-terminal
CHS	Chalcone synthase
CLF	<i>Chain length factor</i>
cMT	C-méthyltransférase
CoA	Coenzyme A
CYC	Cyclase
CYP	Cytochrome P450
Da	Dalton
DD	Domaine de <i>docking</i>
DE	Élément de dimérisation
DH	Déshydratase
DHD	Déshydratase <i>docking</i>
DLS	Diffusion dynamique de la lumière
EM	Microscopie électronique
$\Psi$ ER	Pseudo-énoyle réductase
ER	Énoyle réductase
FAS	Synthase d'acides gras
FAD	Flavine adénine dinucléotide
FMN	Flavine mononucléotide
FMO	Monooxygénase flavine-dépendante
FPLC	Chromatographie protéique rapide en phase liquide

GNAT	<i>N</i> -acétyltransférase type GCN5
HMG-CoA	3-hydroxy-3-méthylglutaryl-coenzyme A
HMGS	3-hydroxy-3-méthylglutaryl-CoA synthase
HR	Hautement réductrice
iPKS	PKS de type I itérative
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
$\Psi$ KR	Pseudo-kétoréductase
KR	Kétoréductase
KS	Kétosynthase
LB	<i>Lysogeny broth</i>
MAD	Diffusion anormale à plusieurs longueurs d'onde
MAT	Malonyl-/acétyltransférase
MIR	Remplacement isomorphe multiple
MS	Spectrométrie de masse
$\Psi$ MT	Pseudo-méthyltransférase
MT	Méthyltransférase
MCAT	Malonyl-CoA:ACP transacylase
MCS	Site de clonage multiple
MDR	<i>Medium-chain dehydrogenase/reductase</i>
MSAS	Acide 6-méthylsalicylique synthase
NADPH	Nicotinamide adénosine dinucléotide phosphate
<sup>N</sup> DD	Domaine de <i>docking</i> N-terminal
NR	Non réductrice
NRP	Peptide non ribosomal
NRPS	Synthétase de peptide non ribosomal
PCP	Protéine porteuse de peptide
PCR	Réaction en chaîne par polymérase
PDB	<i>Protein Data Bank</i>
PKS	Polycétide synthase
Ppant	Phosphopantéthéine
PPTase	4'-Phosphopantéthéinyl transférase
PR	Partiellement réductrice



PTM	Macrolactame tétramate polycyclique
PUFA	Acide gras polyinsaturé
RMN	Résonance magnétique nucléaire
SAD	Diffusion anormale à une longueur d'onde
SAT	Unité d'initiation-ACP transacylase
SAXS	Diffusion des rayons X aux petits angles
SDR	<i>Short-chain dehydrogenase/reductase</i>
SEC	Chromatographie d'exclusion de taille
SNAC	S-N-acétylcystéamine
STS	Stilbène synthase
SUMO	<i>Small Ubiquitin-like Modifier</i>
TAL	Acide triacétique lactone
TB	<i>Terrific broth</i>
TIM	Triose-phosphate isomérase
UHPLC	Chromatographie liquide ultra haute performance
ULP1	Protéase <i>ubiquitin-like-specific 1</i>
WT	Sauvage



---

# *État de l'art*

---

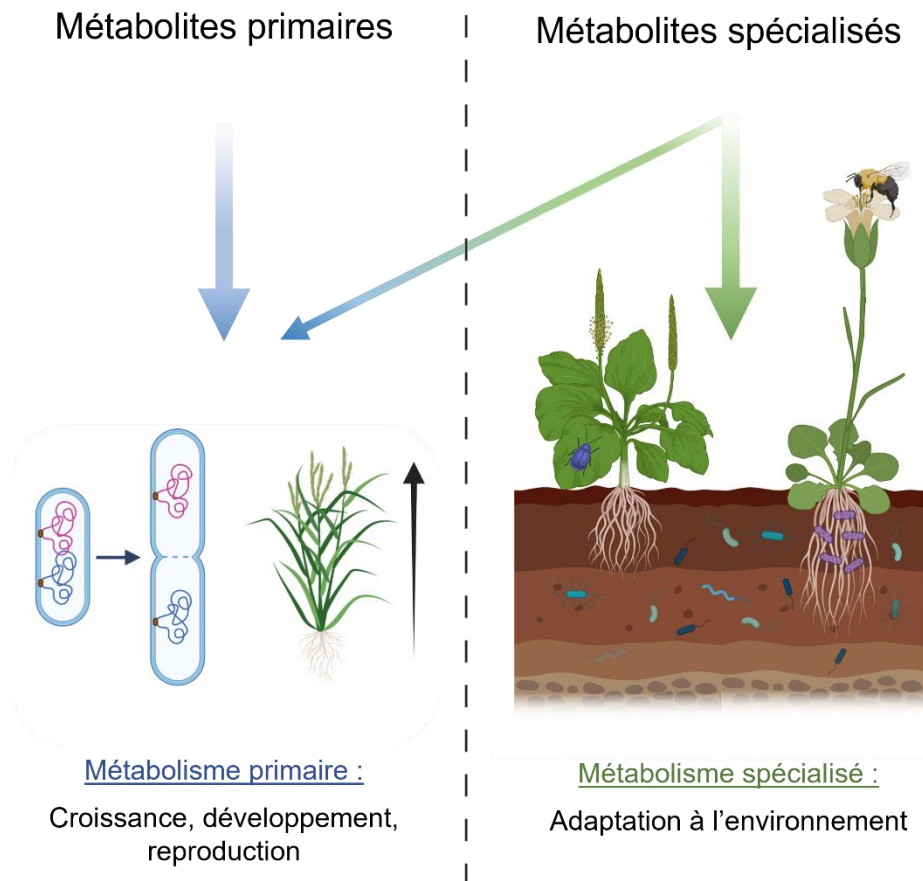


## I. Les polycétides, des métabolites spécialisés aux propriétés variées

### 1. Métabolites spécialisés

Chez tout organisme, qu'il soit uni- ou pluricellulaire, eucaryote, archée ou bactérien, un grand nombre de molécules organiques de faible poids moléculaires, les métabolites, est produit, issu d'un ensemble de réactions biochimiques faisant partie des métabolismes primaire et spécialisé. Les métabolites dits primaires permettent d'assurer les fonctions essentielles à la survie et à la pérennité de l'organisme, à savoir la croissance, le développement et la reproduction de celui-ci. Ces métabolites comprennent notamment les nucléotides et les acides aminés protéinogènes, qui sont les unités de base des acides nucléiques et des protéines, respectivement. Au contraire, les métabolites spécialisés ne sont pas indispensables à la survie des organismes mais leur confèrent un avantage présumé vis-à-vis de leur environnement (**Figure 1**), bien que la fonction écologique de ces molécules soit souvent méconnue. Ces métabolites peuvent jouer un rôle de défense envers d'autres organismes hostiles ou encore servir à la communication entre l'hôte et son environnement <sup>1</sup>. Des exemples notables sont les flavonoïdes et les terpènes qui en fonction du contexte écologique, vont jouer un rôle défensif contre des prédateurs ou au contraire permettre des interactions avec des symbiotes ou animaux pollinisateurs (la plupart des pigments floraux sont des flavonoïdes appartenant aux classes des anthocyanines, aurones et chalcones) <sup>2-4</sup>.

Le terme « spécialisé » rend compte de l'obsolescence de l'opposition entre métabolites primaires et secondaires et décrit plus justement la fonction adaptative essentielle de ces molécules <sup>5</sup> (**Figure 1**). L'implication de métabolites à la fois dans les fonctions primaires et spécialisées a par ailleurs été décrite chez les végétaux <sup>6</sup> et peut être illustrée par le cas de la L-canavanine <sup>7</sup>. Cet acide aminé non protéinogène est retrouvé en très grandes quantités dans les graines de certaines légumineuses (jusqu'à 5% du poids sec des graines de *Canavalia ensiformis*), faisant de ce métabolite la source de stockage d'azote principale, une fonction de toute évidence essentielle de l'organisme. De plus, son analogie de structure avec la L-arginine lui confère des propriétés antimétaboliques, provoquant la synthèse de protéines aberrantes chez les prédateurs ingérant ces graines, un second rôle de défense, une fonction adaptative.



**Figure 1 :** Implication des métabolites primaires et spécialisés dans les fonctions biologiques des organismes qui les synthétisent. Si les métabolites primaires sont essentiels aux fonctions primordiales du vivant, les métabolites spécialisés, eux, permettent l'adaptation à l'environnement et les interactions avec les organismes qui le composent. Dans certains cas, ces derniers peuvent s'avérer également impliqués dans le métabolisme primaire (figure créée avec BioRender.com).

Parmi les divers métabolites spécialisés synthétisés dans les différents règnes du vivant, la classe des polycétides se distingue par la variété remarquable de structures et de propriétés biologiques de ses composés.

## 2. Les polycétides

Lors de ce doctorat, les efforts de recherche se sont concentrés exclusivement sur les enzymes de synthèse de polycétides de type I. Par conséquent, la section qui suit se focalise spécifiquement sur cette classe. La notion de polycétide de type I sera définie ultérieurement (voir la section III.4.).

### a. Une grande diversité fonctionnelle

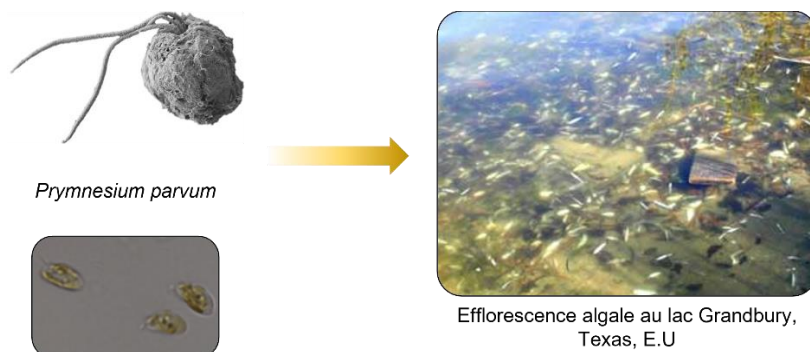
Les polycétides sont des métabolites spécialisés synthétisés par une grande variété

d'organismes, à savoir les bactéries <sup>8</sup>, les champignons <sup>9</sup>, les végétaux <sup>10</sup> et certains animaux <sup>11</sup>. Leurs rôles et synthèses étant étroitement liés à leurs niches écologiques, leurs fonctions pour l'hôte restent difficiles à identifier dans des conditions de laboratoire. Néanmoins, plusieurs travaux ont permis d'éclairer cette zone d'ombre et ont démontré qu'à l'instar des autres classes de métabolites spécialisés évoquées précédemment, les polycétides s'avèrent impliqués dans une grande variété de mécanismes biologiques permettant ainsi de faire face aux diverses pressions environnementales.

### i. Toxines

Certains polycétides sont de puissantes toxines qui permettent la défense chimique de l'organisme face aux menaces prédatrices qui l'entourent. C'est notamment le cas des bryostatines. Ces molécules sont retrouvées en concentrations élevées dans les larves du bryzoaire *Bugula neritina*, bien qu'il ne les synthétise pas lui-même (voir la section I.2.a.v. sur les relations symbiotiques). Ces larves possèdent un corps mou et représentent un apport en protéines important pour leurs prédateurs mais ces fortes concentrations en bryostatines les rendent non-comestibles, les protégeant ainsi. Le bryzoaire va progressivement évoluer vers un stade adulte dans lequel la concentration en bryostatines va fortement diminuer, une diminution corrélée avec des concentrations internes grandissantes en chitine et carbonate, suffisantes pour assurer une défense physique <sup>12</sup>.

Les polycétides toxiques peuvent également permettre aux organismes qui les synthétisent de s'attaquer à d'autres organismes. C'est le cas de la prymnésine, polycétide biosynthétisée par l'algue unicellulaire *Prymnesium parvum*. L'excrétion de cette toxine permet d'immobiliser d'autres organismes planctoniques dont l'algue se nourrit. Lors des événements d'efflorescence algale, c'est-à-dire lorsque la concentration en *P. parvum* devient très importante, l'excrétion massive de prymnésine peut même causer la mort de poissons par milliers <sup>13</sup> (**Figure 2**).



**Figure 2 :** La microalgue *Prymnesium parvum* visualisée par microscopie électronique à balayage (en haut) et microscopie optique (en bas) (images issues de Wagstaff *et al.*, 2018 <sup>14</sup>) provoque la mort de nombreux poissons due à l'excrétion de prymnésine. Cette excrétion massive survient lors d'événements d'efflorescences algales à l'origine de la couleur jaunâtre caractéristique de l'eau (à droite), d'où son surnom d' « algue dorée » (photo par

Joan Glass, Texas Parks & Wildlife).

Finalement, les antibiotiques constituent une autre catégorie de toxines et représentent une part importante des polycétides identifiés à ce jour. Des exemples notables incluent l'érythromycine et la rifamycine dont les propriétés antibactériennes ont entraîné leur utilisation pour le traitement de maladies infectieuses <sup>15</sup>. Néanmoins, ces propriétés pharmacologiques sont considérées d'un point de vue anthropocentré et ne reflètent pas nécessairement le rôle écologique de ces polycétides. En effet, une hypothèse suggère que la sécrétion des antibiotiques dans l'environnement ne permet pas d'atteindre des concentrations entraînant l'inhibition de croissance ou la mort des microorganismes compétiteurs <sup>16</sup>. Il est proposé qu'à des concentrations sous-inhibitrices, ces molécules jouent un rôle de messagers chimiques, provoquant des changements métaboliques et physiologiques chez les organismes ciblés <sup>17,18</sup> (voir la section I.2.a.ii.).

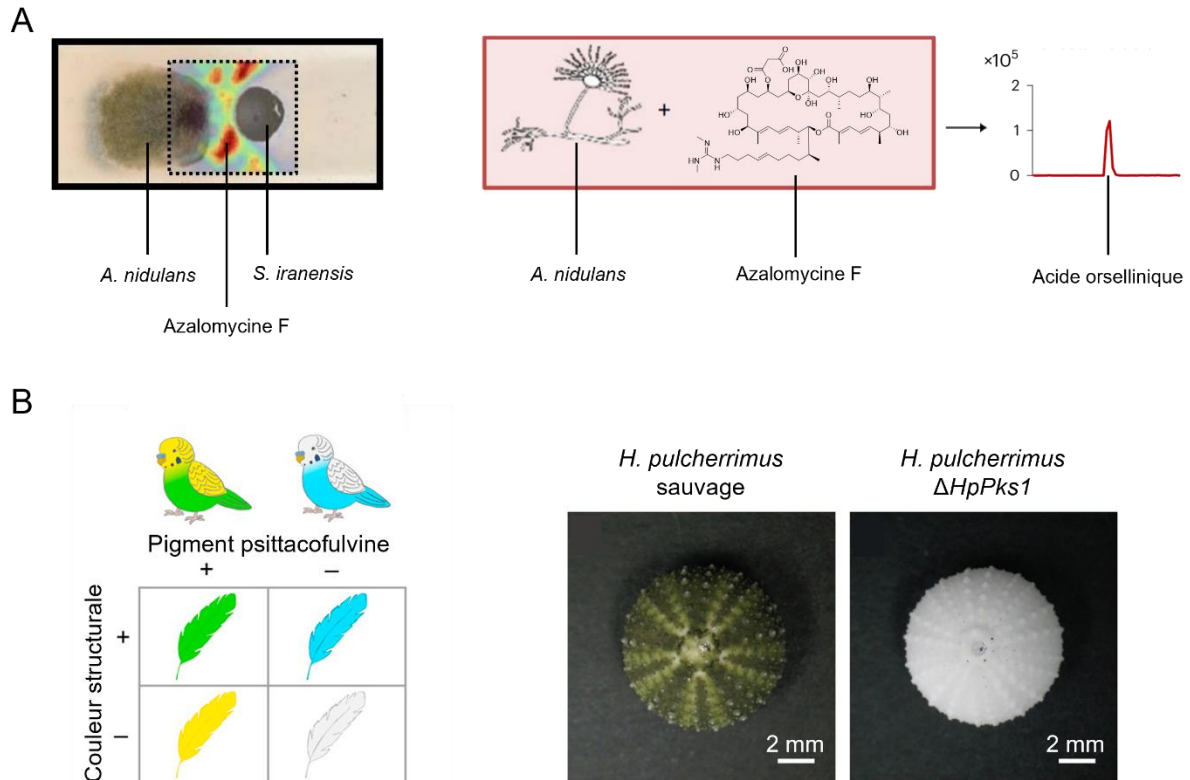
### ii. Communication

Si les polycétides jouent parfois un rôle de défense, ils peuvent également jouer un rôle antagoniste de communication au sein des communautés de microorganismes. Une classe de polycétides, les arginocétides, décrit particulièrement bien cette fonction. Ces molécules sont produites par des bactéries filamenteuses du genre *Streptomyces*, et il a été montré que ces polycétides servent de messagers moléculaires inter-organismes. En effet, l'azalomycine F, un arginocétide synthétisée par *S. iranensis*, est sécrétée en grandes quantités en présence du champignon *Aspergillus nidulans* et fait office de signal pour ce dernier, activant la transcription des gènes impliqués dans la synthèse d'acide orsellinique et de ses dérivés acide lecanorique, F-9775A et F-9775B <sup>19</sup> (**Figure 3**). Cette relation médiée par les arginocétides a été observée entre des *Streptomyces* divers (*S. rapamycinicus*, *S. libani*, *S. mashuensis*, ...) et des champignons phylogénétiquement distants (*A. nidulans*, *A. fumigatus*, *Penicillium* spp.) démontrant à quel point ce langage chimique est répandu et son importance dans l'organisation des écosystèmes microbiens. De plus, *S. rapamycinicus* et *A. nidulans* partagent le même environnement ce qui rend pertinente la communication chimique mise en évidence en laboratoire entre ces deux microorganismes <sup>20</sup>.

La fonction pigmentaire de certains polycétides peut également être considérée comme un moyen de communication inter-organisme, jouant un rôle de signal visuel. Chez la perruche ondulée (*Melopsittacus undulatus*), par exemple, la couleur jaune du plumage est due à la synthèse de pigments polycétidiques psittacofulvines, une synthèse inactivée chez la perruche ondulée bleue (une couleur qualifiée ici de « structurale ») en raison d'une mutation génomique d'un seul nucléotide <sup>21</sup> (**Figure 3**). Ce rôle de pigment a également été démontré chez l'oursin pourpre *Strongylocentrotus purpuratus* et l'oursin *Hemicentrotus*



*pulcherrimus*<sup>22,23</sup>. La délétion des clusters de gènes responsables de la synthèse du polycétide echinochrome dans ces deux espèces, respectivement *SpPks* et *HpPks1*, a empêché l'accumulation de ce pigment provoquant la croissance d'embryons et d'individus albinos (**Figure 3**).



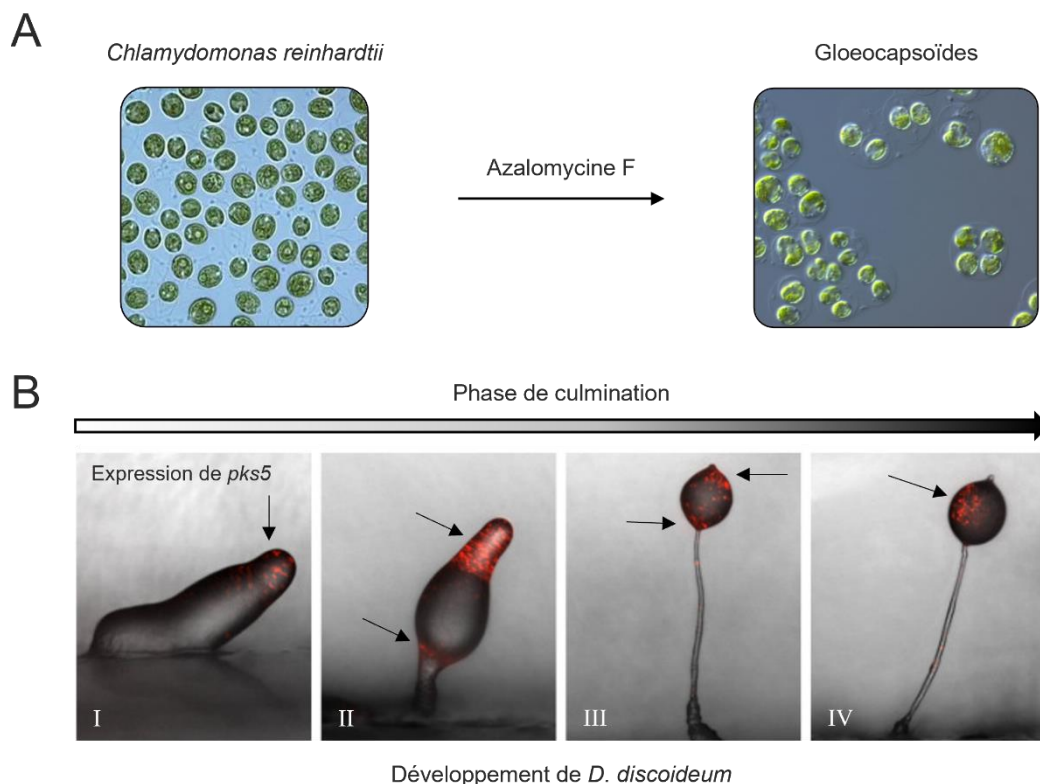
**Figure 3 :** Mécanismes de communication médiés par les polycétides. A) La coculture de *A. nidulans* et *S. iranensis* provoque la synthèse d'azalomycine F chez ce dernier (à gauche) et l'excrétion de ce polycétide entraîne la synthèse d'acide orsellinique chez *A. nidulans* (à droite) (figure adaptée de Krespach *et al.*, 2023<sup>19</sup>). B) La synthèse de psittacofulvine est responsable des couleurs jaunes et vertes de la perruche ondulée (à gauche) (figure reproduite de Cooke *et al.*, 2017<sup>21</sup>) tandis que chez l'oursin *H. pulcherrimus*, l'abolition de la synthèse du polycétide echinochrome provoque la croissance d'individus albinos (figure reproduite de Liu *et al.*, 2019<sup>23</sup>).

### iii. Régulation de processus biologiques

Certains processus biologiques complexes sont déclenchés ou contrôlés par la présence de polycétides. L'azalomycine F évoquée précédemment est toxique pour l'algue unicellulaire *Chlamydomonas reinhardtii*. À des concentrations sous-léthales, elle provoque l'agrégation de l'algue en structures protectrices appelées gloeocapsoides<sup>24</sup> (**Figure 4**), la multicellularité facilitant la protection de l'organisme aux stress environnementaux. De manière similaire, la linearmycine est un arginocétide toxique pour la bactérie *Bacillus subtilis*. Néanmoins, certaines souches de *B. subtilis* ont développé une résistance à ce polycétide en adoptant également un comportement multicellulaire, sous la forme cette fois-ci de biofilm<sup>25</sup>. La pertinence de ces deux mécanismes reste toutefois à démontrer dans l'environnement natif

de ces microorganismes puisqu'en présence du champignon *A. nidulans*, l'algue *C. reinhardtii* forme non pas des gloeocapsoïdes mais une association de type lichen avec cette espèce, se protégeant ainsi des bactéries potentiellement dangereuses <sup>26</sup>. De plus, comme évoqué dans le cas des antibiotiques, il est difficile d'affirmer que les concentrations des polycétides dans l'environnement sont compatibles avec la mise en place de ces mécanismes de pluricellularité observés en conditions de laboratoire.

Dans d'autres cas de régulations de processus biologiques, le polycétide est bénéfique pour l'organisme et est directement synthétisé par celui-ci. C'est le cas des dictyodènes produits par l'amibe *Dictyostelium discoideum*. Ces molécules agissent comme supprimeurs de germination des spores permettant de contrôler précisément l'évènement de dissémination des spores, essentiel à la reproduction de l'organisme <sup>27</sup>. La synthèse de ces composés est par conséquent précisément régulée dans le temps et l'espace (**Figure 4**). Dans une autre amibe, *Polysphondylium violaceum*, l'abolition de la synthèse du polycétide 4-méthyl-5-pentylbenzene-1,3-diol provoque une diminution de la capacité d'agrégation cellulaire nécessaire à la croissance de l'amibe, démontrant l'implication de ce polycétide dans ce processus biologique <sup>28</sup>.



**Figure 4 :** Implication de polycétides dans le déclenchement de processus biologiques complexes. A) L'arginocétide azalomycine F induit la formation de gloeocapsoïdes de *C. reinhardtii* pour se protéger de sa toxicité (d'après Krespach *et al.*, 2021 <sup>24</sup>). B) La biosynthèse de dictyodènes est régulée de manière spatiotemporelle pour permettre le développement de *D. discoideum*. L'expression du gène *pks5* codant l'enzyme de synthèse des dictyodènes est observée à différents stades de développement de l'amibe lors de la phase de culmination. À l'état

de limace (I), ces polycétides sont produits majoritairement à l'extrémité de l'organisme. Leur expression est ensuite localisée dans les régions inférieures et supérieures du sore contenant les futurs spores (II). Lorsque la fructification devient mature (III), l'expression de *pks5* reste localisée dans les deux mêmes régions avant de gagner l'ensemble du sore lors de son vieillissement (IV). Les différentes régions d'expression de *pks5* sont indiquées par des flèches noires (figure adaptée de Günther *et al.*, 2022<sup>27</sup>).

#### iv. Composants membranaires

Les membranes cellulaires sont constituées d'un mélange complexe de biomolécules, avec notamment différents lipides organisés en bicouche. L'incorporation d'acides gras polyinsaturés (PUFA) dans ces membranes est essentielle au maintien de leur fluidité, propriété cruciale pour leurs fonctions biologiques. L'origine biosynthétique des PUFA chez les animaux, plantes et eucaryotes unicellulaires est connue de longue date. Ces lipides sont synthétisés via la voie de synthèse classique des acides gras conjuguée à l'action de désaturases et élongases. Néanmoins, il a été mis en évidence que certains microorganismes tels que des myxobactéries<sup>29</sup>, des bactéries psychrophiles<sup>30</sup> et des algues unicellulaires<sup>31</sup> étaient également capables de cette synthèse en ayant recours à une autre voie biosynthétique semblable à celle des polycétides<sup>32</sup>, étendant de ce fait le catalogue des fonctions assurées par les polycétides.

#### v. Relations symbiotiques

Il est à noter que si les organismes pour lesquels la production des polycétides est bénéfique synthétisent fréquemment eux-mêmes ces métabolites, il est également commun d'observer des relations symbiotiques qui permettent à l'hôte de tirer parti des polycétides produits par l'organisme symbiote. Hill *et al.* avancent même que l'un des environnements les plus propices à la découverte de nouveaux polycétides bactériens serait les eucaryotes (notamment les insectes) auxquels s'associent ces bactéries symbiotiques<sup>33</sup>.

En effet, dans le cas du bryzoaire *Bugula neritina* (voir la section 1.2.a.i.), les bryostatines ne sont pas directement synthétisés par l'organisme défendu mais par l'endosymbiote bactérien *Candidatus Endobugula sertula*<sup>12</sup>. Une symbiose similaire qui décrit parfaitement le mutualisme de ce type de relation est observée entre le tunicier *Lissoclinum* sp. et la bactérie *Candidatus Didemnitutus mandela*. Cette dernière bénéficie des ressources de son hôte, ce qui a entraîné une importante réduction de la taille de son génome, n'ayant plus besoin de produire la majorité de ses métabolites elle-même. Néanmoins, son génome sommaire ne contient pas moins de sept copies du cluster de gènes biosynthétiques (BGC) impliqué dans la synthèse des polycétides cytotoxiques mandelalides, constituant ainsi une défense chimique importante pour son hôte<sup>34</sup>.

## b. Une grande diversité structurale

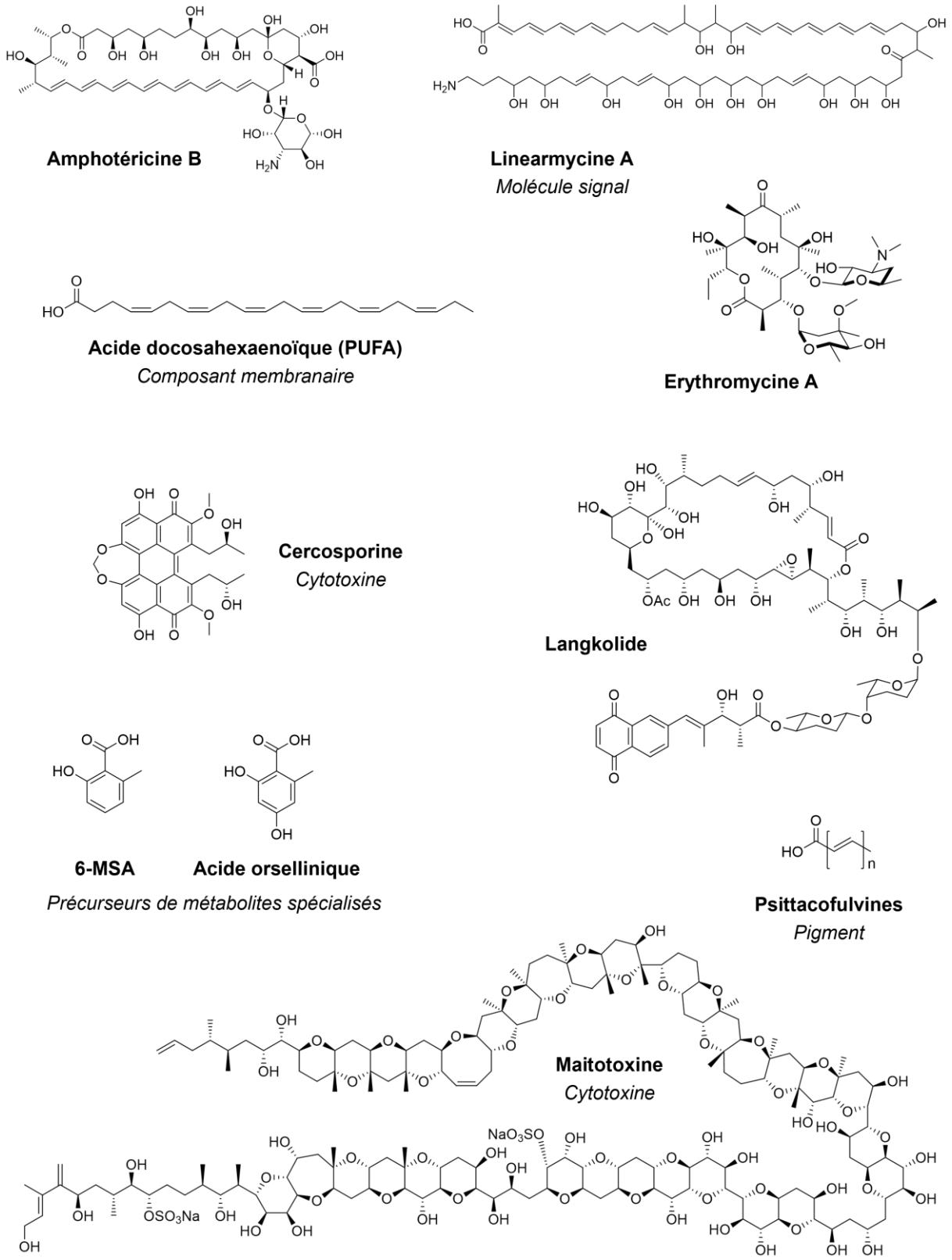
Cet éventail de propriétés biologiques est le reflet de la grande hétérogénéité de propriétés physicochimiques des polycétides de type I <sup>35</sup> (**Figure 5**).

La taille des polycétides peut varier de moins de dix atomes à plusieurs milliers de daltons (Da). Le premier polycétide étudié d'un point de vue biosynthétique chez *Penicillium griseofulvum*, l'acide 6-méthylsalicylique (6-MSA) <sup>36</sup>, ne contient que 8 atomes de carbone. De façon similaire, l'acide orsellinique, synthétisé entre autres par le champignon *Aspergillus nidulans* <sup>20</sup>, et l'acide triacétique lactone (TAL) synthétisé par la plante *Gerbera hybrida* <sup>37</sup> contiennent uniquement huit et six atomes de carbone, respectivement. Au contraire, la microalgue *Gambierdiscus toxicus* est responsable de la biosynthèse d'une des toxines les plus puissantes connues à ce jour, la maitotoxine, composée de pas moins de 32 cycles et possédant une impressionnante masse de 3422 Da <sup>38</sup>.

Un grand nombre de polycétides présente une structure aromatique très rigide, à l'image de la cercosporine (isolée à partir de *Cercospora kikuchii* <sup>39</sup>). À l'inverse, d'autres polycétides réduits possèdent une structure bien moins contrainte telle que celle de la linearmycine, constituée d'une chaîne carbonée linéaire, ou la structure des membres de la classe des macrolides, tels que la langkolide <sup>40</sup>, formés par cyclisation de leurs précurseurs. Le macrolide amphotéricine B (isolé à partir de *Streptomyces nodosus* <sup>41</sup>) illustre également la diversité chimique des polycétides par sa structure polyinsaturée, caractéristique des polyènes.

Les PUFA sont des polycétides fortement réduits, chimiquement simples puisque uniquement constitués d'une fonction carboxylate et d'une chaîne carbonée polyinsaturée <sup>32</sup>. À l'inverse, d'autres polycétides possèdent une richesse chimique importante et sont structurellement complexifiés par des modifications installées après synthèse de la structure principale du polycétide comme dans le cas de l'érythromycine A, isolée à partir de *Saccharopolyspora erythraea* <sup>42</sup>.

Cette multitude de caractéristiques structurales est à l'origine de la diversité des phénomènes biologiques décrits précédemment dans lesquels sont impliqués les polycétides et en font des molécules bioactives particulièrement attractives d'un point de vue pharmaceutique.



**Figure 5 :** Exemples de polycétides illustrant la grande diversité structurale et chimique de ces métabolites. Leurs fonctions écologiques sont indiquées lorsqu'elles sont connues.

### c. Propriétés biologiques et intérêt pharmacologique

En plus de leurs propriétés écologiques indispensables à la structuration des

écosystèmes, les polycétides, et plus globalement les métabolites spécialisés, ont démontré leur grande utilité pharmaceutique et continuent toujours à solliciter les efforts de recherche dans la course à la découverte et à la conception de nouvelles molécules bioactives <sup>43,44</sup>.

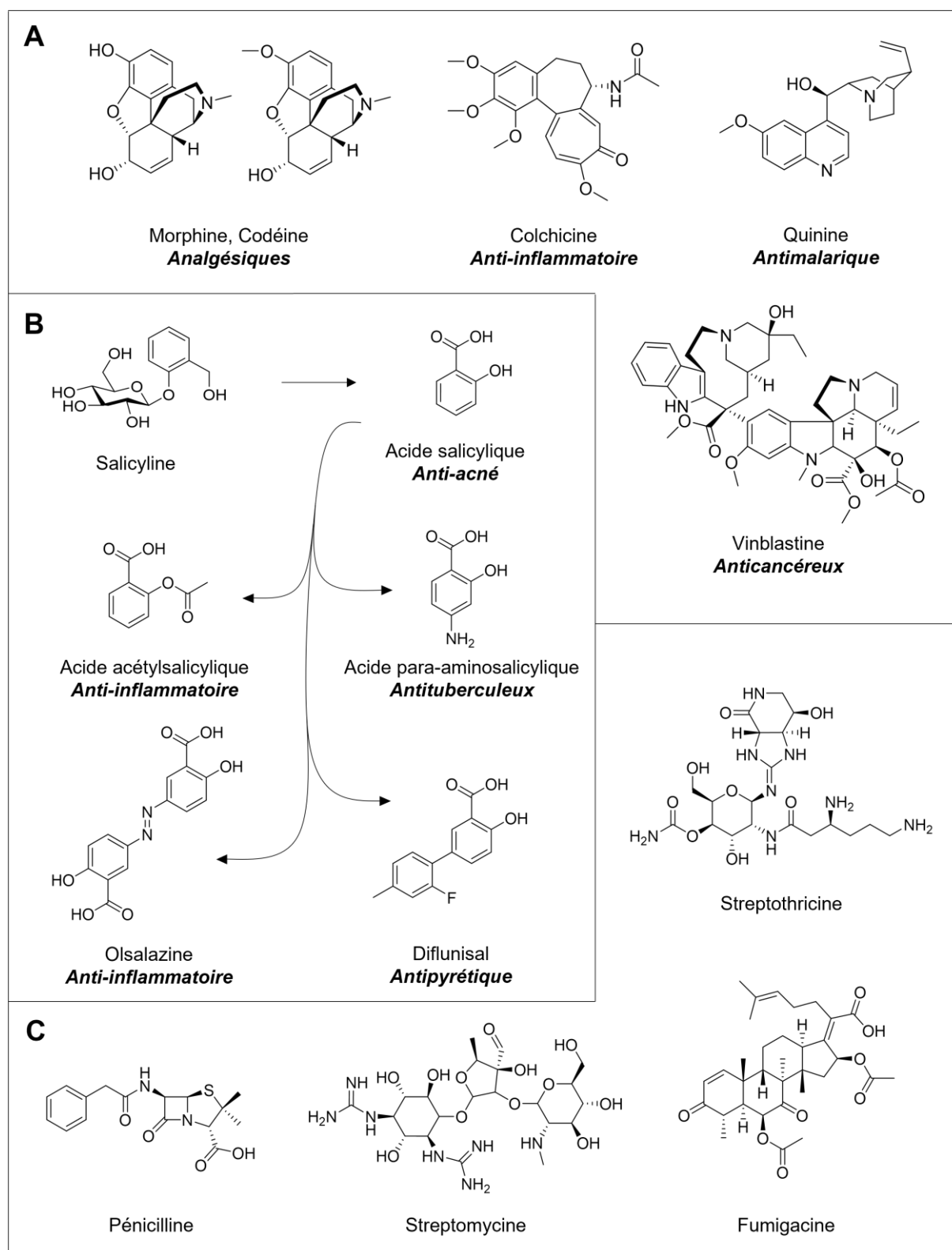
### i. Les métabolites spécialisés, source de composés bioactifs

L'origine de l'exploitation de métabolites spécialisés à des fins médicales est difficile à dater tant ils ont été utilisés par de nombreuses civilisations au cours des siècles. Les écrits les plus anciens décrivant ces pratiques sont datés de 2600 ans avant notre ère et documentent l'utilisation de myrrhe, de pavot, de réglisse et d'huiles de cyprès et de cèdre, substances d'origine végétale encore employées de nos jours en cas de rhume, inflammation ou infection parasitaire, par exemple <sup>45</sup>. Les médecines traditionnelles chinoises et indiennes ont largement eu recours aux métabolites spécialisés végétaux pendant plusieurs milliers d'années par le biais de plantes médicinales <sup>46,47</sup>, également omniprésentes dans les médecines traditionnelles de nombreuses cultures à travers le globe et les âges <sup>48-52</sup>.

À partir du XIX<sup>e</sup> siècle, les premières études impliquant la découverte, l'isolation et la caractérisation de métabolites bioactifs marquèrent des étapes importantes et des avancées majeures dans le domaine pharmacologique <sup>53</sup>. En 1805, Friederich Wilhelm Sertürner isolait pour la première fois le composé actif du pavot à opium *Papaver somniferum* avant d'en caractériser plusieurs propriétés chimiques et biologiques une dizaine d'années plus tard <sup>54</sup>. Il nomma cette molécule « morphine » (**Figure 6**) et découvrit ainsi le premier alcaloïde, dont la liste de composés ne tarda pas à s'agrandir lors des décennies suivantes <sup>55</sup>. Près de cent ans plus tard, en 1897, les chimistes allemands Felix Hoffmann et Arthur Eichengrün accomplissaient la synthèse d'acide acétylsalicylique à partir de salicyline (**Figure 6**) extraite de l'écorce du saule *Salix alba*, molécule ensuite commercialisée pour ses propriétés anti-inflammatoires et analgésiques sous le nom « Aspirine » <sup>56</sup>.

Au début du siècle suivant, l'un des accomplissements les plus importants dans le domaine des produits naturels est incontestablement l'heureuse découverte de la pénicilline (**Figure 6**) par Alexander Fleming en 1928 <sup>57</sup>. Les propriétés antibiotiques de cette molécule et l'origine fongique de sa biosynthèse (*Penicillium notatum*) ouvrirent alors la voie à la recherche de métabolites bioactifs microbiens. Le décernement du prix Nobel de médecine à Sir Fleming en 1945 souligna encore davantage l'importance fondamentale de cette découverte pour les domaines de la médecine et de la microbiologie. Cette effervescence provoquée par l'isolation de molécules bioactives extraites de microorganismes conduisit notamment quelques années plus tard l'équipe du Dr. Selman Waksman à contribuer de manière importante à l'isolation de nouveaux composés, marquant le début d'une période d'une vingtaine d'années de recherche prolifique reconnue comme l'âge d'or de la découverte

d'antibiotiques <sup>58</sup>. Cette équipe fut à l'origine de la découverte de nombreux antibiotiques <sup>59</sup> (**Figure 6**) isolés de différents champignons et *Streptomyces* (le nom de ce genre d'actinomycètes ayant d'ailleurs été attribué par Waksman lui-même <sup>60</sup> en collaboration avec le Dr. Arthur Henrici). De 1939 à 1943, ses doctorants isolèrent ainsi successivement la tyrothricine <sup>61</sup> (René Dubos), l'actinomycine <sup>62</sup> (Boyd Woodruff), la fumigacine et la patuline <sup>63</sup> (Elizabeth Horning), la streptothricine <sup>64</sup> (Boyd Woodruff) et la streptomycine <sup>65</sup> (Albert Schatz). Si les propriétés antibiotiques des cinq premiers composés évoqués furent clairement démontrées, leur toxicité animale empêcha toutefois leur utilisation pharmacologique. La streptomycine en revanche se révéla sûre pour la santé humaine et efficace notamment dans le traitement de la tuberculose <sup>66</sup>, ce qui vaudra en 1952 le prix Nobel de médecine à Waksman.

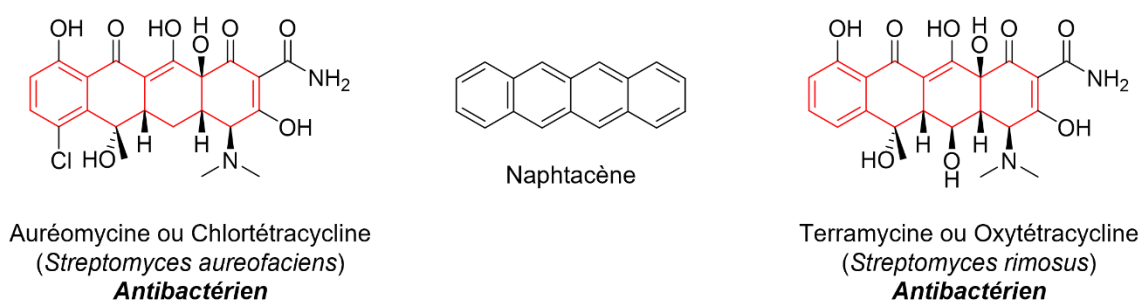


**Figure 6 :** Quelques métabolites spécialisés exploités pour leurs propriétés biologiques, indiqués en gras et italique. A) Exemples d'alcoïdes, métabolites d'origine végétale, et leurs fonctions biologiques d'intérêt. B) Acide salicylique et son précurseur biologique, la salicyline, ainsi que plusieurs dérivés synthétiques bioactifs développés au cours du XX<sup>e</sup> siècle<sup>67</sup>. C) Divers antibiotiques d'origine microbienne identifiés lors de la période de recherche intensive de nouveaux composés initiée par la découverte de la pénicilline.



## ii. Les polycétides en médecine

Les polycétides ne tardèrent pas à dévoiler également leur grand potentiel pharmacologique. En 1948, Benjamin Minge Duggar, alors à la tête du département de criblage des sols de la société pharmaceutique *Lederle Laboratories*, publia ses résultats sur l'identification de colonies bactériennes d'une couleur jaune inhabituelle et au potentiel antibiotique prometteur <sup>68</sup>. Il isola le composé responsable de ces propriétés qu'il nomma « aureomycine » en référence à la couleur dorée des souches de *Streptomyces* d'intérêt (*S. aureofaciens*), composé qui fut approuvé la même année par la *Food and Drug Administration* américaine pour un usage thérapeutique en raison de son action antibiotique puissante efficace contre de nombreuses maladies infectieuses <sup>69</sup>. Deux ans plus tard, un composé aux propriétés similaires mais présentant une meilleure bioactivité était extraite d'une autre bactérie isolée d'échantillons de sols (*S. rimosus*), la terramycine <sup>70</sup>. Deux années supplémentaires furent nécessaires pour élucider la structure de ces deux molécules <sup>71</sup>, révélant ainsi leur caractéristique commune, un cœur naphtacène, structure tétracyclique qui donnera son nom à cette première classe d'antibiotiques polycétidiques, les tétracyclines (**Figure 7**).



**Figure 7 :** Exemples de polycétides tétracyclines exploités par l'industrie pharmaceutique. Structures chimiques de l'auréomycine (à gauche) et de la terramycine (à droite). Leur structure tétracyclique commune est représentée en rouge et le squelette naphtacène des tétracyclines est représenté au centre.

La découverte des macrolides constitue une deuxième étape clef dans l'utilisation médicale des polycétides. Cette classe de molécules se caractérise par un cycle macrolactone (**Figure 8**) et fut également mise en lumière lors de l'âge d'or de la découverte des antibiotiques. Le premier macrolide identifié fut la pikromycine en 1950 <sup>72</sup> mais la découverte de l'érythromycine A deux ans plus tard <sup>42</sup> marqua réellement un progrès substantiel dans l'utilisation pharmaceutique des macrolides. Isolée à partir de cultures de *Saccharopolyspora erythraea* (anciennement *Streptomyces erythraeus*), l'érythromycine A est constituée d'un cœur macrolactone à 14 atomes substitué par deux monosaccharides, un mycarose en position C3 et un cladinose en position C5. L'activité antibactérienne importante de cette classe d'antibiotiques contre les espèces Gram-positives mais également contre certaines espèces

Gram-négatives firent et font encore actuellement de l'érythromycine et de ses dérivés (voir la section II.2.) des molécules de choix dans le traitement de nombreuses infections incluant pneumopathies <sup>73,74</sup>, sinusites ou encore certaines maladies sexuellement transmissibles.

Le grand intérêt pharmacologique des macrolides est également dû à la large gamme de leurs propriétés biologiques allant bien au-delà de l'unique activité antibactérienne <sup>15</sup>. En effet, en 1950 déjà, Rachel Brown et Elizabeth Hazen découvrirent le premier macrolide polyène antifongique, la nystatine (hexaène) <sup>75</sup>, suivi 3 années plus tard par la découverte des amphotéricines. L'amphotéricine B (heptaène) (**Figure 8**) retint plus particulièrement l'attention en raison de son activité antifongique plus performante que celle de la nystatine ou de l'amphotéricine A, justifiant que ce macrolide soit encore aujourd'hui un traitement de référence en cas d'infection fongique <sup>76</sup>.

En 1973, l'équipe du biochimiste et microbiologiste Satoshi Ōmura, isola plusieurs macrolides nommés avermectines (**Figure 8**) présentant une activité antiparasitaire anthelminthique hautement spécifique <sup>77</sup>. L'ivermectine associant deux dérivés hydrogénés d'avermectines (B<sub>1a</sub> et B<sub>1b</sub>), a montré son utilité dans les domaines agricoles et vétérinaires avant de se montrer particulièrement efficace en médecine humaine, notamment dans le traitement de l'onchocercose <sup>78</sup>, ce qui vaudra en 2015 le prix Nobel de médecine au professeur Ōmura.

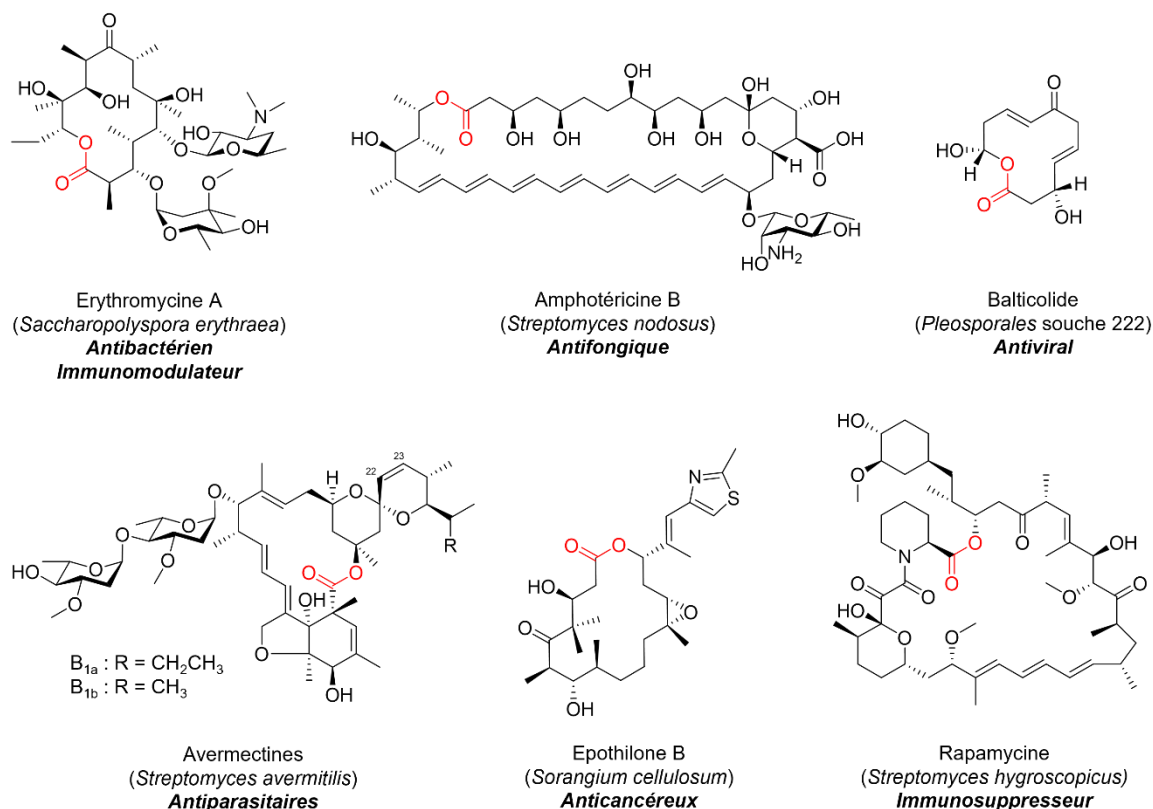
Par la suite, d'autres macrolides ont montré des propriétés immunosuppressives à l'image de la rapamycine (**Figure 8**) (aussi appelée sirolimus) isolée à partir de *Streptomyces hygroscopicus* collecté sur l'île de Pâques <sup>79</sup> et sont par conséquent utilisés dans la prévention de rejet de greffe <sup>80</sup> en raison de leur capacité à inhiber la prolifération des lymphocytes T.

À la fin du XX<sup>e</sup> siècle, Gerth *et al.* identifièrent les épothilones A et B (**Figure 8**) à partir d'une culture de *Sorangium cellulosum* (myxobactérie) initialement pour leurs propriétés antifongiques <sup>81</sup>. Cependant, il fut rapidement mis en évidence que ces macrolides possédaient une puissante activité anticancéreuse, agissant par stabilisation des microtubules <sup>82</sup> à l'image du paclitaxel, molécule utilisée en chimiothérapie. L'épothilone B montra par la suite une activité antiproliférative des cellules cancéreuses plus importante que l'épothilone A et le paclitaxel, et une efficacité maintenue contre des lignées cellulaires paclitaxel-résistantes<sup>83</sup>. Son dérivé amide hémisynthétique, l'ixabepilone, est aujourd'hui utilisé dans le traitement de cancers du sein avancés voire métastatiques <sup>84</sup>.

Plus récemment, en 2011, un macrolide à 12 atomes appelé balticolide (**Figure 8**) a été identifié et a démontré une activité antivirale efficace contre le virus de l'herpès simplex de type 1 <sup>85</sup>.

Par ailleurs, l'érythromycine et ses dérivés sont capables d'influer sur le système immunitaire et sont ainsi qualifiés d'immunomodulateurs, possédant une activité anti-inflammatoire exploitée dans le cadre d'infections provoquées par des bactéries multi-

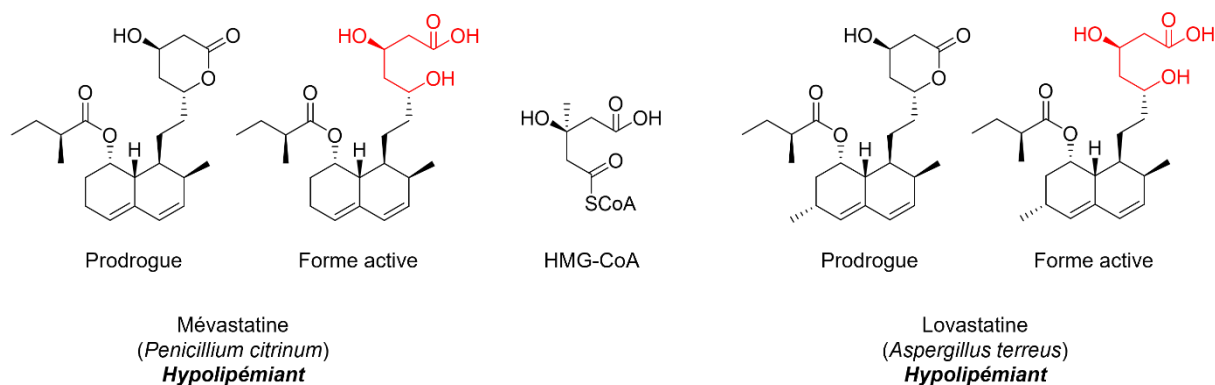
résistantes<sup>86,87</sup> (*P. aeruginosa*) ou dans le cadre d'inflammations chroniques<sup>88</sup> (asthme, mucoviscidose).



**Figure 8 :** Exemples de polycétides macrolides exploités par l'industrie pharmaceutique. La fonction ester caractéristique des macrolactones est représentée en rouge. Les microorganismes dont ont été isolés ces polycétides sont indiqués et leurs fonctions pharmacologiques d'intérêt sont indiquées en gras et italique. Dans le cas des avermectines, la réduction de la double liaison entre les positions 22 et 23 conduit à la formation des ivermectines.

Parmi les métabolites spécialisés d'origine fongique, une autre classe de polycétides se distingue par ses propriétés hypolipémiantes. Ce sont les statines (**Figure 9**). La première statine identifiée, isolée de *Penicillium citrinum* en 1976, est la mévastatine<sup>89</sup> qui montra une importante capacité d'inhibition de la biosynthèse de cholestérol. Cette propriété est due à l'inhibition compétitive de l'enzyme 3-hydroxy-3-méthylglutaryl-coenzyme A (HMG-CoA) réductase vis-à-vis de son substrat. En effet, le HMG-CoA, substrat de la HMG-CoA réductase, présente une similarité structurale avec le cycle lactone de la mévastatine lorsque celui-ci est métabolisé en sa forme acide dans le foie<sup>90</sup> (**Figure 9**). Cette enzyme étant impliquée dans la voie de synthèse du cholestérol, son inhibition entraîne une diminution du taux de lipides circulant dans le sang<sup>91</sup>. Cette découverte incita à la recherche d'autres métabolites inhibiteurs spécifiques de la HMG-CoA réductase, ce qui conduisit quelques années plus tard à l'identification et la caractérisation de la lovastatine (ou monacoline K), issue d'*Aspergillus terreus*<sup>92</sup> et *Monascus* spp.<sup>93</sup>, commercialisée en 1987 pour sa forte capacité à réduire l'hypercholestérolémie.

## État de l'art



**Figure 9 :** Premiers polycétides hypolipémiants identifiés, inhibant la HMG-CoA réductase. La métabolisation de la forme lactonique « prodrogue » des statines engendre la forme acide capable d'inhiber compétitivement l'enzyme. La similarité de structure entre les formes acides « actives » des deux statines et le HMG-CoA est indiquée en rouge. Les microorganismes dont ont été isolés ces polycétides sont indiqués et leurs fonctions pharmacologiques d'intérêt sont indiquées en gras et italique.

## II. Stratégies de découverte/synthèse de nouveaux polycétides

La grande variété de structures et fonctions chimiques des polycétides et le large spectre d'activités biologiques qui en découle représentent un intérêt fondamental pour la médecine humaine et vétérinaire et le secteur agricole. Cependant, les années fructueuses de découverte de nouveaux antibiotiques arrivèrent peu à peu à leur terme lors de la deuxième moitié du XX<sup>e</sup> siècle. Les méthodes classiques de criblage d'extraits microbiens en fonction de leur bioactivité couplées à l'incapacité à cultiver de nombreux microorganismes, conduisirent progressivement à une redécouverte fréquente de composés connus. Dans le même temps, l'apparition de résistances bactériennes<sup>94</sup> et la nécessité de développer des molécules avec une plus grande spécificité et de meilleurs paramètres pharmacocinétiques et -dynamiques, ont accru le besoin de mettre au point de nouvelles approches permettant de répondre à ses problématiques. Des méthodologies faisant appel et combinant la bioinformatique, la microbiologie, la chimie de synthèse et l'ingénierie génétique ont ainsi été développées, conduisant à la découverte et à la synthèse de nombreuses molécules bioactives<sup>95,96</sup>.

### 1. La biosphère

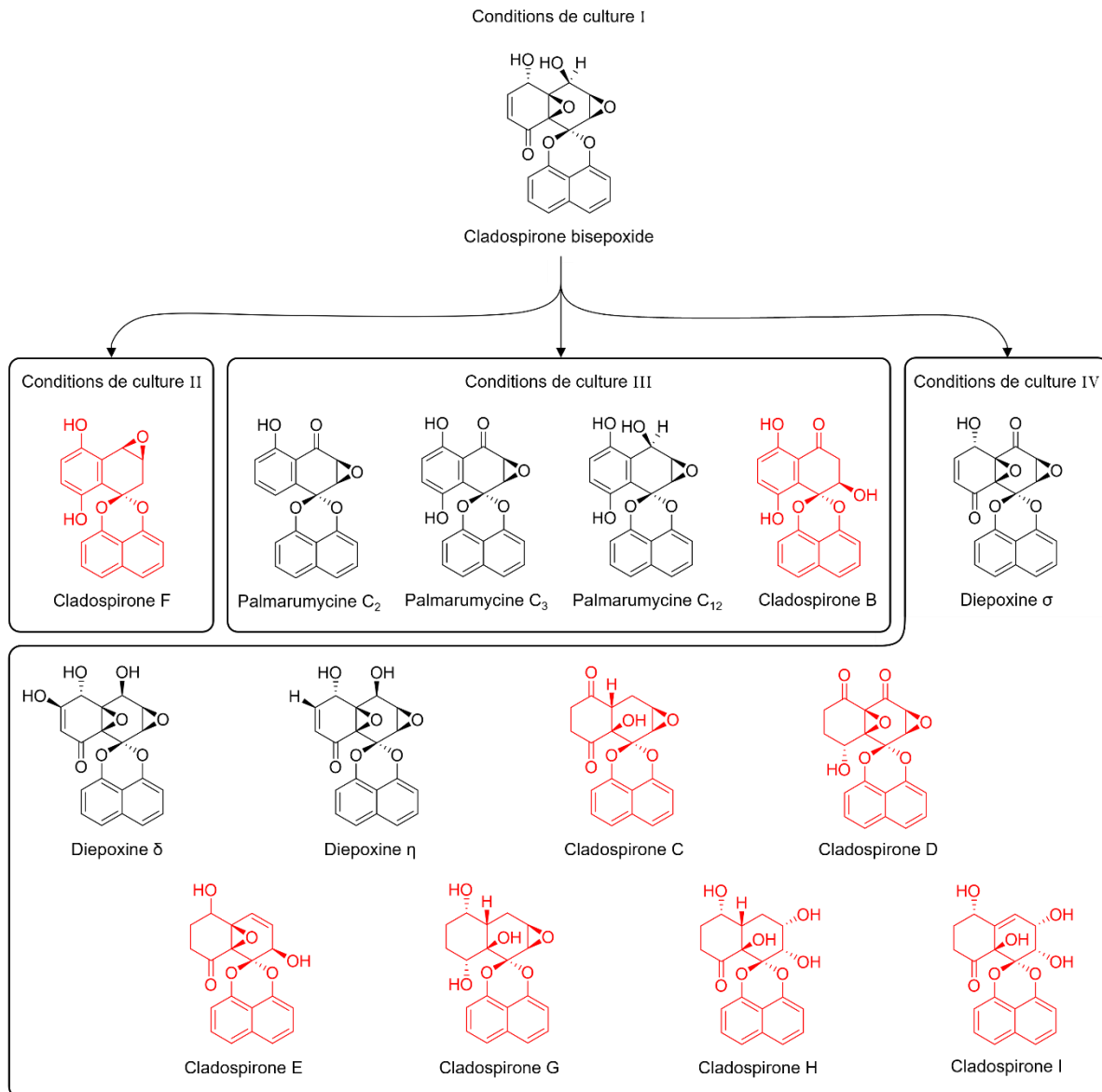
#### a. Exploiter la diversité naturelle

L'approche qui semble la plus évidente et la plus directe pour agrandir la bibliothèque de polycétides disponibles, et plus largement de métabolites spécialisés, est l'identification de ces composés directement au sein des organismes qui les synthétisent. Cette approche a été la première employée et a permis la découverte de tous les polycétides bioactifs décrits en partie I.2.c.ii. La méthodologie, initialement développée par René Dubos<sup>61</sup>, consiste à prélever

des échantillons de sols de diverses provenances, à en cultiver les microorganismes puis à cribler les extraits pour différentes propriétés biologiques d'intérêt. Toutefois, la production de métabolites spécialisés est fortement dépendante de l'environnement des organismes étudiés puisque le rôle de ces molécules de faible poids moléculaire est précisément de fournir un avantage compétitif dans un contexte biologique spécifique (voir la section 1.2.a.). Cela implique que tous les polycétides potentiellement synthétisables ne sont pas produits simultanément, et que la production de l'un ou l'autre est la réponse à des signaux internes et externes à l'organisme <sup>97</sup>. Dans les cas où les conditions d'activation de voies de biosynthèse ne sont pas réunies, il est nécessaire de fournir le stimulus externe approprié.

#### i. Une souche, beaucoup de composés

Variation des conditions de culture peut permettre d'induire ou d'optimiser la production de polycétides par activation ou suractivation de leurs voies de biosynthèse. En effet, les conditions de culture appropriées peuvent apporter les signaux manquants pour provoquer une stimulation de ces voies. L'approche OSMAC ou **One Strain–Many Compounds** repose sur ce principe en essayant de déclencher des changements métaboliques par modifications de paramètres de culture tels que le pH, la température, la composition du milieu de culture, l'oxygénation et le type de contenant (culture sur surface statique, Erlenmeyer, fermenteur,...) <sup>98</sup>. L'application de cette stratégie à la culture de champignons et de bactéries a ainsi permis de caractériser de nouveaux polycétides qui n'étaient pas produits dans des conditions standards de culture ou dans des rendements trop faibles pour en permettre la caractérisation <sup>99–102</sup>. La souche fongique *Sphaeropsidales* sp. F-24'707, par exemple, était décrite comme productrice du polycétide antifongique cladospironone bisepoxide, un métabolite synthétisé dans des conditions de laboratoire riches en oxygène et avec de longs temps d'incubation. En modifiant les paramètres de cultures de ce champignon (composition du milieu et méthode d'incubation), Bode *et al.* furent capables de détecter la production de quatorze polycétides supplémentaires, huit d'entre eux étant des cladospironones identifiées pour la première fois <sup>102</sup> (**Figure 10**). La stratégie OSMAC continue à être utilisée actuellement dans le but de découvrir des métabolites bioactifs, à l'image de la découverte en 2017 de trois nouveaux polycétides de la classe des fusarielines (J–L) produits par *Fusarium tricinctum* <sup>103</sup>. La cultivation de cette espèce dans un milieu à base de riz supplémenté en jus de fruits et légumes a permis de mimer davantage son habitat naturel, *F. tricinctum* étant un champignon endophyte <sup>103</sup>. Toutefois, si l'OSMAC est une approche simple à mettre en place, elle présente une capacité limitée à mimer de manière efficace les niches écologiques des souches cultivées <sup>95</sup>.



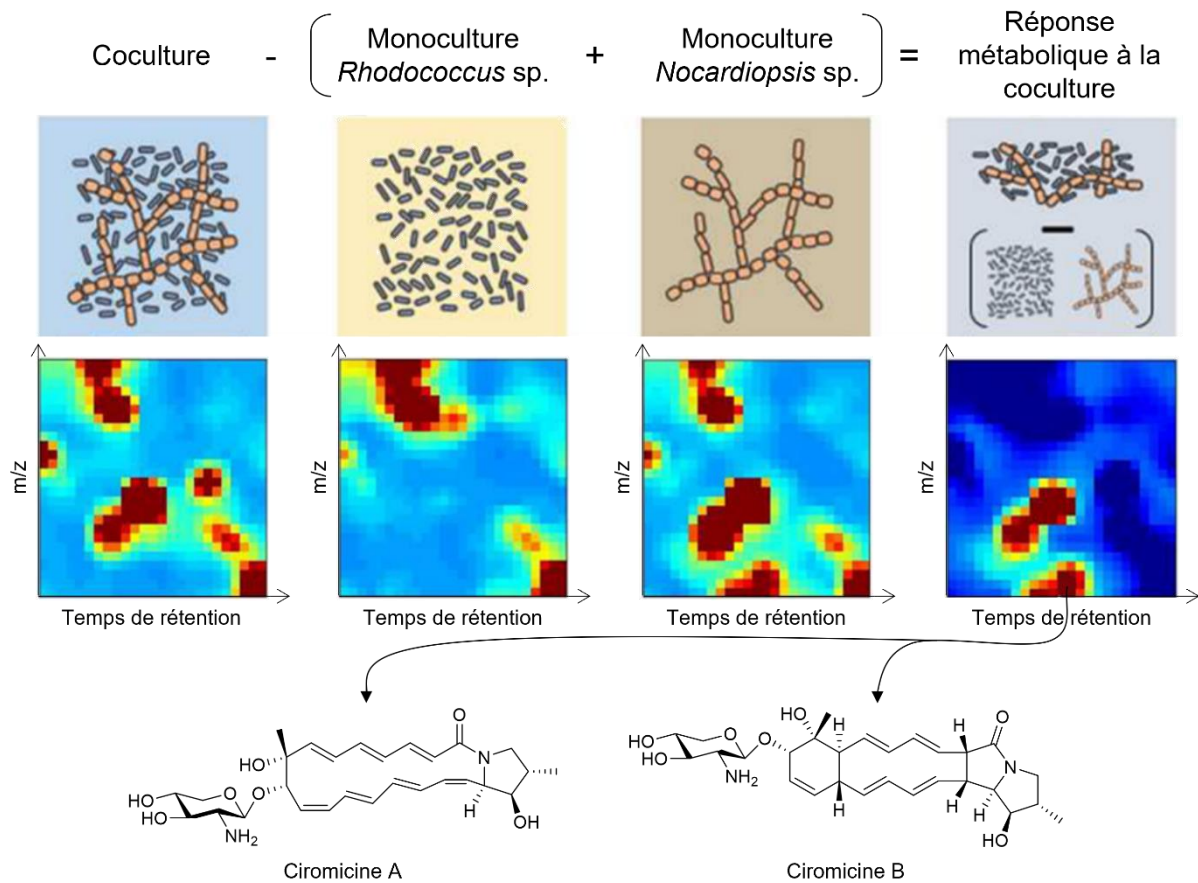
**Figure 10 :** Application de l'approche OSMAC à la culture de *Sphaeropsidales* sp. F-24707<sup>102</sup>. Dans les conditions de culture initialement mises en place, le seul polycétide identifié était la cladospirone bisepoxide. En modifiant le milieu de culture, la méthode et le temps d'incubation, 14 polycétides supplémentaires ont été identifiés dont 8 polycétides caractérisés pour la première fois (représentés en rouge). Conditions de culture I : erlenmeyer à baffles de 350 mL, agitateur rotatif, culture de 100 mL de glucose 2%, flocons d'avoine 2%, farine de soja dégraissée 2%, eau déionisée, incubation 3 jours à 28 °C. Conditions de culture II : erlenmeyer à baffles de 1 L, agitateur unidirectionnel, culture de 150 mL de glycérol 2%, extrait de malt 1%, extrait de levure 0.4%, eau déionisée, incubation 3 jours à 28 °C. Conditions de culture III : flasques de 4 L, culture en surface statique de 1 L de glucose 2%, flocons d'avoine 2%, farine de soja dégraissée 2%, eau déionisée, incubation 28 jours à 28°C. Conditions de culture IV : flasques de 4 L, culture en surface statique de 250 g d'avoine imbibés d'eau déionisée puis inoculés avec 5 mL de préculture, incubation 28 jours à 28°C.

## ii. La coculture

Les microorganismes dans leur environnement naturel ne sont pas des entités isolées mais font au contraire partie de communautés microbiennes complexes au sein desquelles la

communication inter-espèces est permise par les métabolites spécialisés. La reproduction de cette communication en laboratoire par co-culture de différentes espèces microbiennes apparait donc comme une façon d'activer les voies de biosynthèse de ces messagers chimiques <sup>97,104</sup>.

En effet, Derewacz *et al.* ont démontré la puissance de la coculture pour la découverte de métabolites spécialisés en comparant des monocultures bactériennes de *Nocardioopsis* et *Rhodococcus* et la coculture des deux espèces. Ces études ont mis en évidence une différence de 14% entre les métabolomes sommés des monocultures et le métabolome issu de la coculture, une différence concrétisée par la caractérisation de deux nouveaux polycétides, les ciromicines A et B <sup>105</sup> (**Figure 11**). Une telle synergie inter-espèce a également été observée entre un champignon du genre *Pestalotia* isolé de la surface d'algues brunes et une bactérie marine non-identifiée, résistante à divers antibiotiques. La compétition bactérienne induite par la coculture de ces deux souches a conduit à la découverte d'un nouveau polycétide fongique aux propriétés antibiotiques, la pestalone, tandis que la culture individuelle du champignon ou de la bactérie ne permettait pas de détecter ce composé <sup>106</sup>. Cette approche peut éventuellement provoquer des changements métaboliques chez les deux microorganismes cocultivés à l'instar de l'influence mutuelle de *Streptomyces iranensis* et *Aspergillus nidulans*. La présence de *A. nidulans* induit la production d'azalomycine F par *S. iranensis* qui va elle-même induire la production d'acide orsellinique chez *A. nidulans* <sup>19</sup> (voir la section I.2.a.ii.).



**Figure 11 :** Influence de la coculture de *Rhodococcus* sp. et *Nocardioopsis* sp. sur le profil métabolique (figure adaptée de Derewacz *et al.*, 2015<sup>105</sup>). L'étude du métabolome issu uniquement de la coculture a conduit à l'isolation des ciromicines A et B, caractérisées pour la première fois. Le jeu d'images inférieur correspond à des *heat maps* de données de spectrométrie de masse (MS) obtenues par analyse du métabolome des différentes cultures. Ces données correspondent à des ions définis par leurs temps de rétention et leurs valeurs de masse/charge et sont organisées informatiquement de sorte que les ions détectés possédant des profils d'intensité similaires sont localisés dans la même zone.

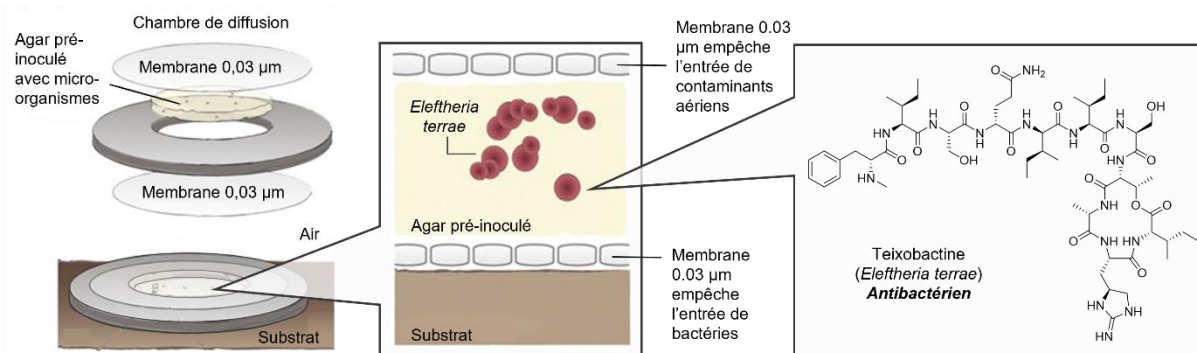
### iii. Culture de microorganismes « non-cultivables »

La co-culture permet de réinstaurer en laboratoire une communication entre organismes mais ne mime qu'une infime partie de la complexité des consortia microbiens naturels et nécessite par essence de pouvoir cultiver les organismes d'intérêt. Or, malgré les améliorations apportées aux méthodes de culture, il a été estimé que seul 1% des bactéries peut être cultivé en laboratoire par des méthodes standards<sup>107</sup>. Cette réalité transparait au travers d'une observation faite il y a plus de 120 ans par Heinrich Winterberg lorsqu'il décrit une incohérence entre le nombre important de cellules dans ses échantillons et le nombre plus restreint de colonies formées sur milieu nutritif à partir de ces mêmes échantillons<sup>108</sup>, un phénomène connu sous le nom de *Great Plate Count Anomaly*. Cette biosphère non-cultivable représente donc une source de nouveaux métabolites spécialisés en apparence inaccessible. Différentes stratégies ont donc été mises au point afin de répondre à cette problématique,



toutes visant à reproduire le plus fidèlement possible l'environnement natif de ces microorganismes qualifiés de non-cultivables <sup>95,109</sup>.

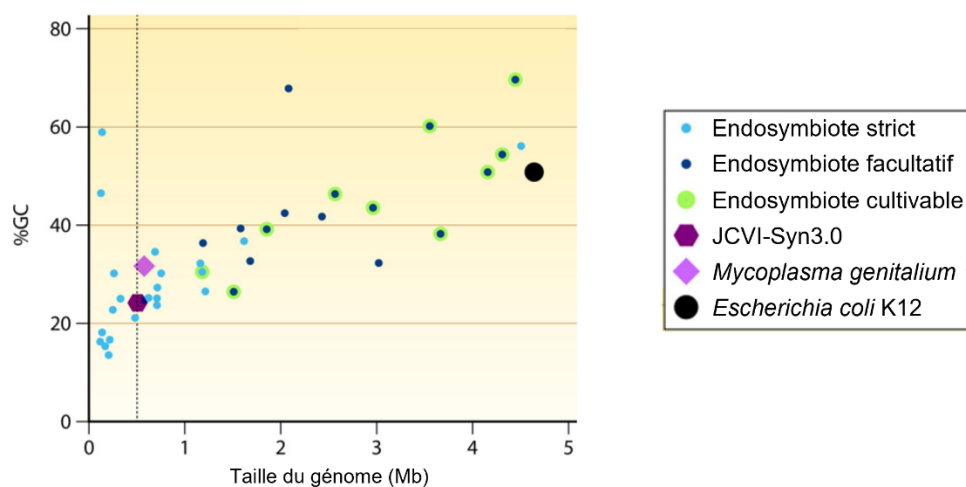
Kaerberlein *et al.* ont ainsi développé une chambre de diffusion dans laquelle un échantillon dilué de divers microorganismes prélevés dans leur habitat naturel et mélangé avec de l'agar est déposé entre deux membranes scellées entre elles par une rondelle en acier. Ce dispositif inoculé est ensuite placé dans l'environnement duquel proviennent les microorganismes et y est incubé pendant plusieurs jours. La taille des pores membranaires (0.03  $\mu\text{m}$ ) ne permet pas aux cellules d'entrer ou sortir de la chambre de diffusion mais est suffisante pour laisser les nutriments et métabolites diffuser de part et d'autre des membranes (**Figure 12**). En utilisant une telle approche, les premières expériences ont révélé que 22% des espèces microbiennes initialement prélevées étaient capables de former des colonies, soit bien plus que par cultivation en boîte de Petri sur milieu synthétique <sup>110</sup>. Une version haut-débit de la chambre de diffusion a ensuite été développée, la *iChip* (*Isolation Chip*), pour simplifier l'isolation et l'inoculation de colonies individuelles <sup>111</sup>. Cette puce constituée de centaines de chambres de diffusion est incubée dans une suspension de cellules diluée permettant l'isolation d'une cellule en moyenne par chambre. La puce est ensuite scellée par des membranes semi-perméables puis replacée dans l'environnement d'origine à la manière d'une chambre de diffusion simple. Cette méthode haut-débit a montré sa puissance avec la croissance de 50% des microorganismes capturés et la concrétisation des efforts de criblage de nouvelles bactéries « non-cultivables » par la découverte d'un puissant antibiotique, la teixobactine, actif principalement contre les pathogènes Gram-positifs <sup>112</sup> (**Figure 12**).



**Figure 12 :** Conception de la chambre de diffusion et application (figure adaptée de Kaerberlein *et al.* <sup>110</sup>). L'utilisation de la stratégie de chambre de diffusion à plus haut-débit au travers de la *iChip* a permis l'isolation et la cultivation de la bactérie « non-cultivable » *Eleftheria terrae* produisant un métabolite aux propriétés antibactériennes, la teixobactine. Il est à noter que la teixobactine n'est pas un polycétide mais un peptide synthétisé de manière non ribosomale (NRP).

La problématique des microorganismes « non-cultivables » représente un défi plus important encore lorsque l'on considère les relations endosymbiotiques puisque le développement de la bactérie ou du champignon d'intérêt va non seulement nécessiter la reproduction de son environnement chimique (nutriments, facteurs de croissance, ...) mais

également des conditions physiques caractéristiques de l'hôte <sup>113</sup>. La taille du génome des endosymbiotes est un facteur pouvant être considéré afin d'estimer la propension de ces microorganismes à pouvoir être cultivés sans leur hôte. En effet, la coévolution d'un endosymbiote avec l'organisme qui l'abrite conduit au fil du temps à une érosion génomique du microorganisme puisque celui-ci n'a plus besoin de produire lui-même tous ces métabolites, à l'image de la bactérie *Candidatus* Didemnitutus mandela et du tunicier *Lissoclinum* sp. <sup>34</sup> (voir la section I.2.a.v.). Il existe donc une corrélation entre la capacité d'un endosymbiote à croître *in vitro* et la taille de son génome, ou autrement dit, le stade d'avancement de l'érosion de son génome indique le niveau de dépendance à son hôte <sup>113</sup> (**Figure 13**). L'information de la taille du génome peut donc faire office de premier crible lors du processus de culture de microorganismes endosymbiotiques (**Figure 13**).



**Figure 13 :** Tailles de génomes de bactéries endosymbiotiques d'insecte (figure reproduite de Masson & Lemaitre, 2020 <sup>113</sup>). Le caractère strict ou facultatif des symbioses est indiqué ainsi que la capacité à cultiver ces symbiotes. Le génome d'*Escherichia coli* K12 est indiqué comme point de comparaison avec une bactérie au développement hôte-indépendant. *Mycoplasma genitalium* est l'endosymbiote cultivable possédant le plus petit génome et JCVI-Syn3.0 représente la limite empirique de taille génomique en dessous de laquelle aucune bactérie n'a été cultivée, indiquée par une ligne pointillée (données allant jusqu'au 11 novembre 2020).

Ces différentes approches permettant la découverte de polycétides dans la biosphère sont des méthodes empiriques qui misent sur le potentiel de divers microorganismes issus de d'environnements variés à synthétiser des composés bioactifs, sans avoir au préalable d'information sur la réelle capacité biosynthétique de ces espèces. Le séquençage et l'analyse des génomes de ces producteurs de métabolites spécialisés est une stratégie complémentaire palliant ce problème.

### **b. Dévoiler le potentiel biosynthétique par *genome mining***

L'étude des voies de biosynthèse des métabolites spécialisés a permis d'en comprendre les particularités tant sur le plan enzymatique que génétique. Notamment dans le

cas des polycétides, l'identification des gènes impliqués dans leurs synthèses est facilitée par le regroupement en cluster de ces gènes. De plus, les enzymes de synthèse qu'ils codent, les polycétide synthases (PKS), sont des protéines possédant souvent un poids moléculaire élevé et une architecture modulaire conservée (voir la section III.4.). Ces caractéristiques permettent ainsi de détecter les BGC d'intérêt au sein des génomes microbiens. Le nombre de génomes séquencés augmentant de manière exponentielle, cette approche d'analyse génomique a mis en lumière une inégalité entre le nombre de BGC identifiés et la quantité restreinte de métabolites spécialisés détectés pour un microorganisme donné <sup>104</sup>. Par conséquent, même si la diversification et l'affinement des conditions de culture permet l'isolation et la caractérisation de polycétides bioactifs <sup>103</sup>, de nombreux gènes biosynthétiques demeurent non- ou sous-exprimés et paraissent être une source inaccessible de composés bioactifs par des méthodes expérimentales empiriques. C'est pourquoi la prédiction et l'isolation de métabolites spécialisés basées sur l'analyse des génomes, ou *genome mining*, constituent une approche puissante pour agrandir la bibliothèque de polycétides caractérisés <sup>114</sup>.

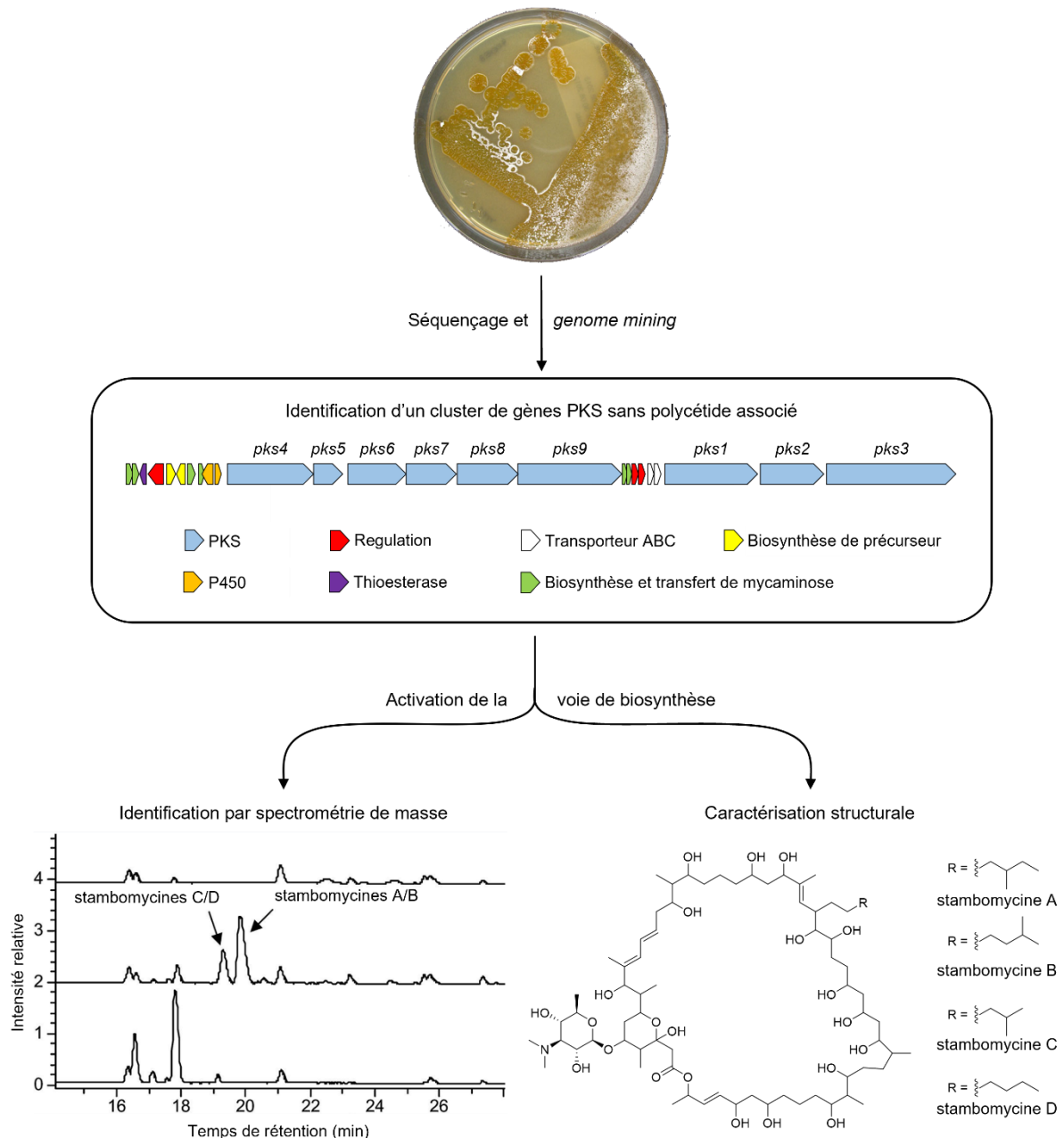
Les données métagénomiques d'échantillons environnementaux permettent de contourner le problème de cultivation rencontré avec de nombreux microorganismes. L'analyse de ces séquences complexes provenant d'une multitude d'espèces différentes peut conduire à l'identification de clusters de gènes impliqués dans la synthèse de métabolites spécialisés sans passer par une étape préalable de culture <sup>114</sup>. La nature de ces métabolites peut être déterminée par comparaison de leurs voies de biosynthèse à des BGC homologues et prédiction de la structure du produit de synthèse <sup>96</sup>. Ils peuvent ensuite être expérimentalement caractérisés par clonage et expression hétérologue de leurs voies de biosynthèse dans des hôtes dont la manipulation génétique a été optimisée <sup>115</sup>. Par exemple, l'analyse du métagénome d'échantillons du spongiaire *Mycale hentscheli* a permis d'identifier trois BGC impliqués dans la synthèse de polycétides au sein des génomes de symbiontes bactériens non cultivables <sup>116</sup>. Tous les clusters de gènes présents dans ce microbiome ont été identifiés et leurs produits de synthèse prédits par l'outil de détection antiSMASH <sup>117</sup>. Les clusters de gènes PKS ont ensuite été analysés plus en détail par l'outil de prédiction TransATor <sup>118</sup>, permettant d'identifier les voies de biosynthèse de la mycalamide A, la patéamine A et la péloruside A <sup>116</sup>.

L'identification par *genome mining* de voies de biosynthèse polycétidiques silencieuses chez une espèce cultivable peut également conduire à l'isolation du polycétide associé par activation de cette voie. Il est alors nécessaire de manipuler génétiquement le BGC en question pour surexprimer ou introduire un élément activateur approprié <sup>95,119</sup>. L'analyse du génome de *Streptomyces ambofaciens* a par exemple révélé la présence d'un cluster de gènes PKS auquel aucun polycétide connu n'était associé. Des analyses transcriptionnelles ont démontré que ces gènes n'étaient pas exprimés dans des conditions de culture de

## État de l'art

laboratoire, ce qui était en accord avec l'absence de produit de synthèse de la PKS encodée. Pour pallier ce problème, Laureti *et al.* ont exprimé de manière constitutive un régulateur transcriptionnel présent au sein du cluster, déclenchant ainsi l'expression des gènes biosynthétiques silencieux et conduisant finalement à l'identification de quatre macrolides à 51 atomes, les stambomycines A–D <sup>120</sup> (**Figure 14**).

*Streptomyces ambofaciens* ATCC23877



**Figure 14 :** Identification de la voie de biosynthèse des stambomycines chez *Streptomyces ambofaciens* par *genome mining*. Le séquençage et l'analyse du génome de cette bactérie ont permis d'identifier un cluster de gènes PKS sans qu'aucun polycétide ne soit associé à l'activité des enzymes encodées. L'activation de l'expression de ce cluster a entraîné la production de stambomycines, isolées et caractérisées par spectrométrie de masse (à gauche) et résonance magnétique nucléaire (à droite). Les chromatogrammes (à gauche) indiquent les pics les plus intenses détectés lors de l'élution des espèces ioniques extraites du mycélium de différentes souches : la

souche contrôle (en haut), la souche dans laquelle le régulateur transcriptionnel est surexprimé (au milieu) et la souche dans laquelle un gène codant une partie de la PKS a été délété (en bas) (figure adaptée de Laureti *et al.*<sup>120</sup>).

## 2. L'hémisynthèse

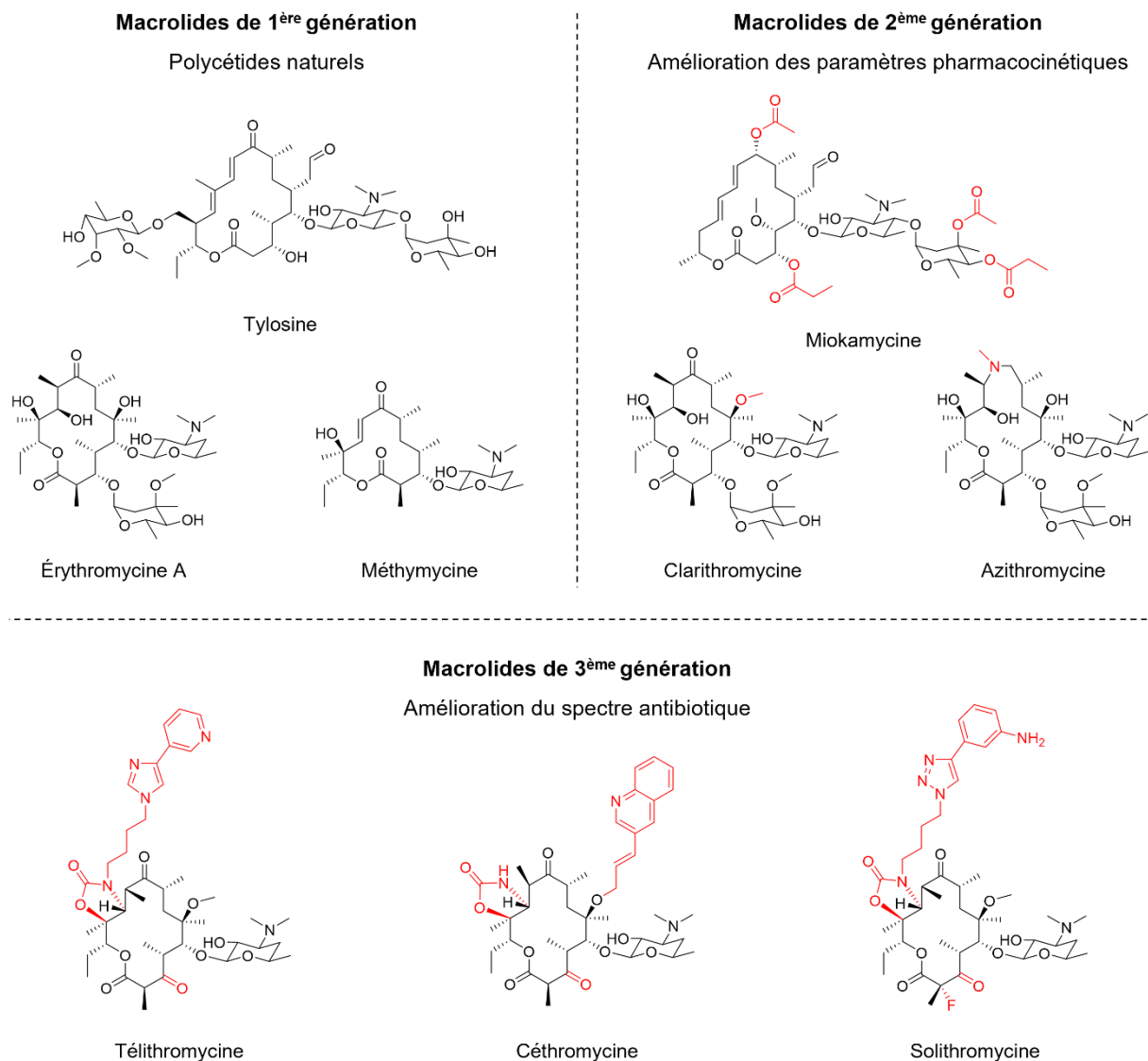
Les différentes méthodologies décrites ci-dessus permettent la découverte de nouveaux polycétides, possédant parfois une bioactivité d'intérêt pharmaceutique. Cependant, le rôle écologique de ces molécules est rarement lié à leur utilisation médicale ce qui explique que sous leur forme native, non-modifiée, celles-ci présentent souvent des propriétés physicochimiques, d'absorption, de distribution, de métabolisation, d'élimination et de toxicité (ADMET<sup>121</sup>) sous-optimales. Couplée à cette limitation, l'utilisation massive de ces composés aux propriétés antibiotiques entraîne inévitablement l'apparition de résistances microbiennes qui constituent un second problème de grande importance. L'hémisynthèse permet d'accéder à des dérivés aux propriétés nouvelles afin de surmonter ces problématiques en prenant pour point de départ la structure chimique de composés naturels, et en les modifiant chimiquement à certaines positions spécifiques.

Une telle approche fut utilisée dans la quête d'antibiotiques aux propriétés pharmacocinétiques améliorées dans les années 1980 et début 1990, qui conduisit au développement par hémisynthèse de plusieurs dérivés d'érythromycine A, appelés macrolides de seconde génération. Deux d'entre eux eurent un grand succès pharmaceutique, la clarithromycine<sup>122</sup> et l'azithromycine<sup>123</sup> (**Figure 15**), commercialisés sous les noms Biaxin (Abbott) et Zithromax (Pfizer), respectivement. Ces deux polycétides hémisynthétiques ont pour point commun une résistance améliorée à l'inactivation en milieu acide due à la modification de leurs positions C6 ou C9, augmentant de fait leur temps de demi-vie dans l'organisme ce qui rend leur posologie moins contraignante. Leurs structures leur confèrent également une plus grande lipophilie ce qui améliore leur capacité de pénétration tissulaire, leur conférant une efficacité dans le traitement d'infections provoquées par des pathogènes intracellulaires<sup>124</sup>. Le succès encore plus retentissant de l'azithromycine s'explique par des effets secondaires moins sévères comparés à l'érythromycine et ses dérivés, à savoir notamment un risque moindre d'arythmie et de problèmes gastrointestinaux et peu d'interactions avec les autres molécules thérapeutiques, permettant sa coadministration<sup>125</sup>.

Afin de trouver une solution au problème grandissant de résistances bactériennes aux antibiotiques, l'approche d'hémisynthèse fut également appliquée à la conception de macrolides de troisième génération aussi appelés kétolides. Ces dérivés sont caractérisés par la substitution du L-cladinose par un groupement cétone en position 3 du cycle macrolactone ainsi que, pour la plupart des kétolides, par l'addition d'un carbamate cyclisé entre les positions 11 et 12 et d'une chaîne alkyle-aryle (**Figure 15**). Le seul kétolide ayant atteint le stade de la

## État de l'art

mise sur le marché est la télithromycine commercialisée sous le nom Ketek (Aventis) et obtenue en huit étapes chimiques à partir de la clarithromycine <sup>126</sup>. Ces kétolides présentent une forte activité antibiotique envers les bactéries Gram-positives, notamment des souches de *Streptococcus pneumoniae* connues pour leur résistances aux macrolides <sup>127</sup>, répondant ainsi à la problématique qui avait initié les efforts de recherche à l'origine de leur conception. En effet, l'un des mécanismes de défense de ces bactéries fait intervenir les enzymes de la classe des *Erythromycin resistance methylases* (Erm) méthylant le ribosome pour empêcher les macrolides de s'y fixer. Or, l'échange du L-cladinosose pour une fonction cétone en C3 des kétolides induit faiblement ce système de défense, expliquant la sensibilité de bactéries Gram-positives envers ces polycétides, pourtant résistantes à d'autres macrolides comme l'érythromycine et l'azithromycine <sup>128</sup>.



**Figure 15 :** Exemples de macrolides naturels et dérivés hémisynthétiques de deuxième et troisième générations. Les groupements chimiquement ajoutés aux polycétides natifs sont indiqués en rouge.

La génération de molécules d'intérêt pharmaceutique hémisynthétiques requiert

néanmoins de pouvoir accéder au composé naturel de départ de manière durable et économiquement viable, et se limite généralement à la modification des fonctions décorant le squelette polycétidique, pas du squelette lui-même.

### 3. La synthèse chimique totale

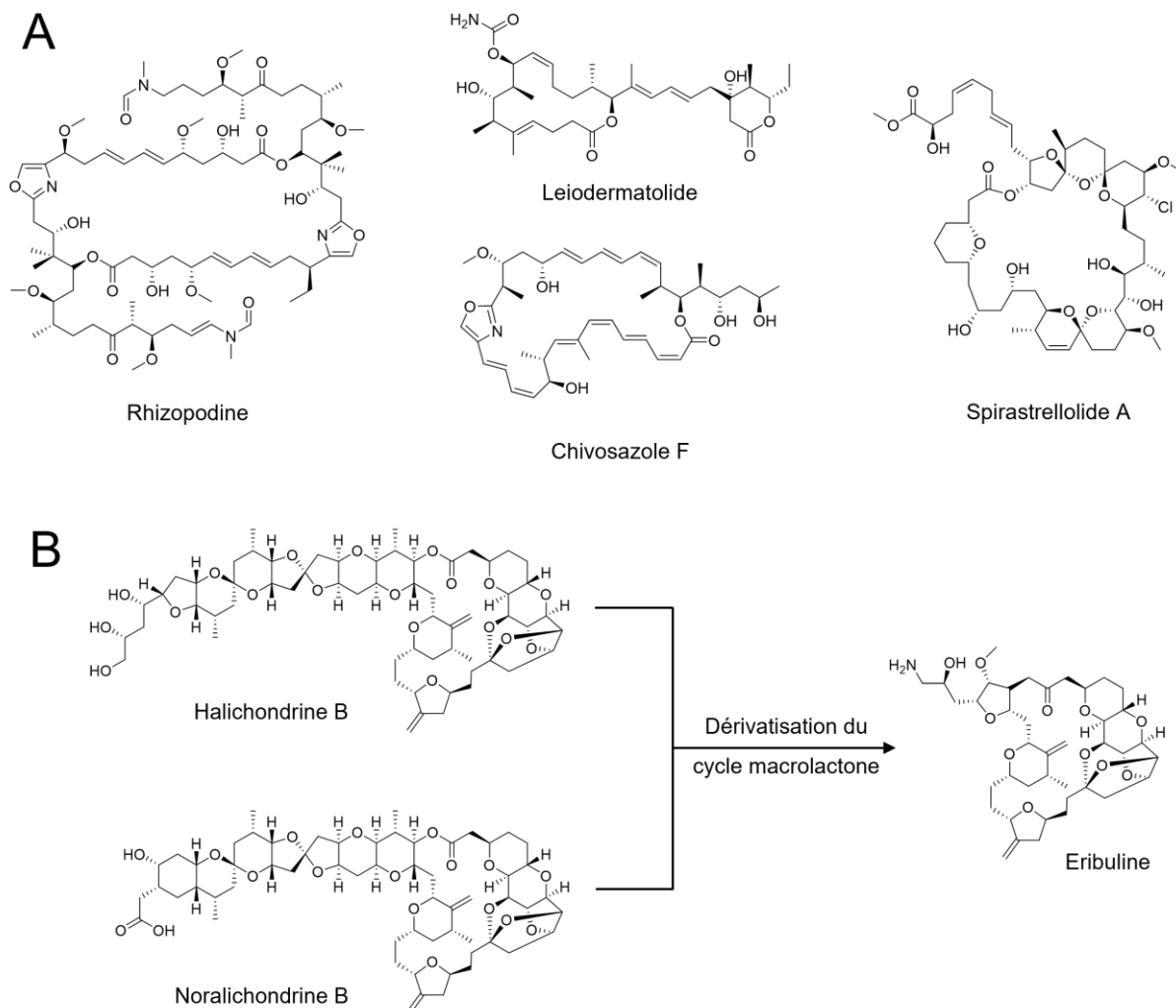
La production en quantités suffisantes de métabolites est indispensable à la caractérisation de leurs structures et de leurs propriétés biologiques or l'obtention de rendements compatibles avec ces études peut s'avérer difficiles par biosynthèse. La synthèse chimique totale de polycétides naturels ou dérivés constitue une stratégie alternative de production de ces molécules tout en permettant la fonctionnalisation rationnelle de positions spécifiques pour générer de nouveaux analogues <sup>129</sup>. Par exemple, le groupe de Ian Paterson a ainsi accompli avec succès la synthèse chimique des macrolides spirastrellolide A, leiodermatolide, rhizopodine et chivosazole F possédant des structures variées et complexes (**Figure 16**) et présentant une activité anticancéreuse prometteuse <sup>129</sup>.

Les macrolides de la classe des halichondrines ont également focalisé les efforts de synthèse chimique totale. L'équipe de Yoshito Kishi notamment, s'est particulièrement intéressée à l'halichondrine B, se distinguant des autres polyéthers par sa puissante activité antiproliférative de cellules cancéreuses. Motivée par les faibles rendements obtenus par extraction de sources naturelles, cette équipe est parvenue à synthétiser chimiquement ce polycétide et son analogue norhalichondrine B <sup>130</sup> (**Figure 16**). Dans le but d'identifier son pharmacophore minimum, le potentiel anticancéreux des intermédiaires de synthèse chimique de l'halichondrine B a été évalué et a révélé une activité d'inhibition de croissance de cellules cancéreuses portée par le fragment macrolactone du polycétide. Les modifications du macrocycle n'ayant pas provoqué d'amélioration majeure de l'activité, les efforts de synthèse se sont portés sur le système de cycles pyrano-pyranes C27-C38 <sup>131</sup>. Un analogue simplifié, l'éribuline (**Figure 16**), a ainsi été identifié par son activité d'inhibition de croissance de différentes lignées cellulaires cancéreuses humaines <sup>132</sup> et approuvé en 2010 comme agent anticancéreux <sup>133</sup>.

Par ailleurs, une nouvelle stratégie de découverte de métabolites bioactifs a récemment été mise au point par le groupe de Sean Brady, grâce à l'utilisation conjointe des approches de *genome mining* et de synthèse chimique. Lors d'une première phase bioinformatique, Wang *et al.* ont analysé plus de 10 000 génomes bactériens dans le but d'identifier des BGC associés à la synthèse d'antibiotiques polymyxines <sup>134</sup>. Un cluster de gènes a particulièrement attiré leur attention du fait de sa divergence de séquence avec les autres clusters de cette famille de métabolites. En se basant sur la prédiction de spécificité de substrat des enzymes encodées, la structure putative du NRP associé a été synthétisée chimiquement, contournant l'étape

## État de l'art

fastidieuse de manipulation microbiologique pour isoler le composé naturel. Cette approche de synthèse chimique totale guidée par la génomique a ainsi conduit à la découverte de la macolacine, antibiotique efficace contre des bactéries pathogènes multirésistantes <sup>134</sup>.



**Figure 16 :** Exemples de synthèse chimique totale de polycétides. (A) Par cette approche, quatre macrolides naturels aux tailles de cycles et fonctions chimiques variées ont pu être synthétisés <sup>129</sup>. (B) La synthèse chimique des macrolides halichondrine B et noralichondrine B a conduit à l'identification de la structure du pharmacophore macrolactonique minimum. Sa modification chimique a ensuite permis la synthèse de l'agent anticancéreux éribuline <sup>131</sup>.

Une multitude de composés potentiellement bioactifs peut être générée en employant cette stratégie tant que les structures d'intérêt sont chimiquement synthétisables et que les rendements de synthèse sont suffisants pour la caractérisation structurale et fonctionnelle voire l'utilisation médicale.



## 4. L'ingénierie génétique

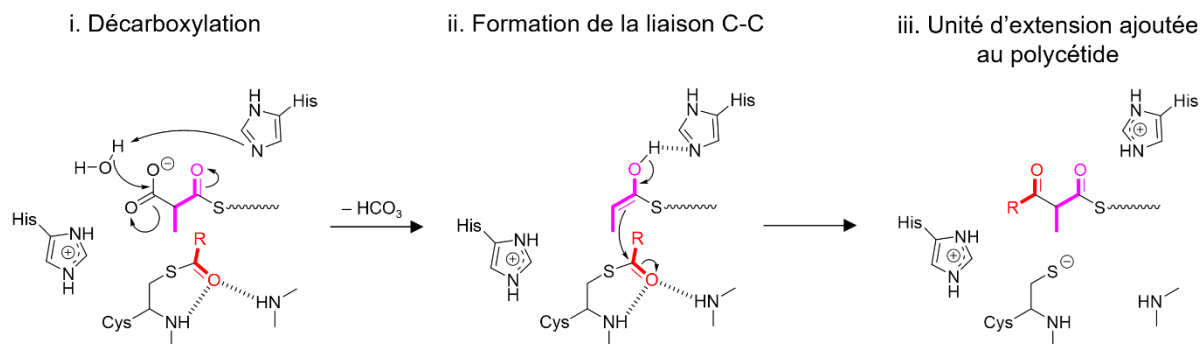
Au début des années 1990, l'analyse de la voie de biosynthèse de l'érythromycine A a révélé une architecture de PKS sans précédent, constituée d'un méga-complexe de multienzymes au sein duquel les différentes activités catalytiques nécessaires à la synthèse du polycétide sont réparties tout au long de la synthase<sup>135,136</sup> (**Figure 27**). Cette découverte allait ouvrir la voie à une nouvelle approche de diversification des structures polycétidiques en modifiant non pas le polycétide lui-même mais la PKS responsable de sa synthèse par ingénierie génétique. L'objectif est alors de reprogrammer la suite d'évènements biosynthétiques catalysée par une PKS d'intérêt par exemple en inactivant, en échangeant ou en rajoutant certaines activités enzymatiques, (voir la section III.4.c.). La machinerie catalytique extrêmement complexe des PKS est ainsi exploitée pour la création de polycétides inédits.

Cette approche est une stratégie très prometteuse pour générer des analogues potentiellement bioactives puisqu'elle permet d'accéder à de nouvelles structures complexes parfois inaccessibles par des approches de synthèse chimique. Cependant, modifier les voies de biosynthèse de polycétides nécessite de comprendre en détail les mécanismes catalytiques et interactions qui en régissent le fonctionnement.

## III. Les polycétides synthèses

### 1. Un principe de synthèse commun

La multitude de polycétides caractérisés à ce jour présente une hétérogénéité structurale dont nous ne connaissons toujours pas les limites. Pourtant, ces métabolites sont tous synthétisés selon un processus enzymatique commun. En effet, les polycétides sont générés à partir de différents précurseurs cétoniques qualifiés d'unités d'initiation et d'unités d'extension, sous leur forme active acyl-coenzyme A (CoA), principalement acétyl- et malonyl-CoA, condensés les uns à la suite des autres à la manière de « blocs de construction » pour former une chaîne polycétidique en élongation. Les PKS catalysent ces réactions de condensations successives par décarboxylation des unités d'extension suivie de la formation d'une liaison C-C entre l'unité d'extension décarboxylée et la chaîne polycétidique<sup>137</sup> (**Figure 17**). Des réactions de réduction et de déshydratation des fonctions  $\beta$ -cétones catalysées par ces mêmes enzymes permettent ensuite de générer une plus grande diversité chimique. La cyclisation du polycétide linéaire suivie de l'intervention d'enzymes de modification du produit après sa libération de la PKS (enzymes post-PKS) complexifient encore davantage les structures polycétidiques.

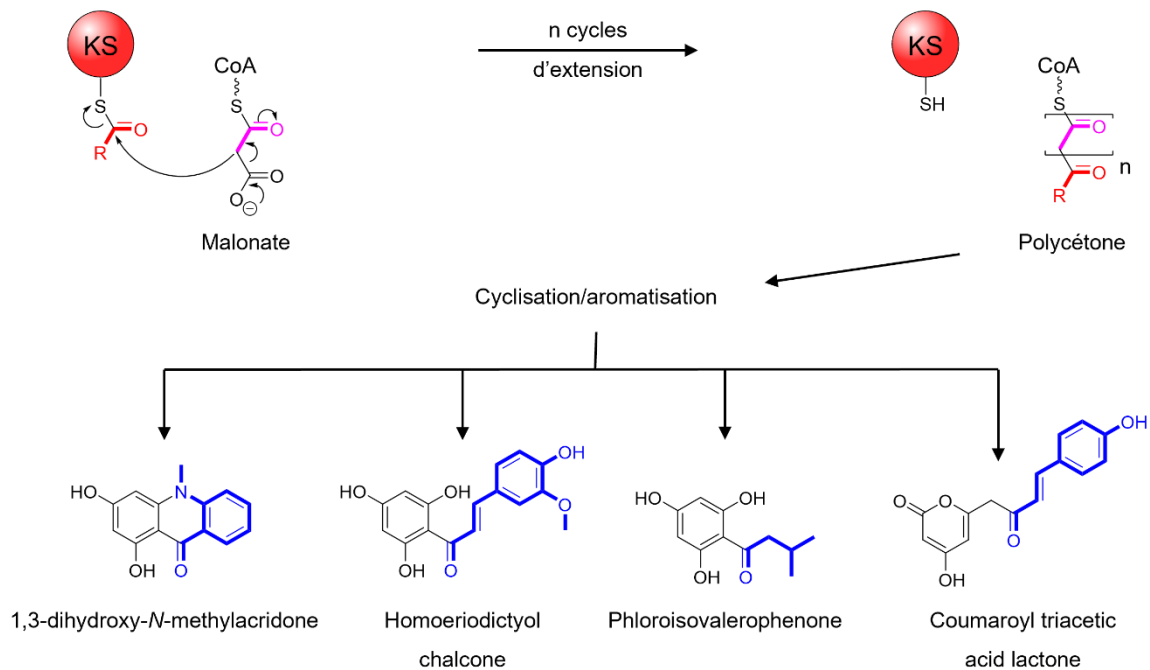


**Figure 17 :** Mécanisme de condensation décarboxylative de Claisen, extrapolé du mécanisme d'enzymes homologues impliquées dans la synthèse d'acide gras proposé par Witkowski *et al.*<sup>137</sup>. La décarboxylation de l'unité d'extension (i) (ici, un méthylmalonate) conduit à la formation d'un énolate nucléophile qui attaque ensuite le thioester électrophile de la chaîne polycétidique à prolonger générant une nouvelle liaison C-C (ii). Le polycétide possède alors deux carbones supplémentaires (iii), un carbonyle formant un thioester avec l'enzyme et un groupement variable en position  $\alpha$ , dépendent de l'unité d'extension ajoutée (dans cet exemple, un groupement méthyle). Les acides aminés impliqués dans la catalyse enzymatique sont représentés en noir, la chaîne polycétidique en rouge, et l'unité d'extension en rose (figure adaptée de Weissman, 2017<sup>138</sup>).

Le mécanisme enzymatique des différentes PKS caractérisées repose toujours sur ce principe de condensation-modification. Toutefois, l'organisation protéique de ces enzymes ainsi que la répartition des activités catalytiques permet de distinguer trois classes de PKS.

## 2. Les PKS de type III

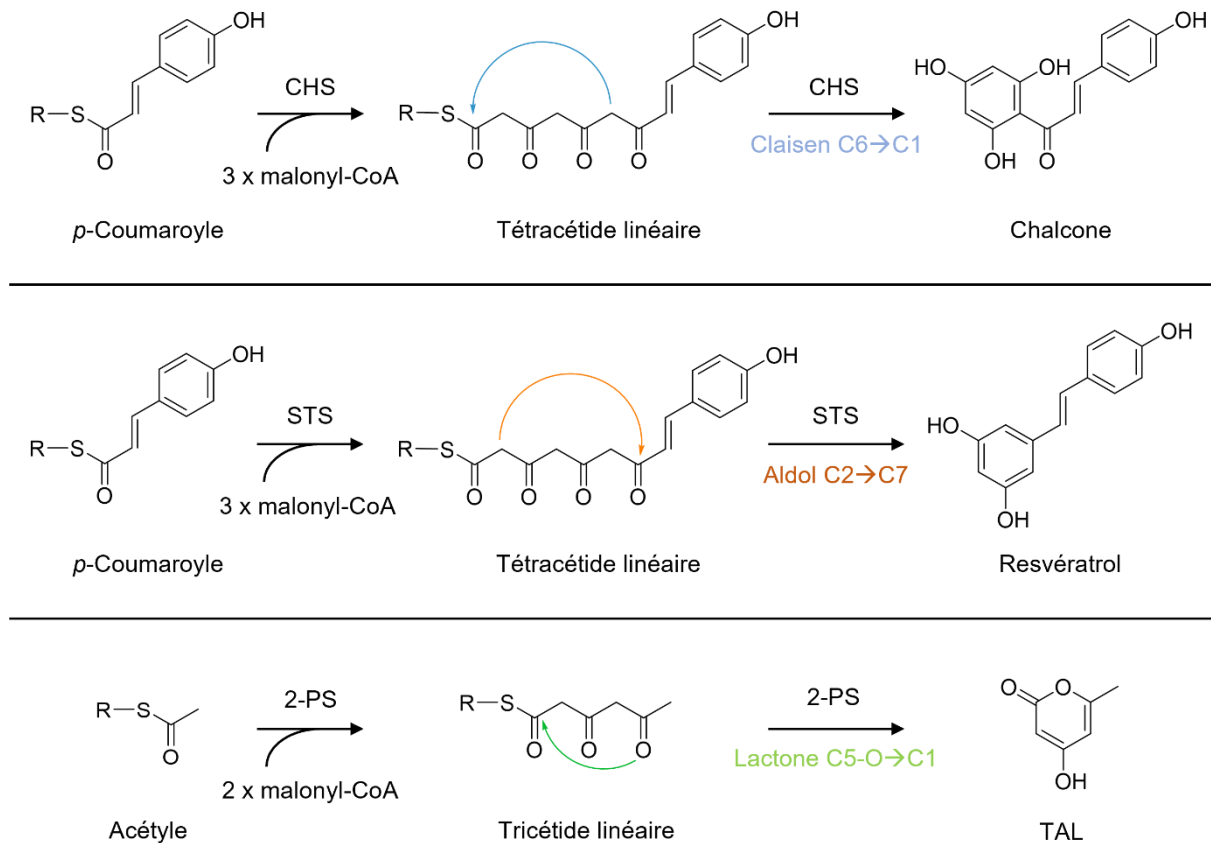
Identifiées chez les plantes<sup>139</sup>, les bactéries<sup>140</sup>, les champignons<sup>141</sup> et certains protistes<sup>142</sup>, les PKS de type III également qualifiées de superfamille chalcone synthase/stilbène synthase (CHS/STS) sont des enzymes à l'organisation simple responsables de la synthèse de polycétides de faible poids moléculaire servant fréquemment d'intermédiaire à la synthèse de polycétides plus complexes et d'autres métabolites spécialisés. Un unique complexe homodimérique de kétosynthase (KS) catalyse la condensation d'unités acyl-CoA sur une unité d'initiation souvent cyclique et porte également dans le même site actif les activités de cyclisation interne et d'aromatisation de l'intermédiaire linéaire pour conduire à la formation d'un polycétide typiquement mono- ou bicyclique<sup>143</sup> (**Figure 18**). Ce mécanisme catalytique étant commun à toutes les PKS de type III, leur diversité fonctionnelle provient des différentes unités d'acyl-CoA d'initiation sélectionnées (*p*-coumaroyl-CoA, benzoyl-CoA, acétoacétyl-CoA, ...), du nombre de cycles d'extension (1 à 8) et du mécanisme de cyclisation intramolécule (pas de cyclisation, condensation de Claisen, aldolisation, lactonisation, etc.)<sup>144</sup>. Diverses enzymes post-PKS peuvent ensuite complexifier davantage les squelettes polycétidiques biosynthétisés.



**Figure 18 :** Mécanisme catalytique des PKS de type III. Une KS homodimérique catalyse les condensations itératives d'acyl-CoA pour former une chaîne polycétidique puis cyclise et aromatise cet intermédiaire pour générer le polycétide final. La chaîne polycétidique est représentée en rouge, l'unité d'extension en rose et les unités d'initiation des différents exemples en bleu (figure adaptée de Weissman K, 2009<sup>35</sup>). KS = kétosynthase.

La CHS est une enzyme ubiquitaire chez les végétaux supérieurs. Étant la première PKS de type III identifiée<sup>145</sup>, son mécanisme enzymatique a été amplement étudié. La chalcone, produit généré, constitue la « charpente » de nombreux flavonoïdes. Sa biosynthèse est le résultat de la condensation successive de trois unités d'extension malonyl-CoA sur une unité d'initiation *p*-coumaroyl-CoA suivie de la cyclisation par condensation de Claisen entre les positions C1 et C6 et de l'aromatisation du tétracétide linéaire<sup>143</sup> (**Figure 19**). De façon intéressante, la PKS de type III STS synthétise un intermédiaire linéaire identique à celui généré par la CHS. Cependant, le tétracétide est cette fois cyclisé par aldolisation et perte du carbonyle C1 sous forme de dioxyde de carbone, formant le resvératrol après aromatisation<sup>143</sup> (**Figure 19**). Le recours à une autre unité d'initiation et un nombre d'extension différent conduit à la formation de nouvelles structures polycétidiques. C'est le cas par exemple de la PKS de type III 2-pyrone synthase (2-PS), qui utilise une unité d'initiation plus simple, un acétyl-CoA, et synthétise un tricétide par condensation de deux unités malonyl-CoA puis cyclise celui-ci par lactonisation (**Figure 19**) pour former le polycétide TAL<sup>146</sup>.

## État de l'art



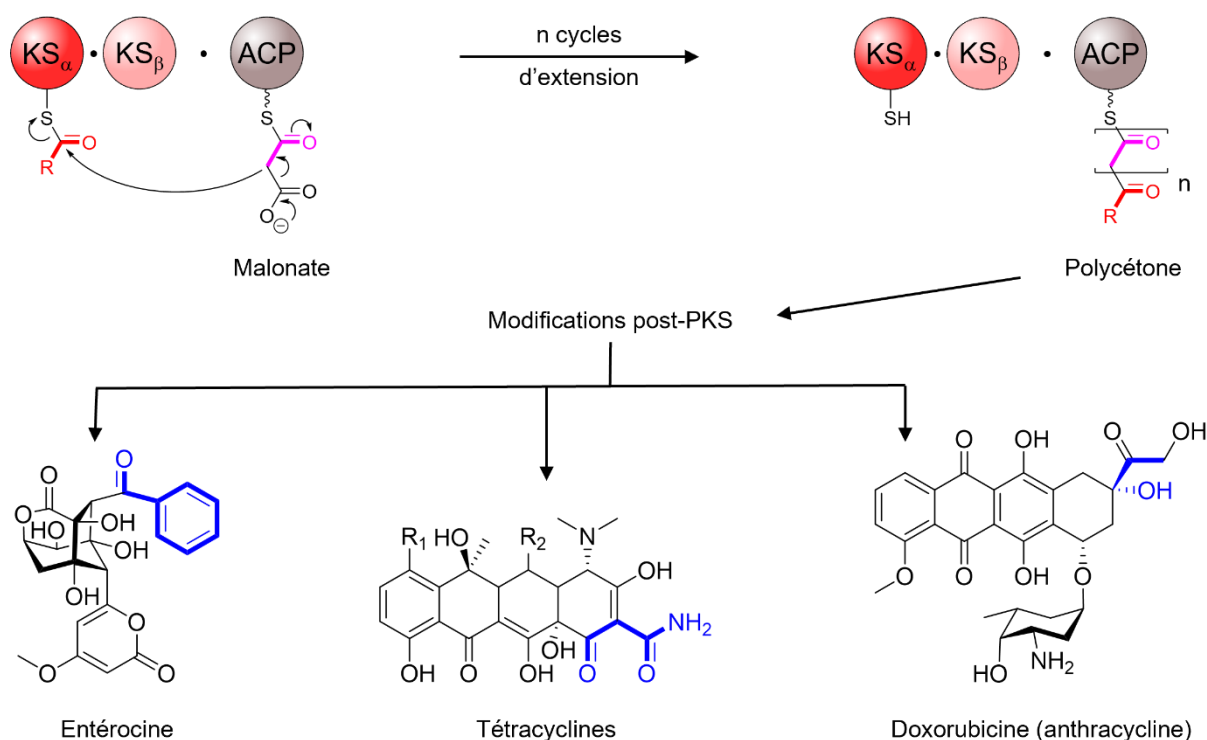
**Figure 19 :** Synthèses de la chalcone, du resvératrol et du TAL par les PKS de type III CHS, STS et 2-PS, respectivement. La variété de structure chimique générée est la conséquence de l'utilisation de différentes unités d'initiation, nombres d'extension et mécanismes de cyclisation (figure adaptée de Austin et Noel, 2003<sup>143</sup>). R = CoA ou KS.

### 3. Les PKS de type II

Identifiées uniquement dans des microorganismes bactériens, les PKS de type II sont responsables de la synthèse de nombreux polycétides exploités pour leurs propriétés biologiques tels que les antibiotiques resistomycine, nanaomycine A et tétracyclines (**Figure 7**) ou les puissants anticancéreux anthracyclines. Ces PKS sont constituées de plusieurs protéines discrètes exécutant chacune une activité spécifique et itérative. L'initiation de la synthèse du polycétide et son élongation sont assurées par trois composants protéiques formant un complexe enzymatique qualifié de PKS minimale. Celle-ci consiste en un hétérodimère de KS, désignées  $KS_{\alpha}$  et  $KS_{\beta}$ , qui s'associe à une protéine porteuse d'acyle (ACP) servant de support à la chaîne polycétidique en élongation<sup>147</sup>. La  $KS_{\beta}$  ne possède pas la cystéine catalytique nécessaire à l'activité de condensation des unités d'extension mais reste capable de décarboxyler une unité malonate liée à l'ACP pour générer un acétate qui est ensuite transféré au site actif de la  $KS_{\alpha}$ . Un nouveau malonate peut ensuite être chargé sur l'ACP, permettant à la  $KS_{\alpha}$  de catalyser une extension de chaîne<sup>147</sup>. De plus, la  $KS_{\beta}$  joue un rôle majeur dans le contrôle de la taille de la chaîne polycétidique, d'où sa deuxième

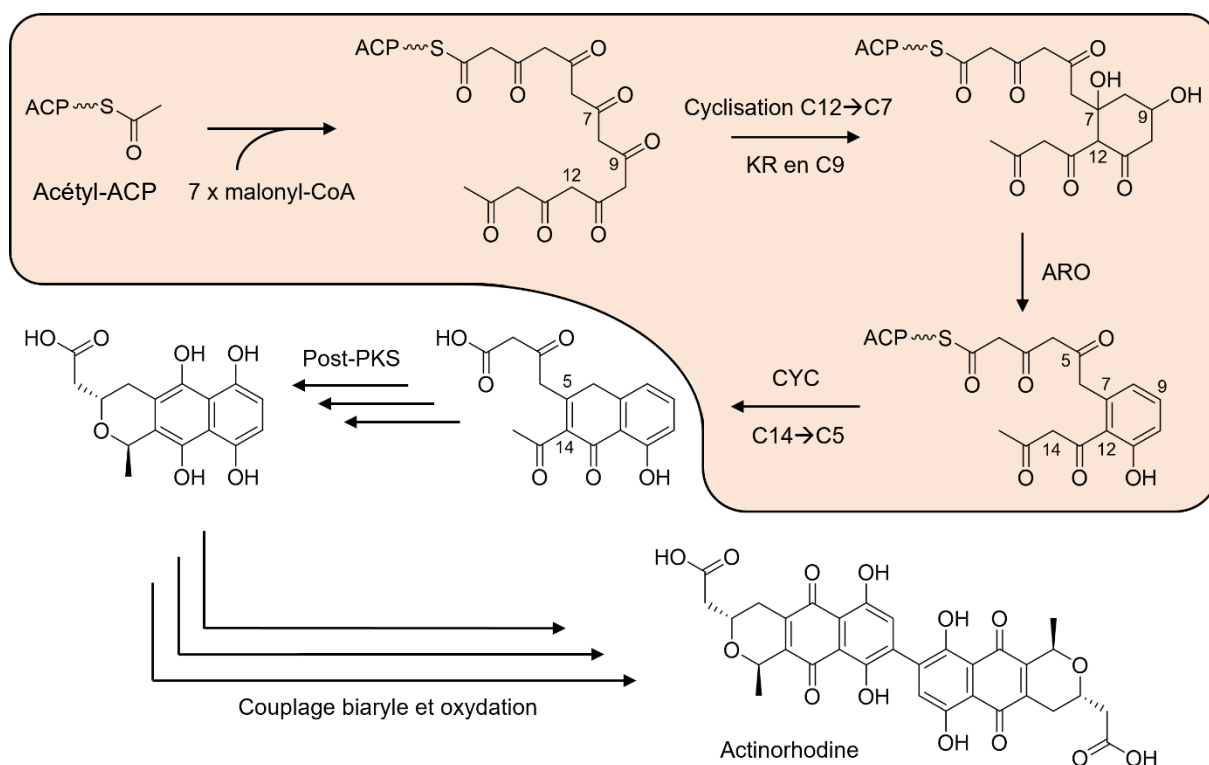
qualification de *chain length factor* (CLF) <sup>148</sup>.

Le processus de synthèse débute par l'acylation de l'ACP avec une unité d'initiation, réaction catalysée par une malonyl-CoA:ACP transacylase (MCAT) provenant vraisemblablement de synthèses d'acides gras (FAS), puisque les BGC codant les différentes protéines de PKS de type II ne contiennent aucun gène de MCAT et que FAS et PKS présentent un mécanisme enzymatique similaire (voir partie IV.2.b.ii.) <sup>149</sup>. Cette acylation est réalisée sur un groupement prosthétique pantéthéine transféré post-traductionnellement d'un CoA vers un résidu sérine de l'ACP par une 4'-phosphopantéthéinyl transférase (PPTase), convertissant ainsi l'ACP en sa forme active *holo*-ACP <sup>150</sup>. L'unité d'initiation est ensuite décarboxylée par la  $KS_{\beta}$  puis transférée de l'ACP vers la  $KS_{\alpha}$ , permettant à l'ACP d'être acylé avec une unité d'extension. La  $KS_{\alpha}$  catalyse ensuite la condensation décarboxylative des deux unités et répète ce procédé d'acylation-condensation jusqu'à formation d'un polycétide linéaire de pleine taille <sup>149</sup> (**Figure 20**). Des enzymes post-PKS telles que des cyclases (CYC) et des aromatasés (ARO) participent alors au repliement de cet intermédiaire linéaire. Le polyphénol généré peut finalement être complexifié par l'action d'autres enzymes post-PKS, telles que des oxygénases, glycosyl- et méthyltransférases <sup>151</sup>.



**Figure 20 :** Mécanisme catalytique des PKS de type II synthétisant des polycétides aromatiques. Une hétérodimère  $KS_{\alpha}$ - $KS_{\beta}$  s'associe à un ACP et catalyse les condensations itératives d'acyl-CoA sur cet ACP pour former un intermédiaire polycétone. Cet intermédiaire linéaire subit ensuite des modifications chimiques conduisant à la formation du polycétide final. La chaîne polycétidique est représentée en rouge, l'unité d'extension en rose et les unités d'initiation des différents exemples en bleu (figure adaptée de Weissman, 2009 <sup>35</sup>).  $KS_{\alpha}$  = kétosynthase  $\alpha$  ;  $KS_{\beta}$  = kétosynthase  $\beta$  ; ACP = protéine porteuse d'acyle.

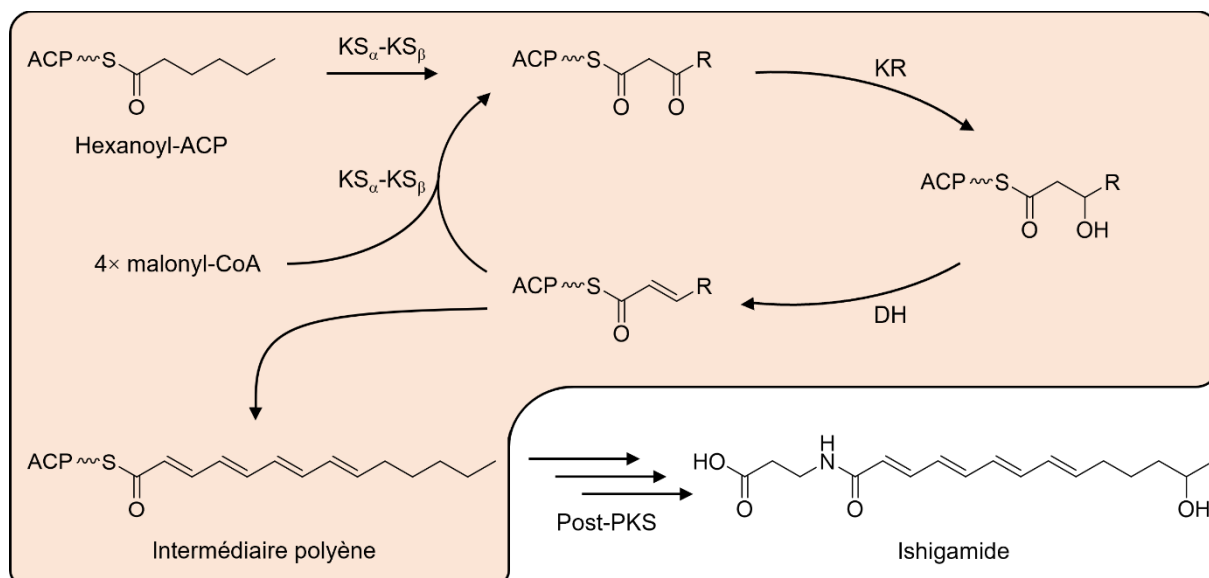
Le mécanisme de biosynthèse des polycétides de type II est illustré par le modèle d'étude de cette classe de composé, l'actinorhodine (**Figure 21**). Ici, la condensation décarboxylative de sept malonates sur l'unité initiatrice acétate conduit à la formation d'un intermédiaire octacétone. Une première cyclisation a alors lieu entre les positions C12 et C7 par aldolisation, suivie de la réduction du carbonyle en C9 par une kétéductase (KR) et de l'aromatisation de ce premier cycle par une ARO. Une seconde cyclisation est catalysée par une CYC entre les positions C14 et C5 qui libère dans le même temps le polycétide de l'ACP du fait de sa bifonction thioestérase (TE) <sup>152</sup>. Cet intermédiaire bicyclique subit ensuite plusieurs modifications post-PKS jusqu'au couplage biaryle final de deux intermédiaires pour former l'actinorhodine <sup>153</sup> (**Figure 21**).



**Figure 21 :** Voie de biosynthèse proposée de l'actinorhodine impliquant une PKS de type II (figure adaptée de Chen *et al.*, 2018 <sup>149</sup>). L'intermédiaire linéaire de pleine taille est issu de la condensation de sept unités d'extension malonates sur une unité d'initiation acétate. Trois réactions successives de cyclisation, kétéduction et aromatisation forment le premier cycle aromatique. L'intermédiaire subit ensuite une seconde cyclisation suivie de sa libération de la PKS, catalysée par une CYC bifonctionnelle. Deux réductions et deux hydroxylations conduisent à la formation d'un intermédiaire tétrahydroxynaphtalène dont l'oxydation et la dimérisation par couplage biaryle génère l'actinorhodine. La zone rose indique les intermédiaires de synthèse liés à la PKS.

Les PKS de type II ont longtemps été considérées comme des synthèses non- ou partiellement réductrices, conduisant uniquement à la formation de polycétides aromatiques. Cependant, des PKS de type II hautement réductrices ont été identifiées plus récemment. Ces PKS ne sont plus itératives puisque chaque cycle d'extension fait alors intervenir une KR discrète qui réduit le carbonyle en position  $\beta$  nouvellement installé puis une déshydratase (DH),

discrète également, forme une liaison  $\alpha,\beta$ -insaturée conduisant à la formation de polyènes<sup>149</sup>. Cette seconde catégorie de PKS de type II est illustrée par la voie de biosynthèse de l'ishigamide, formée par condensation d'un hexanoyl-CoA et de quatre malonyl-CoA. Le produit final de synthèse est alors linéaire et constitué de quatre doubles liaisons conjuguées<sup>154</sup> (**Figure 22**).



**Figure 22 :** Voie de biosynthèse proposée du polyène ishigamide impliquant une PKS de type II hautement réductrice. Chaque élongation est suivie d'une kétoréduction et d'une déshydratation<sup>154</sup>. La zone rose indique les intermédiaires de synthèse liés à la PKS. R correspond au reste de la chaîne carbonée des intermédiaires de synthèse.

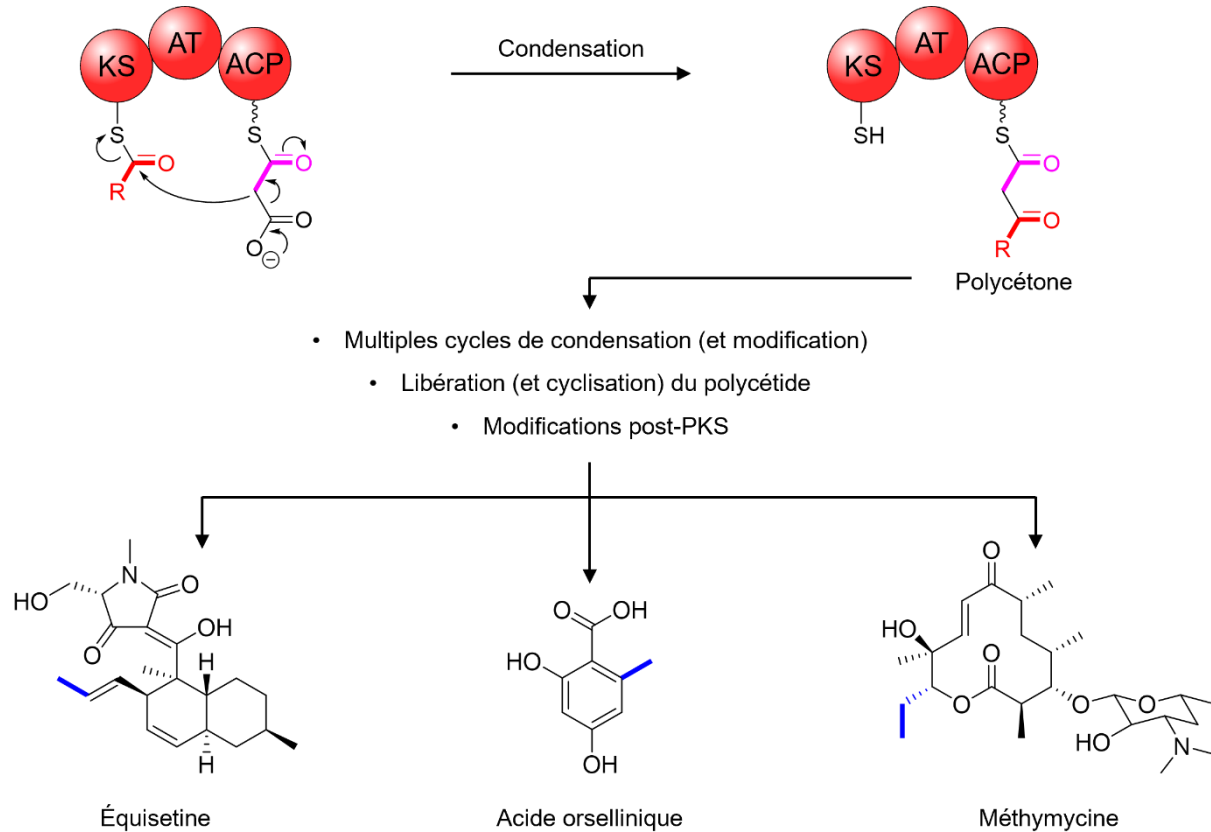
#### 4. Les PKS de type I

Les PKS de type I, retrouvées chez les bactéries, champignons et animaux<sup>155-157</sup>, présentent l'architecture la plus complexe des trois classes de ces synthèses. Bien qu'encore divisées en plusieurs sous-catégories, ces PKS ont toutes pour point commun la présence de plusieurs sites actifs au sein de chaque chaîne polypeptidique, chacun assurant une fonction essentielle à l'élongation ou la modification chimique du polycétide<sup>35</sup>. Ainsi, à l'inverse des PKS de type II et III, les PKS de type I sont constituées de protéines de haut poids moléculaire repliées en domaines catalytiques successifs. L'ensemble des domaines requis pour un cycle complet d'élongation du polycétide et modification de l'unité d'extension ajoutée est qualifié de module. Un module fonctionnel minimal doit donc être composé d'un domaine ACP pour porter l'intermédiaire de synthèse et les unités d'extension, un domaine acyltransférase (AT) qui sélectionne spécifiquement les unités d'initiation et d'extension et les transfère à l'ACP et un domaine KS qui catalyse les condensations d'unités d'extension sur la chaîne polycétidique (**Figure 23**). Des domaines optionnels peuvent s'ajouter à ces domaines essentiels pour modifier la fonction  $\beta$ -cétone nouvellement ajoutée. Les plus fréquemment retrouvés au sein

## État de l'art

des modules sont les domaines KR, DH et énoylréductases (ER), qui génèrent un intermédiaire  $\beta$ -hydroxylé,  $\alpha,\beta$ -insaturée et totalement saturée, respectivement <sup>155</sup> (**Figure 23**).

L'action itérative d'un unique module ou l'intervention successive de différents modules pour synthétiser le polycétide permet de distinguer les PKS de type I en deux catégories.



**Figure 23 :** Mécanisme catalytique général des PKS de type I synthétisant des polycétides aromatiques ou réduits. Une chaîne polypeptidique est repliée en domaines catalytiques permettant la synthèse du polycétide. Un module fonctionnel minimal est constitué des domaines AT, KS et ACP assurant la sélection des unités acyl-CoA, leur condensation et la liaison des unités et intermédiaires de synthèse, respectivement. Des domaines additionnels peuvent modifier les positions  $\alpha$  et  $\beta$  des unités d'extension ajoutées, formant un polycétide réduit en fin de synthèse. La chaîne polycétidique en extension est représentée en rouge, l'unité d'extension en rose et les unités d'initiation des différents exemples en bleu (figure adaptée de Weissman, 2009 <sup>35</sup>). KS = kétosynthase ; AT = acyltransférase ; ACP = protéine porteuse d'acyle.

### a. Les PKS de type I itératives

Majoritairement retrouvées chez les champignons <sup>158</sup> mais également présentes chez les bactéries <sup>8</sup> et certains eucaryotes photosynthétiques <sup>159</sup>, les PKS de type I itératives (iPKS) catalysent la formation de polycétides par action répétée d'un seul module. Les iPKS fongiques possèdent une capacité variable de réduction de leurs produits de synthèse, permettant de les catégoriser en PKS hautement réductrices (HR), partiellement réductrices (PR) et non réductrices (NR) <sup>158</sup>. Les NR-iPKS ne contiennent aucun domaine permettant la réduction des intermédiaires polycétone, générant ainsi des produits de synthèse mono- ou polycycliques



aromatiques. À l'inverse, les PR- et HR-iPKS possèdent divers domaines de modifications des intermédiaires de synthèse, à savoir KR et DH pour les PR-iPKS et KR, DH, ER et C-méthyltransférase (cMT) dans le cas des HR-iPKS <sup>160</sup>. Les voies de synthèse de l'aflatoxine B<sub>1</sub>, du 6-MSA et de la lovastatine illustrent les différences enzymatiques entre ces différentes classes d'iPKS.

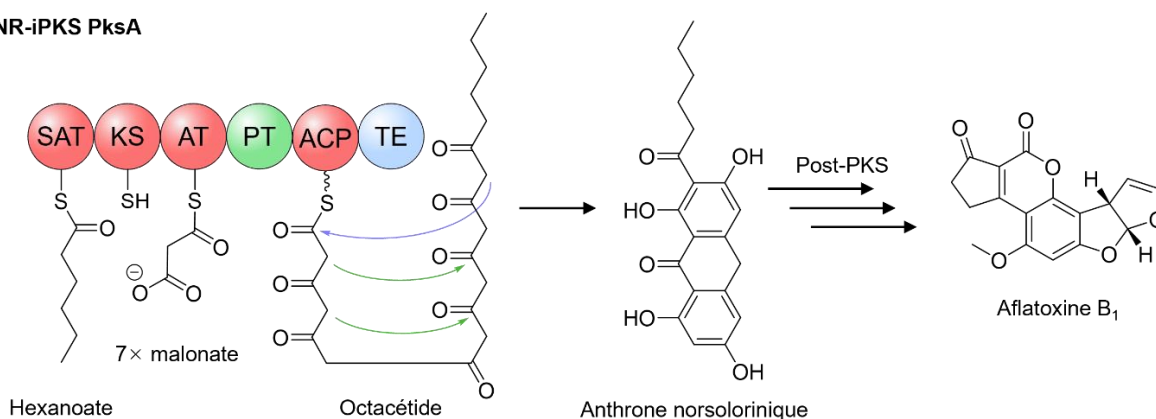
La NR-iPKS PksA est responsable de la synthèse d'anthrone norsolorinique, précurseur de l'aflatoxine B<sub>1</sub>, par condensation décarboxylative de sept malonates sur une unité d'initiation hexanoate. Étant non réductrice, cette PKS ne contient pas de domaine KR, DH, ER ou cMT. Elle est constituée d'un premier domaine « unité d'initiation-ACP transacylase » (SAT) qui sélectionne l'unité hexanoate, d'une KS, d'une AT qui sélectionne les unités d'extension, d'un domaine *product template* (PT) qui catalyse les deux premières cyclisations et aromatisations de l'intermédiaire linéaire, d'un ACP et finalement d'un domaine TE qui libère l'intermédiaire bicyclique par une troisième cyclisation formant l'anthrone norsolorinique <sup>160</sup> (**Figure 24**).

La PR-iPKS MSAS synthétise le polycétide 6-MSA qui sert de précurseur à la synthèse de métabolites plus complexes, par condensation de trois acétates issus de la décarboxylation de malonates sur une unité d'initiation acétyle. Cette PKS est constituée de cinq domaines comprenant les essentielles KS, KR et ACP ainsi qu'un domaine TE et un domaine KR. L'activité KR justifie la qualification de PR-iPKS puisque celle-ci permet de réduire la fonction β-cétone introduit par condensation de la deuxième unité d'extension en β-hydroxyle, conduisant après libération et cyclisation de l'hexacétide à la formation de 6-MSA <sup>161</sup> (**Figure 24**).

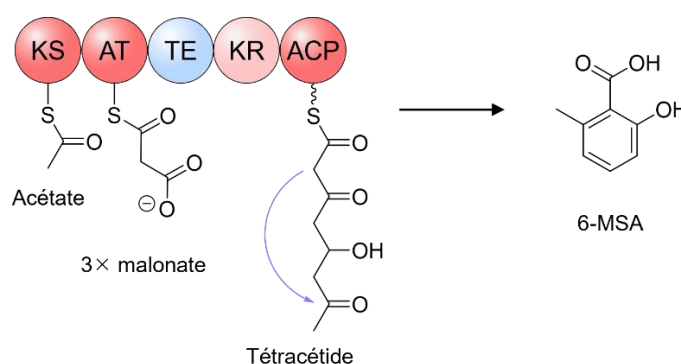
Le polycétide hypolipémiant lovastatine est synthétisé à partir d'acide dihydromonacoline L, produit de synthèse de la HR-iPKS LovB. Cette PKS est formée de l'extrémité N-terminale vers C-terminale des domaines KS, AT, DH, cMT, d'une pseudo-KR inactive (ΨKR), d'une pseudo-ER inactive (ΨER) qui contribue vraisemblablement à la dimérisation de la PKS <sup>162</sup>, d'une KR, d'un ACP et finalement d'un domaine de condensation (C) possédant une fonction analogue aux KS et retrouvé classiquement dans les synthétases de peptides non ribosomiaux (NRPS). Sa fonction dans LovB reste cependant indéterminée. Ce système est complété en *trans* par la protéine discrète LovC qui apporte l'activité ER que le domaine ΨER ne peut pas assurer <sup>162</sup>. LovB catalyse huit cycles de condensations décarboxylatives pour former un intermédiaire nonacétide fortement réduit. Cet état avancé de réduction est dû aux activités KR, DH, MT et ER qui interviennent spécifiquement lors de certains cycles après condensation conduisant à la formation, après libération, du polycétide d'acide dihydromonacoline L <sup>160</sup> (**Figure 24**).

## État de l'art

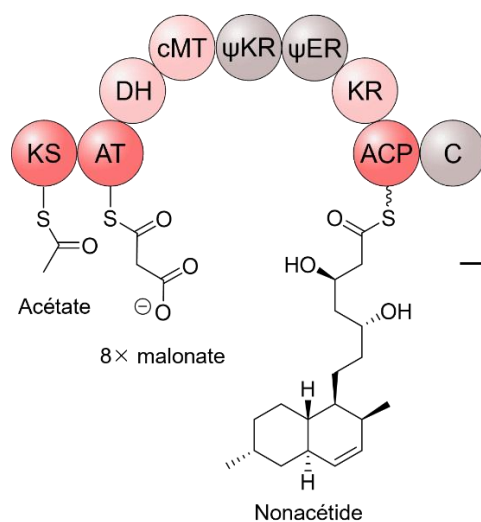
### NR-iPKS PksA



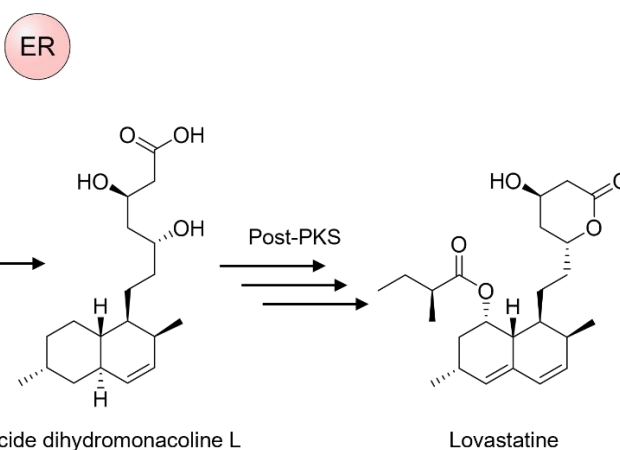
### PR-iPKS MSAS



### HR-iPKS LovB



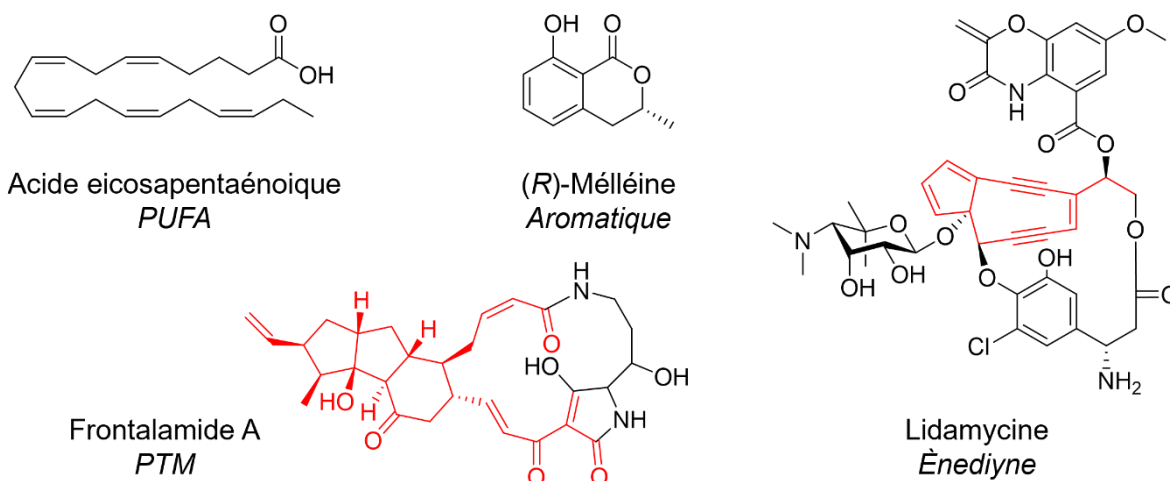
### LovC



**Figure 24:** Voies de biosynthèse de l'aflatoxine B<sub>1</sub>, de la 6-MSA et de la lovastatine impliquant une NR-, PR- et HR-iPKS, respectivement (d'après Weissman, 2010<sup>160</sup> et Wang et al., 2021<sup>162</sup>). Les réactions de cyclisation sont indiquées par une flèche de la couleur des domaines qui les catalysent. Les domaines colorés en gris sont inactifs (ΨKR, ΨER) ou leur rôle est inconnu (C).

SAT = unité d'initiation-ACP transacylase ; KS = kétosynthase ; AT = acyltransférase ; PT = *product template* ; ACP = protéine porteuse d'acyle ; TE = thioestérase ; KR = kétoréductase ; DH = déshydratase ; cMT = C-méthyltransférase ; C = condensation ; ER = énoylréductase.

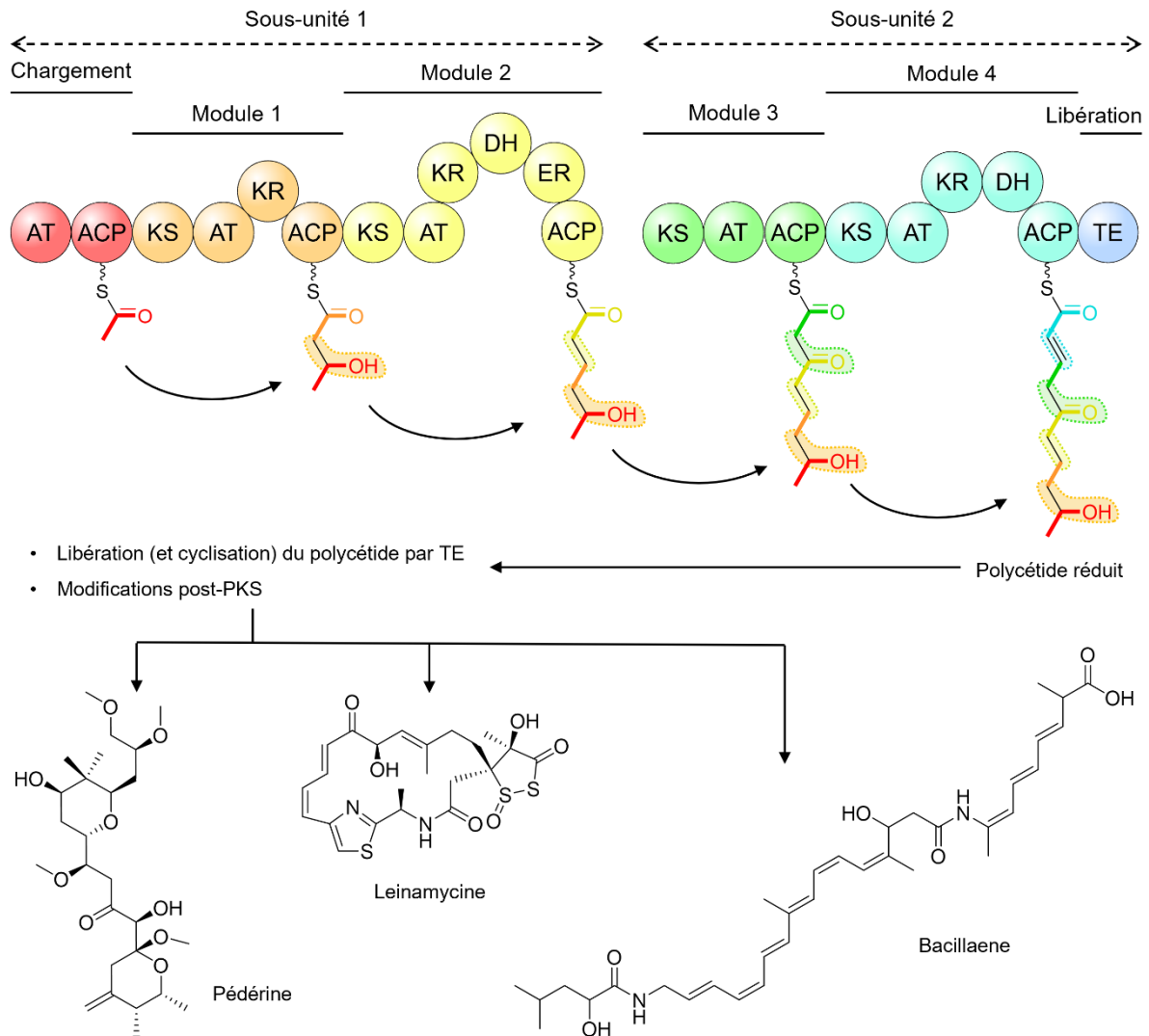
L'étude des voies de biosynthèse de polycétides bactériens a par ailleurs permis d'identifier dans ces microorganismes des iPKS impliquées dans la synthèse de polycétides aromatiques <sup>163</sup>, de PUFA <sup>32</sup>, d'ènediynes <sup>164</sup> et de macrolactames tétramates polycycliques (PTM) <sup>165</sup> (**Figure 25**), révélant encore davantage le potentiel biosynthétique de ces synthèses itératives.



**Figure 25 :** Métabolites appartenant aux classes des PUFA, composés aromatiques, PTM et ènediynes illustrant la diversité structurale des polycétides synthétisés par les iPKS bactériennes. Les éléments de structure synthétisés par une iPKS sont indiqués en rouge pour la frontalamide A et la lidamycine.

## b. Les PKS de type I modulaires

Contrairement aux PKS itératives, les PKS de type I modulaires sont composées de plusieurs modules, chacun responsable d'un unique cycle de condensation-modification et par conséquent possédant sa propre composition de domaines enzymatiques. Chaque module va ainsi sélectionner une unité d'extension, la condenser sur l'intermédiaire linéaire et la modifier en fonction des domaines optionnels présents avant de transférer le polycétide au module suivant à la manière d'une ligne d'assemblage (**Figure 26**). Cette synthèse par modules successifs rend prédictible la structure du polycétide généré si l'on a connaissance de la composition modulaire de la PKS puisque chaque domaine enzymatique catalyse une étape biosynthétique, une particularité connue sous le nom de « règle de colinéarité » <sup>166</sup>. L'architecture modulaire de ces PKS nécessite inéluctablement des protéines de hauts poids moléculaires qui atteignent fréquemment plusieurs mégadaltons d'où la qualification de mégasynthèses, les différents modules étant bien souvent répartis sur différents polypeptides appelés sous-unités <sup>155</sup>. Les PKS modulaires sont notamment responsables de la synthèse des macrolides, largement exploités pour leurs propriétés antibiotiques, anticancéreuses et immunosuppressives (voir partie I.2.c.ii.).

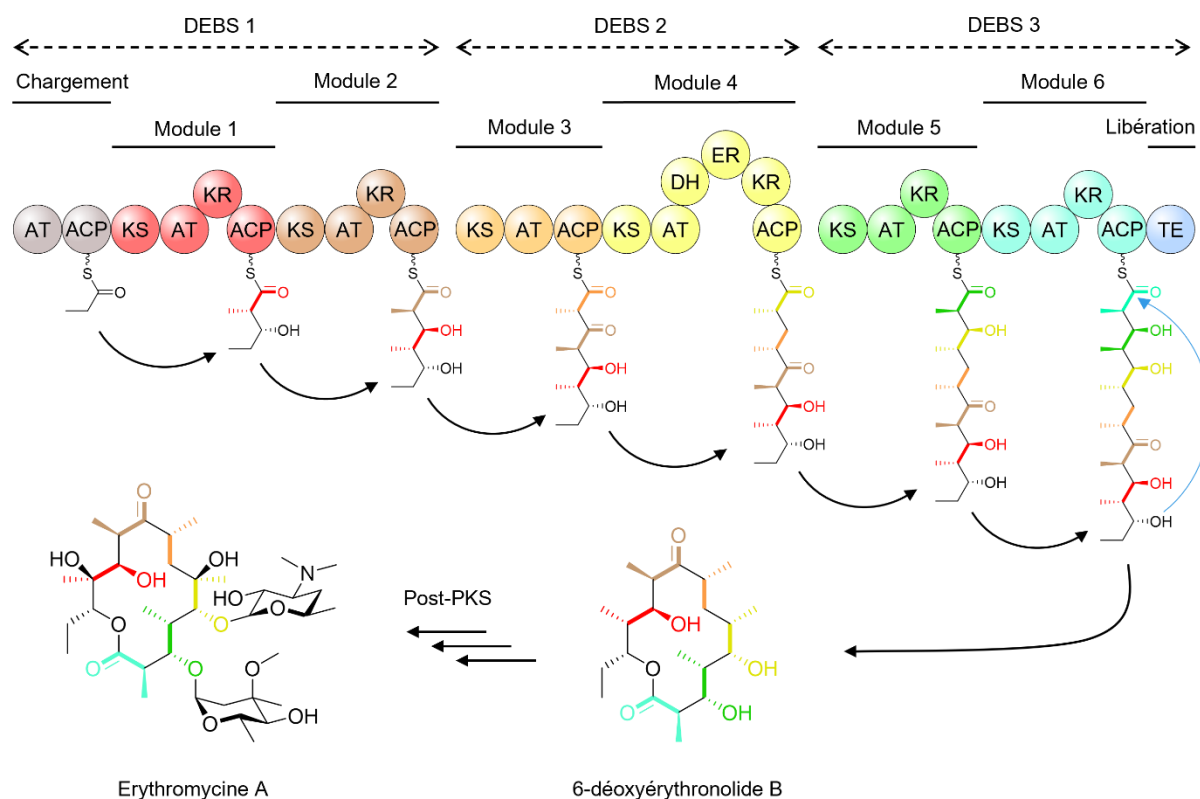


**Figure 26 :** Mécanisme catalytique des PKS de type I modulaires synthétisant des polycétides réduits, illustré par une PKS fictive. Ces PKS sont constituées de plusieurs chaînes polypeptidiques appelées sous-unités, composées de modules assurant chacun un cycle d'élongation et de modification de l'intermédiaire polycétidique linéaire. Chaque module est replié en domaines portant les activités enzymatiques. Les unités d'extension ajoutées et les positions  $\alpha/\beta$  de chaque unité ajoutée sont colorées et entourées en accord avec la couleur du module qui les condense et les modifie, respectivement (figure adaptée de Weissman, 2009<sup>35</sup>).

AT = acyltransférase ; ACP = protéine porteuse d'acyle ; KS = kétosynthase ; KR = kétoréductase ; DH = déshydratase ; ER = énoylréductase ; TE = thioestérase.

La 6-déoxyérythronolide B synthase (DEBS) biosynthétise le précurseur de l'érythromycine A, le 6-déoxyérythronolide B, et représente le modèle d'étude des PKS de type I modulaires puisqu'elle est la première de cette classe de synthase à avoir été découverte<sup>135,136</sup>. DEBS est constituée de sept modules répartis sur trois sous-unités nommées DEBS 1, DEBS 2 et DEBS 3 (**Figure 27**). Le premier module permet le recrutement de l'unité d'initiation propionate et est donc constitué uniquement d'un domaine AT et d'un ACP. Le monocétide est ensuite transféré au premier des six modules d'élongation successifs dont les domaines AT présentent une spécificité pour le méthylmalonate. La chaîne

polycétidique est ainsi acheminée de module en module où elle est systématiquement allongée par condensation d'un propionate issu de la décarboxylation des unités d'extension méthylmalonates, générant des intermédiaires linéaires méthylés en position  $\alpha$  et plus ou moins réduits en position  $\beta$  dépendamment de la présence de domaines KR, DH et ER. L'intermédiaire heptacétide est finalement libéré de la PKS par le dernier domaine TE et cyclisé par lactonisation pour former le 6-déoxyérythronolide B. L'intervention d'enzymes discrètes post-PKS va ensuite entraîner deux hydroxylations, deux glycosylations et une méthylation du macrolide, générant le produit de synthèse bioactif final, l'érythromycine A <sup>167</sup> (**Figure 27**).



**Figure 27 :** PKS de type I modulaire DEBS synthétisant le cœur aglycone de l'érythromycine A (figure reproduite de Weissman, 2016 <sup>167</sup>). DEBS catalyse la formation de 6-déoxyérythronolide B par condensation et modification chimique d'un propionyl-CoA et sept méthylmalonyl-CoA. L'heptacétide réduit est libéré de la PKS par le domaine terminal TE qui le cyclise également par une lactonisation indiquée par une flèche bleue. Le macrolactone résultant est ensuite hydroxylé et glycosylé par des enzymes discrètes post-PKS pour former le polycétide bioactif érythromycine A.

AT = acyltransférase ; ACP = protéine porteuse d'acyle ; KS = kétosynthase ; KR = kétoréductase ; DH = déshydratase ; ER = énoylréductase ; TE = thioestérase.

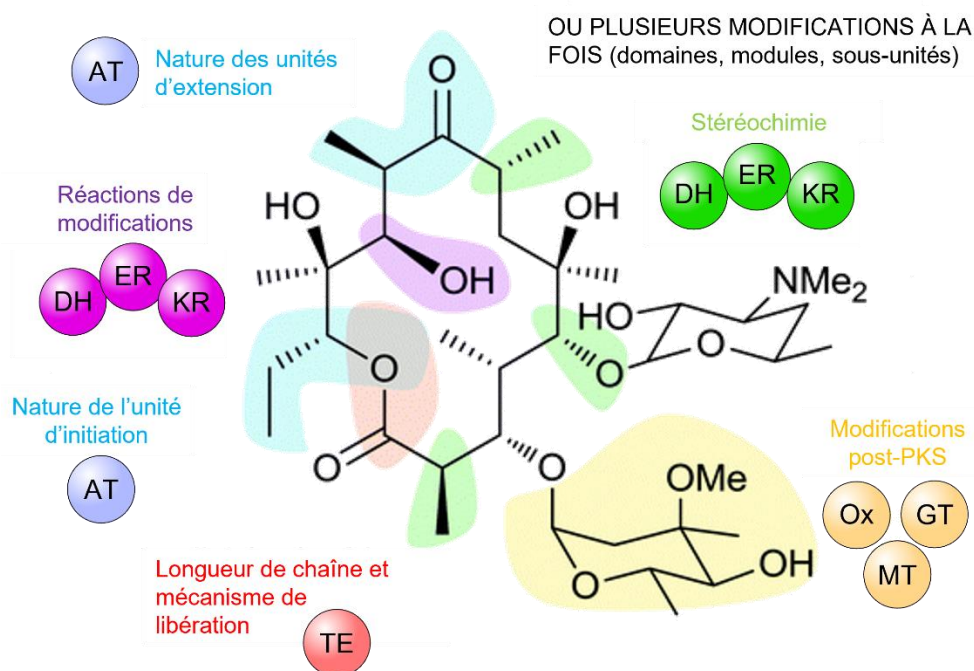
De manière surprenante, dès 1993, des systèmes PKS modulaires ne possédant pas de domaines AT au sein de leurs modules ont été identifiés <sup>168-170</sup>. En 2002, l'un de ces systèmes fut associé pour la première fois à la production d'un polycétide, la pédérine, prouvant que ces PKS sans AT pouvaient être fonctionnelles et deux gènes isolés ont été identifiés dans le même BGC codant potentiellement deux AT discrètes pouvant compléter

la PKS<sup>171</sup>. Ce mécanisme d'acylation en *trans* fut ensuite prouvé *in vitro* dans la voie de biosynthèse de la leinamycine<sup>172</sup> donnant ainsi naissance à une nouvelle classe de PKS modulaires dites *trans*-AT, par opposition aux PKS modulaires prototypiques *cis*-AT telles que DEBS. Dans ces PKS *trans*-AT, les enzymes AT discrètes interagissent de manière itérative avec chaque module pour en acyler l'ACP. Du fait de la spécificité des AT, la même unité d'extension est sélectionnée pour chaque module, un malonyl-CoA dans la plupart des cas. Des AT aux spécificités différentes sont par exemple présentes dans les voies de biosynthèse de la kirromycine (éthylmalonyle)<sup>173</sup>, de l'oxazolomycine (méthoxymalonyle)<sup>174</sup> et de l'étnangiène (succinyle)<sup>175</sup>. En plus de la localisation de l'activité acyltransférase, les PKS *trans*-AT présentent une diversité architecturale bien plus importante que les systèmes *cis*-AT. Cette diversité inclut des modules avec un ordre de domaines inhabituel ou des domaines à l'activité atypique, des modules NRPS permettant d'incorporer des acides aminés, des modules scindés sur deux sous-unités ou encore des activités enzymatiques apportées en *trans* autres que AT. À cela s'ajoutent des particularités fonctionnelles avec l'action itérative de certains modules ou au contraire l'inactivité d'autres<sup>176</sup>. Toutes ces particularités rendent la prédiction du produit de synthèse des PKS *trans*-AT moins évidente, de sorte que la règle de colinéarité des systèmes *cis*-AT ne s'applique pas. La voie de biosynthèse de la batumine illustre la complexité de cette classe de synthases<sup>176</sup> (**Figure 28**). Toutefois, des analyses phylogénétiques de PKS *trans*-AT ont révélé que les domaines KS forment des clades qui corrélerent fortement avec la structure de leurs substrats<sup>177</sup>. Basés sur ces « clades kétides », de puissants outils bioinformatiques ont été développés, permettant la prédiction de structures de polycétides (TransATor<sup>118</sup>), et l'identification de modules consécutifs identiques dans différentes PKS et par conséquent, de sous-structures chimiques communes à différents polycétides (*trans*PACT<sup>178</sup>).



### c. Pertinence de la modularité pour l'ingénierie génétique

La répartition du labeur enzymatique en différents domaines, modules et sous-unités rend les PKS modulaires particulièrement attrayantes pour des approches d'ingénierie génétique visant à modifier ces voies de biosynthèse pour générer de nouvelles structures polycétidiques potentiellement bioactives. En effet, Il est possible, en principe, de manipuler un domaine spécifique afin de l'inactiver, de l'échanger pour un autre domaine ou de l'ajouter à un module pour en étendre la chimie. Les modules et sous-unités entières peuvent être également la cible des efforts d'ingénierie puisque les déléter peut permettre de réduire la taille du polycétide, les ajouter peut produire l'effet inverse et les échanger modifie plusieurs positions du polycétide simultanément. Les enzymes post-PKS peuvent être ciblées de la même manière. Ainsi, en modifiant chaque caractéristique des PKS modulaires, toutes les positions du produit de synthèse peuvent théoriquement être modifiées de manière spécifique et prédictible <sup>166</sup> (**Figure 29**).



**Figure 29 :** Modifications potentielles de polycétides par ingénierie génétique. L'érythromycine A est prise pour exemple. Les modifications sont colorées en accord avec les domaines ou enzymes discrètes qui les catalysent (figure reproduite de Weissman, 2016 <sup>167</sup>).

AT = acyltransférase ; DH = déshydratase ; ER = énoylréductase ; KR =kétoréductase ; TE = thioestérase ; Ox = oxygénase ; GT = glycosyltransférase ; MT = méthyltransférase.

Ce concept de biosynthèse combinatoire a permis la synthèse de nombreux analogues de polycétides naturels. Il est par conséquent impossible de dresser une liste exhaustive des composés générés par cette approche mais un cas concret illustrant chaque type de stratégie est décrit ci-après. Par exemple, le remplacement des domaines DH et KR du module 2 de la



PKS de l'avermectine par le trio réducteur DH-ER-KR provenant de la PKS de la rapamycine a permis la biosynthèse de l'analogue réduit ivermectine, traditionnellement obtenu par hémisynthèse <sup>180</sup> (**Figure 30**). La suppression d'un module dans une autre PKS *cis*-AT, celle du tétraène néoaureothine, a raccourci le processus de biosynthèse d'un cycle d'extension-modification, conduisant ainsi à la formation du triène homoaureothine <sup>167</sup> (**Figure 30**). Quant à l'ingénierie des sous-unités entières, elle a par exemple été appliquée pour remplacer la troisième sous-unité de la PKS de la pikromycine par celle de DEBS conduisant à la synthèse d'un nouveau macrolide <sup>181</sup> (**Figure 30**).

Les PKS modulaires *trans*-AT représentent un intérêt important pour ces approches d'ingénierie génétique du fait de leur énorme diversité fonctionnelle et de la nature discrète de certains de ces composants, les rendant potentiellement plus facilement modifiables que des domaines enzymatiques au sein de polypeptides de haut poids moléculaire <sup>182</sup>. Les sous-unités et modules *trans*-AT semblent en principe également plus appropriés pour ces approches compte tenu de l'implication intensive proposée d'évènements de transferts de gènes horizontaux dans leur mode d'évolution <sup>177</sup>, faisant de ces lignes d'assemblage des mosaïques biosynthétiques. Au contraire, les PKS *cis*-AT ont vraisemblablement évolué par duplication de gènes <sup>183</sup>. Un modèle alternatif d'évolution a toutefois été récemment suggéré par Nivina *et al.* dans lequel les deux types de systèmes ont tous deux d'abord évolué par recombinaison, avant que les PKS *trans*-AT n'évoluent par perte des domaines AT et les PKS *cis*-AT par conversion de gènes entre domaines KS paralogues <sup>184</sup>. Ce dernier modèle n'explique cependant pas toute la diversité organisationnelle des systèmes *trans*-AT par rapport aux *cis*-AT.

Les performances de la biosynthèse combinatoire ne sont cependant plus aussi prometteuses que ce qui était escompté. Les modifications génétiques apportées aux voies de biosynthèse se révèlent bien souvent très délétères pour les rendements de production des polycétides voire inactivent totalement leurs synthèses <sup>167</sup>. Ces résultats traduisent le fait que les domaines, modules et sous-unités de PKS ne sont pas de simples blocs interchangeables contrairement aux espoirs ayant initié les efforts de biosynthèse combinatoire. En effet, la spécificité des domaines pour le produit de synthèse des modules en amont et des domaines au sein de leurs propres modules, les séquences *linkers* interdomaines et les interfaces entre domaines et entre modules sont des éléments clefs de la machinerie enzymatique des systèmes modulaires, potentiellement impactés par l'ingénierie génétique. Néanmoins, la biologie synthétique des PKS demeure une approche prometteuse pour la génération de dérivés spécifiques de polycétides connus à condition de comprendre les facteurs essentiels à la spécificité des PKS et à l'acheminement du polycétide le long de la synthase ainsi que les diverses activités enzymatiques inhabituelles régulièrement mises en lumière. L'étude de la structure et de la fonction des domaines, modules et sous-unités des PKS apparaît alors

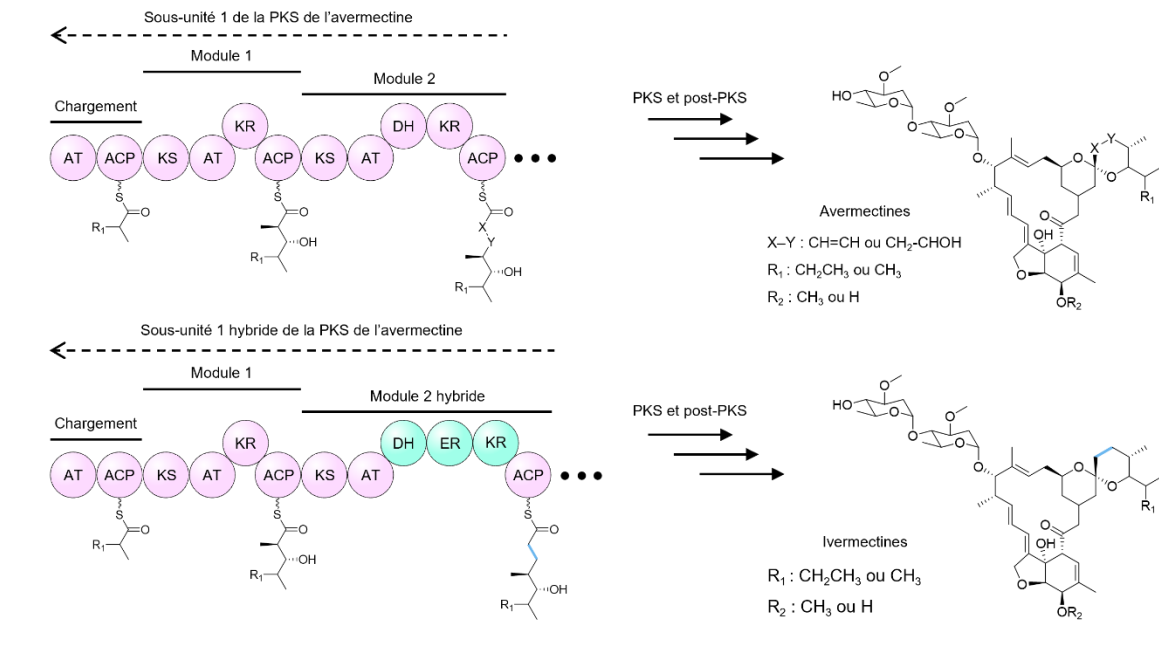
## État de l'art

essentielle pour éclairer ces zones d'ombre, un objectif poursuivi par l'équipe du Pr. Kira Weissman, équipe au sein de laquelle ce doctorat fut effectué.

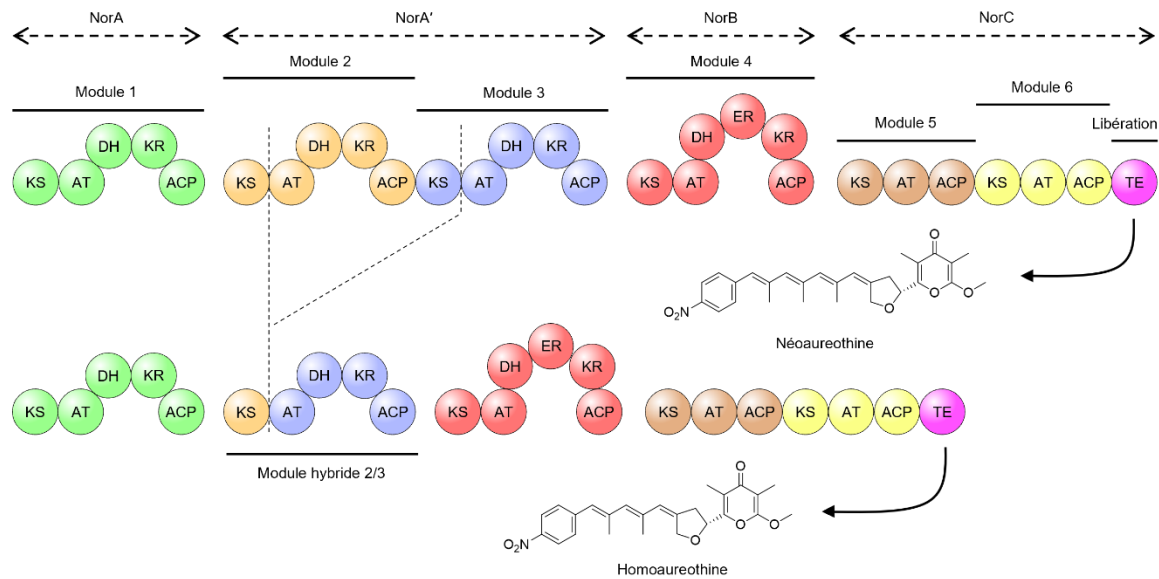
*“PKS synthetic biology holds significant promise for obtaining designer polyketides. However, realizing the true potential of this approach depends on further deepening our understanding of the molecular genetics, structure and function of both cis- and trans-AT PKSs...”*

Kira J. Weissman <sup>167</sup>

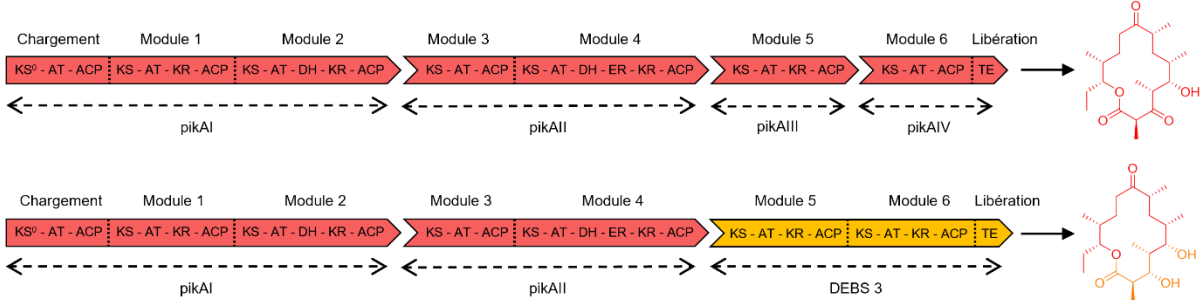
**Ingénierie de domaines (gain de fonction)**



**Ingénierie de modules (raccourcissement de chaîne polycétidique)**



**Ingénierie de sous-unités (hybridation des voies de biosynthèse)**



**Figure 30 :** Exemples d'ingénierie de PKS modulaires pour la synthèse de nouvelles structures polycétidiques par ciblage de domaines, modules ou sous-unités spécifiques (figure reproduite de Weissman et Leadlay, 2005<sup>166</sup> et Weissman, 2016<sup>167</sup>).

## IV. Déterminer la structure pour appréhender la fonction

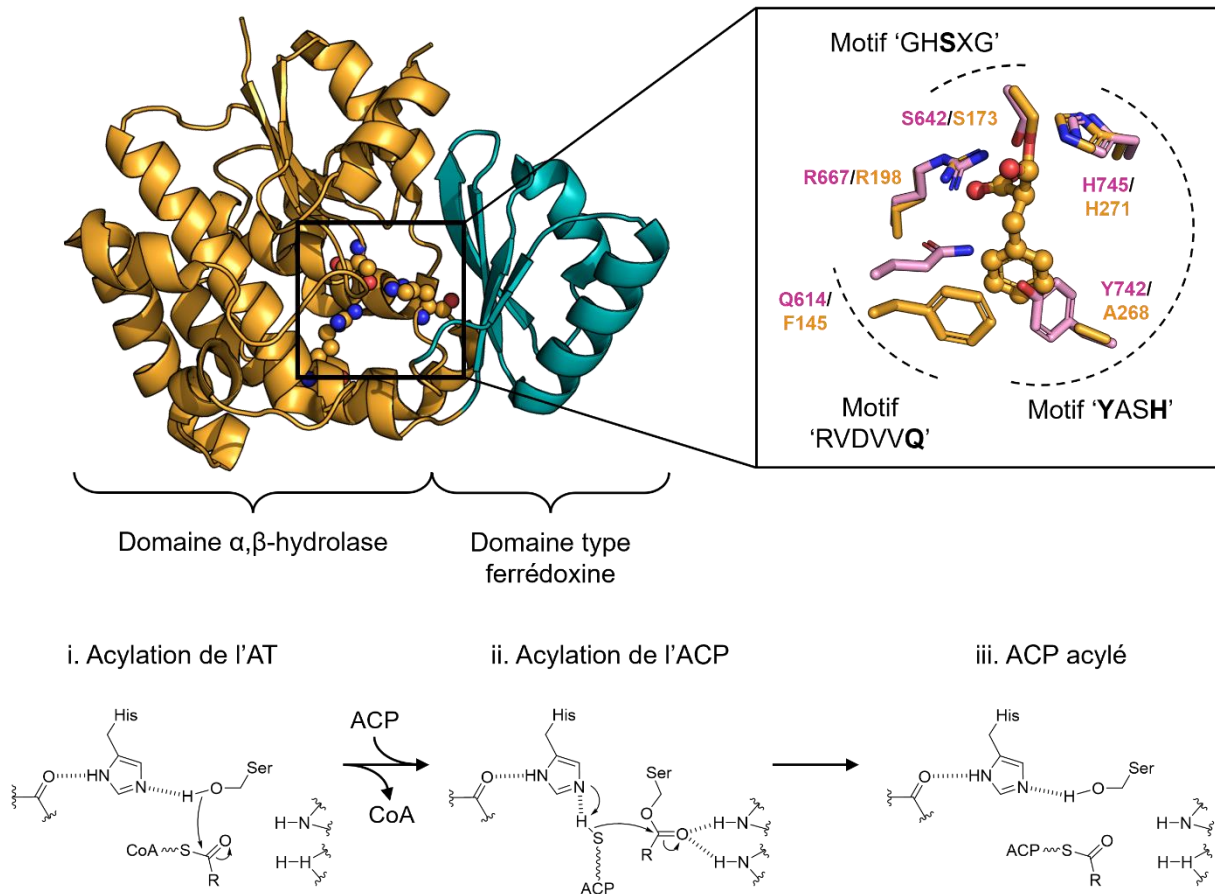
De nombreuses études structurales ont été menées sur les différents domaines PKS afin d'en élucider les relations structure/fonction et de comprendre les interactions établies avec leurs substrats et partenaires protéiques.

### 1. Les domaines canoniques de PKS

#### a. Les domaines acyltransférases

Les domaines AT constituent le premier acteur de chaque cycle d'extension au sein des modules PKS puisqu'ils sont responsables de la sélection spécifique d'une unité d'initiation/d'extension acyl-CoA, et de son transfert à un domaine ACP. Dans les systèmes *cis*-AT, ces domaines sont retrouvés dans chaque module après un domaine KS excepté dans le module de chargement où l'activité de condensation n'est pas requise (bien que certains modules de chargement contiennent tout de même un domaine KS à activité uniquement décarboxylative (voir la section IV.1.c.)). Au contraire dans les systèmes *trans*-AT, l'activité acyltransférase est portée par des enzymes discrètes qui vont interagir avec chaque module. Les structures cristallographiques de plusieurs AT faisant partie de différentes PKS ont été résolues, notamment les didomains KS-AT des modules 3 et 5 de la PKS *cis*-AT prototypique DEBS<sup>185,186</sup>. Les AT sont des domaines monomériques d'environ 300 résidus composés de deux domaines structuraux, un domaine principal au repliement  $\alpha,\beta$ -hydrolase et un domaine d'une soixantaine de résidus au repliement de type ferrédoxine. Le site actif est localisé à l'interface de ces deux domaines et possède une dyade catalytique sérine-histidine portée par le domaine principal (**Figure 31**). Dans le cas d'une unité d'extension carboxylate (malonate par exemple), ce groupement est stabilisé par une arginine conservée. La caractérisation des AT discrètes des PKS *trans*-AT de la vicénistatine en complexe avec un ACP et du disorazole<sup>187-189</sup> ont révélé que ces caractéristiques structurales et fonctionnelles étaient partagées entre les domaines AT des PKS *cis*-AT et les AT discrètes des PKS *trans*-AT.

La réaction enzymatique catalysée par les AT est similaire à celle des protéases à sérine<sup>190</sup>. La sérine catalytique attaque le carbonyle du substrat pour former l'intermédiaire acyl-enzyme et le second produit de réaction (ici, l'acyl-ACP) est généré par attaque d'un nucléophile externe à l'enzyme. L'histidine catalytique permet l'activation des groupements nucléophiles et la stabilisation des groupes partants (**Figure 31**). Cependant, à la différence des protéases à sérine, les deux réactions consécutives catalysées par l'AT impliquent la formation non pas d'intermédiaires tétraédriques mais d'états de transition tétraédriques et l'étape finale d'hydrolyse est dans ce cas une réaction de transacylation de l'AT vers l'ACP (**Figure 31**).



**Figure 31 :** Structure du domaine SpnD-AT de la PKS-NRPS de la splénocine (PDB : 5YDM)<sup>191</sup> et mécanisme d'acylation. La dyade catalytique et l'arginine conservées dans le site actif à l'interface des deux domaines structuraux sont représentées en sphères. La superposition des résidus clés de la spécificité de substrat de SpnD-AT (en orange) et l'AT5 de DEBS (PDB : 2HG4) (en rose) montre que la divergence des motifs 'RVDVVQ' et 'YASH' chez SpnD-AT permet d'accommoder des substrats encombrants tels que le benzylmalonyl-CoA (représenté en sphères) lié à la sérine catalytique. Le mécanisme d'acylation de l'AT puis de transfert de l'unité à l'ACP est adapté de Tsai et Ames, 2009<sup>192</sup> par Kira J. Weissman. Une fonction carbonyle de la chaîne principale de l'enzyme stabilise probablement l'histidine catalytique, par analogie au rôle d'un aspartate conservé dans les protéases à sérine.

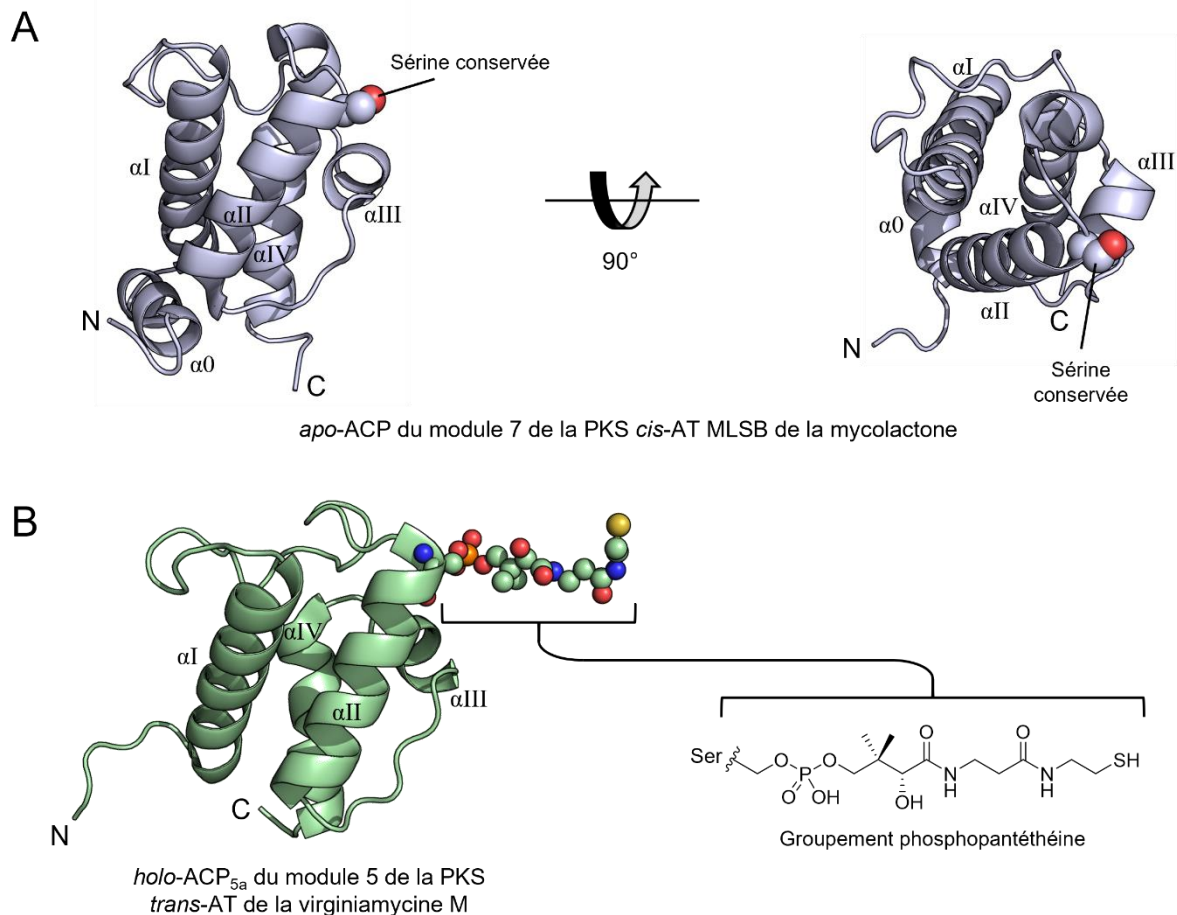
Les domaines AT des modules de chargement ont une spécificité de substrat variable et peuvent dans certains cas accepter une grande variété d'unités d'initiation non natives telles que l'acétyl-, le malonyl-, le butyryl-, le crotonyl-, le phénylacétyl-, et l'hydroxybutyryl-CoA<sup>193</sup> tandis que les domaines AT au sein des modules d'extension présentent une spécificité de substrat importante régit par trois principaux éléments de séquence<sup>194</sup> : 1) un motif situé au sein d'une hélice  $\alpha$  de 35 Å formant le site de fixation du substrat et qui est défini par les six résidus 'RVDVVQ' ou plus largement [RQSED]V[DE]VVQ dans les AT spécifiques pour le méthylmalonyl-CoA et ZTx\$[AT][QE] où Z est hydrophile et \$ est aromatique dans les AT spécifiques pour le malonyl-CoA ; 2) le motif 'GHSxG' incluant la sérine catalytique où la nature du résidu x adjacent participe à la détermination de la spécificité de l'AT. Un acide aminé hydrophobe tel qu'une valine ou isoleucine indique une spécificité pour le malonate tandis

qu'une glutamine ou une méthionine sont retrouvées dans les autres AT<sup>195</sup> ; et 3) un déterminant de spécificité principal, le motif 'YASH' situé 100 résidus en aval de la sérine catalytique. Plus précisément, les AT méthylmalonate-spécifiques possèdent un motif YASH tandis que les AT malonate-spécifiques présentent un motif HAFH<sup>194</sup>. La résolution récente de la structure cristallographique du domaine AT de la synthase hybride PKS-NRPS de la splénocine a permis de confirmer et préciser le rôle des motifs 1) et 3) dans la spécificité de substrat des AT. Contrairement à la plupart des domaines AT de modules d'extension, ce domaine AT, SpnD-AT, accepte une gamme d'unité d'extension plus large que les classiques malonyl- et méthylmalonyl-CoA avec des chaînes aliphatiques contenant jusqu'à sept carbones et l'unité aromatique benzylmalonyl-CoA<sup>196</sup>. Dans cette AT, la glutamine du motif RVDVVQ est une phénylalanine et l'encombrement stérique généré par le motif YASH est remplacé par un motif AAAH (**Figure 31**). La mutagenèse dirigée de l'AT6 de DEBS au niveau de ces trois résidus identifiés dans SpnD-AT a permis de valider leurs rôles et de modifier la tolérance de l'AT6 envers divers substrats non natifs sans pour autant en modifier totalement la spécificité<sup>191</sup>. En effet, les différents motifs détaillés ci-dessus sont impliqués dans la spécificité de substrat des AT et leurs mutations permettent d'étendre la spécificité de ces domaines, mais elles ne permettent pas à elles-seules de permuter complètement cette spécificité<sup>197,198</sup>. Celle-ci semble ainsi être une propriété inhérente au domaine entier, expliquant le succès plus important des approches d'échanges d'AT par rapport aux approches de mutagenèse pour la génération de nouveaux polycétides<sup>199</sup>.

### b. Les domaines porteurs d'acyles

Bien que ne possédant pas d'activité enzymatique propre et n'étant constitué que de 80 à 100 résidus, le domaine ACP constitue la plateforme centrale du module. En effet, les ACP portent les unités d'extension (ou d'initiation selon le module considéré) transacylées par l'AT et les intermédiaires de synthèse générés par action du domaine KS en amont sur l'unité d'extension. Ce domaine d'une dizaine de kilodaltons joue donc le rôle d'intermédiaire entre les différents domaines catalytiques d'un module, qui interagissent séquentiellement avec lui pour modifier la chaîne polycétidique en cours de synthèse.

La structure tridimensionnelle des ACP consiste en un fagot de quatre hélices  $\alpha$ , auquel s'ajoute une courte hélice N-terminale  $\alpha_0$  augmentant la stabilité du domaine et prédite dans 98% des ACP de modules d'extension de PKS *cis*-AT<sup>200</sup> (**Figure 32**). Une chaîne latérale sérine conservée située à l'extrémité N-terminale de l'hélice  $\alpha_{II}$  est modifiée post-traductionnellement par une PPTase qui catalyse l'ajout d'un groupement Ppant à partir de CoA<sup>150</sup>. Ce bras Ppant constitue le point d'ancrage des unités d'initiation/extension et de la chaîne polycétidique grâce à sa fonction thiol terminale (**Figure 32**).



**Figure 32 :** Structures de domaines ACP résolues par résonance magnétique nucléaire. A) Structure de l'ACP sous forme *apo* du module 7 de la PKS *cis*-AT MLSB synthétisant la queue polyène du polycétide mycolactone<sup>200</sup>. Quatre hélices  $\alpha$  forment un fagot ( $\alpha I$ – $\alpha VI$ ) stabilisé davantage par une hélice de dix résidus à son extrémité N-terminale ( $\alpha 0$ ). La sérine sur laquelle est transféré le groupement Ppant est située à l'extrémité N-terminale de l'hélice  $\alpha II$  est représentée en sphères. B) Structure de l'ACP<sub>5a</sub> sous forme *holo* du module 5 de la PKS *trans*-AT synthétisant la virginiamycine M<sup>201</sup>. La sérine conservée porte un groupement phosphopantéthéine, représenté en sphères, provenant d'une molécule de CoA et transféré par une PPTase. Sous cette forme *holo* active, l'ACP est capable de lier une unité d'extension et la chaîne polycétidique après condensation.

### c. Les domaines kétosynthases

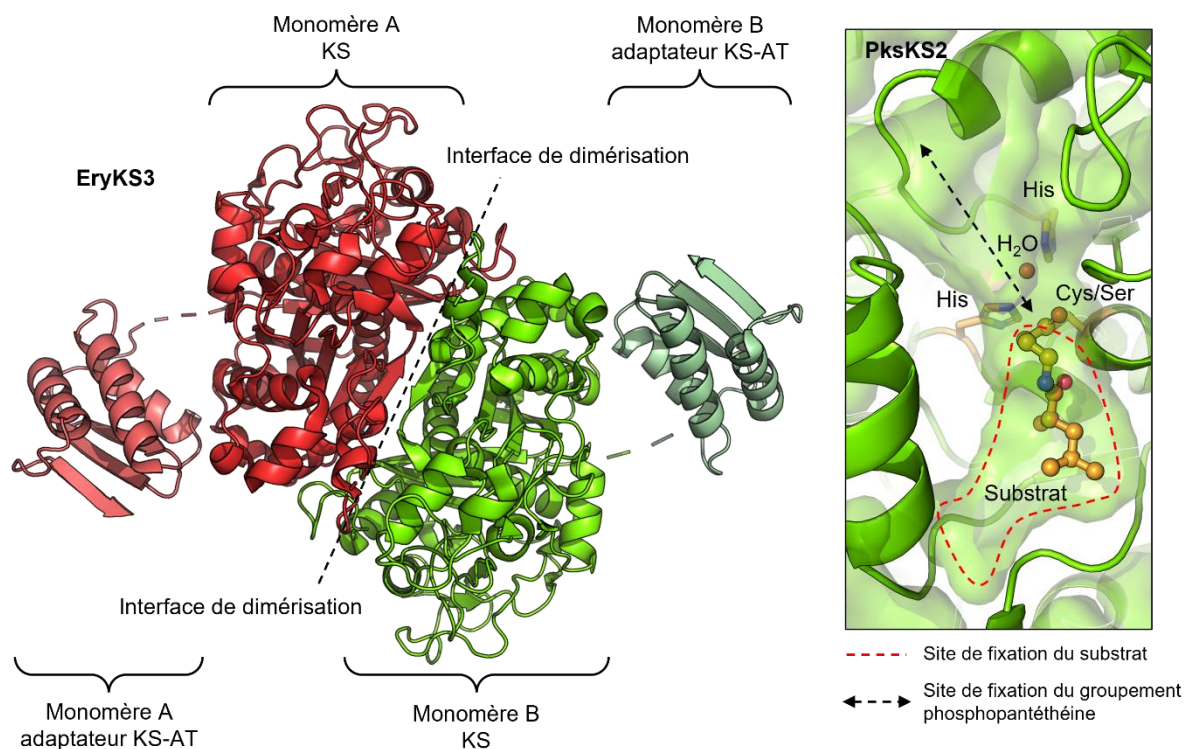
Les domaines KS sont constitués d'un peu plus de 400 acides aminés, ils reçoivent l'intermédiaire de synthèse provenant de l'ACP du module en amont et catalysent la condensation décarboxylative de la nouvelle unité d'extension sur cet intermédiaire (**Figure 17**). La chaîne polycétidique est ainsi transférée de la KS d'un module à l'ACP du même module. Les domaines KS sont donc responsables de l'élongation du polycétide mais également de sa translocation entre les différents modules de la PKS, par transfert d'un ACP au suivant.

Leur structure présente un repliement caractéristique de la famille enzymatique des thiolases, constituée de cinq strates alternant hélices  $\alpha$  et feuillets  $\beta$  pour former un cœur

structuré en  $2\alpha$ - $5\beta$ - $2\alpha$ - $5\beta$ - $2\alpha$ . En C-terminal, un domaine structural additionnel d'une centaine de résidus permet l'interaction entre la KS et l'AT en aval et est qualifié d'adaptateur KS-AT. Malgré l'absence de domaine AT au sein des modules des systèmes *trans*-AT, ce domaine structural post-KS y est conservé<sup>202</sup>. Dans ce cas, le linker C-terminal de l'adaptateur revient au contact de la KS, permettant ainsi de stabiliser l'ensemble. Les KS sont homodimériques et apparaissent comme l'élément principal gouvernant la dimérisation de modules et sous-unités entières (voir partie IV.2.b.). En effet, bien que d'autres domaines soient dimériques (DH, TE), ceux-ci ne sont pas présents dans tous les modules contrairement aux KS.

Les réactions d'acylation et de condensation décarboxylative catalysées par les KS font intervenir une triade catalytique Cys-His-His<sup>137</sup> (**Figure 33**) dont les résidus sont retrouvés au sein des motifs TxCSSS, HGTGT et KSNIGH, respectivement<sup>202</sup>. La cystéine est située à l'extrémité N-terminale d'une hélice  $\alpha$  dont le moment dipolaire globale permet d'augmenter la nucléophilie de sa chaîne latérale<sup>194</sup>. Cette cystéine peut alors attaquer le thioester de l'intermédiaire polycétidique porté par l'ACP en amont pour le transférer à la KS. L'ACP libre quitte alors le site actif pour être remplacé par l'ACP en aval, acylé par l'unité d'extension. La condensation décarboxylative survient ensuite telle que décrit en partie III.3. à l'aide des chaînes latérales des deux histidines conservées (**Figure 17**). Par ailleurs, la charge négative portée par l'intermédiaire tétraédrique formé lors des réactions de transfert du polycétide de l'ACP vers la KS et de condensation de la KS vers l'ACP est stabilisée par l'amide des chaînes principales de la cystéine catalytique et d'un résidu proche formant le trou de l'oxyanion<sup>137</sup>. La région du site actif allant de la surface de la KS à la cystéine catalytique possède une composition en acides aminés conservée entre les différentes KS indiquant son rôle de fixation du groupement phosphopantéthéine<sup>194</sup>. À l'inverse, les résidus qui forment la cavité allant de la cystéine catalytique à l'interface de dimérisation de la KS accommodent l'intermédiaire linéaire et présentent une diversité de résidus plus importante d'une KS à l'autre, indiquant le rôle de cette région dans la fixation spécifique de certains intermédiaires<sup>194</sup> (**Figure 33**).





**Figure 33 :** Structures cristallographiques de domaines KS. La structure de la KS du module 3 de DEBS (à gauche) permet d'observer le caractère homodimérique du domaine ainsi que la présence C-terminale du petit domaine adaptateur KS-AT faisant le lien avec le domaine AT en aval, présent dans la structure cristallographique complète<sup>203</sup> (PDB : 2QO3). Le site actif des KS (à droite) fait intervenir une triade catalytique Cys-His-His pour catalyser le transfert de l'intermédiaire, la décarboxylation de l'unité d'extension et la condensation finale. Un gros plan sur le site actif de la KS du module 2 de la PKS *trans*-AT de la bacillaene<sup>202</sup> (PDB : 4NA2) permet de visualiser la triade catalytique ainsi qu'une molécule d'eau à proximité des deux histidines conservées, en accord avec le mécanisme de décarboxylation de l'unité d'extension<sup>137</sup>. Le substrat naturel de la KS2 est acylé à la cystéine catalytique, qui a préalablement été mutée en sérine pour stabiliser l'intermédiaire covalent.

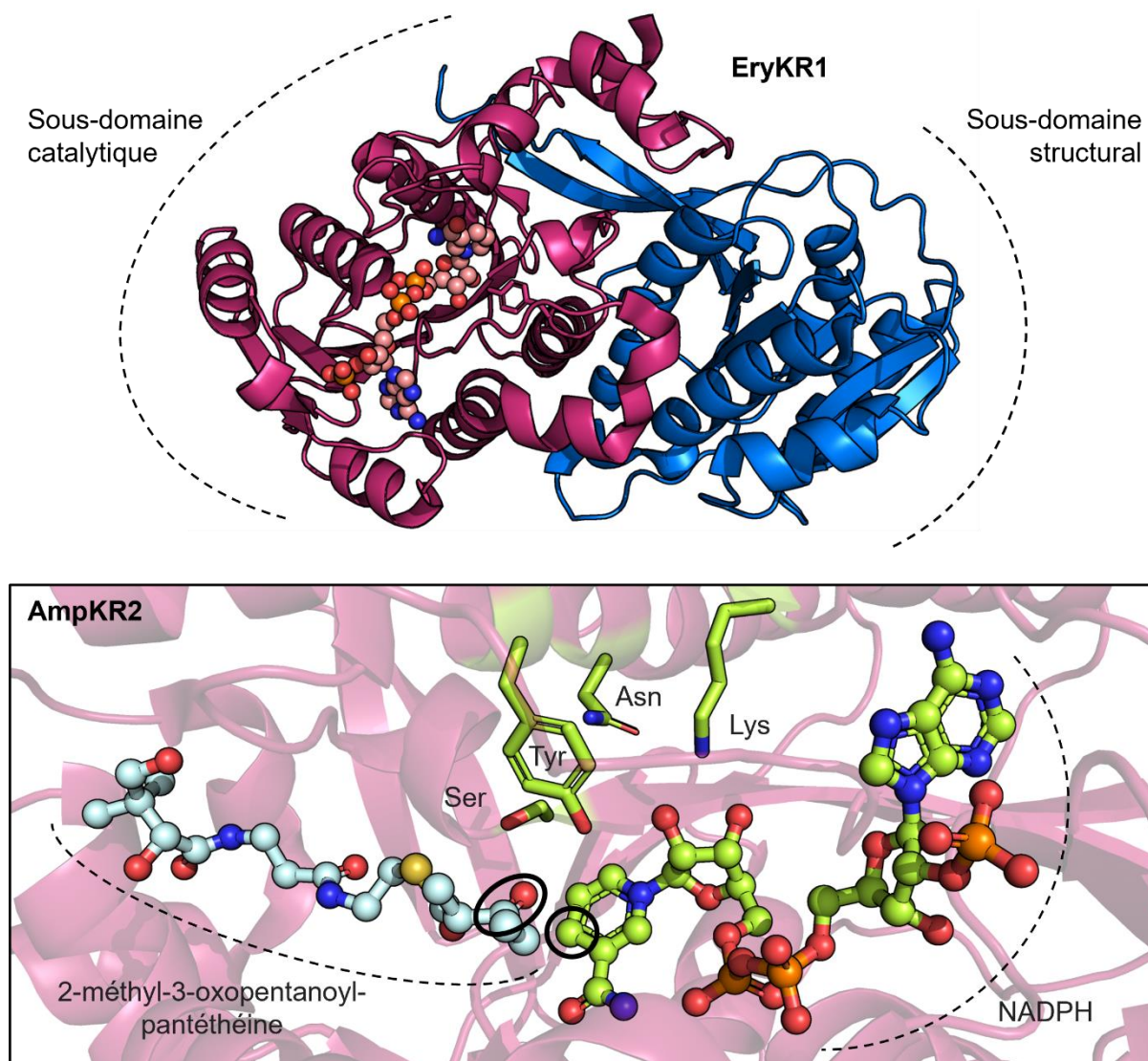
Certaines PKS présentent dans leur module de chargement une KS dont la cystéine catalytique est remplacée par une glutamine. Ces KS<sup>Q</sup> sont incapables de catalyser une condensation mais n'en sont pas pour autant catalytiquement inactives. À l'image des KS<sub>β</sub> des PKS de type II, la KS<sup>Q</sup> d'un module de chargement est capable de décarboxyler une unité d'extension générant de ce fait une unité d'initiation pour démarrer la synthèse du polycétide. La mutation de la cystéine en glutamine aurait pour conséquence de mimer la présence du substrat dans le site actif, induisant une conformation de la KS compétente pour la décarboxylation. Ces domaines KS<sup>Q</sup> jouent donc un rôle de facteur d'initiation de chaîne<sup>204</sup>. On trouve également au sein des PKS *trans*-AT des modules non-élongateurs possédant pourtant un domaine KS. Celui-ci, désigné KS<sup>0</sup>, a pour particularité de ne pas posséder l'histidine conservée du motif HGTGT, ce qui rend impossible la condensation décarboxylative<sup>176</sup>. En revanche, à l'inverse des KS<sup>Q</sup>, les KS<sup>0</sup> possèdent la cystéine catalytique et conservent donc la fonction de translocation de l'intermédiaire d'un ACP au

suivant. La KS<sup>0</sup> inhibe alors l'activité *trans*-acyltransférase sur l'ACP de son module afin de pouvoir transférer l'intermédiaire sur le thiol libre du Ppant<sup>205</sup>.

#### d. Les domaines kétoréductases

Les KR sont à l'origine de la majorité des centres stéréogènes des polycétides. Ce sont des domaines monomériques de plus de 400 résidus composés de deux sous-domaines structuraux présentant chacun un repliement de type Rossmann associé à la fixation de dinucléotides (**Figure 34**). Seul le sous-domaine C-terminal est cependant capable de fixer le cofacteur nicotinamide adénosine dinucléotide phosphate (NADPH) et porte donc l'activité catalytique. Vraisemblablement, le sous-domaine N-terminal joue un rôle structural en stabilisant l'ensemble de la KR<sup>206</sup>. Le premier et le dernier brin  $\beta$  du sous-domaine structural relie les motifs Rossmann des deux sous-domaines, créant un feuillet  $\beta$  continu au sein de la KR. La résolution de la structure de KR issues de différentes PKS a mis en évidence la conservation de l'arrangement structural et du mode de fixation du cofacteur entre les systèmes *cis*- et *trans*-AT<sup>206-208</sup>.

Les domaines KR catalysent la réduction de fonctions cétones en position  $\beta$  des intermédiaires polycétidiques après extension de la chaîne. Ils font partie de la large famille enzymatique des *short-chain dehydrogenase/reductase* (SDR) et possèdent par conséquent une tétrade catalytique Tyr-Ser-Lys-Asn conservée, caractéristique des SDR<sup>209</sup>. La tyrosine et la sérine se lient au groupement  $\beta$ -cétone à réduire par leurs hydroxyles terminaux et la chaîne latérale chargée de la lysine permet de stabiliser la forme négativement chargée de la tyrosine. Le pK<sub>a</sub> de la tyrosine est ainsi baissé, ce qui facilite le transfert de proton à l'oxygène du substrat<sup>210</sup>. Un mécanisme alternatif propose l'implication d'un relais de protons permettant à la lysine de reprotoneur la tyrosine par l'intermédiaire du D-ribose nicotinamidique du cofacteur NADPH<sup>211</sup>. L'asparagine conservée possède quant à elle un rôle organisationnel supposé, du fait de sa proximité avec les autres résidus mentionnés<sup>210</sup> (**Figure 34**). La présence du cofacteur NADPH permet le transfert de l'hydride en position 4'-*pro-S* du nicotinamide au carbone  $\beta$  du substrat<sup>212,213</sup>, conduisant à la formation d'un carbone asymétrique à la stéréochimie bien définie.



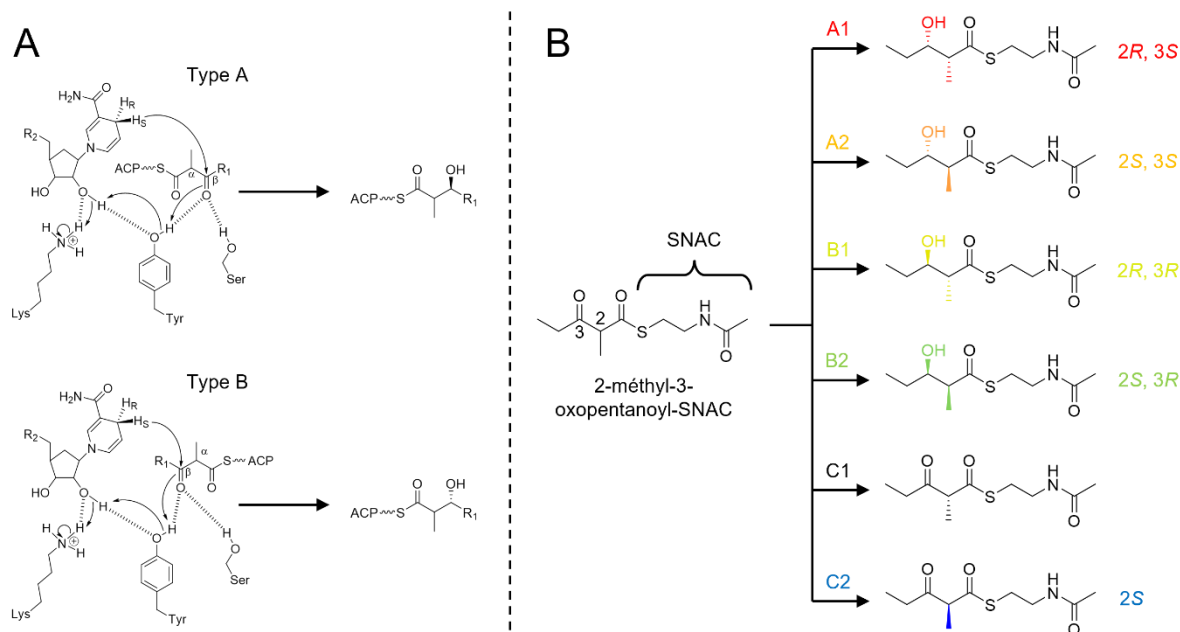
**Figure 34 :** Structure d'un domaine KR et organisation du site actif. La première structure de domaine KR résolue <sup>206</sup> (EryKR1, en haut) révèle le repliement en deux sous-domaines de type Rossmann dont un porte le site actif (en violet) contenant le cofacteur NADPH représenté en sphères (PDB : 2FR0). La résolution de la structure quaternaire de la KR AmpKR2 <sup>214</sup> a permis d'observer l'organisation du site actif d'une KR en présence d'un analogue de substrat  $\beta$ -kétocyle lié à un groupement Ppant. Les résidus de la tétrade catalytique sont présents à l'interface du cofacteur et du substrat permettant leur stabilisation et le transfert de proton(s). Le  $\beta$ -carbonyle ciblé par la kétoréduction et le carbone du nicotinamide portant l'hydride 4'-*pro-S* sont entourés en noir (PDB : 5XWV).

La configuration du  $\beta$ -carbone asymétrique générée dépend de la direction de la kétoréduction et donc de la stéréosélectivité de la KR. On les distingue ainsi en deux classes, A et B <sup>215</sup>, qui génèrent un groupement  $\beta$ -hydroxyle sur un carbone de configuration *S* ou *R*, respectivement. Des motifs de séquence conservés sont associés à l'une ou l'autre de ces classes et permettent d'en prédire la stéréochimie <sup>215</sup>. Le motif Leu-Asp-Asp, parfois couplée à la présence des résidus Pro144 et Asn148, au sein d'une boucle dite « couvercle » près du site actif est caractéristique des KR de type B. Au contraire, l'absence de ce motif et la présence d'un tryptophane en position 141 suggèrent une KR de type A (résidus numérotés

tels que Caffrey, 2003<sup>215</sup>). Malgré ces différences entre types de KR, tous possèdent la tétrade catalytique conservée et fixent le cofacteur NADPH dans la même orientation, suggérant que les deux directions alternatives de kétoréduction résultent de deux modes de fixation opposés du substrat au sein du site actif<sup>216</sup>. Par conséquent, le substrat doit accéder aux résidus catalytiques par un côté ou l'autre de la KS, désignés entrées « sud-est » et « nord-ouest »<sup>217</sup>. Ainsi, la présence conjointe de motifs conservés et d'entrées alternatives a permis de proposer un modèle de positionnement du substrat. Dans les KR de type A, le bras Ppant de l'ACP portant le substrat accède au site actif via l'entrée sud-est et contacte le tryptophane conservé<sup>216</sup>. Dans les KR de type B, cette entrée est obstruée par interaction du motif LDD et de la boucle couvercle, contraignant le groupement Ppant à emprunter l'entrée nord-ouest qui le positionne à proximité de la leucine conservée<sup>217</sup>. Ces deux modes de fixation alternatifs positionnent donc différemment le  $\beta$ -carbonyle cible de la kétoréduction en présentant l'une ou l'autre de ses faces au NADPH, selon la stéréospécificité de la KR.

Un autre modèle de stéréocontrôle suggère que la direction de réduction est régie par la conformation du site actif en présence du cofacteur. La fixation du NADPH par les KR de type A génère un site actif catalytiquement compétent formé par un réseau de liaisons hydrogènes entre les résidus Lys et Tyr de la tétrade catalytique, le cofacteur et la boucle couvercle<sup>217</sup>. L'entrée nord-ouest est bloquée par un résidu méthionine, ne laissant l'accès libre que par l'entrée sud-est où le tryptophane conservé pourrait aider à l'orientation du Ppant<sup>217</sup>. En revanche, la faible densité électronique correspondant au NADPH dans les structures cristallographiques de KR de type B indique une fixation plus lâche du cofacteur, permettant potentiellement l'accès du substrat par les deux entrées. Seule la liaison du substrat par le canal nord-ouest conduit cependant à une conformation compétente du site actif<sup>217</sup>.

De plus, parallèlement aux réactions de  $\beta$ -kétoréduction, certaines KR catalysent l'épimérisation de la position  $\alpha$  adjacente<sup>218</sup>. Ces domaines sont donc classés de manière plus précise, selon leur activité enzymatique vis-à-vis des positions  $\alpha$  et  $\beta$  : les KR spécifiques de substrats sans  $\alpha$ -méthyle (A0 et B0), les KR non-épimérisantes (A1 et B1), les KR réductrices et épimérisantes (A2 et B2), les KR ne catalysant ni épimérisation ni réduction (C1) et les KR épimérisantes non-réductrices (C2)<sup>138</sup> (**Figure 35**).

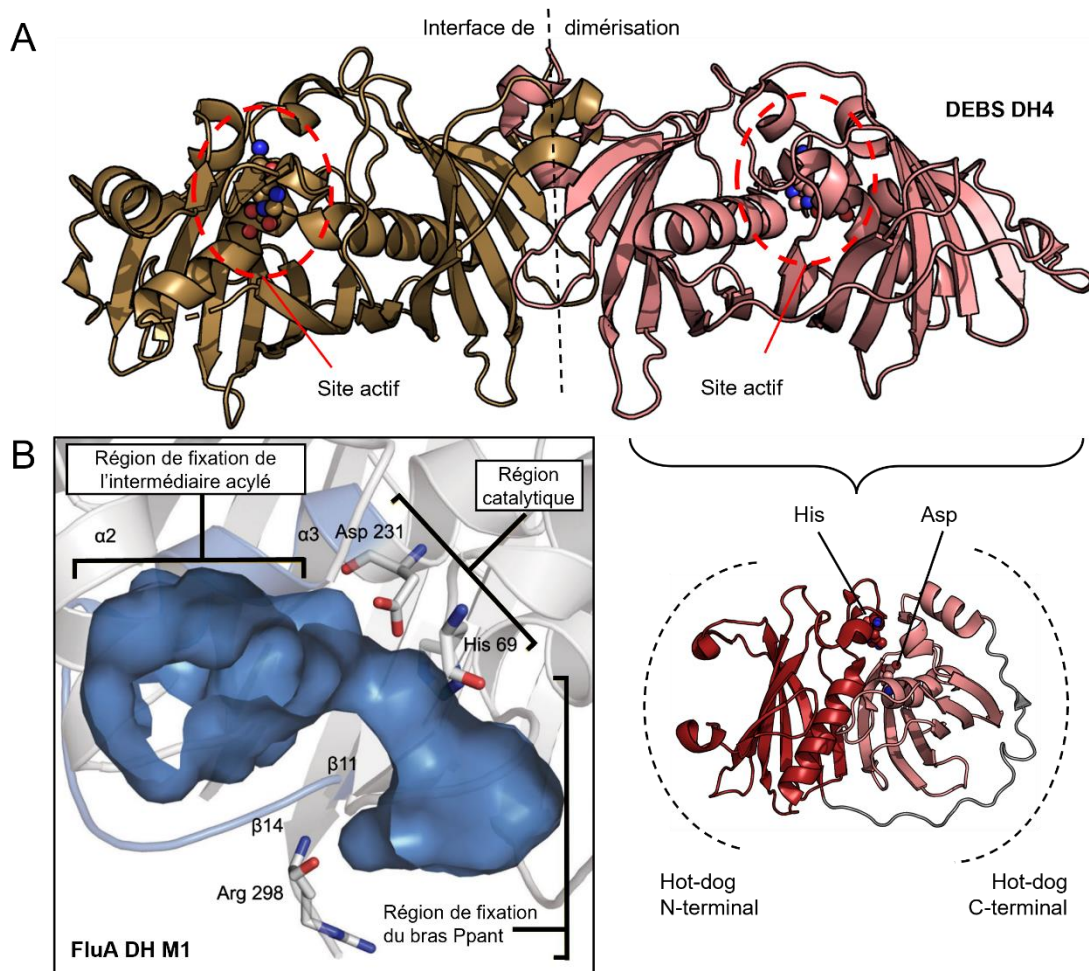


**Figure 35 :** Mécanisme catalytique des domaines KR. A) Mécanisme de réduction de la fonction  $\beta$ -cétone de la chaîne polycétidique en  $\beta$ -hydroxyle, selon le modèle de relais de protons. Les chaînes latérales Tyr et Ser de la tétrade catalytique interagissent avec l'oxygène du groupement  $\beta$ -carbonyle, la Lys est impliquée dans la reprotonation de la Tyr et l'hydride 4'-*pro*-S du cofacteur NADPH est transmis au carbone d'intérêt. L'asparagine conservée n'est pas représentée car son rôle dans la catalyse est incertain. B) Configurations absolues des positions 2 et 3 ( $\alpha$  et  $\beta$ , respectivement) de l'analogue de substrat 2-méthyl-3-oxopentanoïl-SNAC après action des différents types de domaines KR. Les réactions catalysées par les KR de type A0 et B0 ne sont pas représentées puisque ces domaines agissent sur des substrats non méthylés en position  $\alpha$  (figure adaptée de Drufva *et al.*, 2021<sup>219</sup>).

### e. Les domaines déshydratases

Les domaines DH acceptent pour substrats les intermédiaires 2-méthyl-3-hydroxy-acyl-ACP issus de la réduction par un domaine KR du groupement  $\beta$ -cétone en hydroxyle. La réaction réversible<sup>220</sup> de déshydratation qu'ils catalysent entraîne la formation d'une double liaison, la plupart du temps de configuration *trans*, générant une chaîne linéaire  $\alpha,\beta$ -insaturée. Les DH sont des domaines homodimériques dont chaque monomère (~300 résidus) est constitué de deux sous-domaines fusionnés en un motif structural « double hot-dog », qualifié comme tel en raison de la formation d'un long feuillet  $\beta$  enveloppant deux hélices  $\alpha$ <sup>221</sup> (**Figure 36**). À l'image des enzymes de la superfamille double hot-dog, les domaines DH possède une dyade catalytique His-Asp et le site actif est situé à l'interface des deux sous-domaines hot-dog de chaque monomère avec l'histidine localisée sur le hot-dog N-terminal et l'aspartate sur le hot-dog C-terminal<sup>222</sup> (**Figure 36**). Ce site actif est composé de trois zones aux rôles distincts avec une région de fixation du groupement Ppant, une région catalytique et une région d'accommodation de la chaîne polycétidique (**Figure 36**). La région de fixation du Ppant est notamment formée par une arginine conservée située à une distance également hautement

conservée de l'histidine catalytique, cohérente avec la fixation de l'intermédiaire  $\beta$ -hydroxyacyl-Ppant<sup>223</sup>.



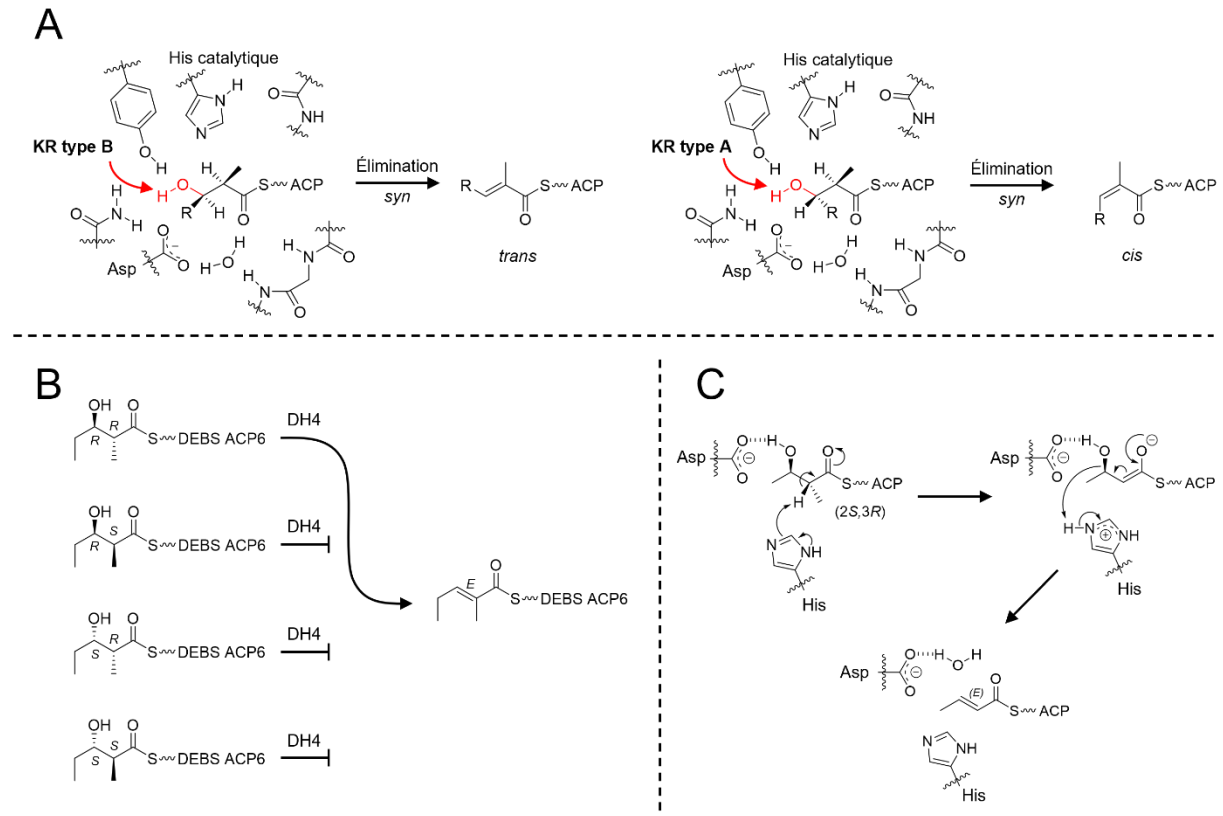
**Figure 36 :** Structures de domaines DH et organisation du site actif. A) L'étude cristallographique de domaine DH révèle une structure homodimérique où chaque monomère présente un repliement double hot-dog. La dyade catalytique His-Asp est répartie sur les motifs hot-dog N- et C-terminaux, respectivement (PDB : 3EL6)<sup>221</sup>. B) Le site catalytique des DH peut être divisé en trois régions responsables de la fixation du groupement Ppant, de la fixation de la chaîne polycétidique et de la catalyse enzymatique (PDB : 6OBV) (figure reproduite de Barajas *et al.*, 2019<sup>223</sup>).

L'étude du mécanisme enzymatique des domaines DH issus de mégasynthèses FAS a permis de proposer un modèle de catalyse pour les domaines DH de PKS. La formation d'une double liaison survient par élimination *syn*-coplanaire de l' $\alpha$ -hydrogène *pro*-(2S) et de la fonction  $\beta$ -hydroxyle. L'abstraction systématique de l'hydrogène *pro*-(2S) implique alors que le carbone  $\alpha$  du substrat doit être de configuration *R* (non épimerisé). Quant à l'élimination *syn*-coplanaire, elle implique que la fonction  $\beta$ -hydroxyle est toujours fixée au sein du site actif de la même manière, indépendamment de sa nature 3*R* ou 3*S*<sup>138</sup>, afin que les deux substituants à éliminer s'éclipsent. Dans le mécanisme proposé, la configuration du  $\beta$ -hydroxyle guide la formation spécifique d'une double liaison *cis* ou *trans*. Ainsi, une fonction (3*R*)-hydroxyle

généralisé par une KR de type B conduirait à la formation d'une double liaison *trans* et une fonction (3S)-hydroxyle conduirait à la formation d'une double liaison *cis*<sup>224</sup> (**Figure 37**). Ce modèle a pu être validé dans le cas des doubles liaisons *trans* avec par exemple l'étude de la spécificité de la DH de DEBS. Cette étude *in vitro* a montré la capacité de ce domaine à former une double liaison *trans* à partir d'un substrat (2R,3R)-2-méthyl-3-hydroxyacyle et l'absence de réaction en présence des diastéréoisomères (2R,3S), (2S,3R) et (2S,3S)<sup>220</sup> (**Figure 37**). En revanche, dans le cas des doubles liaisons *cis*, le lien entre configuration du substrat et configuration du produit n'est pas aussi évident puisque l'implication de domaines additionnels ou d'enzymes post-PKS a été démontrée ou suggérée dans la formation de certaines de ces liaisons<sup>225,226</sup>.

Concernant les résidus impliqués dans la déshydratation catalysée par les DH, le modèle actuellement considéré propose que l'histidine catalytique conservée joue le double rôle d'acide et base générale, déprotonant le carbone  $\alpha$  et stabilisant le  $\beta$ -hydroxyle partant sous forme de molécule d'eau<sup>227</sup>. Quant à l'aspartate de cette dyade, il possède un rôle supposé de liaison et d'orientation du substrat par une liaison hydrogène avec le groupement hydroxyle cible<sup>227</sup> (**Figure 37**). Plus récemment, des bimodules de déshydratation ont été identifiés au sein de PKS *trans*-AT dont le domaine DH catalyse deux déshydratations pour former deux doubles liaisons conjuguées<sup>176</sup>. Ces bimodules sont dits de type B par opposition au bimodules de type A dont la deuxième KS est non-élongatrice et sont retrouvés notamment dans les PKS de la diffidicine<sup>228</sup> et de la gladioline<sup>229</sup> où leur structure et fonction ont été étudiées. Une seconde histidine s'ajoute alors à la dyade catalytique classique, chacune des deux histidines catalysant à la fois la déprotonation d'un carbone et la protonation d'un hydroxyle, tel que proposé dans le mécanisme à une base des DH par Xie et Cane<sup>227</sup>, conduisant à la formation d'une fonction diène conjuguée. Toujours en accord avec ce mécanisme, l'aspartate conservé permet de stabiliser et orienter le substrat en interagissant par son carboxylate terminal avec les deux fonctions hydroxyles d'intérêt en C-3 et C-5<sup>229</sup>.

De manière surprenante, l'étude de la voie de biosynthèse de l'hybride polycétide-NRP FR901464 a conduit à la découverte d'un mécanisme de formation de double liaison catalysé par un domaine TE. Ce domaine possède la triade catalytique conservée des TE (voir la section IV.1.g.) et est phylogénétiquement plus proche des domaines TE canoniques que des DH. Pourtant, il ne catalyse pas l'hydrolyse ou la cyclisation de son substrat, mais sa déshydratation pour former une double liaison de configuration *cis*<sup>230</sup>.



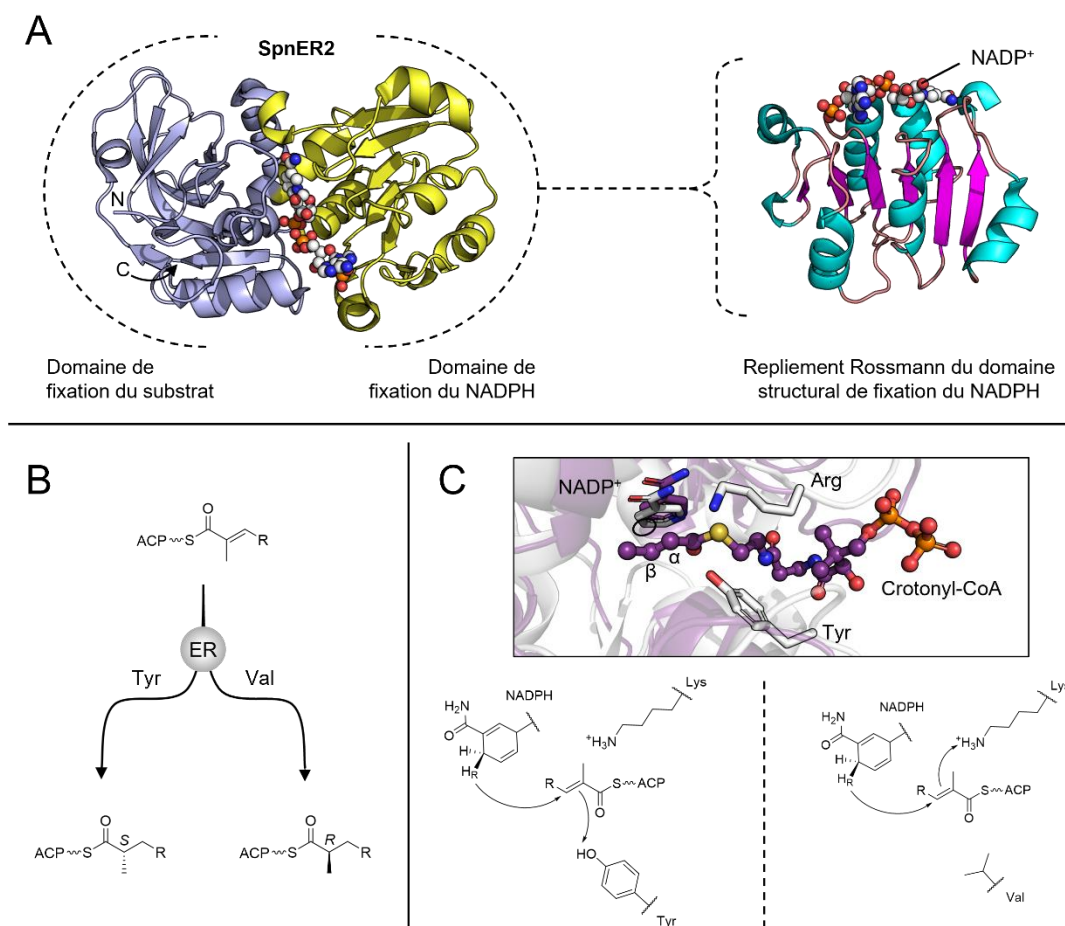
**Figure 37 :** Stéréochimie des domaines DH. A) Modèle de catalyse proposé pour la formation de doubles liaisons *cis* et *trans*. La fonction  $\alpha$ -méthyle du substrat est non-épimérisée (2*R*) et la configuration de la fonction  $\beta$ -hydroxyle détermine la configuration de la double liaison formée par déshydratation (figure reproduite de Weissman, 2017<sup>138</sup>). B) Stéréospécificité de la DH4 de DEBS. Seul le substrat de configuration (2*R*,3*R*) peut être déshydraté par la DH pour former une double liaison *trans*<sup>220</sup>. C) Mécanisme proposé de déshydratation par la dyade catalytique His-Asp des domaines DH. L'histidine déprotone la position C-2 et protone le groupe partant hydroxyle, tandis que l'aspartate stabilise et oriente ce groupement hydroxyle par liaison hydrogène (figure reproduite de Xie et Cane, 2018<sup>227</sup>).

## f. Les domaines énoyle réductases

Les domaines ER catalysent la réduction de double liaisons *trans* pour générer des groupements méthylènes saturés. Les deux structures de domaines énoyle réductases de PKS *cis*-AT résolue à ce jour<sup>231,232</sup> révèlent un repliement typique de la superfamille de protéines *medium-chain dehydrogenase/reductase* (MDR)<sup>209</sup>. Par conséquent, l'ER de module 2 de la PKS de la spinosyne est constituée d'un sous-domaine structural de fixation du substrat et d'un sous-domaine de fixation du cofacteur NADPH adoptant un repliement Rossmann<sup>231</sup> et plaçant le cofacteur au sein d'une cavité à l'interface des deux sous-domaines (**Figure 38**). Ces domaines enzymatiques réduisent les doubles liaisons par attaque spécifique de l'hydride 4'-*pro-R* du NADPH en position C-3 et par protonation de la position C-2 de l'intermédiaire insaturé<sup>138</sup>. Dans le cas d'un intermédiaire  $\alpha$ -méthylé, la réduction génère un centre stéréogène en C-2 dont la configuration est corrélée à la présence de certains résidus.



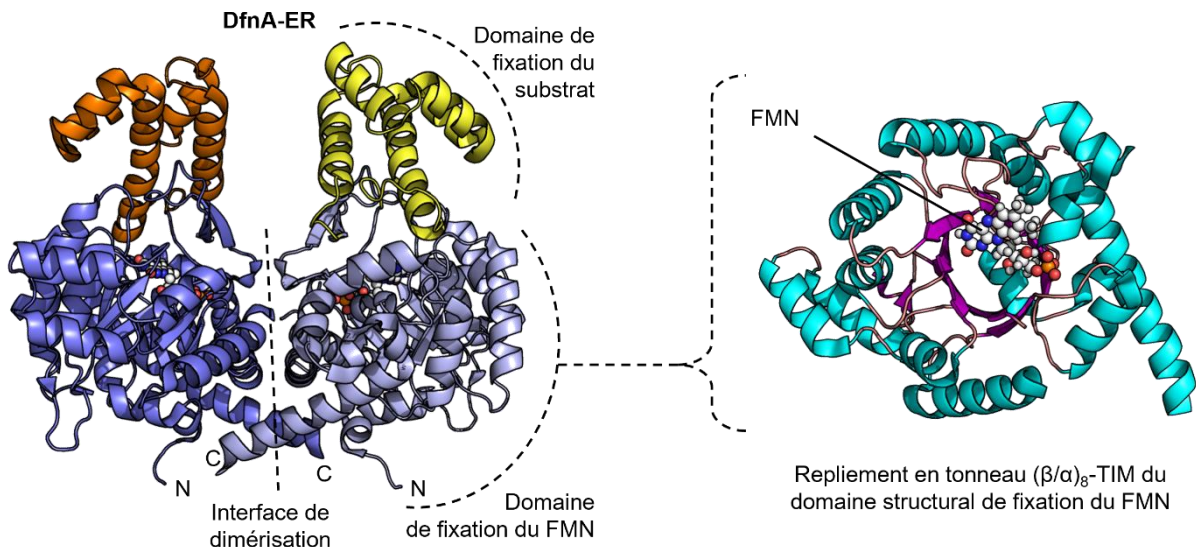
Ainsi, une tyrosine conservée au sein du site actif est associée à la formation d'un carbone asymétrique 2S tandis que ce résidu est substitué dans la plupart des cas par une valine et parfois une alanine ou une phénylalanine dans les ER générant une configuration *R* en position C-2<sup>233</sup> (**Figure 38**). Des expériences de mutagenèse dirigée ont permis de valider le rôle de cette tyrosine en inversant la stéréochimie en C-2 par une substitution Tyr → Val<sup>234</sup> et la mutation d'une lysine conservée également située au sein du site actif a montré son importance fondamentale dans la catalyse<sup>231,234</sup>. Un mécanisme de protonation en C-2 a alors pu être proposé. Lorsque présente, la tyrosine protone le substrat pour générer un carbone 2S et en son absence, la lysine protone le substrat par la face opposée pour générer la stéréochimie inverse (**Figure 38**)<sup>231</sup>. La résolution de la structure du complexe ternaire d'une énoyle-thioester réductase<sup>235</sup> est en accord avec ce mécanisme, avec la tyrosine et la lysine conservées situées de part et d'autre du substrat lorsque crotonyl-CoA et NADPH sont superposés au site actif de SpnER2 (**Figure 38**).



**Figure 38** : Structure d'un domaine ER de PKS *cis*-AT et mécanisme catalytique proposé. A) Le domaine SpnER2 de la PKS de la spinosyne possède un repliement typique des protéines MDR avec un domaine de fixation du NADPH organisé en motif Rossmann (à droite), avec ici le feuillet  $\beta$  en magenta, les hélices  $\alpha$  en cyan, les boucles en rose et le NADP<sup>+</sup> co-cristallisé représenté en sphères (PDB : 3SLK)<sup>231</sup>. B) La stéréochimie des domaines ER peut être prédite par la présence d'une tyrosine conservée corrélée à la formation d'un centre stéréogène 2S, qui est souvent substituée par une valine dans le site actif des ER générant une position 2R (figure reproduite de

Weissman, 2017<sup>138</sup>). C) Le NADP<sup>+</sup> et le crotonyl-CoA co-crystallisés avec une énoyle-thioester réductase (PDB : 4WAS) superposés au site actif de SpnER2 (r.m.s.d = 3.443 par superposition structurale) permettent d'observer la Lys et la Tyr potentiellement impliquées dans la protonation en C-2 de part et d'autre du substrat. La paire Lys-Tyr et le nicotinamide de SpnER2 sont représentés en blanc tandis que le nicotinamide et le substrat de l'énoyle-thioester réductase sont représentés en violet. Le carbone du NADPH portant l'hydride 4'-*pro-R* est indiqué par une ellipse noire. Ce positionnement des chaînes latérales peut expliquer la corrélation entre présence de la tyrosine et stéréochimie de l'ER (figure adaptée de Weissman, 2017<sup>138</sup>).

Ces différentes spécificités structurales et fonctionnelles des ER ne semblent cependant pas s'appliquer aux enzymes ER discrètes des PKS de type *trans*-AT. En effet, la PKS *trans*-AT de la difficidine fait intervenir au cours de sa synthèse une protéine discrète AT-ER dont la structure du domaine ER a été résolue<sup>236</sup>. Celle-ci révèle une organisation homodimérique où chaque monomère présente une structuration en deux sous-domaines, l'un fixant le substrat, l'autre le cofacteur. À l'inverse des ER de PKS *cis*-AT, le domaine de fixation du cofacteur est replié en tonneau triose-phosphate isomérase (TIM) ( $\beta/\alpha$ )<sub>8</sub> et lie spécifiquement la flavine mononucléotide (FMN) (**Figure 39**). Les ER discrètes de PKS *trans*-AT sont ainsi phylogénétiquement plus proches des domaines ER des FAS et des synthases de PUFA<sup>236</sup> tandis que les ER discrètes de HR-iPKS semblent se rapprocher davantage des domaines ER de PKS *cis*-AT, comme le montre la structure MDR de l'ER LovC impliquée dans la synthèse de la lovastatine<sup>237</sup> (**Figure 24**).



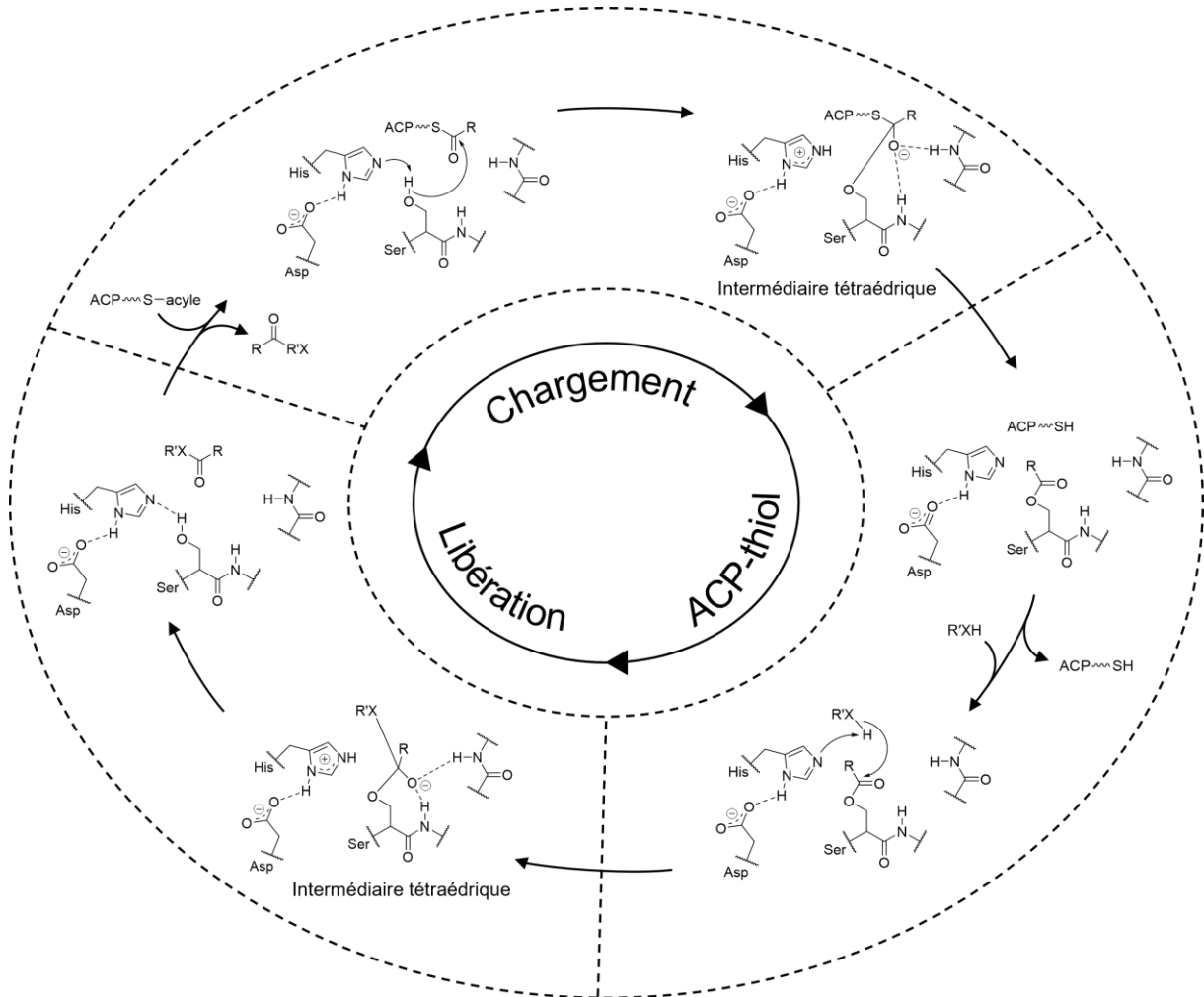
**Figure 39 :** Structure de l'ER discrète homodimérique DfnA-ER impliquée dans la synthèse de la difficidine. Chaque monomère fixe son cofacteur FMN à l'aide d'un sous-domaine organisé en tonneau ( $\beta/\alpha$ )<sub>8</sub>-TIM (à droite), avec ici le tonneau  $\beta$  central en magenta, le tonneau  $\alpha$  externe en cyan, les boucles en rose et le FMN co-crystallisé représenté en sphères (PDB : 4CW5)<sup>236</sup>.

## g. Les domaines thioestérases

Les TE sont des domaines de 260 à 300 résidus distingués en deux classes

fonctionnelles majoritaires, type 1 et type 2. Les TE de type 1 sont classiquement retrouvées dans le module terminal des PKS et sont responsables de la libération du polycétide en fin de synthèse, après que tous les cycles d'extension-réduction aient eu lieu. La réaction de clivage du thioester nécessite un groupement nucléophile pouvant être une molécule d'eau ou un groupement interne du polycétide, ce qui va générer un produit linéaire ou macrocyclique, respectivement. Dans le cas particulier des polycétides macrodiolides, le nucléophile provient d'un second intermédiaire de synthèse, ce qui a pour effet de dimériser le polycétide <sup>238</sup>. Les TE de type 2 quant à elles, sont des protéines discrètes interagissant en *trans* avec la PKS pour cliver certaines unités d'extension et intermédiaires de synthèse. Leur rôle principal est de cliver les unités et intermédiaires aberrants par hydrolyse pour assurer l'extension correcte de la chaîne polycétidique à l'image de la TE de type 2 de la PKS de la tylosine <sup>239</sup>, une activité similaire à celle des domaines AH (**Figure 28**). Par ailleurs, elles peuvent également être impliquées dans le contrôle de la sélection de l'unité d'initiation ou encore la libération d'intermédiaires afin de les rendre disponibles pour d'autres enzymes <sup>240</sup>.

Les deux classes de TE ont un mécanisme enzymatique commun faisant intervenir une triade catalytique Ser-His-Asp au cours de deux étapes successives de chargement et libération de la chaîne. Lors de la phase de chargement du substrat, l'histidine stabilisée par l'aspartate active la sérine catalytique. Celle-ci réalise alors une attaque nucléophile sur le carbonyle du thioester de l'acyl-ACP ce qui lie covalamment la TE à son substrat et libère l'*holo*-ACP sous forme thiol libre <sup>241</sup> (**Figure 40**). La phase de libération est ensuite initiée par attaque d'un groupement nucléophile sur l'ester de l'acyl-TE ce qui clive la liaison covalente formée entre l'enzyme et son substrat et régénère par conséquent la TE <sup>241</sup> (**Figure 40**). L'équipe de Craig Townsend propose que l'ion histidinium généré lors de la formation des deux intermédiaires tétraédriques reprotone le Ppant de l'ACP partant et la sérine catalytique lors de l'effondrement des états de transition <sup>242,243</sup>. La formation d'un produit de synthèse linéaire ou macrocyclique dépend de la nature du groupement nucléophile. L'activation d'une molécule d'eau va provoquer l'hydrolyse de l'ester de l'acyl-TE, libérant de ce fait un polycétide linéaire tandis que l'activation d'un nucléophile interne au substrat va engendrer une cyclisation de celui-ci, générant un macrolactone (le groupement activé étant dans la majorité des cas un hydroxyle) <sup>241</sup> (**Figure 40**). Par ailleurs, les TE présentent une stéréosélectivité envers leurs substrats, la configuration de la position 3-hydroxyacyle (et de l'hydroxyle nucléophile interne) étant importante pour permettre la réaction d'hydrolyse (ou de macrocyclisation) <sup>244,245</sup>.

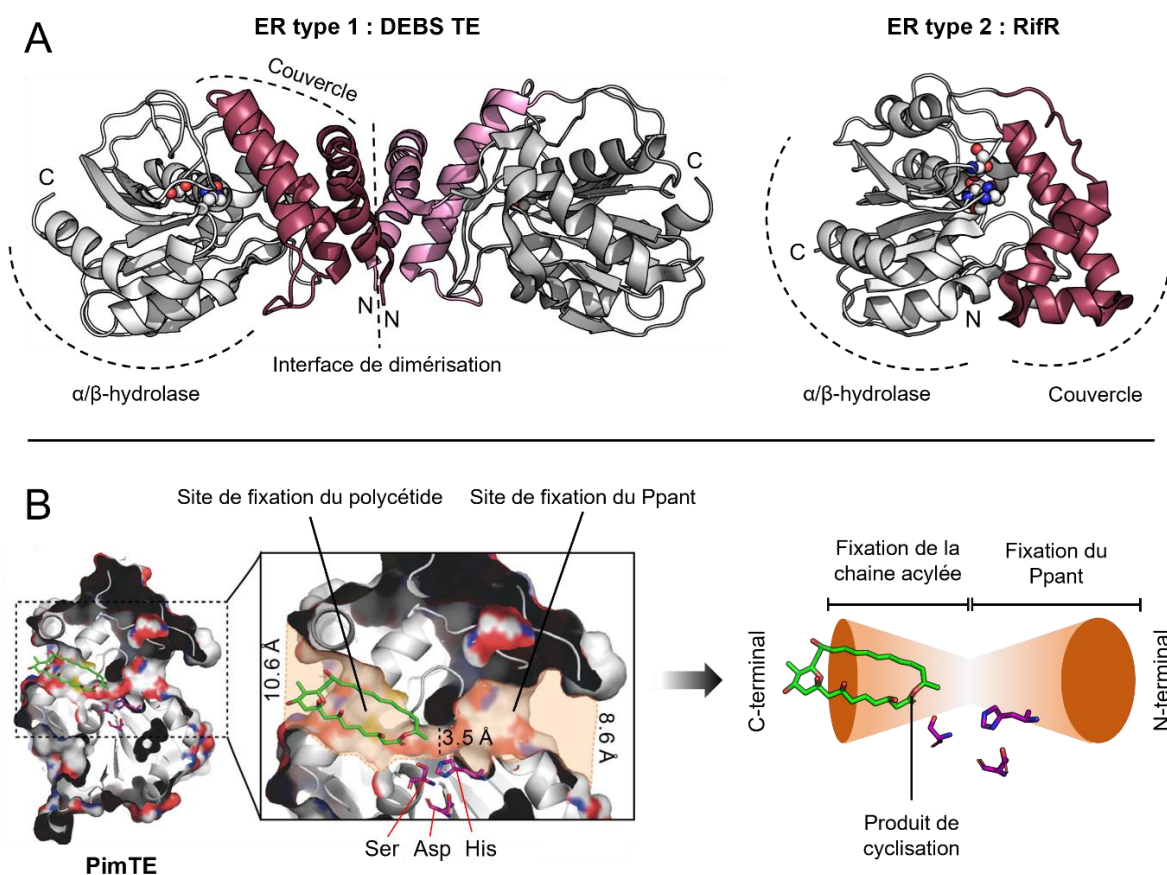


**Figure 40 :** Mécanisme catalytique proposé des TE de type 1 et 2 pour les étapes de chargement et de libération. La sérine catalytique activée par la paire His-Asp attaque le thioester de l'acyl-ACP pour former un intermédiaire covalent type acyl-enzyme. L'ACP partant laisse place à un second nucléophile X potentiellement activé par l'histidine conservée qui attaque l'ester-TE pour libérer le polycétide. R = reste de la chaîne polycétidique, R' = reste du nucléophile. Si X est une molécule d'eau, R' est un hydrogène et le produit de clivage est linéaire et si X est un groupement nucléophile interne, R' correspond au reste de la chaîne polycétidique et le produit de clivage est macrocyclique (figure adaptée de Horsman *et al.*, 2016 <sup>246</sup>).

Au sein des PKS, les TE de type 1 sont des domaines homodimériques dont l'interface de dimérisation est formée par interaction des deux hélices  $\alpha$  N-terminales de chaque monomère <sup>244,247-249</sup>. À l'inverse, les TE de type 2 ne possèdent pas ces deux hélices  $\alpha$  et sont par conséquent monomériques <sup>250,251</sup> (**Figure 41**). Bien que la forme tronquée cristallisée de la TE de type 2 RedJ dimérise par une autre interface, sa forme pleine taille est bien monomérique en solution <sup>252</sup>. Les deux classes arborent néanmoins un repliement similaire avec un cœur  $\alpha/\beta$ -hydrolase surmonté d'un sous-domaine « couvercle » structural composé de trois hélices  $\alpha$  chez les TE de type 2 et de quatre hélices  $\alpha$  chez les TE de type 1 dont les deux les plus exposées au solvant forment l'interface de dimérisation (**Figure 41**).

Ce couvercle et les boucles flexibles qui lui sont associées couvrent le site actif qui se

situé donc à l'interface des deux sous-domaines. Ils déterminent ainsi la forme et la taille du site actif et par conséquent le mode de fixation du substrat. Le site actif est un canal en forme de double entonnoir au travers duquel passe le substrat acyl-Ppant et dont la partie centrale, la plus étroite, contient la triade catalytique. Cette organisation tridimensionnelle décrit deux cavités, l'une N-terminale qui accommode le groupement Ppant lié à l'ACP et l'autre C-terminale qui permet la fixation de la chaîne polycétidique positionnant ainsi le thioester à cliver au niveau de la sérine nucléophile (**Figure 41**). Cette dernière cavité est large dans les TE de type 1 macrocyclisantes ce qui permet de contenir le produit de réaction après lactonisation. Elle est en revanche plus étroite dans les TE de type 2 et de type 1 non-cyclisantes ce qui est cohérent avec un produit de réaction linéaire <sup>251</sup>.



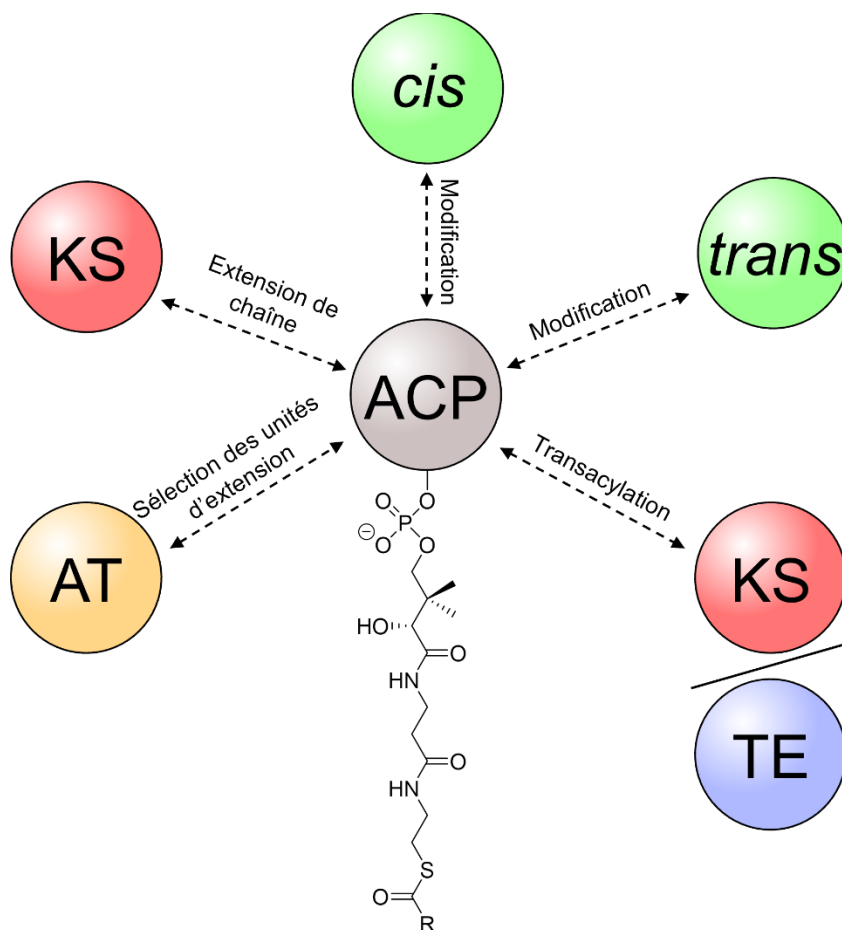
**Figure 41 :** Structures de TE issues de systèmes PKS. A) Les TE de type 1 et 2 partagent une structure tertiaire commune avec un domaine  $\alpha/\beta$ -hydrolase et un domaine couvercle. La triade catalytique représentée en sphères se trouve à l'interface des deux domaines. Les domaines TE de type 1 tels que celui de DEBS (à gauche) (PDB : 1KEZ) <sup>247</sup> dimérisent par l'intermédiaire de deux hélices  $\alpha$  N-terminales à l'inverse des TE discrètes de type 2 monomériques telles que RifR de la PKS de la rifamycine (à droite) (PDB : 3FLA) <sup>250</sup>. B) La co-cristallisation du domaine TE de la PKS de la pimaricine en présence de son substrat naturel linéaire lié à un groupement SNAC a permis d'observer l'accommodation du produit de réaction cyclisé dans le site actif (PDB : 7VO5) <sup>249</sup>. Le macrolactone 2-hydro-3-hydroxyle pimaricinolide est fixé au sein de la cavité C-terminale du site actif et la fonction ester issue de la cyclisation est logiquement localisée à proximité de la sérine catalytique. La cavité N-terminale permet donc la fixation du groupement Ppant auquel le polycétide et l'ACP sont liés (figure reproduite de Zhou *et al.*, 2022 <sup>249</sup>).

## 2. Organisation séquentielle et tridimensionnelle des PKS

### a. Les modules et sous-unités

#### i. L'ACP au centre du module, perspective structurale

L'ACP constitue le point d'ancrage des unités d'initiation/extension et des intermédiaires de synthèse au cours des différentes réactions catalysées au sein de chaque module. Par conséquent, le domaine ACP est l'unité centrale du module et il est nécessaire qu'il interagisse avec tous les domaines de celui-ci ainsi qu'avec le domaine KS du module en aval et finalement avec le domaine TE terminal pour permettre à chacune des enzymes d'agir <sup>253</sup> (**Figure 42**).

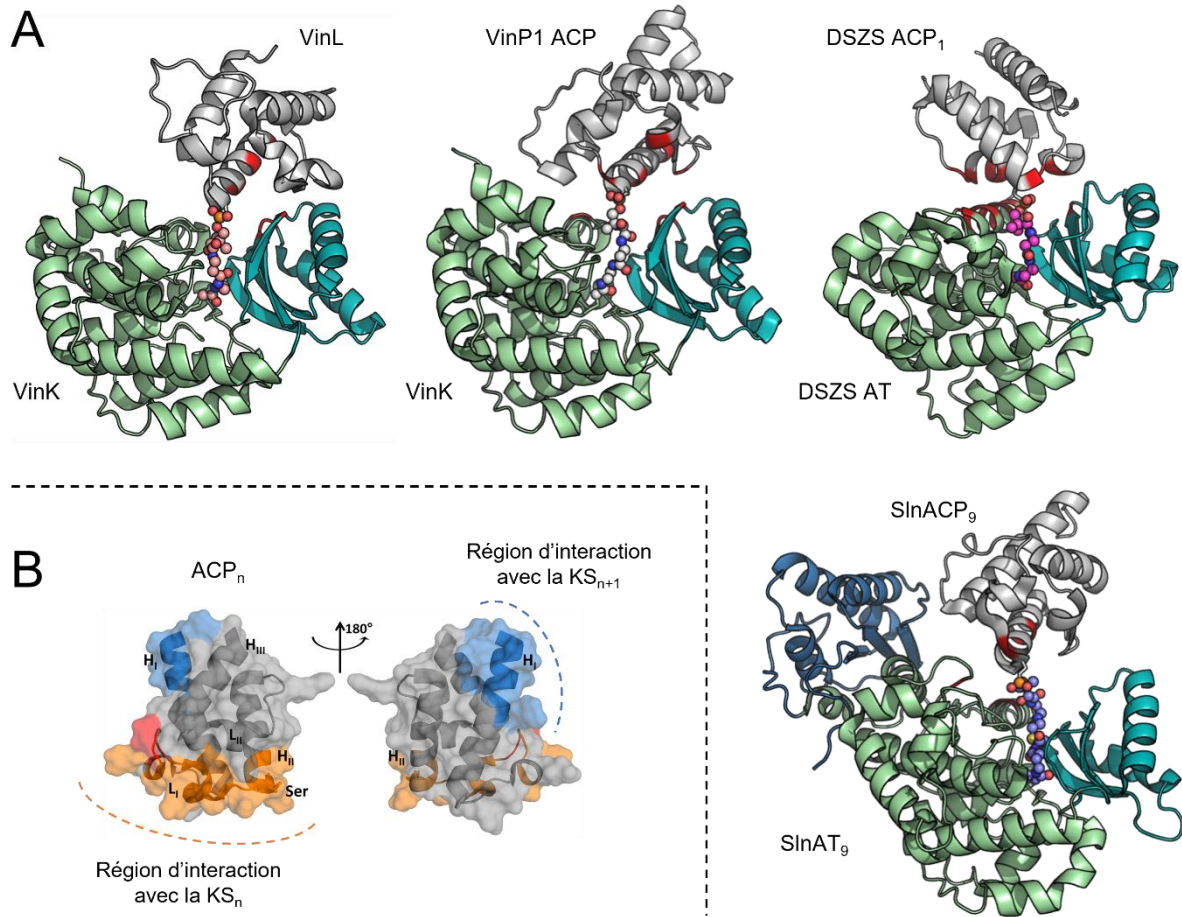


**Figure 42 :** L'ACP est au centre des interactions entre domaines enzymatiques. Le domaine ACP sous forme *holo* joue un rôle de navette qui véhicule les unités d'extension et intermédiaires linéaires entre les domaines et partenaires discrets de son module selon un ordre défini, pour étendre et modifier le polycétide. L'ACP interagit également avec le module directement en aval afin de transmettre l'intermédiaire de synthèse au domaine KS suivant ou au domaine TE final (figure adaptée de Weissman, 2020 <sup>208</sup>).

Des investigations *in vitro* ont permis d'en apprendre davantage sur les interactions permettant l'association transitoire de l'ACP avec les domaines canoniques de PKS.

Concernant les enzymes AT, la cristallisation de l'ACP1 de la PKS du disorazole et des ACP VinL et VinP1ACP<sub>L</sub> de la PKS de la vicénistatine en complexe avec les AT discrètes de leurs voies de biosynthèse respectives<sup>254,187,188</sup> ainsi que du complexe AT-ACP du module 9 de la PKS *cis*-AT de la salinomycine<sup>255</sup> a fourni de précieuses informations sur les structures secondaires et résidus mis en jeu dans l'interaction. Les ACP VinL et VinP1ACP<sub>L</sub> sont tous les deux reconnus par leur deuxième hélice  $\alpha$  portant la sérine modifiée mais présentent un mode d'interaction avec l'AT VinK légèrement différent avec une liaison électrostatique supplémentaire dans le cas de VinL<sup>187,188</sup> (**Figure 43**). L'ACP du premier module de la PKS du disorazole interagit quant à lui avec son AT discrète par le biais de sa première boucle et de sa courte hélice  $\alpha$ III (aussi désignée III')<sup>254</sup> (**Figure 43**). Un mode d'interaction encore différent est observé entre la *cis*-AT et l'ACP du module 9 de la PKS de la salinomycine<sup>255</sup>. Dans ce cas, l'AT adopte avec son adaptateur KS-AT N-terminal une forme de C permettant d'envelopper l'ACP. L'AT interagit avec l'ACP majoritairement par son domaine  $\alpha,\beta$ -hydrolase tandis que son domaine type ferrédoxine et son adaptateur KS-AT contraignent l'ACP de part et d'autre (**Figure 43**), contrairement aux AT précédemment évoquées qui contactent les ACP principalement par leur domaine type ferrédoxine. Il apparaît ainsi que malgré une structure conservée entre AT et entre ACP, l'orientation de l'ACP et les interactions d'interface formant le complexe sont spécifiques à chaque paire AT-ACP.

L'étude de la capacité de transfert de chaîne entre ACP et KS de la PKS DEBS a révélé deux zones d'interaction distinctes au sein de l'ACP<sup>256</sup>. D'une part, la région constituée de la première boucle et de l'extrémité N-terminale de l'ACP (comprenant la sérine portant le groupement Ppant) forme l'interface avec la KS de son propre module lors de l'extension de chaîne. D'autre part, l'interaction avec la KS du module suivant lors de l'étape de translocation de chaîne est médiée par les dix résidus N-terminaux de la première hélice  $\alpha$  de l'ACP (**Figure 43**)<sup>256</sup>.



**Figure 43 :** Informations structurales sur les interfaces entre domaines ACP-AT et ACP-KS. A) Structures cristallographiques de complexes ACP-AT obtenus par *cross-linking* et issus de PKS *trans*-AT (en haut) et *cis*-AT (en bas). Différentes interactions permettent la liaison des deux domaines en fonction de la paire considérée, VinL-VinK (PDB : 7F2R)<sup>257</sup>, VinP1 ACP-VinK (PDB : 8H6S)<sup>188</sup>, DSZS ACP<sub>1</sub>-AT (PDB : 5ZK4)<sup>254</sup> et SlnACP<sub>9</sub>-SlnAT<sub>9</sub> (PDB : 7VRS)<sup>255</sup>. Les différents ACP sont représentés en blanc, les domaines α,β-hydrolase et type ferrédoxine des AT en vert pâle et bleu clair, respectivement, et l'adaptateur KS-AT de SlnAT<sub>9</sub> en bleu foncé. Les analogues de substrat liant la sérine conservée des ACP aux sites actifs des AT sont représentées en sphères et les zones d'interaction ACP-AT sont colorées en rouge. B) Zones d'interaction spécifiques de l'ACP<sub>5</sub> de la PKS DEBS avec la KS de son module et la KS suivante. La région représentée en orange permet d'interagir à la KS de son module pour l'étape d'élongation de chaîne (les deux résidus colorés en rouge sont prédits pour se lier à l'adaptateur KS-AT) tandis que la région représentée en bleu forme l'interface avec la KS en aval lors de la réaction de transfert de chaîne au module suivant (figure reproduite de Kapur *et al.*, 2010<sup>256</sup>).

À propos de la reconnaissance entre ACP et KR, l'investigation de l'affinité entre ces deux types de domaines extraits de la PKS de la mycolactone a révélé une contribution importante du groupement Ppant aux interactions ACP/KR tandis que les deux KR étudiées présentaient une affinité similaire pour leurs propres ACP et un autre ACP de la PKS<sup>258</sup>. Concernant l'interaction entre ACP et domaines DH ou ER, aucune structure n'a jusqu'ici permis d'observer directement les interfaces formées. De plus, la structure du complexe de l'AcpP et de la DH discrète FabA de la FAS de type II d'*E. coli*<sup>259</sup> ne semble pas être

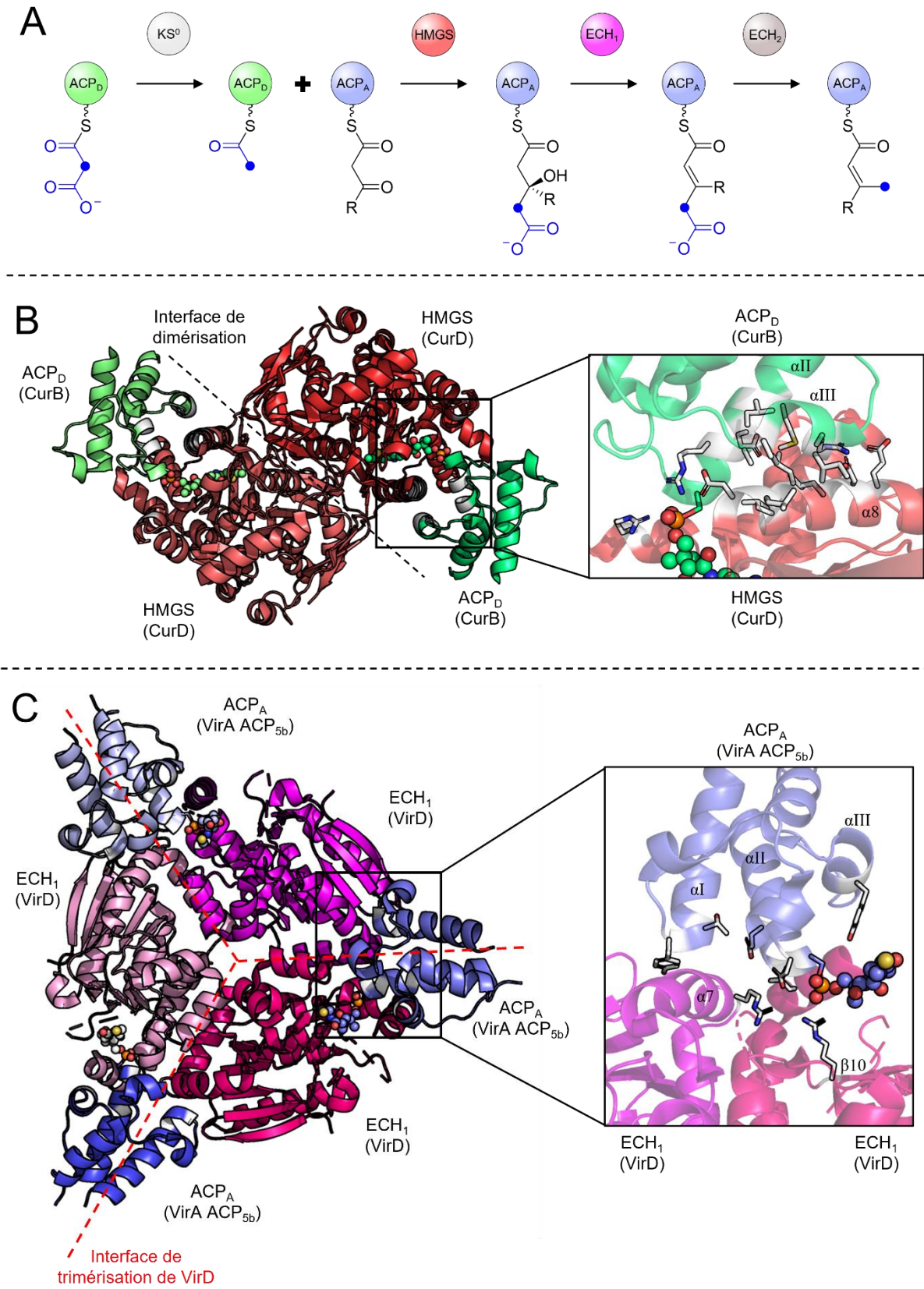


transposable aux complexes ACP-DH de PKS puisque le potentiel électrostatique des surfaces mises en contact entre les deux domaines diffèrent dans les deux systèmes <sup>260</sup>. Finalement, les interactions entre ACP et TE semblent être médiées principalement par le groupement Ppant et le substrat. En effet, des études d'affinité entre le domaine TE de DEBS et l'ACP directement en amont ont montré une absence d'interaction avec l'ACP sous forme *apo* contre une affinité d'environ 20  $\mu$ M pour la forme *holo* et la forme acylée <sup>261</sup>, des interactions facilitées par le fait que les deux domaines sont covalamment liés par un cours *linker* dans le contexte natif de la PKS <sup>262</sup>.

Il est à noter que plusieurs études structurales ont également été menées afin de comprendre les mécanismes de reconnaissance et d'interactions entre ACP et cassette de  $\beta$ -méthylation. Ce mécanisme de  $\beta$ -branchement fait intervenir cinq protéines partenaires (**Figure 44**) : un ACP donneur ou ACP<sub>D</sub> qui apporte l'unité malonate, une KS<sup>0</sup> qui décarboxyle l'unité malonate en acétate, une 3-hydroxy-3-méthylglutaryl-CoA synthase (HMGS) qui catalyse la condensation de l'unité acétate en position  $\beta$  de la chaîne polycétidique portée par un ACP accepteur ou ACP<sub>A</sub>, un premier homologue d'énoyl-CoA hydratase (ECH) qui déshydrate la chaîne pour former un produit  $\alpha,\beta$ -insaturé et finalement un second homologue de ECH catalysant une décarboxylation générant une fonction  $\beta$ -méthyle <sup>201</sup>. La résolution de la structure des complexes ACP<sub>D</sub>/HMGS de la PKS de la curacine <sup>263</sup> et ACP<sub>A</sub>/ECH<sub>1</sub> de la PKS de la virginiamycine M <sup>201</sup> a permis d'en apprendre davantage sur le mode d'interaction d'enzymes avec un ACP discret et sur le recrutement en *trans* de partenaires de domaines ACP, respectivement.

La HMGS CurD présente une structure homodimérique en solution et dans le cristal et chaque monomère lie un ACP<sub>D</sub> CurB de sorte que l'acyl-Ppant puisse accéder au site actif (**Figure 44**). Les surfaces d'interaction de ces deux protéines sont à la fois complémentaires en forme et en charge, avec une cavité fortement électropositive à la surface de l'ACP<sub>D</sub> CurB interagissant avec une zone électronégative de la HMGS CurD <sup>263</sup>. La majorité des interactions est médiée par des résidus de l'ACP localisés sur les hélices  $\alpha$ II et  $\alpha$ III et la moitié N-terminale de l'hélice  $\alpha$ 8 de la HMGS (**Figure 44**). La ECH<sub>1</sub> VirD, quant à elle, forme un hexamère constitué de deux disques homotrimériques superposés dans la structure cristallographique, bien que seule la forme trimérique semble pertinente en solution. Chaque monomère adopte le repliement typique des crotonases en motifs répétés  $\beta\beta\alpha$  <sup>201</sup>. Trois copies de l'ACP<sub>A</sub>, VirA ACP<sub>5b</sub>, se lient à l'interface entre trimères (**Figure 44**). Plus spécifiquement, l'extrémité C-terminale de l'hélice  $\alpha$ I et l'extrémité N-terminale des hélices  $\alpha$ II et  $\alpha$ III de l'ACP interagissent principalement avec le brin  $\beta$ 10 d'un monomère de VirD et l'hélice  $\alpha$ 7 d'un second monomère (**Figure 44**). La comparaison des structures de l'ACP<sub>A</sub> de la PKS de la virginiamycine M avec d'autres ACP<sub>A</sub> non reconnus par VirD semble par ailleurs indiquer que la discrimination entre

différents ACP par les enzymes de la cassette de  $\beta$ -méthylation est médiée majoritairement par le caractère électrostatique des résidus formant leurs surfaces d'interaction <sup>201</sup>.



**Figure 44** : Interactions entre protéines partenaires et mécanisme des cassettes de  $\beta$ -méthylation. A) La cassette

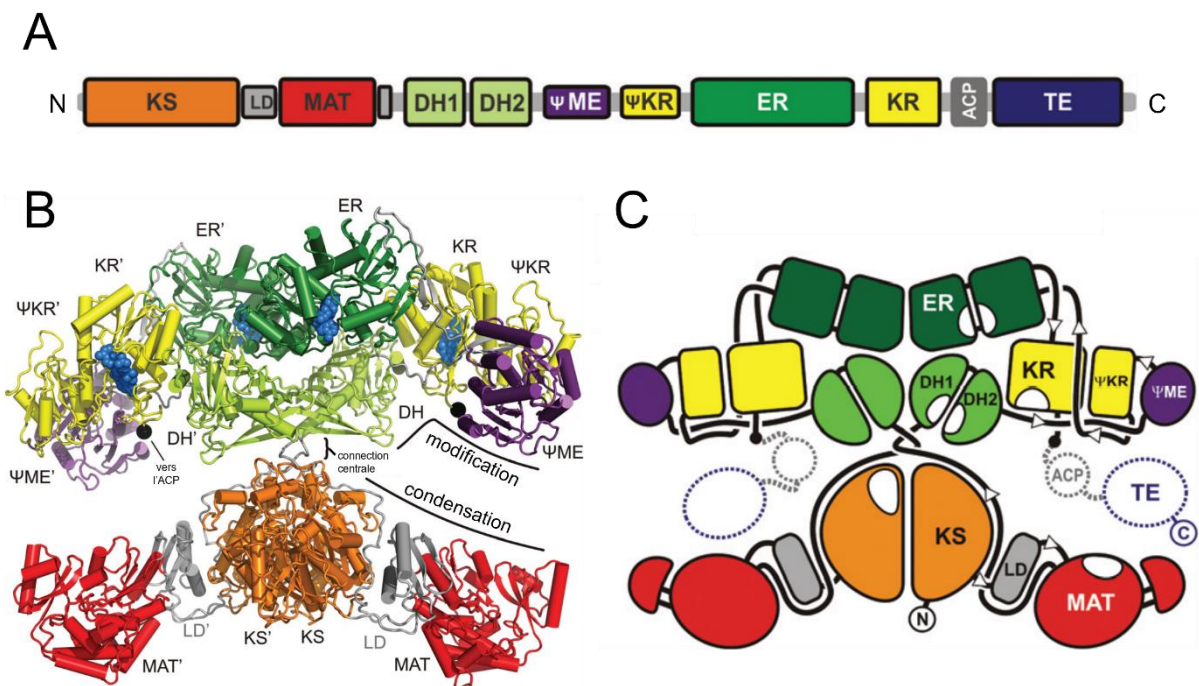
de  $\beta$ -méthylation fait intervenir plusieurs partenaires protéiques permettant de générer un groupement  $\beta$ -méthyle à partir d'une unité malonate et de la chaîne polycétidique  $\beta$ -cétone. Les protéines HMGS, ECH<sub>1</sub> et ECH<sub>2</sub> interagissent en *trans* avec la chaîne polycétidique tandis que les protéines ACP<sub>D</sub> et KS<sup>0</sup> interagissent entre elles et l'ACP<sub>D</sub> avec la HMGS. Quant à l'ACP<sub>A</sub>, il est présent au sein même de la PKS (figure reproduite de Collin *et al.*, 2023<sup>201</sup>). B) Structure du complexe formé entre l'ACP<sub>D</sub> et la HMGS homodimérique de la PKS *cis*-AT de la curacine (respectivement, CurB et CurD). Un zoom est fait sur l'interface entre les deux protéines (PDB : 5KP8). C) Structure du complexe ACP<sub>A</sub>/ECH<sub>1</sub> de la PKS *trans*-AT de la virginiamycine M (respectivement, VirAAC<sub>P5b</sub> et VirD). La ECH<sub>1</sub> forme un disque homotrimérique (s'assemblant en un dimère de trimères dans le cristal) et l'ACP se lie à chaque interface entre deux monomères, montrée en détail à droite de la structure (PDB : 8AHQ). Chaque protéine est colorée en accord avec le code couleur de mécanisme représenté en (A). Les acides aminés participant à la formation des interfaces sont représentés en blanc et les groupements acyl-Ppant ou Ppant libre sont représentés en sphères.

## ii. Structure de modules et sous-unités

Les domaines de PKS extraits de leurs chaînes polypeptidiques, cristallisés seuls ou en complexe avec un domaine partenaire fournissent de précieuses informations sur leurs mécanismes d'interaction et de catalyse. Elles sont en revanche peu informatifs sur la dynamique globale des modules et sous-unités nécessaire à la multitude de réactions catalysées par ces mégasynthèses. Néanmoins, depuis dix ans, le nombre croissant de structures de modules résolues a progressivement mis en lumière l'agencement dans l'espace des domaines les uns par rapport aux autres, et l'évolution de cet agencement au cours du cycle enzymatique.

Les premières informations structurales sur les modules intacts de mégasynthèses furent obtenues de l'étude des FAS de type I retrouvées chez les mammifères. À l'image de leurs homologues PKS de type I, ces mégasynthèses sont composées d'une chaîne polypeptidique homodimérique comprenant tous les domaines nécessaires à l'élongation et la modification d'un intermédiaire de synthèse issu de la condensation d'unités acétates et malonates. Ces FAS sont ainsi constituées d'un domaine KS, d'une malonyl-/acétyltransférase (MAT) et d'un ACP nécessaires à l'extension de la chaîne d'acide gras, ainsi que du trio KR, DH, ER permettant la réduction complète des unités condensées, et d'une TE C-terminale. Un domaine méthyltransférase dépourvu d'activité catalytique ( $\Psi$ MT) est également présent au sein de la mégasynthase (**Figure 45**). Par conséquent, les FAS de mammifères peuvent être considérées comme un unique module de PKS catalysant de manière itérative la formation d'acides gras. À cet égard, la résolution de la structure de la FAS porcine à 4.5 Å<sup>264</sup> puis 3.2 Å<sup>265</sup> fut le point de départ de l'analyse structurale globale de ces mégaenzymes (**Figure 45**). Cette structure révèle une organisation homodimérique en forme de X avec une large interface interpolypeptide médiée par la dimérisation des domaines KS, DH et ER. Une région inférieure dite « condensatrice » est formée des domaines KS et MAT tandis qu'une région supérieure

dite « modifcatrice » comprend les domaines KR, DH et ER ainsi qu'un domaine méthyltransférase inactif ( $\Psi$ MT), les deux régions étant liées par un court *linker* MAT-DH au niveau duquel se croisent les deux chaînes. Cette organisation dans l'espace en deux régions fonctionnelles décrit deux chambres réactionnelles dans lesquelles l'ACP de chaque chaîne navigue afin d'accéder aux sites actifs des différents domaines (bien que les domaines ACP et TE ne soient pas observables dans la structure <sup>265</sup>) (**Figure 45**).

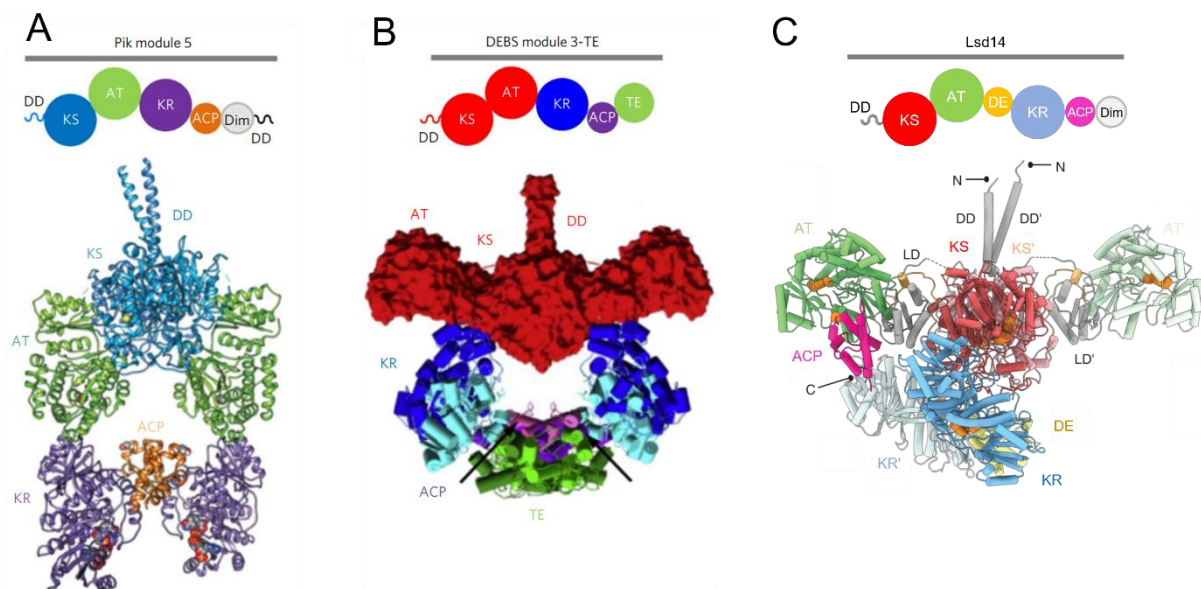


**Figure 45 :** Organisation et structure de la FAS porcine (figure reproduite de Maier *et al.*, 2006 <sup>265</sup>). A) Organisation linéaire des domaines de la FAS. Les rectangles gris entre les domaines KS, MAT et DH1 représentent les *linkers* structurellement organisés. B) Structure cristallographique de la FAS porcine dont les domaines sont colorés tels que dans (A). Les cofacteurs NADP<sup>+</sup> cocrystallisés sont représentés en sphères bleues et les extrémités C-terminales auxquelles sont liés les didomains ACP-TE sont indiquées par une sphère noire. C) Représentation schématique de la structure (B) dont les ACP-TE manquants sont ici représentés dans chaque chambre réactionnelle.

L'étude structurale des domaines isolés de PKS modulaires (voir partie IV.1.) a cependant révélé des différences notables entre FAS et PKS. Contrairement aux observations faites sur les domaines PKS, les ER de FAS sont homodimériques, les TE sont monomériques et bien que les DH soit homodimériques dans les deux types de systèmes, les surfaces de dimérisation mises en jeu diffèrent <sup>266</sup>.

Depuis 2014, plusieurs structures de modules entiers de PKS *cis*- et *trans*-AT ont permis de lever le voile sur l'organisation tridimensionnelle de ces systèmes et de confirmer les doutes initiaux quant à la portabilité du modèle FAS aux systèmes PKS. Les structures du module 5 de la PKS *cis*-AT de la pikromycine, résolue par cryo-microscopie électronique (EM) <sup>267</sup>, et de la fusion du module 3-TE de la PKS *cis*-AT DEBS, résolue par diffusion des

rayons X aux petits angles (SAXS) <sup>268</sup> ont notamment révélé l'organisation des modules KS-AT-KR-ACP (**Figure 46**). Contrairement à la FAS animale, ces modules adoptent une forme d'arche dont la clef de voûte est formée par la KS homodimérique et qui englobe une unique chambre réactionnelle au sein de laquelle les ACP vont se déplacer pour interagir avec chacun des domaines en fonction de l'unité d'extension ou de l'intermédiaire polycétidique acylé <sup>269</sup>. La résolution plus récente de la structure de la PKS *cis*-AT Lsd14 constituée d'un seul module composé également des domaines KS-AT-KR-ACP suggère cependant une organisation différente <sup>270</sup>. Dans ce cas, les deux domaines KR sont à proximité l'un de l'autre en raison de la présence d'un élément de dimérisation pré-KR (DE) et ne sont pas directement en contact avec les domaines AT, créant ainsi deux chambres réactionnelles à l'image de la FAS animale. Ce positionnement sans précédent des KR présente également la particularité d'être asymétrique, provoquant de fait une asymétrie des chambres réactionnelles. Au sein de l'une d'entre elles, l'entrée des sites actifs des trois domaines enzymatiques est orientée vers le centre de la chambre où l'ACP est localisé, en contact avec l'AT. En revanche, dans la seconde chambre, la KR est orientée de manière opposée et l'ACP n'est pas présent dans la structure, vraisemblablement trop mobile (**Figure 46**). Toutes ces différences entre les deux chambres tendent à indiquer un mécanisme catalytique asynchrone où chaque quatuor KS-AT-KR-ACP réalise à tour de rôle un cycle d'élongation/réduction du polycétide <sup>270</sup>. L'étude par cryo-EM du module 1 de DEBS comprenant les mêmes domaines révèle également l'agencement de deux régions catalytiques asymétriques <sup>271</sup>.



**Figure 46 :** Caractérisation structurale de modules entiers de PKS *cis*-AT composés des domaines KS-AT-KR-ACP (figure adaptée de Weissman, 2015 <sup>272</sup> et Bagde *et al.*, 2021 <sup>270</sup>). A) Structure du module 5 de la PKS de la pikromycine dont la forme d'arche définit une seule chambre réactionnelle. La faible densité des éléments de dimérisation post-ACP (Dim) et des <sup>C</sup>DD dans la carte de cryo-EM a empêché leur modélisation <sup>267</sup>. B) Modèle du module 3 de la PKS DEBS fusionné au domaine TE C-terminal basé sur des données de SAXS décrivant à nouveau une forme d'arche. La différence majoritaire avec la structure (A) provient de la localisation des domaines AT,

## État de l'art

vraisemblablement due à la reconstruction du module en corps rigide du didomaine KS-AT <sup>268</sup>. C) Structure cristallographique de la PKS monomodulaire Lsd14, avec les résidus du site actif de chaque domaine indiqués en orange. Deux chambres réactionnelles asymétriques sont cette fois-ci observées. La mobilité importante des deux ACP et des éléments Dim a empêché leur localisation <sup>270</sup>. L'organisation linéaire des modules est indiquée au-dessus de chacun d'entre eux.

Des études structurales supplémentaires s'intéressant à des PKS modulaires variées ont permis de découvrir de nouvelles architectures de modules composés de différents domaines et d'en comprendre le fonctionnement plus en détail. Par exemple, l'étude par SAXS du module 5 <sup>273</sup> et de la fusion ACP<sub>5b</sub>-<sup>C</sup>DD-<sup>N</sup>DD-KS<sub>6</sub>-adaptateur KS-AT <sup>274</sup> de la PKS *trans*-AT de la virginiamycine M a conduit à l'obtention d'une enveloppe à basse résolution de l'interface intersous-unités dans laquelle des modèles des domaines individuels ont été replacés. Le module 5 adopte une conformation très étendue qui permet à son ACP C-terminal et son <sup>C</sup>DD d'entrer en contact avec le <sup>N</sup>DD du module 6. Les deux « bras » du module 5 forment ainsi une large chambre fermée, probablement impliquée dans le recrutement des enzymes discrètes de la cassette de  $\beta$ -méthylation (**Figure 47**).

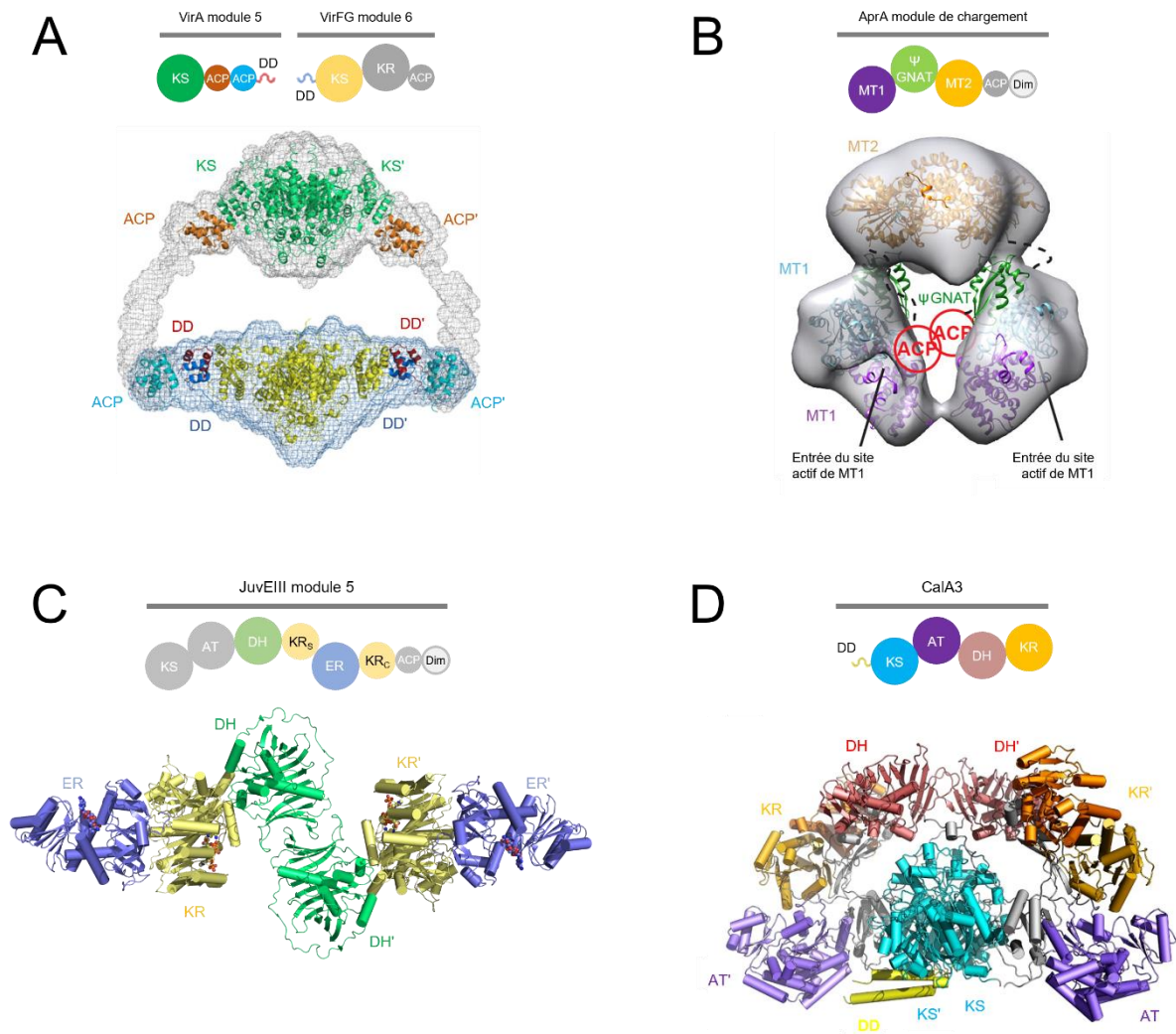
La synthèse de l'aprotaxine A implique un hybride PKS-NRPS qui catalyse la formation d'une fonction *tert*-butyle inhabituelle. Cette chimie est catalysée au sein du module de chargement de la synthase, composé d'un domaine *N*-acétyltransférase type GCN5 (GNAT) catalytiquement inactif flanqué de deux méthyltransférases (MT) et d'un ACP C-terminal. Le premier domaine méthyltransférase (MT1) est une cMT catalysant la diméthylation d'une unité d'initiation malonate <sup>275</sup> et le second (MT2) possède une double activité décarboxylase-cMT conduisant à la formation d'une fonction *tert*-butyle sous la forme d'une unité pivalate <sup>276</sup>. L'étude par cristallographie aux rayons X et cryo-EM du module de chargement a dévoilé une architecture en forme de pince avec au centre la MT2 homodimérique et de part et d'autre, les didomaines monomériques MT1- $\Psi$ GNAT (**Figure 47**). Plusieurs conformations du module se distinguent parmi les différentes classes observées par cryo-EM. Lorsque la pince est totalement fermée, les deux MT1 sont presque en contact l'une de l'autre et la MT2 est inaccessible. Ceci correspond probablement à la conformation permettant la première étape de méthylation et empêchant la décarboxylation prématurée de l'unité d'initiation <sup>276</sup> (**Figure 47**).

Comme évoqué précédemment, la FAS animale possède un domaine ER homodimérique et une surface de dimérisation de DH incompatibles avec les observations faites sur les domaines de PKS isolés, posant la question de l'arrangement de ces domaines dans les PKS modulaires. Des éléments de réponse ont été apportés avec la résolution très récente de la structure cristallographique du tridomaine réducteur DH-KR-ER du module 5 de la PKS *cis*-AT de la jувénimicine <sup>232</sup>. L'arrangement tridimensionnelle de ce fragment de module valide l'hypothèse précédente sur la divergence structurale de ces domaines entre

FAS et PKS modulaires, puisque le domaine ER est dans ce cas monomérique et la dimérisation du tridomaine est médiée uniquement par le domaine DH dont l'interface est semblable à celle observée pour les domaines DH de PKS modulaires isolés (**Figure 47**).

Un autre exemple notable est le module terminal de libération CalA3 de la PKS de la calcimycine. Celui-ci est constitué des domaines KS, AT, DH et ER dont la KS catalyse la réaction inhabituelle de libération de chaîne par amidation et dont les trois autres domaines sont vraisemblablement inactifs puisque les résidus catalytiques de leurs sites actifs ne sont pas conservés<sup>277</sup>. La résolution de la structure du module entier a révélé une organisation de l'homodimère en X à l'image des FAS animales mais dans ce cas, les deux chambres réactionnelles sont symétriques et bien plus restreintes en raison de la proximité spatiale des domaines KR et DH avec la KS. Cet arrangement est cohérent avec la fonction du module puisque celui-ci ne possède pas d'ACP et ne requiert donc pas une large chambre réactionnelle. De plus, les domaines AT, DH et ER ne nécessitent pas l'accès au substrat puisqu'ils sont catalytiquement inactifs et jouent uniquement un rôle structural de stabilisation<sup>277</sup> (**Figure 47**).

Il est à noter que la structure de la HR-iPKS LovB entière en complexe avec l'ER discrète LovC a été résolue par cryo-EM et adopte une fois de plus une conformation en X étendu<sup>162</sup>. Cependant, d'un point de vue structural, l'homodimérisation de son domaine  $\Psi$ ER et l'interface de dimérisation de son domaine DH sont similaires à celles des FAS animales, et donc incompatibles avec les observations faites sur les PKS modulaires.



**Figure 47 :** Modèles et structures de modules entiers de PKS modulaires. L'organisation séquentielle des domaines est représentée au-dessus de chaque module selon le même code couleur et les domaines non-observés sont indiqués en gris. A) Modèle de l'interface entre module 5 et 6 de la PKS de virginiamycine M basé sur des données SAXS. Des enveloppes à basse résolution du module 5 et de la fusion ACP<sub>5b</sub>-<sup>C</sup>DD-<sup>N</sup>DD-KS<sub>6</sub>-adaptateur KS-AT ont été superposées pour obtenir un modèle global (figure adaptée de Dorival *et al.*, 2016 <sup>274</sup>). B) Modèle du module de chargement de la PKS de l'apratoxine A issu de la superposition de la structure cristallographique monomérique du tridomaine MT1-ΨGNAT-MT2 et du modèle homodimérique obtenu par cryo-EM. Les deux côtés de la « pince » formés par les deux MT1 se rapprochent et s'éloignent dépendamment du stade de la catalyse (figure adaptée de Skiba *et al.*, 2018 <sup>276</sup>). C) Structure cristallographique du trio enzymatique réducteur DH-KR-ER du module 5 de la PKS de la juvénimicine. Le domaine ER est monomérique et l'interface de dimérisation du domaine DH est cohérente avec les structures cristallographiques des DH de PKS modulaires isolées (PDB : 8G7W) <sup>232</sup>. D) Modèle atomique du module de libération CalA3 de la PKS de la calcimycine basé sur des données de cryo-EM. Le domaine KS catalysant la libération et l'amidation du polycétide est étroitement lié aux domaines AT, DH et KR catalytiquement inactifs ce qui stabilise le module (figure adaptée de Wang *et al.*, 2023 <sup>277</sup>).

## b. Les domaines de docking

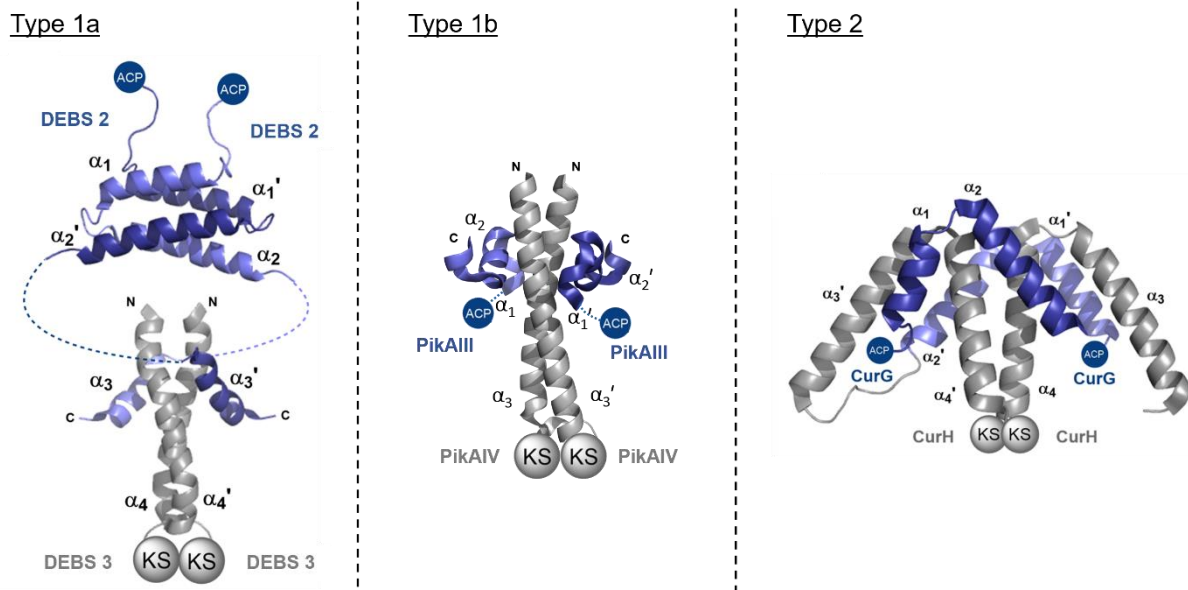
Les différents domaines catalytiques des PKS modulaires *cis*- et *trans*-AT sont souvent répartis sur plusieurs protéines de haut poids moléculaire. Ces sous-unités doivent interagir



entre elles de manière ordonnée et séquentielle afin d'assurer le cheminement et la synthèse fidèle des intermédiaires. Cette fonction est assurée par de courtes séquences présentes aux extrémités N- et C-terminales des sous-unités, qui adoptent un repliement tridimensionnel particulier. Ces domaines, appelés domaines de *docking* (DD), interagissent entre eux, formant les interfaces de communication entre chaînes polypeptidiques, et reconnaissent leur partenaire de manière spécifique pour organiser correctement la mégasynthèse <sup>253</sup>. Les jonctions entre sous-unités au sein des systèmes PKS *cis*-AT sont presque exclusivement intermodulaires, avec des interfaces entre un domaine ACP C-terminal et un domaine KS N-terminal. Dans les systèmes *trans*-AT au contraire, les jonctions intramodulaires sont fréquentes et donnent lieu à des modules scindés avec des interfaces de sous-unités entre domaines KS/KR, KS/DH et DH/KR, entre autres <sup>278</sup>.

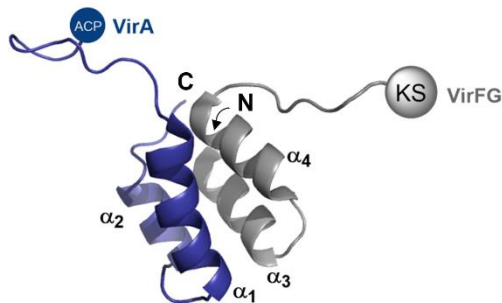
Les différentes paires de DD peuvent être classées selon le type de synthase dont elles proviennent (PKS *cis*-AT, PKS *trans*-AT ou hybride PKS-NRPS) et selon la structure des deux DD complexés. Dans les PKS *cis*-AT, deux classes ont été caractérisées, avec des DD complémentaires en C-terminal d'un domaine ACP (<sup>C</sup>DD) et N-terminal d'un domaine KS (<sup>N</sup>DD). Les DD de type 1 sont définis par un <sup>N</sup>DD constitué d'une unique hélice  $\alpha$  qui forme une superhélice (ou *coiled-coil*) dans le contexte dimérique des PKS (**Figure 48**) <sup>279-281</sup>. Le <sup>C</sup>DD est quant à lui composé d'une (DD de type 1a) ou de deux (DD de type 1b) courtes hélices  $\alpha$  qui s'enroulent autour du <sup>N</sup>DD. En N-terminal du <sup>C</sup>DD, deux hélices supplémentaires forment un fagot de quatre hélice  $\alpha$  qui agit comme élément d'homodimérisation tel qu'observé entre les sous-unités 2 et 3 de la PKS DEBS <sup>279</sup> (**Figure 48**). Il est à noter qu'une paire de DD dite « type 1-related » a été identifiée au sein de l'hybride PKS-NRPS de l'énacyloxine IIa <sup>281</sup>. Celle-ci se distingue des DD de type 1 décrits précédemment par l'absence de l'élément d'homodimérisation en N-terminal de <sup>C</sup>DD.

Le deuxième type de DD caractérisé chez les PKS *cis*-AT forme un complexe dimérique d'hétérodimère composé d'un fagot de huit hélices  $\alpha$ , chaque monomère de DD se repliant en deux hélices  $\alpha$  (**Figure 48**). La seconde hélice  $\alpha$  des deux monomères de <sup>N</sup>DD forment une superhélice au centre du fagot avec laquelle interagissent les deux hélices  $\alpha$  des deux monomères du <sup>C</sup>DD, majoritairement par contacts hydrophobes (**Figure 48**) <sup>281,282</sup>. Tout comme les paires de DD de type 1, une paire qualifiée de « type 2-related » a été identifiée et caractérisée dans l'hybride PKS-NRPS de l'énacyloxine IIa <sup>281</sup>. La structure globale de cette paire est similaire à celle des DD de type 2, mais le mode de dimérisation et la nature des interactions mises en jeu diffèrent <sup>282</sup>.

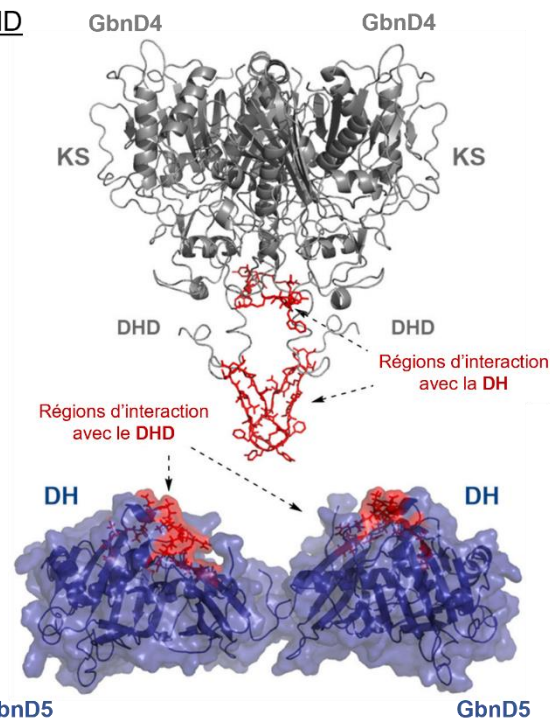


**Figure 48 :** Structures des paires de DD de types 1 et 2. Les DD de type 1a et 1b forment un fagot de quatre et six hélices  $\alpha$ , respectivement. L'élément de dimérisation observé dans la structure NMR de DD de type 1a est également présent en amont des DD de type 1b mais n'a pas été inclus dans la construction protéique cristallisée (PDB : 1PZR, 1PZQ, 3F5H)<sup>279,280</sup>. Les DD de classe 2 forment un fagot plus imposant de huit hélices  $\alpha$ , chaque monomère étant constitué de deux hélices  $\alpha$  (PDB : 4MY)<sup>282</sup> (figure adaptée de Smith *et al.*, 2021<sup>278</sup>).

Dans les PKS *trans*-AT, deux types de paires de DD ont été identifiés à ce jour. La première est constituée d'un fagot de quatre hélices  $\alpha$  mais à la différence des DD de classe 1, le <sup>C</sup>DD et le <sup>N</sup>DD sont dans ce cas monomériques, chacun étant constitué de deux hélices  $\alpha$  (**Figure 49**)<sup>274,283</sup>. Ce type de DD a d'abord été caractérisé à l'interface intermodulaire ACP/KS des sous-unités VirA et VirFG dans la PKS de la virginiamycine<sup>274</sup> puis à l'interface intramodulaire KS/KR des sous-unités MlnD et MlnE de la PKS de la macrolactine<sup>283</sup>. Plus récemment, un nouveau type de DD a été identifié à la jonction intramodulaire KS/DH des sous-unités GbnD4 et GbnD5 de la PKS de la gladioline<sup>284</sup>. Contrairement aux autres classes de DD caractérisées, celle-ci n'est pas composée d'une paire de DD mais d'un seul <sup>C</sup>DD qui interagit directement avec le domaine DH de la sous-unité suivante. Bien qu'aucune structure de ce complexe n'ait été résolue, des études biophysiques ont permis de déterminer que ce DD, qualifié de domaine déshydratase *docking* (DHD), est intrinsèquement désordonné et que le complexe DHD/DH se forme par l'intermédiaire de deux zones d'interaction (**Figure 49**)<sup>284</sup>.

Fagot 4 hélices  $\alpha$ 

## DHD



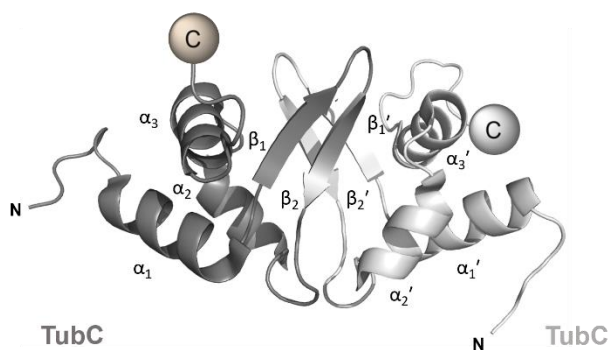
**Figure 49 :** DD caractérisés uniquement au sein de systèmes *trans*-AT. Les paires de DD organisées en fagot de quatre hélices  $\alpha$  sont composées de deux DD monomériques (contrairement aux DD de classe 1 homodimériques), chacun constitué de deux hélices  $\alpha$  (PDB : 2N5D)<sup>274</sup>. Les DHD sont quant à eux intrinsèquement désordonnés et interagissent avec le domaine DH en aval en deux zones distinctes telles que représentées ici sur le modèle structural de la jonction entre les sous-unités GbnD4 et GbnD5<sup>284</sup> (figure adaptée de Smith *et al.*, 2021<sup>278</sup>).

Enfin, l'étude des systèmes hybrides PKS-NRPS a révélé la présence d'un autre type de DD composé d'un court <sup>C</sup>DD intrinsèquement désordonné (motif SLiM) et d'un <sup>N</sup>DD structuré en épingle  $\beta$  (domaine  $\beta$ hD) (**Figure 50**). Ces paires de DD ont été caractérisées à l'interface de domaines NRPS<sup>285–287</sup> et également identifiées au sein de systèmes purement NRPS<sup>288</sup>. Le domaine  $\beta$ hD consiste en un repliement  $\alpha\beta\beta\alpha$  dont les deux brins  $\beta$  forment l'épingle et le <sup>C</sup>DD désordonné se replie en un court brin  $\beta$  lors de son interaction avec le deuxième brin du domaine  $\beta$ hD formant ainsi un feuillet  $\beta$  à trois brins (**Figure 50**). Par ailleurs, la résolution de la structure en solution du domaine N-terminal  $\beta$ hD de la sous-unité TubC de la PKS de tubulyline a révélé une organisation homodimérique en l'absence de son <sup>C</sup>DD partenaire<sup>285</sup> (**Figure 50**). Cette dimérisation est la probable conséquence de son appartenance à une synthase hybride PKS-NRPS, dans laquelle les modules PKS sont homodimériques.

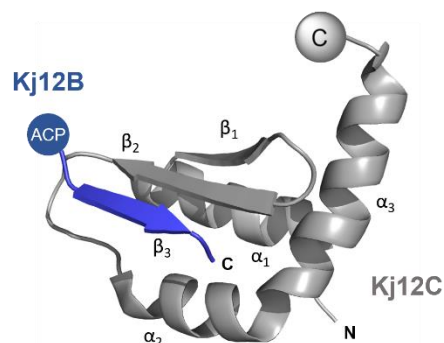
Ces paires de DD ont été caractérisées uniquement entre PCP et domaine C ou PCP et domaine d'hétérocyclisation. Cependant, des analyses bioinformatiques basées sur la séquence des <sup>C</sup>DD de ce type identifiés dans les PKS-NRPS de l'énacyloxine IIa<sup>287</sup>, de la tubulyline<sup>285</sup> et de l'épothilone<sup>286</sup> ont révélé la présence de ces paires de DD dans de nombreuses autres mégasynthèses, formant l'interface entre des domaines variés tels que

des TE, des halogénases ou encore des domaines d'épimérisation <sup>287</sup>.

Domaine βhD N-terminal homodimérique



Complexe <sup>C</sup>DD - βhD



**Figure 50 :** Structures de DD impliquant un domaine N-terminal βhD. La structure du domaine βhD N-terminal de la sous-unité TubC de la PKS-NRPS de la tubulyisine résolue par résonance magnétique nucléaire (RMN) (à gauche) révèle une organisation homodimérique dont l'interface est médiée par interaction des deux épingles β (PDB : 2JUG) <sup>285</sup>. Le rôle de cette dimérisation et son maintien ou non en présence de son <sup>C</sup>DD partenaire restent inconnus puisque les deux DD n'ont pas été étudiés en complexe. Le <sup>C</sup>DD de la sous-unité Kj12B et le <sup>N</sup>DD de la sous-unité Kj12C de la NRPS Kj12ABC sont les seuls DD de ce type dont la structure en complexe a été résolue. L'interaction des deux partenaires induit un repliement du <sup>C</sup>DD en brin β qui se lie à l'épingle β de domaine βhD N-terminal (PDB : 6EWW) <sup>288</sup>.

## V. Caractérisation de domaines non-canoniques

### 1. De nouvelles activités pour étendre la chimie des PKS

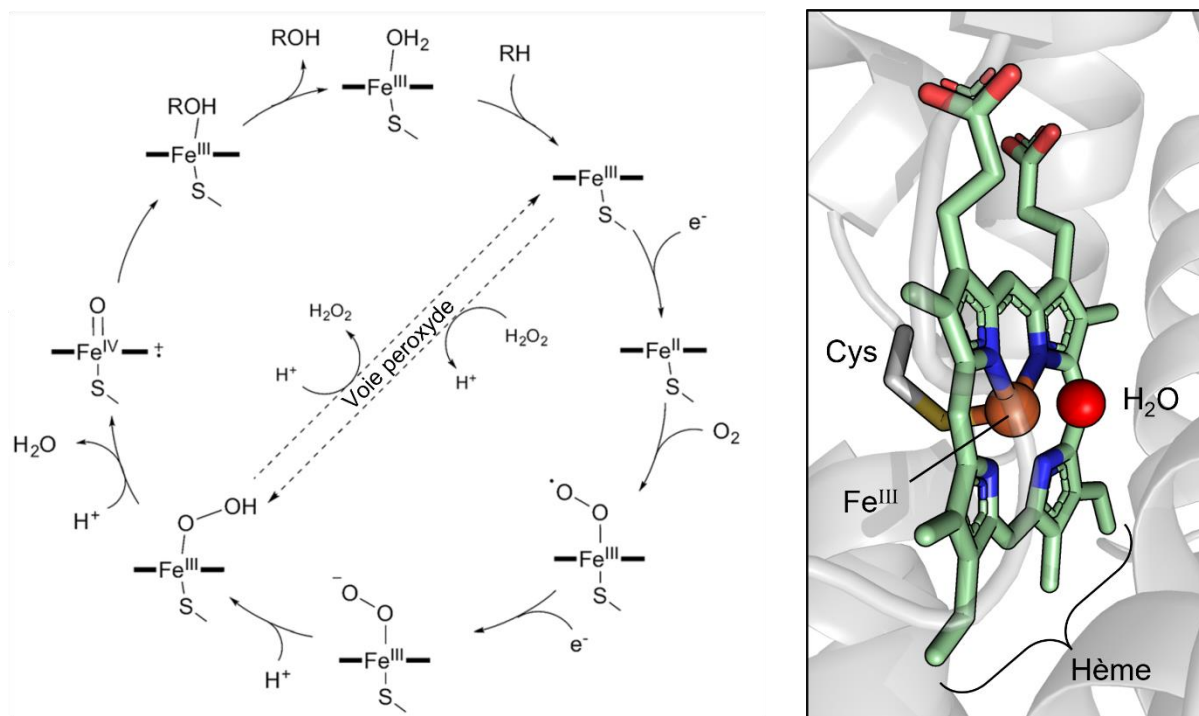
L'étude structurale et fonctionnelle des domaines de PKS a permis de comprendre et manipuler les activités enzymatiques essentielles à l'élongation (AT, ACP, KS), à la réduction des intermédiaires polycétones (KR, DH, ER) et à leur libération (TE). Cependant, le potentiel biosynthétique de ces mégasynthèses est loin de se limiter à ces quelques domaines. En effet, la découverte et la caractérisation d'un nombre croissant de systèmes *cis*- et *trans*-AT met en lumière de nombreuses activités enzymatiques additionnelles, portées par des protéines discrètes ou des domaines intégrés aux modules <sup>176,289</sup>. Ces diverses activités peu communes incluent des C- et O-méthyltransférases <sup>290,291</sup>, des énoyle isomérases catalysant l'isomérisation d'une double liaison <sup>292</sup>, des pyrane synthèses catalysant la formation de cycles à 5 ou 6 atomes <sup>293</sup>, des halogénases <sup>294</sup> ou encore des domaines de β-branchement type Michael <sup>295</sup>. La liste est loin d'être exhaustive et il apparaît essentiel de poursuivre les efforts de caractérisation structurale et fonctionnelle de ces enzymes afin d'étoffer la boîte à outils moléculaire disponible pour l'ingénierie des PKS.

## 2. Activités monooxygénases

La présence de monooxygénases a été rapportée dans différents BGC de PKS *trans*-AT et elles sont responsables d'un large éventail de réactions impliquant l'insertion d'un atome d'oxygène dans la chaîne polycétidique <sup>176</sup>. Ces enzymes peuvent se distinguer en deux catégories selon le cofacteur qu'elles emploient avec d'une part les cytochromes P450 (CYP) et d'autre part les monooxygénases flavine-dépendantes (FMO).

### a. Les cytochromes P450

Les CYP sont des métalloenzymes hème-dépendantes impliquées majoritairement dans des réactions d'hydroxylation catalysées selon le mécanisme illustré en **Figure 51**. Sous sa forme stable, l'atome de fer ferrique  $\text{Fe}^{\text{III}}$  central de l'hème est coordonné par un ion thiolate de l'enzyme et par une molécule d'eau. Lors de la fixation du substrat, cette molécule d'eau est déplacée ce qui rend le fer plus prompt à la réduction. Le gain d'un électron réduit alors le fer ferrique en fer ferreux  $\text{Fe}^{\text{II}}$  qui fixe ensuite une molécule de dioxygène, suivi d'une seconde réduction générant un intermédiaire peroxy-ferrique ( $\text{Fe}^{\text{III}}\text{-OO}^{2-}$ ) (**Figure 51**). Deux protonations successives permettent ensuite la formation d'un intermédiaire oxo-ferryl ( $\text{Fe}^{\text{IV}}=\text{O}$ ) et la libération d'une molécule d'eau. L'hème sous cette forme capte un proton du substrat auquel est finalement transmis la fonction hydroxyle formée (**Figure 51**) <sup>296</sup>.



**Figure 51 :** Cycle catalytique des CYP lors de l'hydroxylation du substrat (à gauche) et structure de l'hème (à droite). La voie peroxyde permet la formation de l'intermédiaire hydroperoxo-ferrique à partir de  $\text{Fe}^{\text{III}}$  en présence de l'enzyme et en réagissant avec une molécule de peroxyde d'hydrogène (figure reproduite de Rudolf *et al.*, 2017 <sup>296</sup>). La structure cristallographique de la CYP AbyV impliquée dans la biosynthèse de l'abyssomicine C

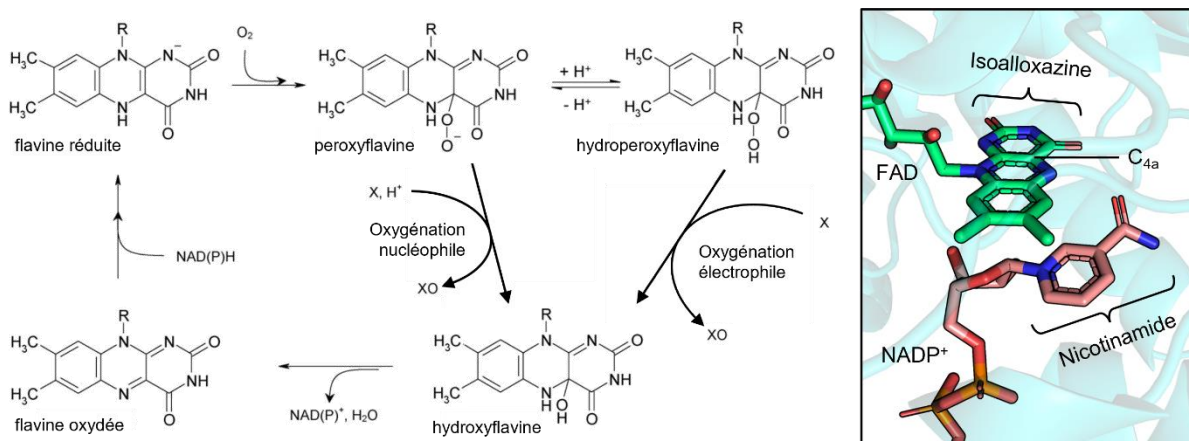
## État de l'art

permet d'observer la conformation de l'hème dans sa forme stable, dont l'atome de fer ferrique au centre est coordonné de part et d'autre de la protoporphyrine IX par une cystéine de la CYP et une molécule d'eau (PDB : 7QAN)<sup>297</sup>. La protoporphyrine IX est représentée en vert, et l'atome de fer et la molécule d'eau sont représentés par une sphère.

L'intervention des CYP dans la biosynthèse de nombreux polycétides<sup>176</sup> a motivé l'étude structurale et fonctionnelle de ces monooxygénases et a permis la découverte et l'analyse de nouvelles activités au-delà de l'hydroxylation classique. Par exemple, la CYP AurH de la PKS de l'aureothine catalyse la formation successive de deux liaisons C-O lui conférant une double fonction d'oxygénation-hétérocyclisation<sup>298</sup>. La CYP BonL quant à elle, interagit avec le précurseur de l'acide bongkreique après sa libération de la PKS et catalyse la formation d'un acide carboxylique sur le carbone terminal de la chaîne pour générer le produit final<sup>299</sup>. Une activité encore différente est celle de la CYP RhiH qui catalyse l'époxydation du polycétide rhizoxine après sa libération par macrolactonisation<sup>300</sup>. Enfin, un dernier exemple de CYP à la chimie inhabituelle sont les enzymes LtmK et BaeS des PKS *trans*-AT de la lactimidomycine et de la bacillaene, respectivement, possédant une activité désaturase générant une double liaison C=C à partir de la liaison saturée correspondante<sup>301,302</sup>. Toutes les CYP impliquées dans la biosynthèse de polycétides de type I identifiées à ce jour ont cependant un point commun, ce sont des enzymes discrètes post-PKS oxygénant leurs substrats après que ceux-ci aient été libérés de leurs PKS modulaires<sup>176</sup>.

### b. Les monooxygénases flavine-dépendantes

Les FMO sont des monooxygénases utilisant un cofacteur flavinique, flavine adénine dinucléotide (FAD) ou FMN, pour activer une molécule de dioxygène par formation d'un intermédiaire covalent C<sub>4a</sub>-(hydro)peroxyflavine dont un atome d'oxygène sera incorporé au substrat. Pour la majorité des FMO, le rôle de donneur d'électron est joué par le NAD(P)H qui permet la réduction du FAD (**Figure 52**)<sup>303</sup>.



**Figure 52 :** Mécanisme catalytique général des FMO (à gauche) et interface FAD/NADP<sup>+</sup> (à droite). L'état de protonation de l'intermédiaire peroxyflavine définit la nature nucléophile ou électrophile de la réaction d'oxygénation

(figure reproduite de van Berkel *et al.*, 2006<sup>304</sup>). L'interface entre le FAD et le NADP<sup>+</sup> peut être observée dans la structure du complexe ternaire de la FMO ZvPNO (PDB : 5NMX)<sup>305</sup>.

Les activités FMO identifiées au sein de BGC de PKS modulaires sont dans certains cas portées par des enzymes post-PKS<sup>306</sup> mais elles peuvent également être catalysées par des enzymes discrètes complétant un module en *trans*, ou directement portées par un domaine catalytique au sein d'un module catalysant de fait la monooxygénation du polycétide en cours d'élongation<sup>176</sup>. Dans ces deux derniers cas, les activités FMO seront qualifiées de modulaires, par opposition aux activités post-PKS, pour souligner leur action concomitante à la synthèse des polycétides au sein des modules. Cependant, contrairement aux CYP, peu d'informations mécanistiques ont été obtenues sur les FMO de PKS modulaires. Ainsi, bien que les FMO catalysent une grande variété de réactions d'oxydation<sup>304</sup>, peu de réactions différentes ont été démontrées au cours de la synthèse de polycétide.

#### i. Oxydation de type Baeyer-Villiger

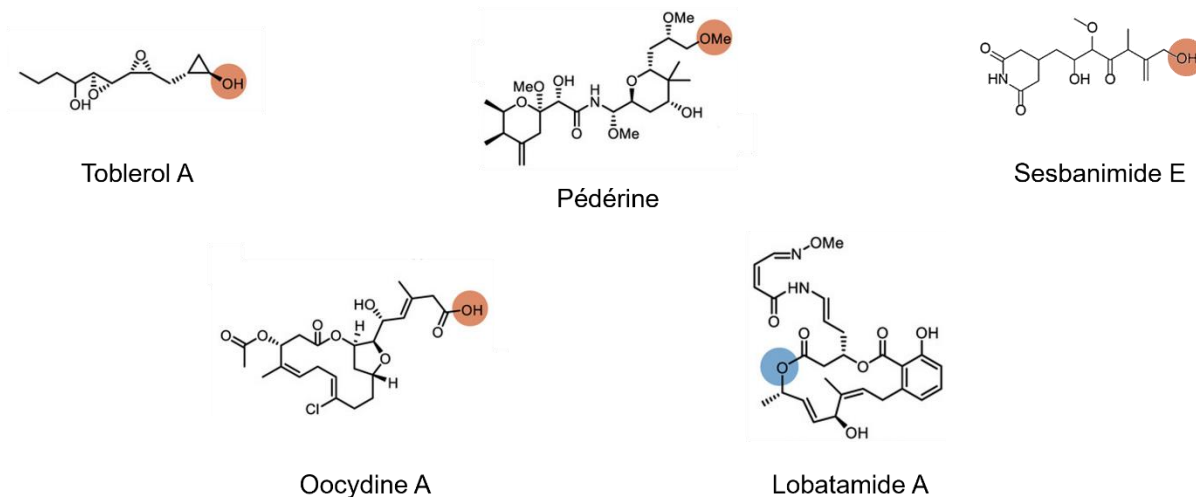
La majorité des FMO de PKS caractérisées catalyse une monooxygénation de type Baeyer-Villiger (BVMO), consistant en la conversion de la fonction  $\beta$ -cétone de l'intermédiaire de synthèse en ester, par insertion d'un atome d'oxygène dans le squelette polycétidique. Cette activité peut être portée un domaine de la PKS, comme c'est le cas du domaine LbmC-BVMO du module 9 de la PKS de la lobatamide<sup>307</sup>, et peut également être apportée en *trans* par une enzyme discrète, telle que les BVMO PedG<sup>308</sup>, OocK<sup>308</sup>, SbnP<sup>309</sup> et TobD<sup>310</sup> des voies de biosynthèse de la péderine, de l'oocydine, de la sesbanimide, et du toblerol respectivement.

Bien que ces enzymes soient toutes des BVMO, l'insertion d'oxygène qu'elles catalysent à des conséquences variées sur la structure finale du polycétide. Dans le cas des BVMO discrètes citées ci-dessus, la fonction ester qu'elles génèrent est absente de la structure du polycétide mature. En effet, le produit de biosynthèse final est obtenu par la synthèse d'un précurseur de plus grande taille dont l'ester interne est hydrolysé, formant ainsi une fonction terminale hydroxyle ou carboxylate<sup>308</sup> (**Figure 53**). Les PKS *trans*-AT de la péderine et de l'onnamide A illustre parfaitement ce phénomène. Malgré une architecture modulaire relativement similaire, ces PKS génèrent deux polycétides aux extrémités bien différentes, l'onnamide A semblant être un homologue élongué de la péderine. L'inspection du BGC de ces deux PKS révèle la présence d'un gène codant une BVMO, PedG, dont la position coïncide avec la formation de l'extrémité méthoxyle de la péderine, du point de vue de la colinéarité. L'absence d'un gène homologue à la même position dans le BGC de la PKS de l'onnamide A explique ainsi le fait que l'extrémité du polycétide ne contient pas de fonction ester, et n'est donc pas hydrolysée<sup>308</sup> (**Figure 53**).

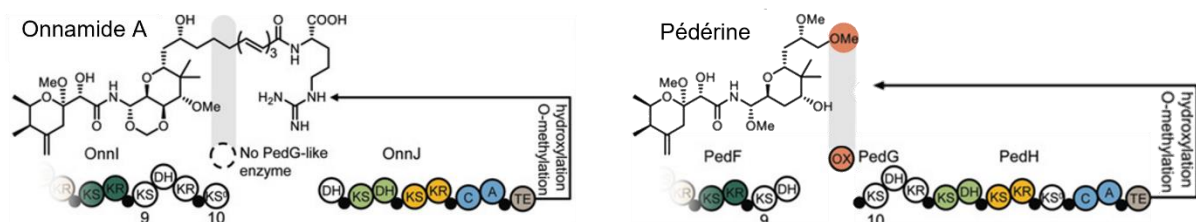
## État de l'art

À l'inverse, l'ester généré par le domaine BVMO de la sous-unité LbmC n'est pas hydrolysé et est donc retrouvé dans la structure de la lobatamide A (**Figure 53**)<sup>307</sup>. Par ailleurs, la régiospécificité de ce domaine diffère de celles des BVMO discrètes, puisqu'il insère un atome d'oxygène en position  $\beta$  (et non  $\gamma$ ) de l'intermédiaire de synthèse linéaire, formant un ester de configuration opposée<sup>307</sup>.

A



B



**Figure 53 :** Fonctions chimiques de polycétides générées par des activités BVMO modulaires. A) Structures de polycétides dont la biosynthèse implique une activité BVMO lors de l'élongation de la chaîne polycétidique. L'ester généré par les BVMO discrètes entraîne la formation d'un hydroxyle ou carboxylate terminal après hydrolyse, surlignés en rouge. Dans le cas de la pédérine, l'hydroxyle terminal est ensuite méthylé, générant une fonction méthoxyle. Le domaine LbmC-BVMO génère quant à lui un ester interne conservé après libération du polycétide, surligné en bleu (figure adaptée de Ueoka *et al.*, 2020<sup>307</sup>). B) Comparaison des PKS de l'onnamide A et de la pédérine. D'après la règle de colinéarité, le point de divergence des deux structures polycétidiques coïncide avec la présence d'une BVMO au sein de la PKS de la pédérine, suggérant l'hydrolyse d'un intermédiaire de plus grand taille similaire à l'onnamide A (figure reproduite de Meoded *et al.*, 2018<sup>308</sup>).

### ii. $\alpha$ -Hydroxylation

L' $\alpha$ -hydroxylation de la chaîne polycétidique consiste en l'ajout d'un groupement hydroxyle en position  $\alpha$  de l'unité d'extension nouvellement condensée. Cette activité a été notamment identifiée dans la voie de biosynthèse de l'oocydine où elle est portée par une



enzyme discrète, OocM, qui catalyse l'hydroxylation de l'intermédiaire de synthèse lié à un module scindé entre deux sous-unités<sup>311</sup>. Cette  $\alpha$ -hydroxylase est aujourd'hui le seul exemple caractérisé fonctionnellement de ce type d'activité, bien que des études bioinformatiques aient prédit la présence d'homologues de cette enzyme dans d'autres PKS<sup>311</sup>.

### iii. Formation d'oxime

Un autre type de monooxygénation catalysé par les FMO dans les voies de biosynthèse polycétidique est la formation d'oxime, impliquant l'oxydation d'un groupement amine. Cette activité a été identifiée au sein d'enzymes post-PKS, telle que la FMO CrmH impliquée dans la biosynthèse de caerulomycine A<sup>312</sup>. De manière intéressante, une telle réaction est impliquée dans la biosynthèse de lobatamide. Dans ce système, l'oxime ne serait pas formée après libération du polycétide mais au cours de sa synthèse, faisant de cette réaction une fonction modulaire de la PKS. La règle de colinéarité suggère que le domaine FMO de la première sous-unité de la PKS, LbmA-Ox, est le responsable vraisemblable de cette formation d'oxime<sup>307</sup>.

### iv. Classes et structures de FMO

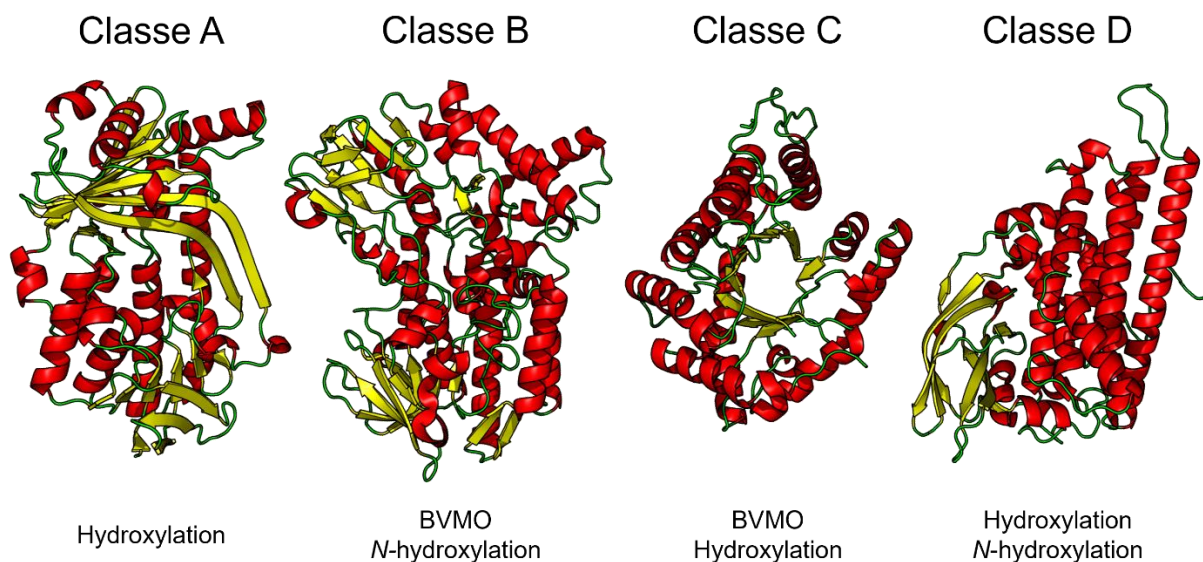
En fonction de leurs propriétés structurales et fonctionnelles, les FMO peuvent être réparties en huit classes<sup>303</sup>. Celles-ci sont définies par la nature du cofacteur flavinique et du donneur d'électron, le repliement tridimensionnel de l'enzyme, et la réaction de monooxygénation catalysée. Bien que très peu de données de ce genre soient disponibles pour les FMO de PKS, de nombreuses enzymes possédant une activité similaire ont été caractérisées. Ces informations permettent donc de déterminer à quelles classes peuvent appartenir les FMO identifiées, et ainsi d'émettre les premières hypothèses sur leurs structures et mécanismes enzymatiques.

Les monooxygénations de type BVMO sont catalysées par des FMO des classes B et C. Au sein de la classe B, les enzymes possèdent un repliement en deux domaines structuraux de type Rossmann (**Figure 54**) leur permettant de fixer d'une part le FAD, et d'autre part le NAD(P)H qui reste fixé durant la catalyse<sup>303</sup>. Au contraire les FMO de classe C présentent une structure en tonneau TIM (**Figure 54**) et lie transitoirement le FMN. Après oxydation de la flavine, celle-ci est à nouveau réduite par l'intermédiaire d'une flavine réductase NAD(P)H-dépendante<sup>303</sup>.

Les réactions d'hydroxylation sont quant à elles catalysées par des FMO de trois classes distinctes, A, C et D. À l'instar de la classe B, les enzymes de classe A sont FAD-dépendantes et utilisent le NAD(P)H en tant que donneur d'électrons. En revanche, le départ du NAD(P)<sup>+</sup> est nécessaire à la fixation du substrat et ces FMO ont un repliement Rossmann de

type glutathion réductase (**Figure 54**). Quant aux enzymes de classe D, elles présentent une structure de type acyl-CoA déshydrogénase (**Figure 54**) et reçoivent le FAD ou le FMN réduits d'une flavine réductase NAD(P)H-dépendante.

Finalement, la formation d'oxime implique une *N*-hydroxylation du substrat, une réaction observée chez des FMO de type B et D.



**Figure 54 :** Structures représentatives des classes de FMO catalysant une réaction retrouvée dans les PKS *trans*-AT. Les classes A, B, C et D sont illustrées par les structures de la 4-hydroxybenzoate 3-hydroxylase (PDB : PBE), la phénylacétone monooxygénase (PDB : 1W4X), la luciférase bactérienne (PDB : 1LUC) et la 4-hydroxyphénylacétate 3-hydroxylase (PDB : 2JBS), respectivement. Les réactions catalysées par ces FMO et identifiées au sein des PKS sont indiquées sous les classes correspondantes.





---

# *Objectifs*

---



La caractérisation des activités enzymatiques de PKS est un élément clef dans le processus d'exploitation de ces enzymes pour en contrôler l'activité et/ou les insérer dans d'autres voies de biosynthèses afin de synthétiser de nouveaux analogues. À cet égard, les FMO identifiées dans des PKS *trans*-AT apparaissent comme un objet d'étude attrayant, étant donné les informations structurales extrêmement lacunaires disponibles à leur sujet. Par conséquent, les travaux présentés dans cette thèse avaient pour objectif de décrypter le mécanisme catalytique de ces FMO ainsi que d'en apprendre davantage sur les interactions avec leur partenaires protéiques au sein des synthèses.

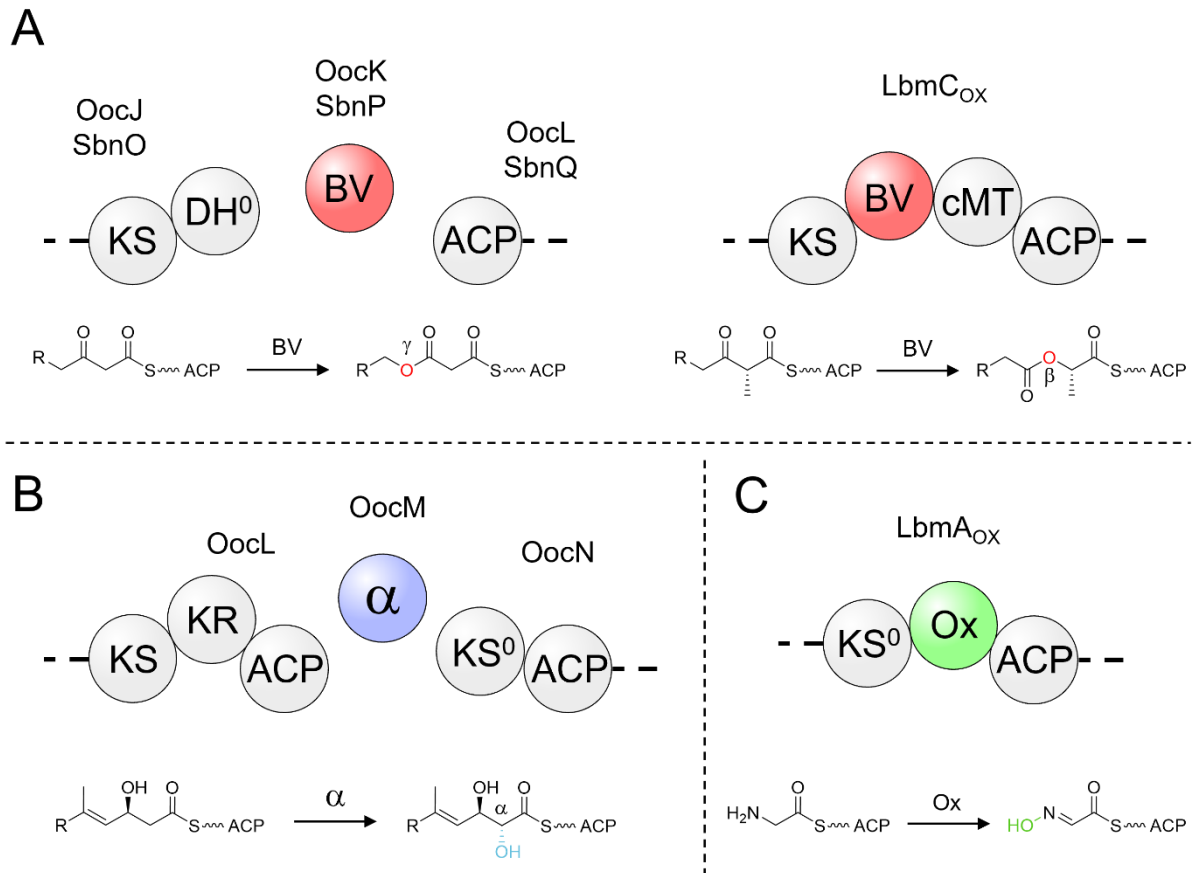
Les modèles d'étude se devaient d'être représentatifs de la diversité des activités de ces FMO et de la diversité des modules dans lesquels elles sont impliquées. Par conséquent, l'étude de l'activité BVMO s'est intéressée aux modules scindées OocJKL et SbnOPQ des PKS-NRPS de l'oocydine et de la sesbanimide, respectivement, et au module LbmC<sub>ox</sub> de la PKS-NRPS de la lobatamide A. Les deux modules scindés d'intérêt présentent une architecture modulaire identique avec la présence sur la sous-unité en amont (OocJ/SbnO) d'une KS suivie d'une DH inactive dont l'histidine catalytique est absente<sup>308,309</sup> et d'un ACP en N-terminal de la sous-unité suivante (OocL/SbnQ) (**Figure 55**). Le module LbmC<sub>ox</sub> est formé quant à lui de quatre domaines liés covalamment, KS-BVMO-cMT-ACP de N- en C-terminal (**Figure 55**). En plus de la divergence de structure des modules desquels elles font partie, les BVMOs discrètes et le domaine BVMO présente une intrigante différence de régiospécificité. En effet, OocK et SbnP insèrent un atome d'oxygène en position  $\gamma$  du polycétide<sup>308,309</sup> tandis que le domaine LbmC<sub>ox</sub>-BVMO insère cet atome en position  $\beta$ , générant un ester de conformation opposée<sup>307</sup>. L'étude de la sélectivité de substrat de ce domaine a démontré que cette chimie inhabituelle se produit uniquement en présence de la fonction  $\alpha$ -méthyle installée par le domaine MT en aval et que son absence engendre la formation d'un ester de conformation identique aux cas de OocK et SbnP<sup>307</sup>.

L'investigation des mécanismes d' $\alpha$ -hydroxylation s'est portée sur la FMO discrète OocM intervenant lors de la synthèse de l'oocydine au sein du module scindée OocLMN. Celui-ci est constitué du tridomaine C-terminal KS-KR-ACP de la sous-unité OocL et du didomaine N-terminal KS<sup>0</sup>-ACP de la sous-unité OocN (**Figure 55**). La reconstitution *in vitro* de l'activité de OocM a montré que l' $\alpha$ -hydroxylation se produit avant la  $\beta$ -kétoréduction et que la KS<sup>0</sup> N-terminale de OocN, bien qu'elle soit incapable de catalyser une extension de chaîne, est indispensable à la monooxygénation<sup>311</sup>.

Finalement, l'investigation des mécanismes de formation d'oxime a été exemplifiée par le module LbmA<sub>ox</sub> de la PKS-NRPS de la lobatamide A, constitué d'une KS<sup>0</sup> suivie d'un domaine monooxygénase et d'un ACP (**Figure 55**). En s'appuyant sur la règle de colinéarité et sur la prédiction de spécificité de substrat du module en amont, il a été supposé que cette

## Objectifs

monoxygénase est responsable de l'oxygénation d'une unité d'initiation glycine, générant une fonction oxime, méthyliée par le module suivant <sup>307</sup>. Cette hypothèse a par la suite été confirmée par reconstitution *in vitro* de l'activité de ce domaine (résultats intégrés à l'article scientifique issu de cette thèse).



**Figure 55 :** Modèles d'étude de FMO modulaires. A) Modules à activité BVMO. Les modules scindés OocJKL et SbnOPQ (à gauche) présentent la même architecture et font intervenir les BVMO discrètes OocK et SbnP, respectivement. Le module LbmCOX de la PKS de la lobatamide contient quant à lui un domaine BVMO encadré par des domaines KS et MT. B) Le module scindé OocLMN catalyse une hydroxylation du polycétide par le biais de l' $\alpha$ -hydroxylase discrète OocM. C) Le module non-élongateur LbmAOX contient un domaine FMO responsable de la formation d'une fonction oxime.

La diversité des réactions et des architectures modulaires impliquant des FMO soulèvent de nombreuses questions. Quelles sont les divergences structurales entre BVMO discrètes et domaines BVMO engendrées par la différence d'architecture modulaire et permettant les interactions avec les partenaires protéiques ? Le domaine DH<sup>0</sup> des modules BVMO scindé est catalytiquement inactif et pourtant conservé entre les PKS de l'oocydine et de la sesbanimide – quel est donc son rôle ? Quelle(s) spécificité(s) du site actif des différentes BVMO permet(tent) d'expliquer la divergence de régiospécificité ? Quel est le rôle de la fonction  $\alpha$ -méthyle du substrat de LbmCOX-BVMO dans la catalyse ? Comment OocM est-elle recrutée par ses partenaires et pourquoi la KS<sup>0</sup> est-elle nécessaire à la monoxygénation ? Joue-t-elle



un rôle d'intermédiaire entre OocM et l'ACP acylé ? Quel mécanisme catalytique permet à LbmA-Ox d'oxygéner une fonction amine ? Comment interagit-elle avec les domaines de son module ?

Pour répondre à ces questions, notre objectif était dans un premier temps de parvenir à produire de manière stable chacune des FMO d'intérêt et de les purifier à homogénéité en ayant recours à diverses constructions protéiques. Les monooxygénases pourraient ensuite être soumises à des études structurales combinant cristallographie aux rayons X et diffusion des rayons X aux petits angles (SAXS) pour en déterminer la structure et le comportement en solution. L'obtention de structures à haute résolution permettraient alors de comprendre l'impact de l'incorporation de ces FMO à des systèmes PKS sur leurs repliements tridimensionnels et d'identifier les acides aminés potentiellement impliqués dans la catalyse enzymatique. Le rôle de ces résidus serait ensuite confirmé ou invalidé par une approche de mutagenèse dirigée couplée à des tests enzymatiques, permis par les compétences de chimie de synthèse et les infrastructures présentes au sein du laboratoire du Pr. Jörn Piel à l'institut de microbiologie de l'ETH Zürich, CH.

La compréhension des mécanismes d'interactions protéine-protéine au sein des modules est une connaissance essentielle pour l'ingénierie génétique puisque ces interactions doivent être maintenues lors des expériences d'échanges de domaines afin de générer un module fonctionnel. Or, peu d'informations sont disponibles sur le recrutement de partenaires discrets pourtant largement présents dans les systèmes *trans*-AT. C'est pourquoi, dans un second temps, nous souhaitons découvrir les différents facteurs nécessaires à l'assemblage des modules BVMO et  $\alpha$ -hydroxylase scindés. Un tel objectif serait permis par une approche alliant la caractérisation des modules en solution par SAXS et l'étude de la capacité d'interaction des différents domaines de ces modules. Cela nous permettrait notamment de découvrir le rôle du domaine DH<sup>0</sup> conservé des modules BVMO scindés et de mettre en lumière les éléments structuraux nécessaires à la formation des complexes protéine-protéine impliquant les FMO.



---

*Matériels et  
Méthodes*

---



## I. Matériel biologique et solutions d'usage courant

### 1. Souches bactériennes

***Escherichia coli* DH5 $\alpha$**  :  $F^-$  *supE44*,  $\Delta$ *lacU169* ( $\phi$ 80 *lacZ* $\Delta$ M15), *hsdR17*, *recA1*, *endA1*, *gyrA96*, *thi-1*, *relA1*. Cette souche a été utilisée pour la propagation de plasmides après clonage d'un insert. Sa caractéristique principale est un taux de transformation élevé.

***Escherichia coli* BL21 (DE3)** :  $F^-$  *ompT gal dcm lon hsdSB* ( $r_B^- m_B^-$ )  $\lambda$ (DE3). Cette souche a été utilisée pour exprimer certaines protéines recombinantes. Elle dispose dans son génome de la séquence codant l'ARN polymérase du phage T7 sous le contrôle d'un promoteur inductible à l'isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). Ceci permet l'expression de protéines dont le gène est inséré dans un plasmide de type pET, et placé sous le contrôle du promoteur T7. De plus, cette souche est déficiente pour certaines protéases, ce qui augmente le rendement de la protéine surexprimée.

### 2. Vecteurs

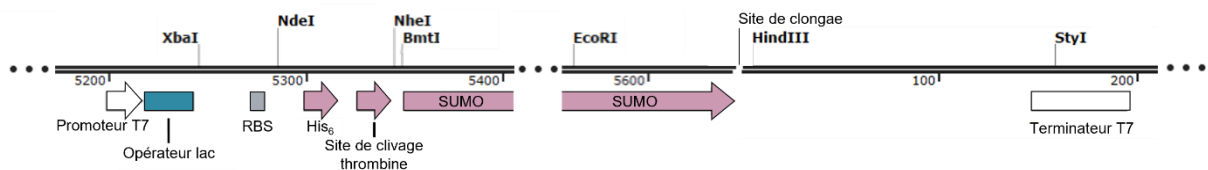
**pBG102** (fourni par le Center for Structural Biology, Vanderbilt University) : ce vecteur d'expression est dérivé du pET27 (Novagen®) et permet d'exprimer une protéine en fusion avec une étiquette N-terminale de six histidines suivie d'une protéine *Small Ubiquitin-like Modifier* (SUMO) qui permettent une meilleure solubilité de la construction protéique et facilite sa purification. Un site de clivage reconnu spécifiquement par la protéase 3C du rhinovirus humain (HRV 3C) est présent entre l'étiquette His<sub>6</sub>-SUMO et le site de clonage multiple. Le plasmide est maintenu dans l'hôte bactérien grâce à un gène de résistance à la kanamycine (**Figure 56**).

**pET-SUMO** (fourni par le Pr. Jörn Piel, ETH Zürich) : ce vecteur d'expression est dérivé du pET28 (Novagen®) et permet d'exprimer une protéine en fusion avec une étiquette N-terminale de six histidines suivie d'une protéine SUMO. À la différence du pBG102, aucune séquence codant un site spécifique de clivage reconnu par une protéase n'est présente entre la séquence SUMO et le site de clonage multiple. En revanche, un site reconnu par la thrombine est présent entre l'étiquette His<sub>6</sub> et l'étiquette SUMO. Ce plasmide est maintenu dans l'hôte bactérien grâce à un gène de résistance à la kanamycine (**Figure 57**).

## Matériels et méthodes



**Figure 56 :** Carte nucléotidique du plasmide d'expression pBG102, fourni par le Center for Structural Biology (Vanderbilt University). Seule la région comprenant le site multiple de clonage (MCS) et codant les diverses étiquettes et sites de clivage est représentée. Le pBG102 permet de fusionner une protéine avec une étiquette N-terminale His<sub>6</sub>-SUMO. Celle-ci peut être séparée de la protéine d'intérêt par action de la protéase HRV 3C sur son site de clivage spécifique.



**Figure 57 :** Carte nucléotidique du plasmide d'expression pET-SUMO, fourni par le Pr. Jörn Piel (ETH Zürich). Seule la région entre le promoteur et le terminateur T7 est représentée. Ce plasmide permet l'expression d'une protéine en fusion N-terminale avec une étiquette His<sub>6</sub>-SUMO. L'étiquette His<sub>6</sub> peut être séparée de la protéine d'intérêt par action de la thrombine sur son site de clivage spécifique. Ce vecteur est commercialisé sous forme linéaire permettant un clonage TA des amplicons de réactions en chaîne par polymérase (PCR).

### 3. Milieux de culture

Tous les milieux ont été stérilisés préalablement à leur utilisation par autoclavage à 120°C sous 1 bar de pression.

**LB (Lysogeny Broth) :** ce milieu riche est composé de tryptone (10 g.L<sup>-1</sup>), d'extrait de levure (5 g.L<sup>-1</sup>) et de chlorure de sodium (5 g.L<sup>-1</sup>) et son pH est ajusté à 7.5 par ajout d'hydroxyde de sodium. Il a été employé pour la croissance des différentes souches d'*E. coli* lors de pré-cultures et de régénérations à la suite de transformation par un plasmide. La sélection des bactéries transformées a été réalisée sur milieu LB gélosé par ajout d'agar (17 g.L<sup>-1</sup>).

**TB (Terrific Broth) :** ce milieu riche est composé de tryptone (12 g.L<sup>-1</sup>), d'extrait de levure (24

g.L<sup>-1</sup>), de glycérol (4% (v/v)), de KH<sub>2</sub>PO<sub>4</sub> (2,31 g.L<sup>-1</sup>) et de K<sub>2</sub>HPO<sub>4</sub> (12,5 g.L<sup>-1</sup>). Il a été utilisé comme milieu de culture pour l'expression des différentes protéines recombinantes puisque ce milieu est recommandé dans le cadre de la production de protéines issues de PKS et NRPS<sup>313</sup>.

#### 4. Solutions d'usage courant

**Bleu de dépôt pour gel d'agarose (6X)** : Tris-HCl 10 mM, pH 8.0 ; EDTA 1 mM ; glycérol 25 % (v/v) ; bleu de bromophénol 0,12 % (p/v).

**Tampon de migration pour électrophorèse en gel d'agarose** : Tris 40 mM ; acétate de sodium 5 mM, pH 8.0 ; EDTA 1 mM.

**Bleu de dépôt pour gel de polyacrylamide** : Tris-HCl 80 mM, pH 6.8 ; SDS 3 % (v/v) ; glycérol 10 % (v/v) ; bleu de bromophénol 0,1 % (p/v) ; DTT 100 mM.

**Marqueur de taille protéique MiMoPA** : T7 ARN polymérase (97 kDa), albumine de sérum bovin (66 kDa), rétinaldéhyde déshydrogénase (53 kDa), ovalbumine glycosylée (44,8 kDa), ovalbumine (42,8 kDa), glyceraldéhyde phosphate déshydrogénase (37,8 kDa), MsrA (24 kDa), thiorédoxine (12 kDa).

**Tampon de migration pour électrophorèse en gel de polyacrylamide dénaturant (SDS-PAGE)** : Tris-HCl 50 mM, pH 8.5 ; glycine 200 mM ; SDS 0,1 % (v/v).

**Solution de déshydratation des gels de polyacrylamide** : éthanol 50 % ; acide acétique 10 %.

**Solution de coloration des gels de polyacrylamide** : éthanol 5 % ; acide acétique 7,5 % ; solution de bleu de Coomassie G250 (bleu de Coomassie G250 0,25 %, éthanol 96 %).

**Tampon A** : Tris 20 mM, NaCl 200 mM, 10 % glycérol, pH 8.0.

**Tampon B** : Tris 20 mM, NaCl 200 mM, 1 M imidazole, pH 8.0

**Tampon C** : Tris 20 mM, NaCl 200 mM, 5 % glycérol, pH 8.0.

## II. Sous-clonage et mutagenèse des séquences codantes d'intérêt

Des plasmides recombinants contenant les séquences codant les différentes protéines d'intérêt ont été fournis par l'équipe du professeur Jörn Piel (ETH Zürich). La séquence codant la sous-unité LbmA de la PKS de la lobatamide a été clonée par le Dr Hannah A. Minas, les séquences codant les module LbmC<sub>Ox</sub>, et les BVMO OocK et SbnP des PKS de la lobatamide, de l'oocydine et de la sesbanimide respectivement ont été clonées par le Dr Amy E. Fraley et la séquence codant l' $\alpha$ -hydroxylase OocM de la PKS de l'oocydine a été clonée par le Dr Franziska Hemmerling.

### 1. Amplification des séquences codantes

Les séquences nucléotidiques codant les diverses FMO d'intérêt ont été amplifiées par PCR à partir des plasmides recombinants fournis par l'équipe du Pr. Jörn Piel. Les amorces oligonucléotidiques utilisées ont été créées de sorte que leurs extrémités 5' incorporent un site de restriction enzymatique aux amplicons pour permettre par la suite de les cloner dans un vecteur d'expression linéarisé. Quant aux extrémités 3', elles permettent l'hybridation spécifique à la séquence codante à amplifier. Les différents couples d'amorces utilisés pour les PCR sont détaillés dans le **Tableau 1**. L'amplification a été réalisée à l'aide de l'ADN polymérase Phusion HOT START II (Thermo Fisher Scientific) selon un programme et avec un mélange réactionnel détaillés en **Figure 58** et les amplicons générés ont été purifiés grâce au kit NucleoSpin *Gel and PCR Clean-up* (Macherey-Nagel) après migration sur gel d'agarose 0,9 %.

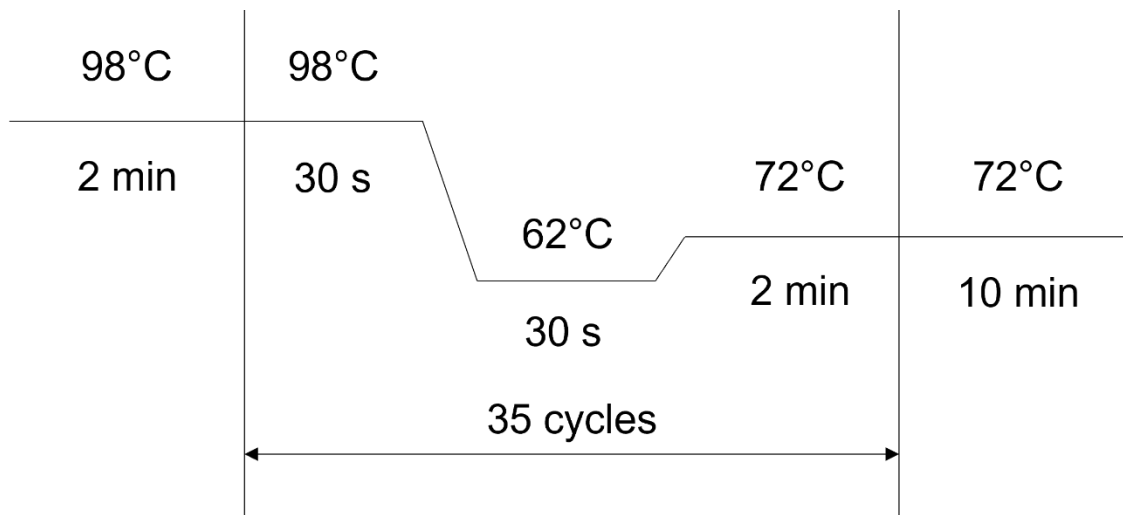
**Tableau 1 :** Séquences nucléotidiques des amorces utilisées pour l'amplification des fragments d'ADN d'intérêt. Les sites de restrictions incorporés sont indiqués en majuscules et les codons d'arrêt de traduction apportés par les amorces antisens sont colorés en rouge. Les séquences sont orientées de 5' vers 3'.

Constructions	Amorces sens	Amorces antisens
LbmA-Ox	tttccGGATCCatgacagaccagaccaggacgaaa	tttccAAGCTTtcaaacagctgaggacgccttttga
LbmA-Ox-ACP	tttccGGATCCatgacagaccagaccaggacgaaa	tttccAAGCTTtcatgtacggcaggctgtgc
LbmC-BVMO	tttccGGATCCagtgcagatgagtcgccgttg	tttccAAGCTTtcaactcatcacagtattcaaactgatattgctgc

Les séquences codant la BVMO SbnP et l' $\alpha$ -hydroxylase OocM n'ont pas été sous-clonées puisqu'elles nous ont été fournies préalablement clonées dans le plasmide d'expression pET-SUMO. La séquence codant la BVMO discrète OocK avait quant à elle été sous-clonée dans le plasmide d'expression pBG102 par Sabrina Collin, ingénieur d'études au



sein de notre équipe, avant le début de ce doctorat.



Composants	Matrice	dNTP	Amorce sens	Amorce antisens	DMSO	Tampon Phusion GC 5X	ADN polymérase	H <sub>2</sub> O
Concentration	4 ng/μL	1,25 mM	0,2 μM	0,2 μM	5 %	1X	1 U.μL <sup>-1</sup>	qsp 25 μL

**Figure 58 :** Programme de PCR et composition du milieu réactionnel employés pour l'amplification des séquences codant les FMO LbmA-Ox, LbmA-Ox-ACP et LbmC-BVMO.

## 2. Restriction enzymatique

Les séquences d'ADN ont été clivées à leurs deux extrémités par les endonucléases de restriction BamHI et HindIII (Thermo Fisher Scientific), dont les sites de clivage ont été apportés par les amorces de PCR. Parallèlement, le plasmide pBG102 a été linéarisé par le même couple d'endonucléases afin de permettre l'association spécifique et orientée du vecteur et des inserts. Les digestions enzymatiques ont été réalisées pendant 1 h à 37 °C en mettant en présence 20 ng.μL<sup>-1</sup> d'ADN et une unité de chaque endonucléase de restriction par μg d'ADN. Les inserts et plasmides digérés ont ensuite été purifiés par migration sur gel d'agarose 0,9 % et extraction à l'aide du kit NucleoSpin *Gel and PCR Clean-up* (Macherey-Nagel).

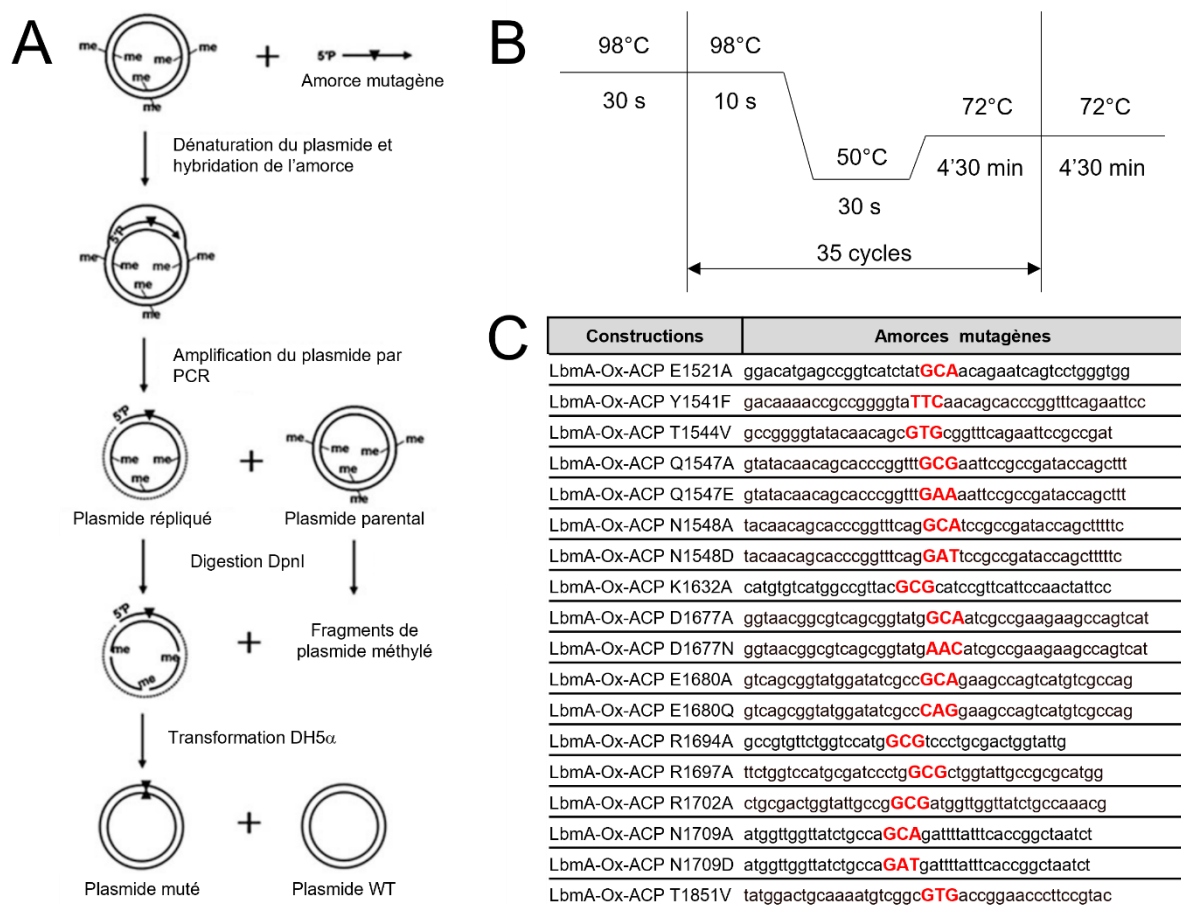
## 3. Clonage en vecteur d'expression

L'ADN ligase du bactériophage T4 (Thermo Fisher Scientific) a permis la formation des liaisons phosphodiesters entre les extrémités 5'-phosphate des inserts et les extrémités 3'-hydroxyle du vecteur linéaire. La réaction de ligation a été réalisée en présence d'un excès molaire d'insert de trois fois par rapport au plasmide et d'une unité de ligase par μg d'ADN.

Les mélanges réactionnels ont été placés à température ambiante pendant une nuit avant qu'une fraction aliquote ne soit prélevée pour une transformation bactérienne.

#### 4. Génération des mutants de LbmA-Ox-ACP

Dans le but de comprendre le mécanisme catalytique de LbmA-Ox, des acides aminés identifiés à partir de la structure cristallographique dans le site actif et dans l'environnement des cofacteurs ont été mutés. Les différents mutants ont été obtenus par mutagenèse dirigée, en utilisant 0,2  $\mu$ M d'une unique amorce <sup>314</sup> s'hybridant de part et d'autre du codon cible (**Figure 59**). Le plasmide pBG102-LbmA-Ox-ACP a servi de matrice à la PCR (**Figure 59**). Suite à l'amplification des plasmides, une unité d'enzyme de restriction DpnI a été ajoutée à 25  $\mu$ L de produits de PCR puis incubé 2 h à 37°C. L'ADN parental méthylé a ainsi été clivé et le mélange d'ADN a été transformé en cellules DH5 $\alpha$ . Les bactéries transformées ont été sélectionnées sur milieu LB gélosé supplémenté en kanamycine puis soumis après culture liquide à une minipréparation d'ADN plasmidique afin de vérifier la présence de la mutation souhaitée sur les différents plasmides par séquençage.



**Figure 59 :** Protocole de mutagenèse dirigée. A) Schéma de la méthode de mutagenèse dirigée à une amorce. L'amplification et la restriction DpnI du plasmide parental génèrent des fragments de plasmide méthylé ainsi que des plasmides héli-méthylés. Le processus de réplication bactérienne permet ensuite de produire des plasmides

non mutés ou portant la mutation sur les deux brins. Le séquençage permet la distinction entre les deux espèces (figure reproduite de Huan et Zhang, 2017<sup>314</sup>). B) Programme de PCR mis en place pour la mutagenèse dirigée. Le mélange réactionnel est identique à celui détaillé en partie II.1. C) Séquences nucléotidiques des amorces mutagènes. Les codons apportant les mutations sont indiqués en majuscules rouges. Les séquences sont orientées de 5' vers 3'.

### III. Transformation bactérienne par les vecteurs recombinants

Afin de pouvoir propager et exprimer les séquences codantes des plasmides recombinants, ceux-ci doivent être insérés dans un hôte approprié, en l'occurrence *Escherichia coli*, et les bactéries transformées doivent être sélectionnées spécifiquement.

#### 1. Préparation des bactéries compétentes

Préalablement au processus de transformation, les souches DH5 $\alpha$  et BL21 (DE3) d'*E. coli* ont été rendues compétentes, c'est-à-dire capables d'intégrer un ADN étranger. Pour cela, les bactéries ont été mises en culture dans un volume de 100 mL de milieu LB à 37 °C sous agitation jusqu'à atteindre 0,5-0,6 de DO<sub>600</sub>. La culture a alors été centrifugée à 4500 g à 4 °C pendant 15 min et le culot repris dans une solution de TfbI (CaCl<sub>2</sub> 10 mM, KCl 100 mM, MnCl<sub>2</sub> 50 mM, glycérol 15% (v/v), pH 7.0). Les cellules ont ensuite été incubées 10 minutes dans la glace puis à nouveau centrifugées 15 min à 4500 g à 4 °C. Le culot a finalement été repris dans 4 mL de solution TfbII (MOPS 10 mM, CaCl<sub>2</sub> 75 mM, KCl 10 mM, glycérol 15% (v/v), pH 7.0) puis aliquoté et conservé à -80 °C.

#### 2. Transformation bactérienne

Les transformations bactériennes ont été réalisées en mettant en présence 100 ng de plasmide recombinant (ou 3  $\mu$ L de mélange de ligation) avec des bactéries compétentes, DH5 $\alpha$  pour la propagation plasmidique après ligation et BL21 (DE3) pour l'expression de protéines recombinantes. Le mélange a été incubé sur glace pendant 30 min puis soumis à un choc thermique par 45 s d'incubation à 42 °C suivies de 5 min d'incubation sur glace. Les cellules ont ensuite été régénérées dans 1 mL de milieu LB pendant 1 h sous agitation à 37 °C avant d'être étalées sur milieu LB gélosé supplémenté en kanamycine pour sélectionner spécifiquement les bactéries transformées. Après une nuit d'incubation à 37 °C, un clone bactérien a été sélectionné et mis en culture en milieu LB liquide une nuit supplémentaire à 37 °C sous agitation. Dans le cas de la sélection de vecteur recombinants en cellules DH5 $\alpha$ , une minipréparation d'ADN plasmidique était ensuite réalisée et la présence de la séquence codante d'intérêt dans le plasmide était validée par séquençage. Dans le cas de transformation en vue d'une expression hétérologue en cellules BL21 (DE3), une fraction aliquote de la culture

était prélevée et conservée à  $-80\text{ }^{\circ}\text{C}$  dans 20 % de glycérol.

## **IV. Production et purification des protéines recombinantes**

Un protocole de production et de purification protéique identique a été mise en place pour les différentes FMO sauvages (WT) de ce projet. Le même protocole de production a été appliqué aux constructions LbmA-Ox-ACP mutantes mais la méthodologie de purification diffère. Elle sera spécifiquement décrite en partie IV.2.e.

### **1. Production des protéines recombinantes**

Les FMO ont été produites dans l'hôte hétérologue *E. coli* BL21 (DE3), transformé par le plasmide recombinant pBG102 ou pET-SUMO. Les protéines d'intérêt ont donc été exprimées en fusion N-terminale avec une étiquette His<sub>6</sub>-SUMO.

Une préculture des cellules transformées a été réalisée en milieu LB kanamycine et incubée sous agitation une nuit à  $37\text{ }^{\circ}\text{C}$ . Une fraction de cette préculture a permis ensuite d'inoculer un plus gros volume de culture en milieu LB kanamycine. La culture a été placée à  $37^{\circ}\text{C}$  jusqu'à obtenir une densité optique à 600 nm d'environ 0,6. La production de la protéine recombinante a alors été induite par ajout de 0,2 mM d'IPTG et 20 mg/L de riboflavine ont également été incorporés afin de favoriser la production du cofacteur flavinique des FMO.

Après une nuit à  $20\text{ }^{\circ}\text{C}$ , la culture a été centrifugée 30 min à 3500 *g*. Les bactéries ont été ensuite resuspendues dans 20 mL par litre de culture de tampon A. Une seconde centrifugation de 30 min à 3500 *g* a été réalisée, le surnageant a été éliminé pour finalement conserver les culots bactériens à  $-80\text{ }^{\circ}\text{C}$ .

### **2. Extraction et purification des protéines recombinantes**

#### **a. Lyse cellulaire**

Chaque culot bactérien issu d'un litre de culture a été repris dans 30 mL de tampon A auxquels ont été rajoutées quatre cents unités de benzonase (Sigma-Aldrich) pour 100 mL de tampon de lyse et 2 mM de MgSO<sub>4</sub> pour éliminer les acides nucléiques. Les cellules ont été lysées par sonication à l'aide d'un désintégrateur de cellules à ultrasons Sonifier 250 (Branson Ultrasonics). Quatre cycles de lyse ont été réalisés dans la glace par alternance de deux minutes de sonication et une minute de repos. Les débris cellulaires ont ensuite été éliminés par centrifugation à 18 000 *g* pendant 30 min.

### **b. Chromatographie d'affinité nickel**

Après filtration à 0,22 mm et ajout de 50 mM d'imidazole, le surnageant de lyse a été injecté sur une colonne Ni-sépharose (Cytiva) d'une capacité de 10 mL couplée à un système de chromatographie protéique rapide en phase liquide (FPLC) de type Äkta Avant (Cytiva). La colonne a été préalablement équilibrée avec du tampon A supplémenté de 50 mM d'imidazole afin de minimiser les risques de fixations aspécifiques. L'étiquette His<sub>6</sub> présente une affinité pour les ions nickel immobilisés sur la résine chromatographique et entraîne donc la rétention de la protéine recombinante sur la colonne. Celle-ci a été ensuite éluée par ajout de 250 mM d'imidazole par mélange des tampons A et B et les fractions contenant la protéine d'intérêt ont été regroupées.

### **c. Protéolyse de l'étiquette His<sub>6</sub>-SUMO et seconde chromatographie d'affinité**

L'échantillon protéique a été soumis à une dialyse en présence de 150 µg de protéase HRV 3C ou *Ubiquitin-like-specific 1* (ULP1) dans 4 L de tampon A pendant une nuit. L'imidazole présent a ainsi été éliminé et l'étiquette His<sub>6</sub>-SUMO séparée de la protéine d'intérêt dans le même temps.

L'échantillon dialysé a de nouveau été injecté sur colonne Ni-sépharose couplée au système FPLC Äkta Avant et préalablement équilibrée en tampon A supplémenté de 50 mM d'imidazole. La FMO, ne possédant plus d'étiquette histidine, n'était pas retenue sur colonne, au contraire de l'étiquette SUMO.

### **d. Chromatographie d'exclusion stérique**

Une ultime étape de chromatographie d'exclusion de taille (SEC) a été réalisée afin d'affiner la pureté de l'échantillon protéique et d'éliminer les éventuels agrégats. La protéine a été concentrée au préalable à l'aide d'un filtre centrifuge Amicon Ultra (Merck Millipore) à 4 500 g jusqu'à un volume inférieur à 5 mL. L'échantillon a été ensuite injecté par l'intermédiaire d'une boucle de 10 mL sur une colonne Superdex 200 26/60 prep grade (Cytiva) équilibrée en tampon C et couplée à un système FPLC Äkta Pure (Cytiva). La protéine a été également éluée dans le tampon C et les fractions collectées ont été analysées par SDS-PAGE. Celles contenant la protéine pure ont été rassemblées et concentrées jusqu'à une valeur compatible avec de futures études structurales et la protéine a été finalement conservée à -80 °C.

### e. Purification des mutants de LbmA-Ox-ACP

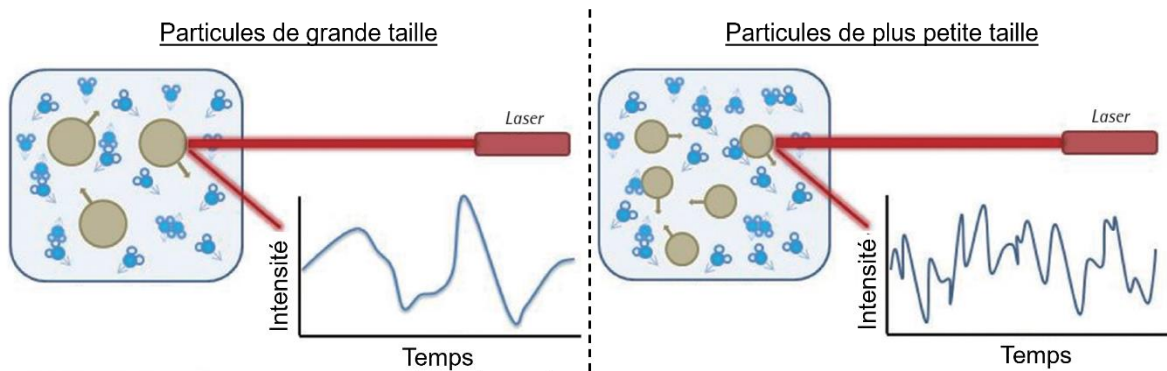
Les mutants de la FMO génératrice d'oxime LbmA-Ox ont été purifiés dans le but d'être soumis à des tests *in vitro* d'activité enzymatique. Par conséquent, la pureté des protéines ne nécessitait pas d'être aussi optimale que celle des FMO WT destinées à des études structurales. Après lyse cellulaire, les mutants ont donc été soumis à une unique étape de chromatographie, par affinité aux ions nickel.

Etant donné le nombre important de mutants générés, une approche indépendante de système FPLC, permettant la purification simultanée de plusieurs protéines a été mise en place. Chaque surnageant de lyse a été filtré à 0,45  $\mu\text{m}$  puis incubé en présence de résine agarose Nickel-NTA (Thermo Fisher Scientific) à 4 °C dans une colonne préparée manuellement. La résine a été ensuite lavée par un volume colonne de tampon A, deux volumes colonne de tampon A avec 50 mM d'imidazole, et un volume colonne de tampon A avec 250 mM d'imidazole pour éluer la protéine étiquetée. La présence et la pureté des protéines ont été vérifiées par SDS-PAGE pour ensuite concentrer les échantillons entre 5 et 10 mg/mL avant de les conserver à -80 °C.

### V. Homogénéité des protéines en solution

Avant de procéder à des tests de cristallisation ou de SAXS, l'état d'agrégation des échantillons protéiques a été évalué par diffusion dynamique de la lumière (DLS). Cette technique permet d'estimer le rayon hydrodynamique des particules en solution par mesure de la variation d'intensité lumineuse qu'elles diffusent au cours du temps. L'intensité diffusée par une particule de petite taille va ainsi varier plus rapidement que celle d'une particule de plus grande taille, en raison de la corrélation entre taille de l'objet et mouvement aléatoire en solution (**Figure 60**).

Avant la mesure, les échantillons ont été centrifugés à 15 000 g pendant 15 min afin d'éliminer les éventuelles poussières qui fausseraient le profil de diffusion. Les mesures ont été effectuées sur un appareil Zetasizer Nano S (Malvern Panalytical). Chaque mesure a été réalisée à l'aide d'une dizaine d'acquisitions de 1 s à 25 °C, permettant d'évaluer la polydispersité des échantillons et d'estimer la taille des espèces qui les composent.



**Figure 60 :** Principe de la technique de DLS. Les particules de l'échantillon d'intérêt sont irradiées par un laser, les transformant chacune en source de lumière ponctuelle. La lumière est alors diffusée dans toutes les directions et son intensité est mesurée tout au long de l'expérience. Les particules en solution se meuvent de manière aléatoire et ce mouvement brownien s'intensifie en même temps que la taille des particules diminue. Ainsi, plus une particule est petite, plus l'intensité lumineuse qu'elle diffuse varie au cours du temps. Cette corrélation permet donc de calculer le rayon hydrodynamique des particules de l'échantillon (figure reproduite de Kim *et al.*, 2014<sup>315</sup>).

## VI. Criblage des conditions de solubilisation de la BVMO Oock

Afin de tester un maximum de conditions de solubilisation en un minimum d'expériences, un plan factoriel incomplet a été créé. Cette stratégie consiste à combiner différentes variables (pH, sels, additifs, ...), chacune définie par plusieurs niveaux (valeurs de pH, concentrations en sels, ...). La combinaison de ces paramètres est calculée de manière aléatoire et non redondante, et garantit que chaque variable et niveau est représenté un nombre de fois identique, permettant donc d'échantillonner les possibilités expérimentales de manière rationnelle. Ce plan a été calculé à l'aide du logiciel SAmBA<sup>316</sup> en spécifiant des variables fréquemment incluses dans les cribles de solubilité protéique<sup>317-321</sup> (**Tableau 2**). Un total de 96 compositions de tampon a ainsi été créé (**Tableau 3**).

**Tableau 2 :** Variables et niveaux de variables utilisés pour la conception du plan factoriel incomplet calculé par le serveur SAmBA<sup>316</sup>. L'ajout des niveaux « rien » et « 0 » a pour conséquence de générer des compositions de tampon dont la moitié ne contiendra pas de détergent, de réducteur et/ou d'additifs, et un sur six ne contiendra pas de sel. Les concentrations en sels sont indiquées en mM.  $\beta$ -OG = 50 mM octyle  $\beta$ -glucoside ; CHAPS = 13 mM 3-[(3-cholamidopropyl)diméthylammonio]-1-propanesulfonate ; NDSB-195 = 200 mM sulfobétaïne non détergente 195 ; NDSB-201 = 200 mM sulfobétaïne non détergente 201 ; TCEP = 10 mM tris(2-carboxyéthyl)phosphine ; diGly = 100 mM diglycine ; triGly = 100 mM triglycine ; Sacc = 100 mM saccharose ; L-Glu = 100 mM L-glutamate ; L-Arg = 100 mM L-arginine ; Glycérol = 10 % glycérol.

Variables	Niveaux											
pH	6.0	7.0	8.0	9.0								
Sels	200 NaCl	400 NaCl	600 NaCl	200 HCOONa	400 HCOONa	0						
Détergents et sulfobétaïnes	$\beta$ -OG	CHAPS	NDSB-195	NDSB-201	rien1	rien2	rien3	rien4				
Réducteurs	TCEP	0										
Additifs	diGly	triGly	Sacc	L-Glu	L-Arg	Glycérol	rien1	rien2	rien3	rien4	rien5	rien6

**Tableau 3 :** Plan factoriel incomplet de 96 conditions, généré par le serveur SAMBA<sup>316</sup> à partir des variables et niveaux détaillés dans le **Tableau 2**. Le pH des solutions tampons a été maintenu par ajout de 50 mM sodium phosphate (pH 6.0 et 7.0), 50 mM Tris (pH 8.0) ou 30 mM CAPSO (pH 9.0).

	pH	Sels	Détergents et sulfobétaïnes	Réducteurs	Additifs
1	6	0	rien2	0	diGly
2	6	0	rien2	0	L-Arg
3	6	0	NDSB-195	TCEP	rien2
4	6	0	NDSB-195	TCEP	rien4
5	6	200 NaCl	rien2	TCEP	triGly
6	6	200 NaCl	rien2	TCEP	L-Glu
7	6	200 NaCl	NDSB-195	0	rien3
8	6	200 NaCl	NDSB-195	0	rien6
9	6	400 NaCl	rien3	TCEP	L-Arg
10	6	400 NaCl	rien3	TCEP	diGly
11	6	400 NaCl	NDSB-201	TCEP	Sacc
12	6	400 NaCl	NDSB-201	TCEP	Glycérol
13	6	600 NaCl	rien1	TCEP	rien1
14	6	600 NaCl	rien1	TCEP	rien5
15	6	600 NaCl	CHAPS	0	Sacc
16	6	600 NaCl	CHAPS	0	Glycérol
17	6	200 HCOONa	β-OG	0	rien3
18	6	200 HCOONa	β-OG	0	rien6
19	6	200 HCOONa	CHAPS	TCEP	triGly
20	6	200 HCOONa	CHAPS	TCEP	L-Glu
21	6	400 HCOONa	rien4	0	rien1
22	6	400 HCOONa	rien4	0	rien5
23	6	400 HCOONa	NDSB-201	0	rien2
24	6	400 HCOONa	NDSB-201	0	rien4
25	7	0	rien4	TCEP	rien3
26	7	0	rien4	TCEP	rien6
27	7	0	CHAPS	0	rien2
28	7	0	CHAPS	0	rien4
29	7	200 NaCl	CHAPS	TCEP	diGly
30	7	200 NaCl	CHAPS	TCEP	L-Arg
31	7	200 NaCl	NDSB-201	TCEP	triGly
32	7	200 NaCl	NDSB-201	TCEP	L-Glu
33	7	400 NaCl	rien1	0	rien3
34	7	400 NaCl	rien1	0	rien6
35	7	400 NaCl	NDSB-195	TCEP	Sacc
36	7	400 NaCl	NDSB-195	TCEP	Glycérol
37	7	600 NaCl	rien2	TCEP	0 diG
38	7	600 NaCl	rien2	TCEP	rien5
39	7	600 NaCl	β-OG	0	rien2
40	7	600 NaCl	β-OG	0	rien4
41	7	200 HCOONa	NDSB-195	0	diGly
42	7	200 HCOONa	NDSB-195	0	L-Arg
43	7	200 HCOONa	NDSB-201	0	rien1
44	7	200 HCOONa	NDSB-201	0	rien5
45	7	400 HCOONa	rien2	TCEP	Sacc
46	7	400 HCOONa	rien2	TCEP	Glycérol
47	7	400 HCOONa	rien3	0	triGly
48	7	400 HCOONa	rien3	0	L-Glu
49	8	0	β-OG	TCEP	rien1
50	8	0	β-OG	TCEP	rien5
51	8	0	NDSB-201	TCEP	rien3
52	8	0	NDSB-201	TCEP	rien6
53	8	200 NaCl	rien3	0	rien1
54	8	200 NaCl	rien3	0	rien5
55	8	200 NaCl	β-OG	0	Sacc
56	8	200 NaCl	β-OG	0	Glycérol
57	8	400 NaCl	rien2	0	0 triG
58	8	400 NaCl	rien2	0	rien4
59	8	400 NaCl	rien4	0	triGly
60	8	400 NaCl	rien4	0	L-Glu
61	8	600 NaCl	rien4	TCEP	diGly
62	8	600 NaCl	rien4	TCEP	L-Arg
63	8	600 NaCl	NDSB-195	0	triGly
64	8	600 NaCl	NDSB-195	0	L-Glu
65	8	200 HCOONa	rien1	TCEP	Sacc
66	8	200 HCOONa	rien1	TCEP	Glycérol
67	8	200 HCOONa	rien3	TCEP	rien2
68	8	200 HCOONa	rien3	TCEP	rien4
69	8	400 HCOONa	rien1	0	diGly
70	8	400 HCOONa	rien1	0	L-Arg
71	8	400 HCOONa	CHAPS	TCEP	rien3
72	8	400 HCOONa	CHAPS	TCEP	rien6
73	9	0	rien1	0	triGly
74	9	0	rien1	0	L-Glu
75	9	0	rien3	0	Sacc
76	9	0	rien3	0	Glycérol
77	9	200 NaCl	rien1	TCEP	rien2
78	9	200 NaCl	rien1	TCEP	rien4
79	9	200 NaCl	rien4	0	Sacc
80	9	200 NaCl	rien4	0	Glycérol
81	9	400 NaCl	β-OG	TCEP	triGly
82	9	400 NaCl	β-OG	TCEP	L-Glu
83	9	400 NaCl	CHAPS	0	rien1
84	9	400 NaCl	CHAPS	0	rien5
85	9	600 NaCl	rien3	TCEP	rien3
86	9	600 NaCl	rien3	TCEP	rien6
87	9	600 NaCl	NDSB-201	0	diGly
88	9	600 NaCl	NDSB-201	0	L-Arg
89	9	200 HCOONa	rien2	0	rien3
90	9	200 HCOONa	rien2	0	rien6
91	9	200 HCOONa	rien4	TCEP	rien2
92	9	200 HCOONa	rien4	TCEP	rien4
93	9	400 HCOONa	β-OG	TCEP	diGly
94	9	400 HCOONa	β-OG	TCEP	L-Arg
95	9	400 HCOONa	NDSB-195	TCEP	rien1
96	9	400 HCOONa	NDSB-195	TCEP	rien5

Pour tester l'influence de ces différents tampons sur la solubilité d'OocK, la BVMO a été exprimée en cellules BL21 (DE3) transformées par le plasmide recombinant pBG102 tel que décrit en partie IV.1. À partir des cultures bactériennes, 96 culots de 10 mL ont été réalisés et chacun d'entre eux a été resuspendu dans 500 µL d'un des tampons du plan factoriel incomplet (**Tableau 3**). Les échantillons ont été disposés dans une plaque 96 puits Deepwell avant d'être lysés par sonication à l'aide d'un désintégrateur de cellules à ultrasons à quadruple sonde (Branson Ultrasonics) selon un programme de 30 s de lyse, 30 s de repos et 30 s de lyse. Les lysats cellulaires ont été ensuite centrifugés 30 min à 15 000 g à 4 °C. Les surnageants et culots de lyse ont été séparés, les culots resuspendus dans 500 µL du tampon correspondant, puis la concentration en protéines de tous les échantillons a été estimée par mesure de la DO<sub>280</sub>. Pour chaque couple surnageant-culot, la même quantité de protéine a été déposée sur SDS-PAGE.

Afin de pouvoir comparer la quantité relative d'OocK présente dans les fractions de surnageant et de culot pour chaque tampon testé, les gels d'électrophorèse ont été numérisés par un scanner Azure 200 (Azure Biosystems) et traités à l'aide du logiciel d'analyse d'image Mesurim. Pour chaque tampon testé, les deux bandes correspondant à la BVMO dans la fraction de culot et la fraction de surnageant ont été encadrées par un rectangle de même aire. L'intensité lumineuse au sein de ces rectangles a ensuite été mesurée, tout en soustrayant le bruit de fond généré par la couleur du gel. Cette intensité étant corrélée à la quantité de BVMO présente, un score de solubilité a été attribué à chaque tampon, représentant le pourcentage



de BVMO situé dans la fraction de surnageant. Sur la base de ces scores, l'impact de chaque variable sur la solubilité a été estimé et a permis de réaliser un second plan factoriel incomplet plus restreint (**Tableaux 4 et 5**) selon la même méthodologie.

**Tableau 4 :** Variables et niveaux de variables utilisés pour la conception du plan factoriel incomplet calculé par le serveur SAmBA <sup>316</sup>. Les différentes concentrations sont indiquées en mM.  $\beta$ -OG = octyl  $\beta$ -glucoside ; diGly = diglycine ; L-Glu = L-glutamate.

Variables	Niveaux			
pH	8.5	9.0	9.5	
NaCl	0	300	400	500
$\beta$ -OG	25	50	rien1	rien2
diGly	50	100	rien1	rien2
L-Glu	50	100	rien1	rien2

**Tableau 5 :** Second plan factoriel incomplet de 20 conditions, généré par le serveur SAmBA <sup>316</sup> à partir des variables et niveaux détaillés dans le **Tableau 4**. Le pH des solutions tampons a été maintenu par ajout de 30 mM CAPSO.

	pH	NaCl	$\beta$ -OG	diGly	L-Glu
1	8.5	0	rien2	rien2	rien1
2	8.5	0	rien1	50	100
3	8.5	300	25	50	50
4	8.5	300	50	100	50
5	8.5	400	25	rien2	rien2
6	8.5	400	50	rien1	rien1
7	8.5	500	rien1	rien1	rien2
8	9.0	0	rien2	rien1	50
9	9.0	0	50	50	rien2
10	9.0	300	rien1	rien2	rien2
11	9.0	300	rien2	rien1	100
12	9.0	400	50	100	100
13	9.0	500	25	50	50
14	9.0	500	rien1	100	rien1
15	9.5	0	25	100	100
16	9.5	300	25	rien1	rien1
17	9.5	400	rien1	rien2	50
18	9.5	400	rien2	50	rien1
19	9.5	500	rien2	100	rien2
20	9.5	500	50	rien2	100

À l'issue de ce deuxième plan factoriel, les gels de polyacrylamide ont été traités informatiquement comme décrit précédemment, pour déterminer la composition optimale du tampon de solubilisation d'Oock.

## VII. Études structurales par cristallographie aux rayons X

La cristallographie aux rayons X est la méthode majoritairement employée pour la résolution de structure protéique à l'échelle atomique. Le processus permettant de passer du stade de la protéine en solution à sa structure tridimensionnelle à haute résolution se divise en trois étapes que sont 1) la cristallisation de la macromolécule ; 2) la collecte et le traitement des données de diffraction des rayons X ; et 3) la résolution et l'amélioration de la structure par cycles de construction du modèle atomique-affinement des données.

## 1. Cristallisation

### a. Principe et techniques de cristallisation protéique

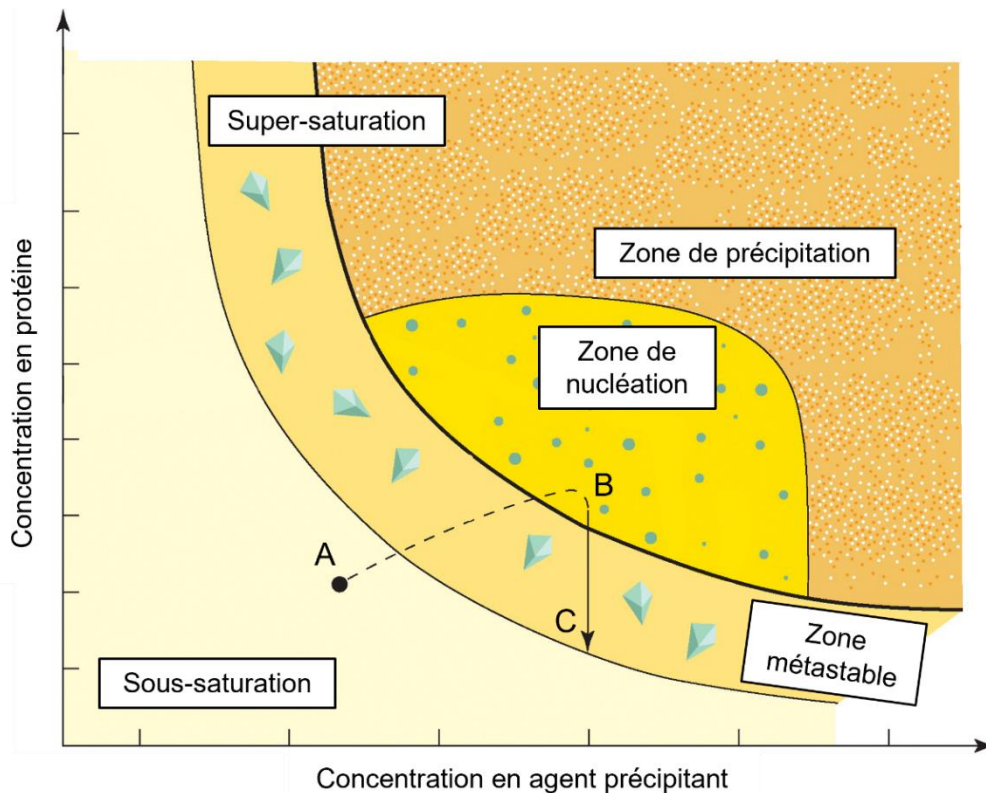
La cristallisation consiste au passage de la protéine d'un état libre en solution à un état solide périodique, l'état cristallin. Cette transition de phase est nécessaire à l'étape de diffraction des rayons X puisque l'agencement ordonné des molécules dans le cristal permet d'amplifier le signal de diffraction, chacune d'entre elles étant orientée et disposée de manière identique dans toutes les mailles. Pour ce faire, la solubilité de la protéine doit être progressivement diminuée par ajout d'agents précipitants afin de favoriser les interactions intermoléculaires.

Au cours des vingt dernières années, plusieurs logiciels ont été développés dans le but d'estimer la propension des protéines à cristalliser <sup>322-325</sup>. Cependant, l'identification des conditions optimales de cristallisation reposent encore aujourd'hui majoritairement sur des approches d'essais-erreurs, en raison du nombre important de paramètres qui influencent la dynamique de ce changement de phase <sup>326</sup>. Les principaux facteurs à prendre en compte sont évidemment la pureté et l'homogénéité de l'échantillon protéique, mais également sa concentration, le pH, la température, la nature et la concentration des agents précipitants, ou encore l'ajout de ligands spécifiques ou d'additifs (détergents, sucres, polymères, ...). La solubilité de la protéine étant permise par ses interactions avec le solvant, la transition vers l'état cristallin nécessite de perturber cette solvation pour provoquer la formation d'interactions protéine-protéine. Divers agents précipitants couramment utilisés permettent d'y parvenir <sup>326</sup> :

- Les sels non chaotropiques (tels que le sulfate d'ammonium et le chlorure de sodium) modifient la force ionique du milieu. Leur présence à faibles concentrations améliore la solubilité des protéines mais à hautes concentrations, ils excluent le solvant ce qui favorise les interactions hydrophobes.
- Les solvants organiques (tels que l'isopropanol et le 2-méthyl-2,4-pentanediol) diminuent la constante diélectrique du milieu.
- Les polymères organiques (tels que les polyéthylènes glycol) excluent le solvant de l'environnement des protéines.

Pour permettre la formation de cristaux, les concentrations en protéine et en agents précipitants doivent augmenter afin que la solution de cristallisation passe d'un état sous-saturé à un état permettant la nucléation de la protéine, c'est-à-dire la formation des premiers germes cristallins. Le passage progressif de la protéine à l'état solide diminue sa concentration en solution, ce qui conduit à un état métastable du système permettant la croissance des

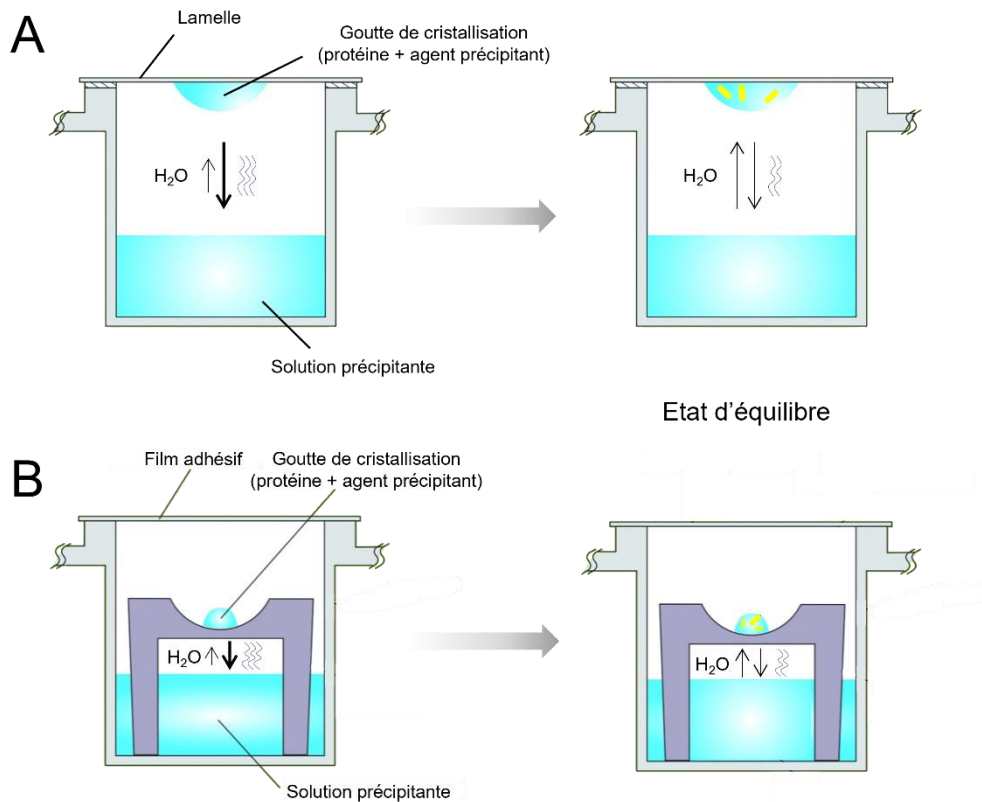
cristaux à partir des germes formés et dans lequel la nucléation n'a plus lieu <sup>327</sup> (**Figure 61**).



**Figure 61 :** Diagramme de phase d'une expérience de cristallisation protéique. L'évolution décrite par la flèche et les points A, B et C correspond à une cristallisation par diffusion de vapeur, détaillée ultérieurement dans cette section. Les concentrations en protéine et en agent précipitant sont progressivement augmentées pour passer d'une solution initiale sous-saturée (A), à une zone de nucléation où les germes cristallins se forment (B). La diminution de la concentration protéique en solution provoque ensuite la transition vers un état métastable permettant la croissance des germes en cristaux (C) (figure adaptée de Chayen, 2004 <sup>327</sup>).

Plusieurs techniques permettent de réaliser ces expériences de cristallisation. Les échanges entre la solution de protéine et la solution précipitante peuvent être réalisés par dialyse, en *microbatch*, ou encore par diffusion sans interface (méthode pouvant être implémentée dans un système microfluidique) <sup>328</sup>. La technique la plus fréquemment utilisée, et celle qui a été employée pour le criblage et l'affinement des conditions de cristallisation des FMO, est la diffusion de vapeur. Dans le cas de la diffusion de vapeur en goutte suspendue, un puits est rempli par un large volume de solution contenant l'agent précipitant, et un microvolume de la solution de protéine est déposé sur une lamelle. À cette goutte sont ajoutés quelques  $\mu\text{L}$  de la solution précipitante puis, la lamelle est retournée et scellée sur le puits, créant un système clos hermétique (**Figure 62**). La solution précipitante disposée sur la lamelle se trouve donc diluée par rapport à la solution du puits. Un échange gazeux va alors se mettre en place de la goutte sur la lamelle vers le puits afin d'atteindre un équilibre. Cette diffusion va provoquer l'augmentation progressive des concentrations en protéine et

précipitants dans la goutte et ainsi permettre, dans des conditions favorables, d'atteindre la zone de nucléation. Une méthodologie alternative de la diffusion de vapeur peut également être réalisée en goutte assise. Dans ce cas, la solution protéique est déposée au fond d'une cupule disposée à côté du puits (**Figure 62**).



**Figure 62 :** Technique de cristallisation par diffusion de vapeur en goutte suspendue (A) et goutte assise (B). Un grand volume de solution d'agent précipitant est placé dans le puits tandis qu'un mélange de solution protéique et solution précipitante est déposé sur une lamelle ou une cupule. Le système est ensuite fermé hermétiquement pour permettre les échanges gazeux de la goutte vers le puits (figure adaptée de McPherson et Gavira, 2014 <sup>326</sup>).

### b. Criblage et affinement des conditions de cristallisation

L'identification de conditions de cristallisation prometteuses a été réalisée à l'aide de cribles commerciaux afin d'identifier l'influence d'un maximum de paramètres sur le comportement des protéines (**Tableau 6**). En effet, ces cribles sont composés d'une importante gamme de précipitants, d'additifs et de pH, associés en conditions ayant déjà permis l'obtention de cristaux de protéines.

**Tableau 6 :** Cribles commerciaux utilisés pour l'identification de conditions de cristallisation prometteuses des FMO.

Cribles	Sociétés	Cribles	Sociétés
SG1	Molecular Dimensions	Wizard Classic	Rigaku Reagents
LMB	Molecular Dimensions	Wizard Precipitant Synergy	Rigaku Reagents
MIDAS	Molecular Dimensions	Index	Hampton Research
MORPHEUS	Molecular Dimensions	PEG/Ion	Hampton Research
Structure Screen	Molecular Dimensions	Crystal Screen	Hampton Research
BCS	Molecular Dimensions		

Ces criblages ont été réalisés par diffusion de vapeur en goutte assise dans des plaques Greiner 96 puits (Greiner Bio-One) dans lesquelles chaque puits est associé à trois cupules autorisant le test de trois ratios protéine/précipitant (en l'occurrence 1/2, 1/1, et 2/1). Le remplissage des plaques a débuté par le transfert dans les 96 puits de 50  $\mu$ L des solutions de cristallisation d'un crible à l'aide d'une station de travail robotisée Freedom EVO 150 (Tecan). Un robot de type Mosquito (SPT Labtech) a ensuite permis de réaliser les gouttes de solution protéique et solution précipitante dans les cupules grâce à sa capacité à pipetter des volumes de l'ordre de quelques centaines de nanolitres, 100 et 200 nL dans le cas présent, en fonction des ratios considérés. Les plaques ont été ensuite scellées par un film adhésif transparent et placées à 20 °C.

Lorsqu'une des conditions testées permettait l'apparition de cristaux, celle-ci était reproduite et affinée dans de plus grands volumes par la méthode de diffusion de vapeur en goutte suspendue. Pour cela, une gamme faisant varier le pH ou la concentration d'un des composants de la condition identifiée était créée et 1 mL de chaque tampon était disposé dans les 24 puits d'une boîte *Linbro* (Greiner Bio-One). Sur une lamelle, trois gouttes étaient réalisées par mélange de la solution de protéine et de la solution précipitante du puits correspondant, en respectant les ratios 1/2, 1/1 et 2/1. Les lamelles étaient ensuite scellées sur leurs puits avant de placer les boîtes à 20 °C.

## 2. Congélation des cristaux

Lors des expériences de diffraction, les cristaux de protéines subissent des dommages causés par l'exposition aux rayons X. Ces dégâts dépendent de deux facteurs, la dose de photons X reçue par le cristal et la durée de l'exposition. Le facteur temps est lié à l'ionisation progressive des molécules sous l'effet des radiations et à la présence de grandes quantités de solvant dans les cristaux de protéines<sup>329</sup>. En effet, l'agencement ordonné de ces macromolécules forme des canaux de solvant qui représentent typiquement entre 30 et 70 % du volume du cristal. Or, les rayons X incidents provoquent la radiolyse progressive de l'eau présente dans ces canaux, ce qui entraîne la formation de radicaux libres, délétères pour la protéine et donc pour le cristal. De plus, l'échauffement provoqué par le faisceau accélère la

diffusion des radicaux libres dans l'échantillon, provoquant d'autant plus de dégâts<sup>330</sup>.

Pour pallier ce problème, le cristal est maintenu à une température cryogénique (généralement 100 K) dans un flux continu d'azote gazeux durant la collecte des données. On parle alors de cryocrystallographie. Cependant, le refroidissement d'un cristal à ces températures extrêmes est accompagné d'une augmentation de 9 % du volume de l'eau, lors de sa transition à l'état de glace<sup>330</sup>. Il est donc nécessaire de protéger les cristaux au préalable par ajout d'une molécule aux propriétés cryoprotectantes. Dans cette optique, les cristaux de LbmA-Ox-ACP et OocM ont été brièvement incubés dans la solution de cristallisation ayant permis leur croissance, supplémentée de 30 % de glycérol ou d'éthylène glycol, avant d'être congelés en flux d'azote à  $-100\text{ °C}$  puis conservés en azote liquide.

### 3. Diffraction et collection des données

Les données de diffraction ont été enregistrées au synchrotron SOLEIL (Gif-sur-Yvette) sur la ligne de lumière PROXIMA-2A. Les cristaux piégés dans une boucle microscopique ont été placés sur une tête goniométrique constamment maintenue sous flux d'azote et alignés manuellement avec la trajectoire du faisceau de rayons X. Les échantillons ont ensuite été irradiés sur un secteur angulaire défini, parcouru selon un pas d'oscillation de  $0,1^\circ$ .

### 4. Traitement des données de diffraction et phasage

#### a. Traitement des données

Les données collectées ont été indexées et intégrées avec le logiciel XDS<sup>331</sup> et mises à l'échelle à l'aide du programme XSCALE<sup>332</sup>. La qualité des jeux de données a ensuite été évaluée par considération de différents facteurs tels que la résolution maximale, la complétude et la redondance du jeu de données, le ratio signal sur bruit de fond ( $I/\sigma I$ ) ou encore la corrélation entre les intensités mesurées de deux moitiés du jeu de données réparties aléatoirement ( $CC_{1/2}$ ).

#### b. Phasage par remplacement moléculaire

Les ondes diffractées peuvent être décrites par un module  $|F_{hkl}|$  et une phase  $\phi_{hkl}$ . L'association de ces deux paramètres permet de définir, pour chaque réflexion, un facteur de structure  $F_{hkl}$  qui dépend du facteur de diffusion atomique  $f_j$  des atomes participant à la diffraction et de la position (xyz) de chaque atome  $j$  dans la maille. Ainsi, ce facteur de structure décrit pour chaque famille de plans réticulaires hkl la contribution de l'ensemble des atomes à la diffraction par l'intermédiaire de leurs électrons (**Équation 1**). D'un point de vue mathématique, le facteur de structure  $F_{hkl}$  dans l'espace réciproque est la transformée de

Fourier du pouvoir diffusant, et donc de la densité électronique, des atomes de l'objet étudié dans l'espace réel. Par conséquent, il est possible de décrire la densité électronique  $\rho$  en tout point (xyz) du cristal par transformée de Fourier inverse des facteurs de structure des ondes diffractées (**Équation 2**). Ce calcul nécessite donc de connaître le module et la phase de ces ondes. Or, lors de l'enregistrement des clichés de diffraction, seul l'amplitude des ondes diffractées, et donc le module des facteurs de structure, peut être déterminée, par mesure de l'intensité des tâches de diffraction. L'information de phase est perdue, et doit être déterminée par une méthode dite indirecte.

$$F_{hkl} = |F_{hkl}| e^{i\phi_{hkl}} = \sum_{j=1}^N f_j e^{[2i\pi(hx_j + ky_j + lz_j)]}$$

**Équation 1 :** Description du facteur de structure des ondes diffractées.  $|F_{hkl}|$  = module de l'onde diffractée ;  $\phi_{hkl}$  = phase de l'onde diffractée ;  $f_j$  = facteur de diffusion atomique de l'atome j ;  $(x_j, y_j, z_j)$  = coordonnées de l'atome j ; hkl = indices de miller des familles de plans réticulaires.

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \exp(i\phi_{hkl}) \exp[-2\pi i(hx + ky + lz)]$$

**Équation 2 :** Relation entre espace réel et espace réciproque par transformée de Fourier.  $\rho(xyz)$  = densité électronique aux coordonnées x,y,z de l'espace réel ; V = volume de la maille ;  $|F_{hkl}|$  = module de l'onde diffractée ;  $\phi_{hkl}$  = phase de l'onde diffractée ; hkl = indices de miller des familles de plans réticulaires.

La stratégie de phasage la plus fréquemment utilisée est le remplacement moléculaire, qui requiert l'utilisation d'un modèle tridimensionnel à la structure proche de celle de la protéine étudiée. Le principe de cette méthode repose sur le fait que si le modèle et la structure à résoudre sont similaires (une identité de séquence supérieure à 25 % est préférable) et sont placés et orientés de manière identique dans une maille présentant les mêmes caractéristiques, alors leurs facteurs de structure dans le réseau réciproque doivent être similaires. Pour assurer l'orientation et le positionnement correctes du modèle, des fonctions de rotation et de translation sont calculées, et la solution est donnée par celles présentant la meilleure corrélation avec les données expérimentales<sup>333</sup>. Ainsi, les phases calculées à partir du modèle peuvent être utilisées avec les modules déterminés expérimentalement pour calculer les facteurs de structure de la protéine d'intérêt.

À l'heure de l'avènement des logiciels de modélisation structurale, le remplacement moléculaire s'impose de plus en plus comme la méthode de phasage de référence. En effet, la puissance de ces programmes permet aujourd'hui d'utiliser avec confiance, non pas une structure homologue, mais un modèle bioinformatique de la structure d'intérêt elle-même comme modèle de recherche. Ainsi, la résolution de la phase des ondes diffractées par LbMA-

Ox-ACP a été réalisée par remplacement moléculaire avec le programme MOLREP<sup>334</sup> de la suite CCP4<sup>335</sup>, à l'aide d'un modèle du domaine Ox seul, généré par AlphaFold2<sup>336</sup>.

Les méthodes alternatives de diffusion anormale (SAD et MAD) et remplacement isomorphe (MIR) ont également été envisagées pour résoudre la phase. Cependant, aucun cristal de LbmA-Ox-ACP sélénée n'a été obtenu, empêchant le phasage par SAD et MAD, et le trempage de cristaux natifs dans des solutions de sels d'atomes lourds (HgCl<sub>2</sub>, NiCl<sub>2</sub>, NaVO<sub>3</sub>) a provoqué la dégradation des cristaux ou n'a pas permis de fixer les atomes lourds, empêchant le phasage par MIR.

### 5. Construction et affinement de la structure

Le programme de remplacement moléculaire MOLREP a permis de générer un premier modèle du domaine LbmA-Ox. La chaîne principale du domaine ACP a ensuite été construite à l'aide du programme Buccaneer<sup>337</sup>, permettant d'identifier les positions des carbones  $\alpha$  potentiellement connectées dans une carte de densité électronique à faible résolution (> 2,5 Å), puis de les lier entre elles pour former une chaîne continue. La structure du didomaine LbmA-Ox-ACP a été ensuite améliorée de manière itérative par construction manuelle avec le logiciel Coot<sup>338</sup> et affinement de la carte de densité avec le programme REFMAC5<sup>339</sup>. Après chaque cycle d'affinement, la qualité du modèle atomique construit a été évaluée au travers des valeurs de déviation moyenne des longueurs et angles de liaisons par rapport à leurs valeurs idéales. De plus, l'accord entre le modèle construit et les données expérimentales a été estimé par le facteur  $R_{\text{factor}}$ , couplé au facteur  $R_{\text{free}}$  basé sur une partie des réflexions non utilisée pendant l'affinement pour éviter de surinterpréter la carte de densité électronique. La qualité de la structure finale a été estimée à l'aide du serveur MolProbity<sup>340</sup>. Celui-ci détermine le nombre de clashes stériques, s'assure que la géométrie de la protéine est correcte, que les angles dièdres  $\Psi$  et  $\Phi$  respectent le diagramme de Ramachandran et attribue un score permettant de comparer les statistiques de la structure à l'ensemble des structures de résolution comparable. La structure du didomaine LbmA-Ox-ACP a finalement été déposée dans la banque de données *Protein Data Bank* (PDB).

### VIII. Tests d'activité de LbmA-Ox-ACP et ses mutants

Les acides aminés potentiellement impliqués dans la catalyse et la fixation des cofacteurs de LbmA-Ox ont été identifiés à partir de la structure du didomaine LbmA-Ox-ACP. Les essais enzymatiques de la monooxygénase WT et mutante ont donc été également effectués sur le didomaine plutôt que sur la FMO seule, afin d'éviter tout changement structural éventuel en l'absence du domaine ACP, qui pourrait modifier la position ou l'orientation des résidus d'intérêt.



Les tests *in vitro* d'activité de LbmA-Ox-ACP WT et de ses mutants ont été réalisés au sein de l'équipe du Pr. Jörn Piel à l'institut de microbiologie de l'ETH, Zürich. Un protocole avait été mis en place au préalable pour la reconstitution *in vitro* de l'activité de LbmA-Ox par le Dr. Hannah A. Minas, et a été adapté pour ces tests.

Dans un volume réactionnel de 100  $\mu$ L, 30  $\mu$ M de LbmA-Ox-ACP WT ou mutante ont été mis en présence de 1 mM de NADPH, 1 mM de FAD, 2,5 % de glycérol et 5 mM de glycine SNAC dans un tampon 50 mM sodium phosphate, pH 6.0. Le glycine SNAC est un analogue du substrat natif glycine Ppant de la monooxygénase, et a été synthétisé par le Dr. Cora L. Dieterich (équipe du Pr. Jörn Piel). Les mélanges réactionnels ont été incubés 60 min à 30 °C puis la phase organique a été extraite par ajout d'acétate d'éthyle, évaporée sous flux d'azote et resuspendue en méthanol. Après 10 min de centrifugation des échantillons, les surnageants ont été injectés sur une colonne de chromatographie en phase inverse de type Kinetex 2.6  $\mu$ m C18 100 Å (Phenomenex) à l'aide d'un gradient d'acétonitrile. La colonne a été connectée à un système de chromatographie liquide ultra haute performance (UHPLC) Dionex UltiMate 3000, couplé à un spectromètre de masse Q Exactive (Thermo Fisher Scientific). Les données collectées ont été ensuite analysées à l'aide du logiciel Thermo Xcalibur 2.2 afin d'identifier le substrat amine ( $[M+H]^+=177.0692$ ) et la formation de produit oxime ( $[M+H]^+=191.0484$ ). La mesure de l'aire sous les pics correspondants à ces deux ions a ensuite permis de déterminer un pourcentage de conversion du substrat et de comparer ce pourcentage entre les différents mutants et l'enzyme WT.



---

# *Résultats*

---



## I. Études structurales et fonctionnelles d'un domaine FMO générateur d'oxime au sein d'un module de PKS *trans*-AT

Ce projet avait pour but d'étudier le mécanisme rare de biosynthèse d'oxime par les PKS *trans*-AT afin d'étendre la chimie connue de ces mégasynthèses, et de potentiellement ajouter cette nouvelle activité enzymatique aux outils de biologie synthétique.

Pour cela, l'équipe de Pr. Jörn Piel a commencé par valider le rôle hypothétique du domaine LbmA-Ox dans la biosynthèse de la fonction oxime du polycétide lobatamide A. La FMO a été exprimée et purifiée à homogénéité avant d'être soumise à un test d'activité en présence de glycine SNAC, un analogue de substrat potentiel. En effet, la structure du polycétide et la prédiction de spécificité du domaine d'adénylation en amont de LbmA-Ox par le logiciel NRPSpredictor2<sup>341</sup> indiquaient tous deux qu'un résidu glycine devait être le substrat de la FMO. L'analyse par MS du mélange réactionnel a ensuite confirmé la formation d'oxime par LbmA-Ox (**Article, Figure 2**).

Des tests de spécificité de substrat ont ensuite établi que la réaction pouvait avoir lieu en présence de glycine SNAC ou alanine SNAC mais que des dérivés SNAC aux chaînes latérales plus encombrantes n'étaient pas reconnus comme substrat par l'enzyme (**Article, Figure 2**). De plus, la réalisation des tests d'activité en présence des acides aminés glycine ou alanine libres n'ont pas non plus permis la formation d'oxime, indiquant la nécessité d'un substrat lié au domaine ACP dans le contexte modulaire de la PKS.

La caractérisation structurale du produit de réaction de LbmA-Ox s'est révélée difficile en raison de l'hydrolyse rapide de sa fonction thioester. L'oxime étant méthylé dans le polycétide final, la reconstitution de la formation du groupement méthyloxime a été entreprise puisque l'ajout du groupement méthyle pourrait jouer un rôle stabilisateur. Aucun produit méthylé n'a été observé lorsque LbmA-Ox et le domaine en aval LbmA-MT ont été extraits de leurs modules, soulignant l'importance des interactions protéine-protéine dans la transmission de l'intermédiaire monooxygéné de la FMO à la MT. Les tests d'activité Ox-MT ont donc été ensuite réalisés avec le trimodule N-terminal de la sous-unité LbmA, permettant en principe l'adénylation de la glycine, sa monooxygénation et sa méthylation subséquente. L'analyse de la réaction par MS a confirmé cette hypothèse et la structure du produit méthyloxime a été confirmée par synthèse et caractérisation MS et RMN d'un standard synthétique (**Article, Figure 3**).

Le domaine LbmA-Ox est le premier cas démontré d'activité de formation d'oxime catalysée au cours de la synthèse du polycétide et portée par un domaine intégré à la PKS. Nous avons donc cherché à comprendre les détails mécanistiques de cette activité sans précédent. Pour cela, nous avons exprimé et purifié à homogénéité le didomaine LbmA-Ox-

## Résultats

ACP<sub>apo</sub> pour le soumettre à des essais de cristallisation, le domaine LbmA-Ox seul n'ayant pas permis l'obtention de cristaux. Les cristaux obtenus étaient de couleur jaune, attestant la présence de FAD, et les données de diffraction ont été enregistrées sur la ligne Proxima-2A du synchrotron SOLEIL. La structure résolue du didomaine a révélé un repliement de la FMO en deux domaines de type Rossmann et un domaine de plus petite taille absent des homologues structuraux de la FMO identifiés par le serveur DALI<sup>342</sup>. Ce troisième domaine était en revanche présent dans les modèles générés par AlphaFold2<sup>336</sup> des FMO homologues impliquées dans la synthèse des fonctions oximes des polycétides nécroxime A et oximine III (**Article, Figure 1, Figure S32**). Il est donc probable que ce domaine structural trouve son importance dans le contexte modulaire des PKS, où il pourrait médier des interactions avec les domaines en amont et en aval.

La co-cristallisation du didomaine avec son cofacteur oxydé NADP<sup>+</sup> et un analogue de substrat éthyle glycinate a permis d'identifier les sites de fixation des cofacteurs et du substrat ainsi que les acides aminés vraisemblablement impliqués dans leur fixation et dans la catalyse (**Article, Figure 4**). Par ailleurs, l'interface entre FAD et NADP<sup>+</sup> révèle que le positionnement relatif de leurs groupements isoalloxazine et nicotinamide respectifs diffère de celui observé dans les structures de FMO discrètes. Dans le cas de LbmA-Ox, les deux groupements sont empilés l'un sur l'autre, permettant au substrat d'accéder au site actif par le canal dans lequel l'éthyle glycinate a été modélisé. En revanche, dans les structures de FMO discrètes, le nicotinamide du cofacteur NADP<sup>+</sup> est localisé au-dessus de l'isoalloxazine du FAD, bloquant l'accès au site actif par le canal évoqué précédemment (**Article, Figure S34**). Cette conformation divergente semble donc être corrélée au fait que LbmA-Ox agit sur un substrat lié à un ACP plutôt que sur un substrat diffus.

Les différents résidus identifiés au niveau du site actif et des sites de fixation des cofacteurs/substrat ont été ensuite mutés et l'impact a été évalué par des tests d'activité suivis d'une caractérisation par UHPLC-HRMS (**Article, Figure 4**). Les effets de ces mutations ont révélé le rôle prépondérant de deux résidus aspartate et arginine, et ont ainsi permis de proposer deux mécanismes potentiels de formation d'oxime catalysés par LbmA-Ox. Le premier mécanisme implique la déshydrogénation d'un intermédiaire monohydroxylé tandis que le second requiert un intermédiaire dihydroxylé généré par deux monooxygénations successives, suivie de sa déshydratation (**Article, Figure 5**). Cependant, seul le deuxième mécanisme nécessite l'intervention d'un couple d'acides/bases générales, ce qui est en accord avec les données de mutagenèse, et représente donc l'alternative la plus probable.

Dans certaines voies de biosynthèse de méthylloxime représentées en **Figure 1 de l'article**, le module MT en aval du module de monooxygénation contient un domaine DH additionnel. Cela suggère que dans ces systèmes, l'intermédiaire dihydroxylé n'est pas déshydraté par le domaine FMO mais est au contraire transmis au module suivant où la

déshydratation a lieu, suivie de la méthylation. Pour tenter de comprendre ce qui distingue LbmA-Ox de ces homologues *N*-hydroxylases, des modèles AlphaFold2<sup>336</sup> ont été générés et comparés à la structure cristallographique de LbmA-Ox (**Article, Figure S35**). La superposition des sites actifs a révélé une conservation des résidus catalytiques, mais également une différence dans l'orientation d'une chaîne latérale tyrosine dont l'implication dans la catalyse avait été démontrée par mutagenèse (**Article, Figure 4**). Ainsi, dans les domaines Ox sans DH en aval, la tyrosine est orientée telle que dans la structure de LbmA-Ox. En revanche, dans les domaines Ox dont le module suivant contient une DH, cette tyrosine n'est plus orientée en direction du site actif et ne peut donc vraisemblablement plus participer à la catalyse.

Finalement, la position relative des domaines Ox et ACP dans la structure cristallographique a été analysée. Deux molécules sont présentes dans l'unité asymétrique et chacune adopte une conformation différente. Toutefois, dans les deux conformations, la distance entre la sérine modifiée de l'ACP et le site actif de l'Ox est incompatible avec la longueur du substrat natif glycyI-phosphopantéthéine (**Article, Figure S36**). Par conséquent, ces deux conformères cristallographiques ne représentent pas un état catalytiquement compétent de l'enzyme mais permettent la formation du réseau cristallin. Des expériences de *docking in silico* ont alors été réalisées avec le serveur HDOCK<sup>343</sup> pour tenter d'identifier une conformation pertinente entre les deux domaines du point de vue de la catalyse. Une contrainte de distance a été imposée entre l'extrémité C-terminale du domaine Ox et l'extrémité N-terminale de l'ACP, basée sur la longueur du linker interdomaine. Cependant, aucune des solutions calculées n'a placé le domaine ACP suffisamment proche du site actif de l'Ox pour permettre une interaction productive (**Article, Figure S37**). Ces résultats nous ont amené à proposer que la réaction de formation d'oxime doit avoir lieu soit sur le domaine PCP du module en amont, soit sur le domaine ACP de la sous-unité LbmA partenaire dans le contexte homodimérique des PKS.

Ces hypothèses ont été vérifiées en reproduisant les tests d'activité du trimodule N-terminal de LbmA, avec des constructions dans lesquelles la sérine modifiée du domaine PCP, du domaine ACP du module Ox, ou du domaine ACP du module MT en aval a été mutée en alanine. Les données de MS ont ensuite révélé que la formation d'oxime a lieu sur l'intermédiaire lié à l'ACP du module Ox, suggérant ainsi que la catalyse de LbmA-Ox se produit sur l'ACP de la deuxième sous-unité de l'homodimère LbmA. Cette observation est cohérente avec les expériences de *docking* entre la structure cristallographique de LbmA-Ox et de l'ACP réalisées sans contrainte de linker interdomaine, dont plusieurs solutions calculées ont révélé un complexe catalytiquement compétent (**Article, Figure S37**).

Ces travaux ont été inclus dans un article scientifique, publié le journal *Angewandte Chemie*

## Résultats

sous le format *Research Article*, et dont je suis co-premier auteur <sup>344</sup> :

Dr. Hannah A. Minas<sup>#</sup>, **Romain M. M. François<sup>#</sup>**, Dr. Franziska Hemmerling<sup>#</sup>, Dr. Amy E. Fraley, Cora L. Dieterich, Dr. Simon H. Rüdiger, Dr. Roy A. Meoded, Sabrina Collin, Prof. Dr. Kira J. Weissman\*, Dr. Arnaud Gruez\*, Prof. Dr. Jörn Piel\*. Modular Oxime Formation by a *trans*-AT Polyketide Synthase. *Angew. Chem. Int. Ed.* **62**, e202304481 (2023).

<sup>#</sup> co-premiers auteurs

Dans ce travail collaboratif entre les équipes des Pr. Kira J. Weissman et Jörn Piel, mon rôle a été de cloner les séquences codant LbmA-Ox et LbmA-Ox-ACP, ainsi que de produire et purifier ces constructions avec l'aide de Sabrina Collin. J'ai également réalisé le travail de cristallographie, de l'étape de cristallisation à l'interprétation de la structure, sous la direction du Dr. Arnaud Gruez et du Pr. Kira J. Weissman. J'ai pu générer la totalité des mutants de LbmA-Ox-ACP dans l'équipe du Pr. Jörn Piel avec l'aide du Dr. Amy E. Fraley, et les tests d'activité des mutants ainsi que la caractérisation des produits de réaction par UHPLC-HRMS ont été réalisés par le Dr. Franziska Hemmerling et moi-même.

Les informations et données supplémentaires de cet article ont été incluses en annexe de cette thèse.



**Polyketide Synthases**

# Modular Oxime Formation by a *trans*-AT Polyketide Synthase

Hannah A. Minas<sup>+</sup>, Romain M. M. François<sup>+</sup>, Franziska Hemmerling<sup>+</sup>, Amy E. Fraley,  
 Cora L. Dieterich, Simon H. Rüdisser, Roy A. Meoded, Sabrina Collin, Kira J. Weissman,<sup>\*</sup>  
 Arnaud Gruez,<sup>\*</sup> and Jörn Piel<sup>\*</sup>

**Abstract:** Modular *trans*-acyltransferase polyketide synthases (*trans*-AT PKSs) are enzymatic assembly lines that biosynthesize complex polyketide natural products. Relative to their better studied *cis*-AT counterparts, the *trans*-AT PKSs introduce remarkable chemical diversity into their polyketide products. A notable example is the lobatamide A PKS, which incorporates a methylated oxime. Here we demonstrate biochemically that this functionality is installed on-line by an unusual oxygenase-containing bimodule. Furthermore, analysis of the oxygenase crystal structure coupled with site-directed mutagenesis allows us to propose a model for catalysis, as well as identifying key protein-protein interactions that support this chemistry. Overall, our work adds oxime-forming machinery to the biomolecular toolbox available for *trans*-AT PKS engineering, opening the way to introducing such masked aldehyde functionalities into diverse polyketides.

## Introduction

Benzolactone enamides are a family of cytotoxic polyketide natural products<sup>[1]</sup> reported from various organisms<sup>[2]</sup> (Figure 1). Many representatives contain an unusual *O*-methyl oxime moiety<sup>[3]</sup> that is present in few other known polyketides. Recently, initial insights into the biosynthetic origin of the oxime-bearing lobatamide A (**1**) and necroxime A (**2**) were obtained for two bacterial producers, the root-

associated *Gyvuella sunshinyii* YC6258<sup>[4]</sup> and the fungal endosymbiont *Burkholderia* sp. HKI-0404.<sup>[5]</sup> Their host-associated lifestyles suggest that **1** and **2** mediate symbiotic interactions in a broad range of organisms.

**1** and **2** are generated by multimodular biosynthetic enzymes termed *trans*-acyltransferase polyketide synthases (*trans*-AT PKSs). In general, PKS modules sequentially elongate and functionalize an acyl carrier protein (ACP)-bound growing polyketide chain. In each homodimeric module, a ketosynthase (KS) domain extends the growing polyketide chain by a ketide unit supplied by an acyltransferase (AT).<sup>[6]</sup> The nascent polyketide can be further processed by optional domains including ketoreductases (KRs), dehydratases (DHs), and enoyl reductases (ERs), to successively convert the  $\beta$ -keto group to  $\beta$ -hydroxy,  $\alpha,\beta$ -double-bond, or fully saturated moieties.<sup>[7]</sup> PKSs can also include modules from non-ribosomal peptide synthetases (NRPS) that incorporate amino acids into the growing chain via the concerted action of adenylation (A) and condensation (C) domains.<sup>[6a,8]</sup>

An intriguing feature of *trans*-AT PKSs is the abundance of non-canonical modules, which include a wide array of module-integrated or isolated enzymes acting in *trans*.<sup>[7,9]</sup> Some of us previously developed a predictive method<sup>[10]</sup> that infers polyketide moieties introduced by *trans*-AT PKS modules from the sequence of coevolving downstream KS domains, as the KS phylogenetic clading correlates with the structure of the incoming intermediate. This method greatly streamlines the prediction and discovery of *trans*-AT PKS products. However, as it relies on pre-correlation of KS clades with substrate structure, it cannot predict novel modular functions.<sup>[9a]</sup> Here we report biochemical studies on the lobatamide A PKS<sup>[11]</sup> (**1**, Figure 1A) that demonstrate oxime formation as a novel, integral function of *trans*-AT PKSs. The modification of an initially installed glycine moiety is accomplished by a bimodular system with oxygenating and methylating components that are conserved in all known benzolactone oxime pathways (Figure 1A–C). These results expand the biosynthetic scope of *trans*-AT assembly lines and offer the potential to introduce oxime tags into engineered complex polyketides.

## Results and Discussion

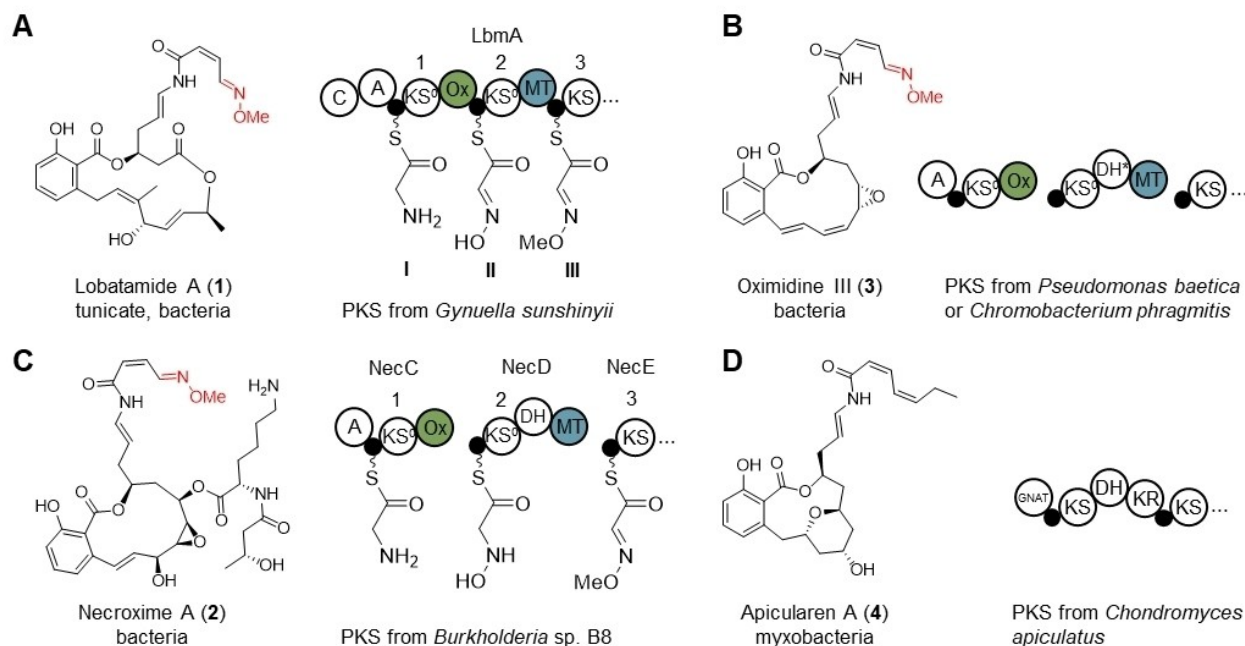
Based on the PKS architecture and the structure of the mature benzolactone polyketides, we and others<sup>[4a,5,12]</sup> hypothesized that the methylated oxime moiety is not

[\*] Dr. H. A. Minas,<sup>+</sup> R. M. M. François,<sup>+</sup> Dr. F. Hemmerling,<sup>+</sup> Dr. A. E. Fraley, C. L. Dieterich, Dr. R. A. Meoded, Prof. Dr. J. Piel  
 Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zürich  
 Vladimir-Prelog-Weg 4, 8093 Zürich (Switzerland)  
 E-mail: jpiel@ethz.ch

R. M. M. François,<sup>+</sup> S. Collin, Prof. Dr. K. J. Weissman, Dr. A. Gruez  
 Université de Lorraine, CNRS, IMoPA  
 54000 Nancy (France)  
 E-mail: kira.weissman@univ-lorraine.fr  
 arnaud.gruez@univ-lorraine.fr

Dr. S. H. Rüdisser  
 Institute of Molecular Biology and Biophysics, Biomolecular NMR Spectroscopy Platform, Eidgenössische Technische Hochschule (ETH) Zürich  
 Hönggerbergring 64, 8093 Zürich (Switzerland)

[†] These authors contributed equally to this work.



**Figure 1.** Selected examples of *trans*-AT PKS-derived benzolactone enamide polyketide natural products and associated biosynthetic gene clusters, with methyl oximes indicated in red. The sources of the metabolites are specified beneath the compound names. A) Putative first steps of lobatamide A (1) biosynthesis. The consecutive action of LbmA-Ox and LbmA-MT is hypothesized to install an oxime moiety which is subsequently methylated. B) Domain representation of the oximidine III (3) PKS (\*the dehydratase (DH) domain is only present in the *C. phragmitis* PKS). C) Domain representation of the necroxime (2) PKS. D) Domain representation of the apicularen A (4) PKS. In agreement with the absence of Ox and MT domains, 4 lacks a methyl oxime moiety. C: condensation domain, A: adenylation domain, KS: ketosynthase, KS<sup>0</sup>: non-elongating KS, Ox: flavin-dependent monooxygenase, MT: methyl transferase, DH: dehydratase, GNAT: GCN5-related *N*-acetyl transferase, black circles: carrier protein, wavy lines: phosphopantetheinyl moieties. The numbers above the KSs indicate their positions in the PKSs. Roman numerals refer to the putative thioester-bound intermediates. Accession numbers are given in Table S2.

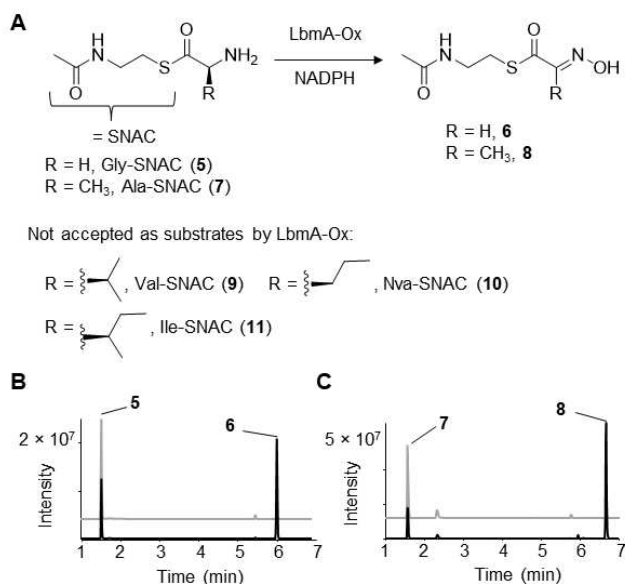
installed by post-PKS enzymes but by the first three modules of the assembly line. Modules of similar domain composition are also found in the PKS-NRPSs of two oximidine III (3) producers (Figure 1B). The biosynthetic assignment is further supported by the absence of the modules in the *trans*-AT PKS of the related compound apicularen A (4), which lacks the oxime unit (Figure 1D).<sup>[5]</sup> In the lobatamide pathway, the first module bears the hallmarks of an NRPS loading module as it comprises an A domain and a peptidyl carrier protein (PCP) domain.<sup>[5,13]</sup> Analysis of the A domain signature by the NRPSpredictor2<sup>[14]</sup> software of antiSMASH<sup>[15]</sup> suggested that glycine is loaded onto the downstream PCP.

Correspondingly, the following module harbors a non-elongating KS<sup>0</sup> domain (LbmA-KS<sup>0</sup><sub>1</sub>) that phylogenetically clades with glycine-accepting KSs.<sup>[4a]</sup> This PKS module further includes a non-canonical domain (LbmA-Ox) with homology to flavin-dependent monooxygenases. This Ox domain was previously hypothesized to hydroxylate the thioester-linked intermediate I (Figure 1A), followed by *O*-methylation by the methyltransferase (MT) of the third module (LbmA-MT). The related *trans*-AT PKS assembly lines for 2 and 3 (from *C. phragmitis*) each harbor an additional DH domain that was postulated to assist in conversion of a hydroxylated glycine to an oxime.<sup>[5]</sup>

In our previous analyses, automated prediction of the substrates for LbmA-KS<sup>0</sup><sub>2</sub> and LbmA-KS<sub>3</sub> using TransATor<sup>[10b]</sup> did not provide conclusive structural suggestions.<sup>[4a]</sup> We therefore inferred an improved phylogenetic tree and found LbmA-KS<sup>0</sup><sub>2</sub> and LbmA-KS<sub>3</sub> in unassigned clades that also contained the KSs from the corresponding modules in the necroxime and oximidine *trans*-AT PKS systems (Figure S1). This prompted us to investigate the enzymology occurring in the respective upstream modules to characterize their potentially new functions.

To test the role of the putative oxygenase LbmA-Ox in oxime formation, we cloned the corresponding region of the *lbm* biosynthetic gene cluster into a derivative of the pET28b expression vector (pET28b-SUMO). Upon expression in *E. coli* and purification as an N-terminally His<sub>6</sub>- and SUMO-tagged protein (Figure S2), yellow protein was obtained indicating bound flavin adenine dinucleotide (FAD). The identity of LbmA-Ox and its bound cofactor were verified by ultra-high performance liquid chromatography/high-resolution mass spectrometry (UHPLC-HRMS, Figure S3) and UV/Vis (Figure S4).

As surrogates for intermediates tethered via 4'-phosphopantetheinyl moieties to the PKS, *N*-acetylcysteamine (SNAC) thioesters were used as substrates to test activity.<sup>[16]</sup> We synthesized glycine SNAC (Gly-SNAC, 5, Figure 2) as a



**Figure 2.** Studying oxime formation in vitro. A) In vitro enzymatic assays with LbmA-Ox and thioesters **5**, **7** and **9–11** as test substrates. B) Extracted ion chromatograms (EICs) from UHPLC-HRMS data of assay mixtures for **5** and expected product **6** (calcd. for  $[M+H]^+$  = 177.0692 and 191.0484, obs. = 177.0688 ( $\Delta$  ppm = 2.3) and 191.0481 ( $\Delta$  ppm = 1.6), respectively); and for C) **7** and expected product **8** (calcd. for  $[M+H]^+$  = 191.0848 and 205.0641, obs. = 191.0845 ( $\Delta$  ppm = 1.6) and 205.0636 ( $\Delta$  ppm = 2.4), respectively). In addition to the reactions with LbmA-Ox (black), negative controls containing boiled enzyme (grey) are shown.

mimic of the proposed intermediate **I** (Figures 1A, S5, S69, and S70). For biochemical assays, **5** was incubated with LbmA-Ox and reduced nicotinamide adenine dinucleotide phosphate (NADPH) as a reducing agent for the bound FAD (Figure 2A).<sup>[17]</sup> To indirectly follow the reaction, NADPH consumption was monitored spectroscopically at 340 nm. As expected, NADPH levels decreased relative to a boiled enzyme control (Figure S6A).

UHPLC-HRMS analyses of the assay containing active enzyme revealed a new product **6** with a mass difference of +14 Da compared to **5**, consistent with the expected addition of one oxygen and loss of two hydrogens during oxime formation (Figure 2B, black trace, Figures S6B and S7A). No such product was detected in the boiled enzyme control (Figure 2B, grey trace, Figure S7B). Further analysis by tandem mass spectrometry (MS/MS) revealed fragment ions in agreement with the expected compound (Figure S8).

To determine whether the installed oxygen originates from water or molecular oxygen, we repeated the assay in an <sup>18</sup>O<sub>2</sub>-enriched atmosphere. Compared to the control experiment, a +2 Da shift corresponding to incorporation of <sup>18</sup>O from molecular oxygen was observed (Figure S6C). In contrast, no <sup>18</sup>O was present in the product of an assay carried out with an H<sub>2</sub><sup>18</sup>O background. The incorporation of molecular oxygen is consistent with LbmA-Ox acting on **5** as a monooxygenase.<sup>[18]</sup>

We next aimed to explore the substrate scope of LbmA-Ox. Four further amino acid SNACs (**7**, **9–11**) were

synthesized and analyzed in the same manner as **5** (Figures 2, S9–S13, S71–S78). In the case of Ala-SNAC (**7**), we observed a product with *m/z* features and MS/MS fragments matching the corresponding monooxygenated compound (**8**, Figure 2C, Figures S11, S12). No putative products were observed for test substrates with bulkier side chains (**9–11**, Figure S13) or free amino acids (Figure S14). The lack of product formation for the latter indicates the need for a carrier protein-bound substrate that is sufficiently mimicked by the SNACs.

Isolation of the glycine thioester-derived product **6** for nuclear magnetic resonance (NMR) characterization was greatly hampered by degradation under purification conditions. To address this issue, we measured the stability of the aminoacyl SNACs in hydrolysis assays using Ellman's reagent<sup>[19]</sup> (Figure S15), which confirmed thioester hydrolysis within minutes, as described previously for similar amino acid SNACs.<sup>[16d]</sup> Relative to **7** and **9–11**, glycine SNAC **5** is particularly unstable. Correspondingly, a comparison of UHPLC-HRMS chromatograms before and after attempted isolation revealed the appearance of new peaks for the free SNAC thiol of hydrolyzed substrates or products (Figure S16). However, no putative glycine- or glyoxylate-like degradation products were detectable in either positive or negative ionization mode.

We therefore synthesized a <sup>15</sup>N-labelled glycine SNAC (**S1**, Figures S12, S17 and S65–S68) as a substrate for assays that were subsequently analyzed by <sup>15</sup>N isotope-edited NMR. Experimental parameters were set to detect <sup>1</sup>H coupled to <sup>15</sup>N within *E*- and *Z*-oxime, imine, or amine moieties. No NMR signals corresponding to any of these functional groups were detectable. Control experiments with <sup>15</sup>N-hydroxyiminoacetic acid (**S2**) synthesized as a standard for hydrolyzed oxime **6** (Figures S18 and S19) indicated a reduced sensitivity of the isotope-edited NMR experiments (Figure S20) compared to the non-edited experiment. This lower sensitivity combined with incomplete conversion may have hindered detection of the oxime.

Likewise, in kinetic NMR experiments, we observed a loss of signals for **S1** and the simultaneous appearance of a putative hydrolysis product (Figures S21 and S22), but no new signals corresponding to an oxime product **6** (Figures S23 and S24). Instead, a new signal at  $\delta(^1\text{H}) = 8.0$  ppm and  $\delta(^{15}\text{N}) = 112.2$  ppm consistent with an amide was observed (Figure S24B). Amide formation might be explained by transformation of the oxime to an amide via Beckmann rearrangement.<sup>[20]</sup>

To overcome these challenges, we investigated the proposed role of LbmA-MT as an oxime *O*-methyltransferase, converting ACP-tethered intermediate **II** to **III** (Figure 1A) with potentially increased stability. The corresponding MT gene region was cloned into pET28b-SUMO and expressed in *E. coli* as an N-terminally His<sub>6</sub>- and SUMO-tagged protein. The MT was subsequently purified (Figure S25) and used in biochemical assays with MgCl<sub>2</sub>, *S*-adenosyl methionine (SAM), and either **6** or **8** as test substrates, the latter obtained from assays with LbmA-Ox. Additionally, we conducted one-pot reactions with LbmA-Ox, LbmA-MT, coenzymes, and either **5** or **7**. However, no

products corresponding to methylated **6** or **8** were detected by UHPLC-MS.

Considering that LbmA-MT is module-embedded, we investigated its role in vivo within an intact PKS environment in which native protein-protein interactions involving LbmA-MT would be maintained. For this, we cloned the gene region encoding the first three modules of LbmA comprising the entire putative methyloxime-generating system—a ca. 3000 amino acid protein terminating with LbmA-KS<sub>3</sub>—into pET28b-SUMO and introduced the plasmid into *E. coli* BAP1.<sup>[21]</sup> This strain harbors the *sfp* gene from *B. subtilis*, which encodes a promiscuous 4'-phosphopantetheinyl transferase (PPTase) for production of active *holo* carrier proteins.<sup>[21]</sup> We relied on the *E. coli* host background to supply glycine and necessary cofactors. After production of the trimodular assembly line, the cells were lysed and intermediates were released from the protein by addition of cysteamine (**12**, Figure 3A).<sup>[22]</sup> Purification and subsequent

in vitro assays with the very large LbmA trimodule (ca. 530 kDa) were not possible due to low protein yield.

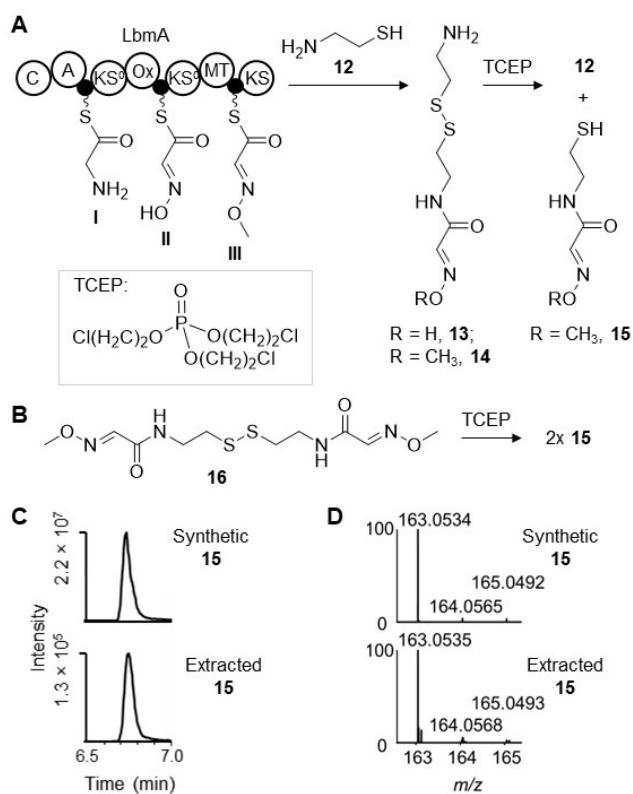
UHPLC-HRMS analysis indicated formation of **II** and **III** observed as **13** (Figure 3A and Figure S26A,  $m/z$  calcd. for C<sub>6</sub>H<sub>12</sub>N<sub>3</sub>O<sub>2</sub>S<sub>2</sub> [M+H]<sup>+</sup> = 224.0522, obs. = 224.0513,  $\Delta$  ppm = 4.0) and **14** (Figure 3A and Figure S26B,  $m/z$  calcd. for C<sub>7</sub>H<sub>14</sub>N<sub>3</sub>O<sub>2</sub>S<sub>2</sub> [M+H]<sup>+</sup> = 238.0678, obs. = 238.0667,  $\Delta$  ppm = 4.6) with predicted molecular formulae consistent with oxygenation and methylation. These compounds were absent in a control experiment using empty pET28b-SUMO (Figure S26). One peak was observed for **13** and two peaks were observed for **14**, that may correspond to the *E* and *Z* isomers of the methylated oxime group. We only observed intermediates arising from conversion of the native glycine but not of alanine.

To further test whether the product contains a methyl group derived from SAM, we repeated the in vivo assay in medium supplemented with <sup>13</sup>C-methyl methionine. Detection of  $m/z$  features corresponding to **14** with a 1 Da mass increase supports incorporation of SAM originating from <sup>13</sup>C-methyl methionine (Figure S27).

To obtain structural information on **14**, we scaled up the in vivo assay for HPLC purification. However, despite promising amounts of product in the crude extracts, it appeared to disintegrate over the course of purification. Due to the observed instability of the methyl oxime **14**, we decided to generate its disulfide dimer **16** as a suitable synthetic standard (Figures 3B, S28, and S79–S83). The corresponding free thiol **15** was obtained by treatment with the mild reducing agent tris(2-carboxyethyl)phosphine (TCEP, Figure 3A). We followed the reaction in situ by <sup>1</sup>H NMR and observed characteristic shifts<sup>[23]</sup> of the triplet at 3.45 ppm to 3.28 ppm as well as of the triplet at 2.87 ppm to 2.54 ppm, indicating that **15** was generated (Figures S29 and S83).

Furthermore, UHPLC-HRMS analysis of the sample displayed a product peak at 6.74 min corresponding to an ion with the characteristic sulfur isotope pattern and the expected mass (Figures 3C and D, upper,  $m/z$  calcd. for C<sub>5</sub>H<sub>11</sub>N<sub>2</sub>O<sub>2</sub>S [M+H]<sup>+</sup> = 163.0536, obs. = 163.0535,  $\Delta$  ppm = 0.6). To validate the product obtained from the in vivo assay, we treated the cell lysate analogously and compared the UHPLC-HRMS data to the synthetic standard. The MS and MS/MS data were congruent (Figures 3C, D and S30), which leads us to conclude that **III** is the product of the starting modules of the lobatamide PKS assembly line.

In organic extracts of lobatamide-producing *Gyvuella sunshinyii*, there was no evidence of an alanine-containing lobatamide congener. Accordingly, in the in vivo assay in which both LbmA-A and LbmA-KS<sup>0</sup> were present, only glycine-derived intermediates were observed. Since free LbmA-Ox accepts the non-native alanine thioester as a substrate in vitro, we hypothesize that the upstream A domain and perhaps the KS<sup>0</sup> domains act as gatekeepers to prevent this reaction on the PKS. Indeed, the strong preference of the LbmA-A domain for glycine combined with action of the downstream LbmA-KS<sup>0</sup> which phylogenetically matches KSs accepting glycine-containing intermediates, could ensure installation of exclusively glycine-



**Figure 3.** Studying methyl oxime formation in vivo. A) In vivo assay expressing the first three modules of LbmA in *E. coli* BAP1. After cell disruption, the cell lysate was treated with **12** and the offloaded intermediates **13** and **14** were converted to thiols with tris(2-carboxyethyl)phosphine (TCEP) and analyzed by UHPLC-HRMS (Figure S26). B) In situ generation of **15** through reduction of dimer **16**. C) Structural verification of **15**. EICs with identical retention times are shown from UHPLC-HRMS data of synthetic **15** (upper panel, calcd. for C<sub>5</sub>H<sub>11</sub>N<sub>2</sub>O<sub>2</sub>S [M+H]<sup>+</sup> = 163.0536) and **15** extracted from an in vivo assay (lower panel). D) MS spectra of synthetic **15** (upper panel, obs. [M+H]<sup>+</sup> = 163.0534,  $\Delta$  ppm = 1.2) and **15** extracted from an in vivo assay (lower panel, obs. [M+H]<sup>+</sup> = 163.0535,  $\Delta$  ppm = 0.6).

derived oxime moieties in the elongated polyketide. Such a gatekeeping function has previously been attributed to amino acid-accepting KSs<sup>[24]</sup> and KS<sup>0</sup>s.<sup>[25]</sup> LbmA-KS<sup>0</sup> may act similarly, protecting the unstable intermediate **II** while channeling it to LbmA-MT for stabilizing methylation.

We next turned our attention to the mechanistic details of oxime formation. Reaction mechanisms for oxime-forming enzymes in PKS, NRPS and PKS-NRPS hybrid pathways have been proposed previously, most commonly involving *N*-hydroxylation by a cytochrome P450<sup>[26]</sup> or flavin-dependent monooxygenase.<sup>[27]</sup> However, all of these examples involve post-assembly line enzymes, not on-line oxime formation.<sup>[26a,28]</sup>

To gain direct mechanistic insight into this atypical biochemistry, we aimed to solve the crystal structure of a ternary complex of the didomain monooxygenase-ACP (Ox-ACP<sub>1</sub>) with NADP<sup>+</sup> and substrate analogue ethyl glycinate (**17**). The 3.07 Å resolution structure (PDB ID: 8ODW) revealed that the asymmetric unit comprises two Ox-ACP didomains, each displaying a distinct orientation of the ACP relative to the Ox domain (Figure 4B). Chain A exhibits a large Ox/ACP interdomain interface (1582 Å<sup>2</sup>), while that in chain B is substantially smaller (137 Å<sup>2</sup>), indicating that the chain B conformation, which had to be refined with a lower occupancy, is less stable in the crystal packing than chain A (Figure S31).

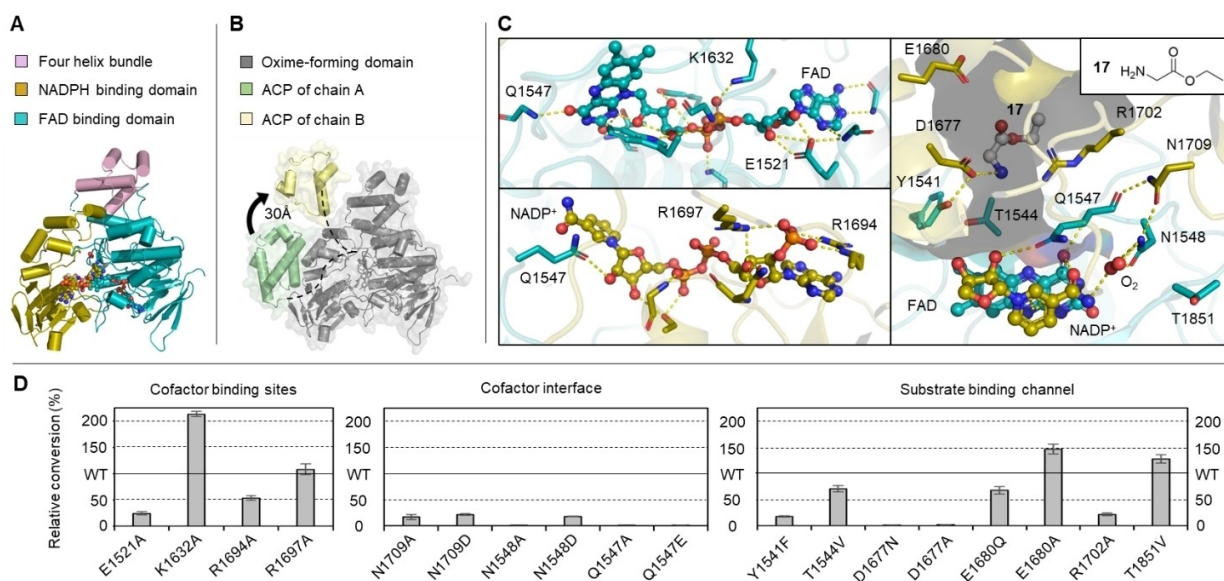
The monomeric Ox belongs to the group B flavin-dependent monooxygenases, which consist of two conserved Rossmann-like subdomains<sup>[29]</sup> that bind FAD and NADPH.

Both domains exhibit classical GxGxxG motifs (residues 1498–1503, 1670–1675) associated with cofactor binding. A third 4 α-helix bundle structural domain (Hα<sub>16</sub> 1913–1921, Hα<sub>19</sub> 1959–1973, Hα<sub>20</sub> 1975–1984 and Hα<sub>21</sub> 1989–1997) is located near the interface between the two nucleotide-binding domains (Figure 4A).

Comparison of the monooxygenase structure to its closest structural homologues identified using the Dali server,<sup>[30]</sup> as well as a selection of other *N*-hydroxylase structures, shows that the Rossmann-fold domains are common to the enzymes, but that only LbmA-Ox includes the additional 4 α-helix bundle (Figure S32). AlphaFold2<sup>[31]</sup> models of the *N*-hydroxylation domains from the oximidine and necroxime PKSs (Figure 1) also include this subdomain (Figure S32), indicating its potential significance in the multimodular context where it could mediate interactions with neighboring domains.

While **5** was too unstable for co-crystallization, we did capture a ternary complex with substrate analogue **17** along with FAD and NADP<sup>+</sup>, revealing the location of the active site and respective cofactor binding domains. Notably, the channel accommodating **17**, which is formed by helix α5 (residues 1673–1685) and the loops β2-Hα2 (1522–1569), β10-Hα5 (1694–1713) and Hα8-Hα9 (1756–1764), opens directly onto the isoalloxazine C4a position which is the site of the redox chemistry (Figure 4C).<sup>[32]</sup>

This substrate-binding pocket encompasses the side chains of residues Tyr1541, Thr1544, Gln1547, Asp1677, Glu1680, and Arg1702, of which the side chain of Asp1677



**Figure 4.** Crystal structure of LbmA-Ox-ACP and enzymatic conversion assays. A) Overall structure of LbmA-Ox domain with NADPH binding domain in gold, FAD binding domain in blue, the 4 α-helix bundle in lilac, and ligands shown as spheres. B) Superposition of the two LbmA-Ox-ACP chains of the asymmetric unit. The relative positioning of the ACPs from chain A (green) and chain B (yellow) differs by 30 Å, reflecting the high flexibility of the interdomain linker (dotted lines). C) Zooms into the FAD (upper left) and NADPH (lower left) binding sites as well as the active and substrate binding channel (right). The residues that interact with the cofactors are shown as sticks, hydrogen bond/electrostatic interactions are represented as dashed yellow lines, and the surface of the substrate binding channel is shown. The structure of **17** is inset. D) Percent conversion in the oxime formation reaction of LbmA-Ox active site mutants relative to the wild-type (WT). The formation of oximated product was monitored at least in triplicate (the error bars indicate the standard deviations).

and Arg1702 are close enough to form hydrogen bonds with the amino group of the substrate (3.1 Å and 3.3 Å respectively, Figure S33). The amine is oriented toward the flavin C4a position and a water molecule is positioned between these two sites (3.1 Å and 4.1 Å, respectively), likely at the location where the hydroperoxide intermediate would form, stabilized by interaction of the backbone carbonyl of Arg1545 with the N5 atom of the isoalloxazine.

The structural similarity with a reconstructed ancestral FMO (PDB ID: 6SE3) and the presence of a corresponding positive electron density peak, led us to tentatively model molecular oxygen within hydrogen-bonding distance (3.0 Å) of the carboxamides of NADP<sup>+</sup> and Asn1548 (Figure S33). Analysis of the cofactor binding sites shows that both FAD and NADP<sup>+</sup> adopt an elongated conformation and interact with each other via stacking of their respective isoalloxazine and nicotinamide moieties, although the electron density for the nicotinamide rings is relatively weak. Overall, cofactor binding arises from an extensive hydrogen-bond network. Specifically, Arg1694 and Arg1697 contact the 2'-phosphate and 3'-hydroxyl groups of the adenosyl moiety of NADP<sup>+</sup>, while Glu1521 and Lys1632 hydrogen-bond to the ribose moiety and the phosphate A group of FAD, respectively (Figure 4C).

Importantly, the reactive C4 of NADP<sup>+</sup> points away from the FAD N5 position (4.7 Å), making this conformation unsuitable for hydride transfer. Consequently, the observed cofactor conformations are incompatible with reduction of FAD by NADPH, but are consistent with the next stage of catalysis in which oxidized NADP<sup>+</sup> stabilizes the (hydro)peroxyflavin intermediate, as is typically observed in FMO crystals.<sup>[33]</sup> Nonetheless, this FAD-NADP<sup>+</sup> non-reducing conformation differs from those characterized previously in stand-alone FMOs, as the nicotinamide is not stacked against the isoalloxazine but is instead located above it, which allows its carboxamide group to interact with the flavin N5 (Figure S34).

This interaction involves a highly conserved Gln/Asn (Gln1547 in LbmA-Ox), as observed previously, which hydrogen bonds with the 2'-OH group of the NADP<sup>+</sup> ribose moiety, potentially stabilizing the C4a-(hydro)peroxyflavin intermediate.<sup>[34]</sup> Such stabilization would only require that the Gln1547 side chain moves by 1.0 Å relative to its position in the crystal, which is reasonable given the highly dynamic behavior of FMO active sites in response to cofactor redox states.<sup>[35]</sup> Although the conserved residue at position 1547 is often also involved in binding molecular oxygen,<sup>[17b]</sup> the proximity of Asn1548 to the bound dioxygen rather suggests that it plays this critical role. Finally, we note that the divergent conformation of NADP<sup>+</sup> with respect to FAD may correlate with the fact that LbmA-Ox acts on a carrier protein-bound rather than a freely diffusible substrate. Indeed, although the LbmA-Ox substrate binding channel is present in stand-alone FMOs, the positioning of the nicotinamide above the flavin N5 position sterically occludes access to the active site via this channel (Figure S34).

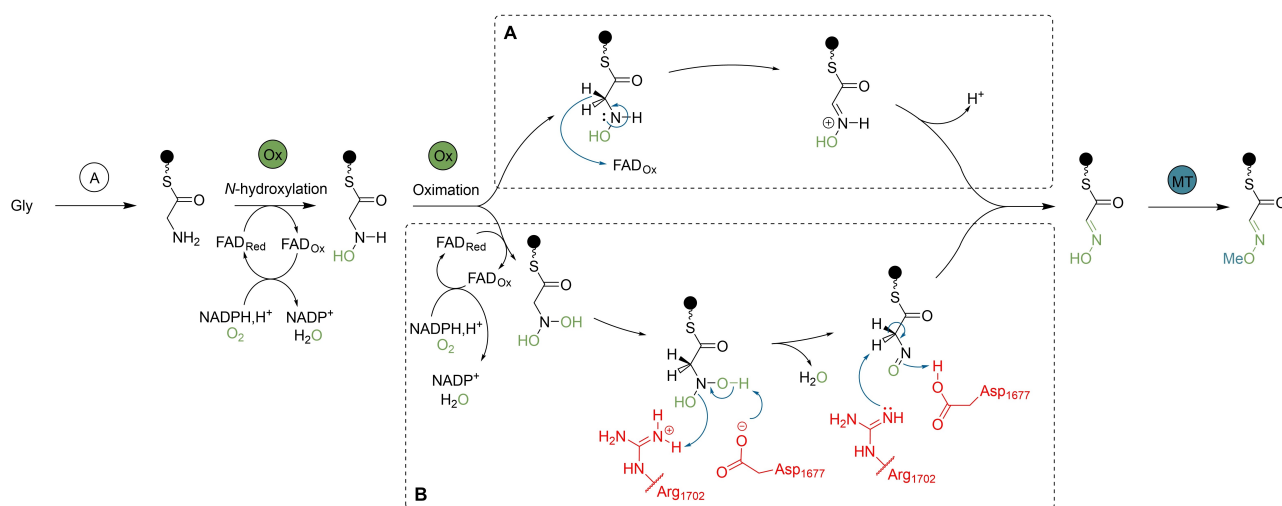
We next probed the role of specific amino acids in catalysis by site-directed mutagenesis coupled with UHPLC-

HRMS analysis. For each freshly purified mutant, the oxidized FAD (FAD<sub>Ox</sub>) occupancy was quantified relative to the wild-type content (55 %) (Table S4), and the activity measured in the presence of excess FAD to ensure reliable product detection (Figure 4D). Thus, any observed effects of the mutations on activity are independent of those on FAD<sub>Ox</sub> binding.

Globally, we observed variable effects of the mutations on FAD<sub>Ox</sub> occupancy (Table S4). N1548D, N1709D, D1677N/A, and R1702A substitutions exhibited reduced FAD<sub>Ox</sub> binding (19–32 %), while that by E1521A was almost completely abolished (2 %), consistent with the high conservation of this amino acid in flavin monooxygenases. The other substitutions involve residues close to the cofactor isoalloxazine and nicotinamide groups or the surrogate substrate, suggesting that the low FAD<sub>Ox</sub> content is due either to diminished FAD binding or to the accumulation of reduced FAD (FAD<sub>Red</sub>). In contrast, FAD<sub>Ox</sub> occupancy in the Q1547E/A, N1548A, Y1541F and E1680A mutants was higher than that of the wild type (84–96 %). Thus, these mutations either increase FAD binding or decrease the efficiency of hydride transfer from NADPH to FAD, leading to accumulation of FAD<sub>Ox</sub>. Nonetheless, the increased FAD<sub>Ox</sub> binding by the E1680A mutant is unexpected, as although this amino acid is located in the substrate binding pocket, it is distant from the cofactors and substrate analogue.

Concerning the effects on activity, alanine substitutions of the highly conserved Glu1521 and Arg1694 reduced conversion by 2- and 4-fold, respectively, relative to the wild-type enzyme (Figures S39 and S40). As expected, mutation of Gln1547 and Asn1548 substantially decreased turnover (Figures S52 and S53). Asn1548 could be substituted by the negatively charged Asp, albeit with a loss in activity (18 % of WT, Figure S48). Substitution of Asn1709 also decreased the activity (21 % and 17 % for N1709D and N1709A, respectively, Figures S45 and S50), which is consistent with a role in ensuring the correct orientation of the Asn1548 side chain necessary for binding molecular oxygen. The Asn or Ala mutants of Asp1677 lost all activity (Figures S49 and S57). Similarly, mutation of Tyr1541 and Arg1702 located close to Asp1677 to Phe and Ala, respectively, reduced turnover (18 % and 22 %, Figures S56 and S41). On the contrary, the T1544V and E1680Q mutants maintained 70 % and 68 % oxime-forming activity, respectively (Figures S61 and S41), in keeping with only a minor role in catalysis, while turnover by E1680A increased (146 %). The latter mutation likely facilitates access to the active site, given the solvent-exposed nature of its side chain. Thr1851 was substituted by a valine to assess the potential role of another channel leading to the cofactor interface, but the resulting mutant exhibited essentially wild type properties (Figure S42).

Oxime formation could occur by one of two mechanisms (Figure 5). In the first, the hydroxylamine intermediate is directly dehydrogenated by the enzyme (Figure 5A). Alternatively, a second monooxygenation yields a *N*-dihydroxylated intermediate which then undergoes dehydration (Figure 5B), a mechanism which would require retention of the



**Figure 5.** Proposed mechanisms of methyl oxime formation in lobatamide biosynthesis. Two different pathways are hypothesized for the oxime formation catalyzed by LbmA-Ox, via dehydrogenation (A), or dehydration following a second mono-oxygenation (B). LbmA-Ox residues putatively involved in the mechanism are indicated in red.

monohydroxylated intermediate in the active site during the  $\text{NADP}^+/\text{NADPH}$  cofactor exchange. Classical FAD-dependent monoamine oxidases such as MAO-A and MAO-B<sup>[36]</sup> do not require a general acid/base to catalyze dehydrogenation. Thus, the strong dependence of LbmA-Ox activity on both Asp1677 and Arg1702, favors the dehydration mechanism (Figure 5B).<sup>[26a,27,37]</sup> Here, both Asp1677 and Arg1702 would act consecutively as general acids/bases. The  $\text{pK}_a$  of Asp1677 may be elevated from its solution value via its proximity to aromatic Tyr1541, increasing its efficiency as a general base. We also note that Arg1702 contacts Asp1763 (3.1 Å), and while somewhat counterintuitive, the role of Asp in adapting the  $\text{pK}_a$  of Arg residues to carry out general acid/base catalysis has experimental support.<sup>[38]</sup>

Unfortunately, our attempts to detect any putative *N*-dihydroxylated intermediate were unsuccessful, as only the oxime-containing product was observed by MS analysis *in vitro*.

Nonetheless, it is notable that the downstream MT-containing module of LbmA lacks a DH domain to catalyze the dehydration reaction, in contrast to the oximidine III and necroxime A PKSS (Figure 1). Taken together, these results argue that the two steps are catalyzed by LbmA-Ox itself, and therefore that LbmA-Ox is a bifunctional *N*-hydroxylase/oxime-forming enzyme.

We next aimed to understand what distinguishes the putatively bifunctional LbmA-Ox from its *N*-hydroxylase counterparts (Figure 1). For this, models of the Ox domains of two oximidine and two necroxime A systems were calculated using AlphaFold2.<sup>[31]</sup> The models and LbmA-Ox structure are closely similar, and reveal a highly conserved double dyad Asp1763-Arg1702/Asp1677-Tyr1541 (for clarity, we use the LbmA-Ox numbering to refer to the equivalent residues in the homologues). Nonetheless, the orientation of the Tyr1541 side chain differs between the models. Indeed, in the Ox domains lacking a downstream DH, the Tyr1541 side chain superimposes with that of

LbmA-Ox and points toward Asp1677. In contrast, in the Ox domains followed by a DH, the Tyr1541 side chains are oriented towards that of a conserved alanine residue (Figure S35) (this Ala is a Ser in LbmA-Ox-type domains). Based on this observation, we propose that this positioning prevents Tyr1541 from interacting with Asp1677 to adapt its  $\text{pK}_a$  for effective general acid/base catalysis.<sup>[39]</sup>

Lastly, we investigated the potential physiological relevance of the two ACP conformations observed in the LbmA-Ox-ACP crystal structure. The serine side chain (Ser2070) that carries the phosphopantetheinyl-bound substrate was identified by comparison with the *holo*-ACP structure of *Pseudomonas putida* (PDB ID: 7PDI), revealing its location at a distance of 40 Å and 60 Å from the Ox active site in the two ACP<sub>1</sub> conformations (Figure S36). Given the 25 Å length of the glycylophosphopantetheine substrate, these distances are incompatible with insertion of the chain into the active site, and thus, the observed Ox-ACP conformers do not represent catalytically competent snapshots of the enzyme.

To try to identify a catalytically relevant interaction between the Ox and ACP<sub>1</sub> domains located on the same subunit, we carried out *in silico* docking experiments using the HDock server.<sup>[40]</sup> Based on the length of the 15-amino acid linker between the domains, we set a maximum distance restraint of 55 Å between the C-terminus of the Ox domain and the N-terminus of the ACP. However, none of the calculated solutions placed the ACP close enough to the active site of the Ox to allow for productive interaction (Figure S37). From this result, we deduced that either oxidation occurs on Gly-ACP<sub>1</sub> furnished by the partner subunit of the LbmA homodimer, or on the upstream PCP domain (species I, Figure 1A). To distinguish between these possibilities, we constructed LbmA carrier protein mutants in which Ser819 (LbmA-PCP), Ser2070 (LbmA-ACP<sub>1</sub>), and/or Ser3053 (LbmA-ACP<sub>2</sub>) were replaced with alanine. We then repeated the *in vivo* cysteamine chain release assays on

strains expressing these constructs and screened the UHPLC-HRMS data for ions corresponding to either the oxime **13** or the *O*-methylated product **14** (Figures 3, S62, and S63).

As expected, **13** was not observed when LbmA-PCP was inactivated, corroborating its role in the loading of the glycine starter unit. Production of **13** was not detected for the LbmA-ACP<sub>1</sub> or LbmA-ACP<sub>1</sub>+LbmA-ACP<sub>2</sub> mutants (Figure S63), ruling out its formation on the LbmA-PCP domain. Consistent with the reaction occurring instead on LbmA-ACP<sub>1</sub>, formation of **13** but not **14** was observed when LbmA-ACP<sub>2</sub> was inactivated. Thus, catalysis by the Ox domain of one subunit most likely occurs on ACP<sub>1</sub> of the paired subunit, a cooperation model consistent with previous results obtained on homodimeric *cis*-AT PKSs.<sup>[41]</sup> Indeed, in silico docking between the Ox and ACP<sub>1</sub> domains in the absence of the constraint imposed by the intervening linker revealed several catalytically relevant poses (Figure S37). These results establish *O*-methyloxime formation as a reaction series occurring on PKS modules.

## Conclusion

Overall, our results support oxime installation as an integrated, modular function of *trans*-AT PKSs. Based on in vitro and in vivo assays, we propose the following order of events: after carrier protein loading by a glycine-specific A domain, the bifunctional LbmA-Ox domain performs an *N*-hydroxylation of the carrier protein-bound amino acid via the incorporation of molecular oxygen. Subsequent oxime formation may be achieved either by direct dehydrogenation or by a second *N*-hydroxylation followed by dehydration, the latter hypothesis being more consistent with the mutagenesis data. The intermediate is then transferred to the next module where LbmA-MT *O*-methylates the oxime. Due to the instability of oximes in aqueous solution, methylation may serve as protection from degradation.

To our knowledge, this is the first characterization of modular oxime formation for a PKS or NRPS system. Our work also establishes the functional investigation of modules upstream of unassignable KSs as a strategy for expanding the toolset for *trans*-AT PKS biosynthetic diversification. Indeed, the discrete, N-terminal architecture of the oxime-forming modules and their functionality in *E. coli* make the machinery a promising candidate for synthetic biology purposes.

## Acknowledgements

We thank Prof. M. Groll for derivatives of the pET28b expression plasmid, Prof. C. Khosla for the *E. coli* BAP1 strain, Dr. S. Leopold-Messer for *G. sunshinyii* cultures, Dr. S. I. Probst for help with HPLC purification, C. Padhi for help with MS experiments, Dr. A. O. Brachmann for <sup>18</sup>O-experiments, the B2S facility (UMS2008/US40 IBSLor) for crystallization and x-ray equipment, Dr. M. Savko and Dr. W. Shepard for help with diffraction data acquisition on the

beamline Proxima-2 A (SOLEIL, synchrotron), and Dr. E. Lakis for helpful discussion. Work in J.P.'s lab was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 742739, SynPlex), and the SNF grants 205321L\_197245 and 205320\_185077 to J.P. Funding for K.J.W.'s lab was provided by the Agence Nationale de la Recherche (ANR-20-CE93-0002-01 PKSOx to K.J.W. (funded jointly with SNF grant 205321L\_197245)), the Université de Lorraine (including a DrEAM travel grant to R.M.M.F.), and the Centre National de la Recherche Scientifique (CNRS).

## Conflict of Interest

The authors declare no conflict of interest.

## Data Availability Statement

The data that support this work are available in the Supporting Information of this article. The LbmA-Ox-ACP crystal structure has been deposited in the PDB under the accession number 8ODW.

**Keywords:** Bacterial Natural Products • Biosynthesis • Oxime • Polyketides • X-Ray Crystallography

- [1] M. R. Boyd, C. Farina, P. Belfiore, S. Gagliardi, J. W. Kim, Y. Hayakawa, J. A. Beutler, T. C. McKee, B. J. Bowman, E. J. Bowman, *J. Pharmacol. Exp. Ther.* **2001**, *297*, 114–120.
- [2] a) K. A. Dekker, R. J. Aiello, H. Hirai, T. Inagaki, T. Sakakibara, Y. Suzuki, J. F. Thompson, Y. Yamauchi, N. Kojima, *J. Antibiot.* **1998**, *51*, 14–20; b) K. L. Erickson, J. A. Beutler, J. H. Cardellina, M. R. Boyd, *J. Org. Chem.* **2001**, *66*, 1532–1532; c) K. L. Erickson, J. A. Beutler, J. H. Cardellina, M. R. Boyd, *J. Org. Chem.* **1997**, *62*, 8188–8192; d) D. L. Galinis, T. C. McKee, L. K. Pannell, J. H. Cardellina, M. R. Boyd, *J. Org. Chem.* **1997**, *62*, 8968–8969; e) B. Kunze, R. Jansen, F. Sasse, G. Höfle, H. Reichenbach, *J. Antibiot.* **1998**, *51*, 1075–1080; f) B. Kunze, H. Steinmetz, G. Höfle, M. Huss, H. Wiczorek, H. Reichenbach, *J. Antibiot.* **2006**, *59*, 664–668; g) Y. Hayakawa, T. Tomikawa, K. Shin-ya, N. Arao, K. Nagai, K. Suzuki, K. Furihata, *J. Antibiot.* **2003**, *56*, 905–908; h) J. W. Kim, K. Shin-ya, K. Furihata, Y. Hayakawa, H. Seto, *J. Org. Chem.* **1999**, *64*, 153–155; i) K. Suzumura, I. Takahashi, H. Matsumoto, K. Nagai, B. Setiawan, R. M. Rantiatmodjo, K. Suzuki, N. Nagano, *Tetrahedron Lett.* **1997**, *38*, 7573–7576.
- [3] M. Huss, H. Wiczorek, *J. Exp. Biol.* **2009**, *212*, 341–346.
- [4] a) R. Ueoka, R. A. Meoded, A. Gran-Scheuch, A. Bhushan, M. W. Fraaije, J. Piel, *Angew. Chem. Int. Ed.* **2020**, *59*, 11698–11698; b) R. Ueoka, R. A. Meoded, A. Gran-Scheuch, A. Bhushan, M. W. Fraaije, J. Piel, *Angew. Chem. Int. Ed.* **2020**, *59*, 7761–7765.
- [5] S. P. Niehs, B. Dose, S. Richter, S. J. Pidot, H. M. Dahse, T. P. Stinear, C. Hertweck, *Angew. Chem. Int. Ed.* **2020**, *59*, 7766–7771.
- [6] a) M. A. Fischbach, C. T. Walsh, *Chem. Rev.* **2006**, *106*, 3468–3496; b) C. Hertweck, *Angew. Chem. Int. Ed.* **2009**, *48*, 4688–716; c) J. Piel, *Nat. Prod. Rep.* **2010**, *27*, 996–1047.
- [7] E. J. N. Helfrich, J. Piel, *Nat. Prod. Rep.* **2016**, *33*, 231–316.



- [8] a) D. E. Cane, C. T. Walsh, *Chem. Biol.* **1999**, *6*, R319–R325; b) H. Wang, D. P. Fewer, L. Holm, L. Rouhiainen, K. Sivonen, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 9259–9264.
- [9] a) E. J. N. Helfrich, R. Ueoka, M. G. Chevrette, F. Hemmerling, X. W. Lu, S. Leopold-Messer, H. A. Minas, A. Y. Burch, S. E. Lindow, J. Piel, M. H. Medema, *Nat. Commun.* **2021**, *12*, 1422–1436; b) A. T. Keatinge-Clay, *Angew. Chem. Int. Ed.* **2017**, *56*, 4658–4660; c) F. Hemmerling, R. A. Meoded, A. E. Fraley, H. A. Minas, C. L. Dieterich, M. Rust, R. Ueoka, K. Jensen, E. J. N. Helfrich, C. Bergande, M. Biedermann, N. Magnus, B. Piechulla, J. Piel, *Angew. Chem. Int. Ed.* **2022**, *61*, e202116614.
- [10] a) T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225–233; b) E. J. N. Helfrich, R. Ueoka, A. Dolev, M. Rust, R. A. Meoded, A. Bhushan, G. Califano, R. Costa, M. Gugger, C. Steinbeck, P. Moreno, J. Piel, *Nat. Chem. Biol.* **2019**, *15*, 813–821.
- [11] M. Pérez-Sayáns, J. M. Somoza-Martín, F. Barros-Angueira, J. M. G. Rey, A. García-García, *Cancer Treat. Rev.* **2009**, *35*, 707–713.
- [12] H. Büttner, S. P. Niehs, K. Vandelannoote, Z. Cseresnyes, B. Dose, I. Richter, R. Gerst, M. T. Figge, T. P. Stinear, S. J. Pidot, C. Hertweck, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2110669118.
- [13] R. D. Süßmuth, A. Mainz, *Angew. Chem. Int. Ed.* **2017**, *56*, 3770–3821.
- [14] M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch, O. Kohlbacher, *Nucleic Acids Res.* **2011**, *39*, W362–W367.
- [15] K. Blin, S. Shaw, A. M. Kloosterman, Z. Charlop-Powers, G. P. van Wezel, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2021**, *49*, W29–W35.
- [16] a) J. Franke, C. Hertweck, *Cell Chem. Biol.* **2016**, *23*, 1179–1192; b) S. Yue, J. S. Duncan, Y. Yamamoto, C. R. Hutchinson, *J. Am. Chem. Soc.* **1987**, *109*, 1253–1255; c) D. E. Cane, C. C. Yang, *J. Am. Chem. Soc.* **1987**, *109*, 1255–1257; d) D. E. Ehmann, J. W. Trauger, T. Stachelhaus, C. T. Walsh, *Chem. Biol.* **2000**, *7*, 765–772.
- [17] a) D. Holtmann, F. Hollmann, *ChemBioChem* **2016**, *17*, 1391–1398; b) S. Eswaramoorthy, J. B. Bonanno, S. K. Burley, S. Swaminathan, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9832–9837; c) E. Romero, M. Fedkenheuer, S. W. Chocklett, J. Qi, M. Oppenheimer, P. Sobrado, *Biochim. Biophys. Acta Proteomics* **2012**, *1824*, 850–857.
- [18] V. Massey, *J. Biol. Chem.* **1994**, *269*, 22459–22462.
- [19] P. W. Riddles, R. L. Blakeley, B. Zerner, *Methods Enzymol.* **1983**, *91*, 49–60.
- [20] M. Boero, T. Ikeshoji, C. C. Liew, K. Terakura, M. Parrinello, *J. Am. Chem. Soc.* **2004**, *126*, 6280–6286.
- [21] a) L. E. N. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber, C. T. Walsh, *Biochemistry* **1998**, *37*, 1585–1595; b) B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, C. Khosla, *Science* **2001**, *291*, 1790–1792.
- [22] K. Belecki, C. A. Townsend, *J. Am. Chem. Soc.* **2013**, *135*, 14339–14348.
- [23] M. G. J. Baud, T. Leiser, P. Haus, S. Samlal, A. C. Wong, R. J. Wood, V. Petrucci, M. Gunaratnam, S. M. Hughes, L. Buluwela, F. Turlais, S. Neidle, F. J. Meyer-Almes, A. J. P. White, M. J. Fuchter, *J. Med. Chem.* **2012**, *55*, 1731–1750.
- [24] C. Kohlhaas, M. Jenner, A. Kampa, G. S. Briggs, J. P. Afonso, J. Piel, N. J. Oldham, *Chem. Sci.* **2013**, *4*, 3212–3217.
- [25] H. Y. He, M. C. Tang, F. Zhang, G. L. Tang, *J. Am. Chem. Soc.* **2014**, *136*, 4488–4491.
- [26] a) W. L. Kelly, C. A. Townsend, *J. Am. Chem. Soc.* **2002**, *124*, 8186–8187; b) O. Sibbesen, B. Koch, B. A. Halkier, B. L. Moller, *J. Biol. Chem.* **1995**, *270*, 3506–3511; c) B. M. Koch, O. Sibbesen, B. A. Halkier, I. Svendsen, B. L. Moller, *Arch. Biochem. Biophys.* **1995**, *323*, 177–186.
- [27] Y. G. Zhu, Q. B. Zhang, S. M. Li, Q. H. Lin, P. Fu, G. T. Zhang, H. B. Zhang, R. Shi, W. M. Zhu, C. S. Zhang, *J. Am. Chem. Soc.* **2013**, *135*, 18750–18753.
- [28] a) A. Becerril, I. Perez-Victoria, S. H. Ye, A. F. Brana, J. Martin, F. Reyes, J. A. Salas, C. Mendez, *ACS Chem. Biol.* **2020**, *15*, 1541–1553; b) I. Garcia, N. M. Vior, A. F. Brana, J. Gonzalez-Sabin, J. Rohr, F. Moris, C. Mendez, J. A. Salas, *Chem. Biol.* **2012**, *19*, 399–413; c) I. Garcia, N. M. Vior, J. Gonzalez-Sabin, A. F. Brana, J. Rohr, F. Moris, C. Mendez, J. A. Salas, *Chem. Biol.* **2013**, *20*, 1022–1032.
- [29] W. J. H. van Berkel, N. M. Kamerbeek, M. W. Fraaije, *J. Biotechnol.* **2006**, *124*, 670–689.
- [30] L. Holm, *Methods Mol. Biol.* **2020**, *2112*, 29–42.
- [31] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohli, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589.
- [32] J. Sucharitakul, M. Prongjit, D. Haltrich, P. Chaiyen, *Biochemistry* **2008**, *47*, 8485–8490.
- [33] A. Alfieri, E. Malito, R. Orru, M. W. Fraaije, A. Mattevi, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 6572–6577.
- [34] C. Kubitzka, A. Faust, M. Gutt, L. Gath, D. Ober, A. J. Scheidig, *Acta Crystallogr. Sect. D* **2018**, *74*, 422–432.
- [35] a) D. P. Ballou, B. Entsch, L. J. Cole, *Biochem. Biophys. Res. Commun.* **2005**, *338*, 590–598; b) B. Entsch, L. J. Cole, D. P. Ballou, *Arch. Biochem. Biophys.* **2005**, *433*, 297–311.
- [36] H. Gaweska, P. F. Fitzpatrick, *Biomol. Concepts* **2011**, *2*, 365–377.
- [37] R. B. Hamed, J. R. Gomez-Castellanos, L. Henry, C. Ducho, M. A. McDonough, C. J. Schofield, *Nat. Prod. Rep.* **2013**, *30*, 21–107.
- [38] Y. V. Guillén Schlippe, L. Hedstrom, *Arch. Biochem. Biophys.* **2005**, *433*, 266–278.
- [39] T. K. Harris, G. J. Turner, *IUBMB Life* **2002**, *53*, 85–98.
- [40] Y. M. Yan, D. Zhang, P. Zhou, B. T. Li, S. Y. Huang, *Nucleic Acids Res.* **2017**, *45*, W365–W373.
- [41] C. M. Kao, R. Pieper, D. E. Cane, C. Khosla, *Biochemistry* **1996**, *35*, 12363–12368.

Manuscript received: March 29, 2023

Accepted manuscript online: May 22, 2023

Version of record online: July 11, 2023

## II. Etudes préliminaires d'enzymes discrètes et domaines BVMO : solubilisation et essais de cristallisation

Les activités BVMO identifiées au sein de modules de PKS présentent des caractéristiques différentes selon les systèmes, incluant leur régiospécificité et l'architecture de leurs modules respectifs. En effet, cette activité est parfois présente directement au sein des modules, comme dans le cas de la PKS *trans*-AT de la lobatamide dont le module LbmC<sub>ox</sub> contient un domaine BVMO, et peut également être apportée en *trans* par une enzyme discrète qui va agir sur un intermédiaire de synthèse lié à un module scindé. Ces *trans*-BVMO sont notamment retrouvées dans les voies de biosynthèse des polycétides oocydine, sesbanimide, toblerol, péderine et mycalazole, et catalysent la formation de fonctions esters par l'insertion d'un atome d'oxygène en position  $\gamma$  de l'unité d'extension nouvellement incorporée. À l'inverse, le domaine LbmC-BVMO génère une fonction ester de structure opposée puisqu'il catalyse une monooxygénation en position  $\beta$  (**Figure 55**). Les différentes compositions en domaines des modules impliqués constituent également une caractéristique intéressante. En plus des domaines KS et ACP, les modules scindés de l'oocydine, sesbanimide et péderine contiennent un domaine DH vraisemblablement inactif entre la KS et le site de césure du module, tandis que ce domaine n'est pas retrouvé dans les modules équivalents du mycalazole et toblerol, qui catalysent pourtant la même réaction d'oxydation. Le module LbmC<sub>ox</sub>, quant à lui, possède un domaine cMT additionnel. Ces multiples particularités ont attiré notre attention et motivé cette étude, visant à déterminer les bases structurales de ces différences mécanistiques.

Les efforts de recherche se sont d'abord focalisés sur la BVMO discrète OocK de la PKS de l'oocydine puisqu'elle a été la première caractérisée fonctionnellement par l'équipe du Pr. Jörn Piel<sup>308</sup> et par conséquent, a été la première dont la séquence codante nous a été fournie. Son homologue SbnP de la PKS de la sesbanimide, ainsi que le domaine LbmC-BVMO ont ensuite également fait l'objet d'expériences d'expression et de purification, suivies de tests de cristallisation.

### 1. Les BVMO discrètes OocK et SbnP

#### a. Criblage des conditions de solubilisation de OocK

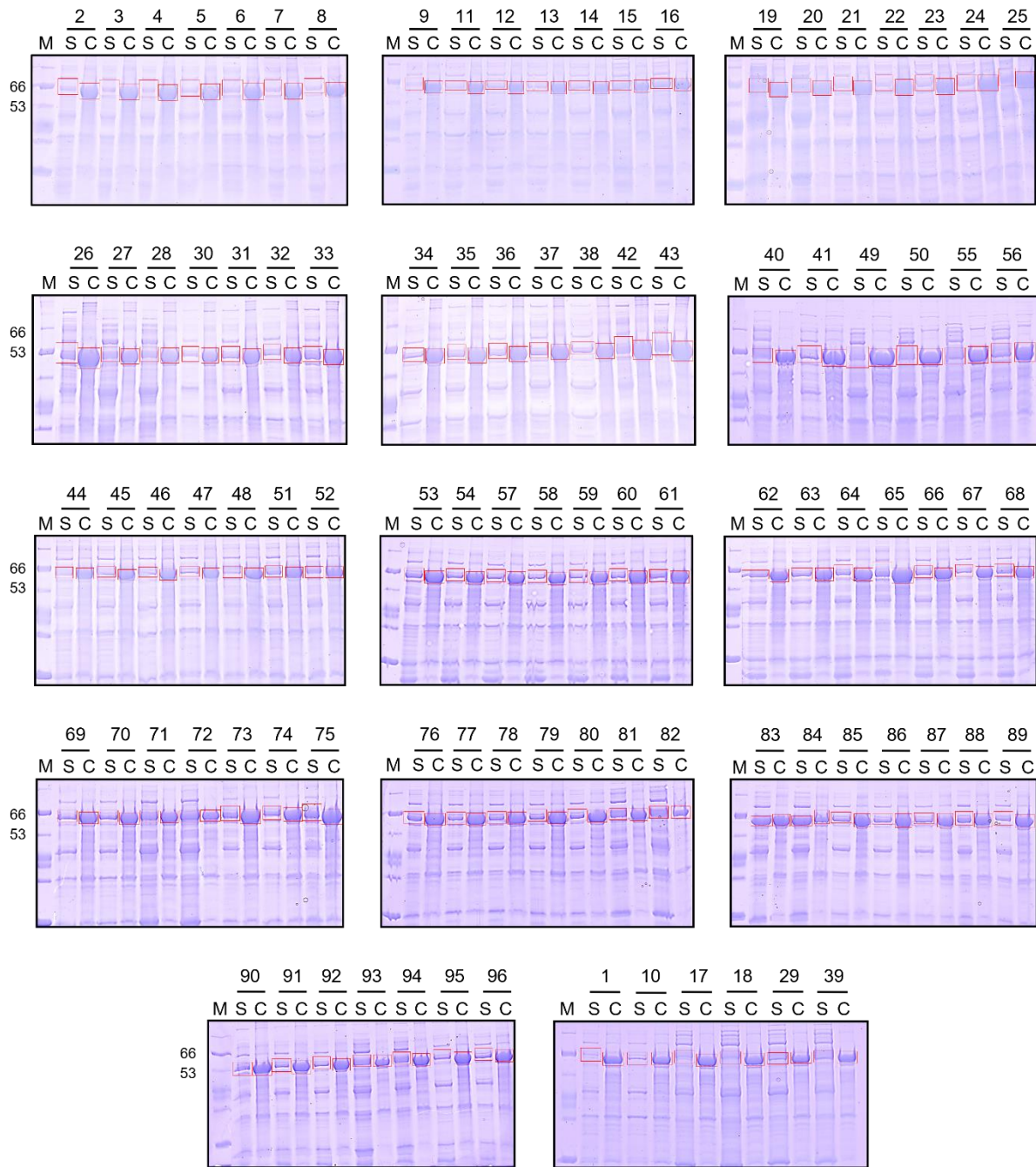
La séquence codant la BVMO discrète OocK a été préalablement sous-clonée en vecteur d'expression pBG102 et transformée en cellules *E. coli* BL21 (DE3) par Sabrina Collin. Lors des premiers tests de purification, les rendements protéiques étaient affectés par une importante précipitation de la BVMO, notamment lors de l'étape de dialyse sur la nuit permettant le clivage de l'étiquette et l'élimination de l'imidazole.

Afin d'envisager des études structurales, il était impératif de pallier ce problème. Pour

obtenir des quantités de protéines compatibles avec des essais de cristallisation, nous avons entrepris d'optimiser la composition de tampon de solubilisation d'OocK par le biais d'un plan factoriel incomplet. Les surnageants et culots d'échantillons obtenus après lyse par sonication dans chacun des 96 tampons du premier plan ont été déposés sur SDS-PAGE et révélés au bleu de Coomassie (**Figure 63**).

La proportion d'OocK soluble a ensuite été estimée en fonction de l'intensité relative des bandes protéiques correspondantes dans les fractions de surnageant et de culot. L'influence de chaque composant du tampon sur la solubilité de la protéine a pu alors être estimée, ainsi que l'impact de la présence simultanée de plusieurs composants (**Tableau 7**). Ainsi, il est apparu que la protéine était plus soluble à un pH élevé (~9.0) et qu'à ce pH, l'ajout de détergents, de diglycine et de L-glutamate était bénéfique. Concernant les sels, une concentration d'environ 400 mM a permis les meilleurs résultats avec un effet positif plus notable du chlorure de sodium à pH 9.0. Réciproquement, en présence d'octyle  $\beta$ -glucoside ( $\beta$ -OG), une forte amélioration de la solubilité d'OocK a été observée lorsque le tampon possédait un pH de 9.0 ou contenait de la diglycine, du L-glutamate ou 400 mM de NaCl, bien que la combinaison de diglycine et 400 mM de NaCl semblait diminuer la solubilité d'environ 45 %. L'ajout de détergent CHAPS a été également bénéfique à pH 9.0 et à 400 mM de NaCl, mais sa présence combinée à celle d'additifs a révélé un impact minime sur la solubilité. Finalement, le réducteur TCEP n'a pas influencé la solubilité d'OocK à pH 9.0 ou combiné au chlorure de sodium, mais a provoqué une légère amélioration en présence de  $\beta$ -OG, diglycine, triglycine et L-glutamate.

## Résultats



**Figure 63 :** SDS-PAGE du premier crible de solubilité de OocK. À chaque paire d'échantillons de surnageant (S) et de culot (C) est associé un tampon dont le numéro correspond à celui attribué dans le **Tableau 3**. Après migration des échantillons et révélation par coloration au bleu de Coomassie, l'intensité de chaque bande correspondant à OocK (délimitée par un rectangle rouge) a été calculée avec le logiciel Mesurim. La quantité de protéine déposée et l'aire des rectangles sont identiques pour chaque couple surnageant-culot afin de permettre une quantification relative des deux fractions. La première piste de chaque gel correspond au marqueur de taille (M) dont les bandes correspondant aux masses 66 et 53 kDa sont indiquées. La protéine SUMO-OocK se situe entre ces deux bandes avec une masse moléculaire de 62 kDa.

**Tableau 7 :** Influence du pH et des composants de tampon sur la solubilité d'OocK. Pour un niveau de variable donné (ex : 200 mM NaCl), un score de solubilité moyen est calculé puis comparé au score moyen des tampons contenant ce niveau en présence d'un autre (ex :  $\beta$ -OG). L'influence sur la solubilité protéique est ensuite évaluée telle que : (- -) = diminution de plus de 50 % par rapport au score moyen ; (-) = diminution entre 50 et 20 % ; (-) = diminution entre 20 et 10 % ; 0 = variation entre -10 et +10 % ; (+) = augmentation entre 10 et 20 % ; (++) = augmentation entre 20 et 50 % ; (+++) = augmentation de plus de 50 %. Les concentrations en sels sont en mM.

Sels	200 NaCl	400 NaCl	600 NaCl	200 HCOONa	400 HCOONa
Scores moyens	28,86	33,03	32,86	31,37	35,49
$\beta$ -OG	37,80	52,18	38,56	30,00	44,02
Effet	(++)	(+++)	(+)	0	(++)
CHAPS	28,32	54,35	37,23	43,34	49,91
Effet	0	(+++)	(+)	(++)	(++)
NDSB-195	22,46	29,93	27,87	28,81	30,32
Effet	(- -)	0	(-)	0	(-)
NDSB-201	27,75	21,86	34,10	32,84	32,08
Effet	0	(- -)	0	0	0
pH 6.0	24,14	19,61	31,42	36,67	30,01
Effet	(-)	(- -)	0	(+)	0
pH 7.0	28,04	30,49	35,42	30,83	33,24
Effet	0	0	0	0	0
pH 8.0	34,34	28,74	32,13	28,81	41,54
Effet	(+)	(-)	0	0	(+)
pH 9.0	28,91	53,27	32,46	29,17	37,17
Effet	0	(+++)	0	0	0
Diglycine	35,60	18,26	37,45	24,65	41,60
Effet	(++)	(- -)	(+)	(- -)	(+)
Triglycine	27,18	37,15	31,34	44,36	33,54
Effet	0	(+)	0	(++)	0
Saccharose	32,38	25,02	37,19	26,94	33,50
Effet	(+)	(- -)	(+)	(-)	0
L-Glu	26,40	43,16	24,40	42,31	31,66
Effet	0	(++)	(- -)	(++)	(-)
L-Arg	21,03	16,46	33,05	32,98	35,60
Effet	(- -)	(- -)	0	0	0
Glycérol	34,14	28,82	37,27	30,16	34,26
Effet	(+)	(-)	(+)	0	0
TCEP	27,75	30,33	31,28	32,22	39,53
Effet	0	0	0	0	(+)

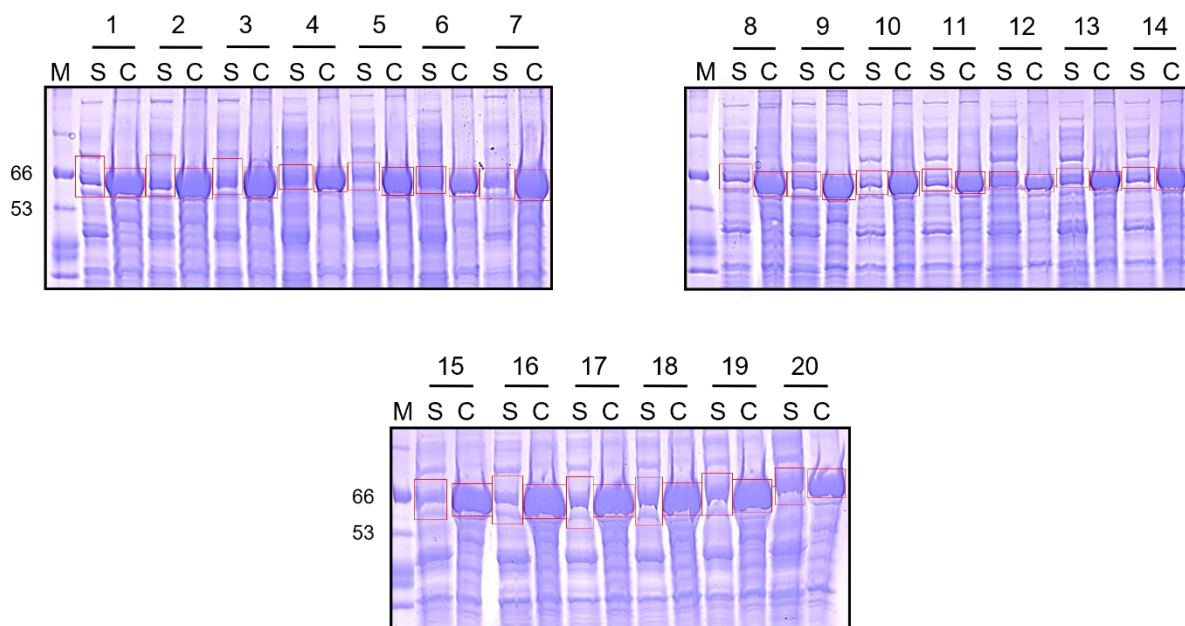
Détergents et NDSB	$\beta$ -OG	CHAPS	NDSB-195	NDSB-201
Scores moyens	38,60	42,44	26,08	30,67
200 NaCl	37,80	28,32	22,46	27,75
Effet	0	(- -)	(-)	0
400 NaCl	52,18	54,35	29,93	21,86
Effet	(++)	(++)	(+)	(- -)
600 NaCl	38,56	37,23	27,87	34,10
Effet	0	(-)	0	(+)
200 HCOONa	30,00	43,34	28,81	32,84
Effet	(- -)	0	(+)	0
400 HCOONa	44,02	49,91	30,32	32,08
Effet	(+)	(+)	(+)	0
pH 6.0	30,00	40,28	19,76	26,97
Effet	(- -)	0	(- -)	(-)
pH 7.0	38,56	34,92	29,37	30,30
Effet	0	(-)	(+)	0
pH 8.0	33,41	49,91	27,87	35,38
Effet	(-)	(+)	0	(+)
pH 9.0	48,10	54,35	30,32	34,10
Effet	(++)	(++)	(+)	(+)
Diglycine	50,00	35,60	24,65	34,43
Effet	(++)	(-)	0	(+)
Triglycine	44,09	44,36	31,34	29,39
Effet	(+)	0	(++)	0
Saccharose	36,87	37,19	31,98	18,06
Effet	0	(-)	(++)	(- -)
L-Glu	60,28	42,31	24,40	26,12
Effet	(+++)	0	0	(-)
L-Arg	38,04	21,03	32,98	33,77
Effet	0	(- -)	(++)	(+)
Glycérol	38,72	37,27	27,89	25,65
Effet	0	(-)	0	(-)
TCEP	41,74	40,52	25,77	28,33
Effet	0	0	0	0

pH	6.0	7.0	8.0	9.0
Scores moyens	26,71	32,51	32,96	35,59
200 NaCl	24,14	28,04	34,34	28,91
Effet	0	(-)	0	(-)
400 NaCl	19,61	30,49	28,74	53,27
Effet	(- -)	0	(-)	(++)
600 NaCl	31,42	35,42	32,13	32,46
Effet	(+)	0	0	0
200 HCOONa	36,67	30,83	28,81	29,17
Effet	(++)	0	(-)	(-)
400 HCOONa	30,01	33,24	41,54	37,17
Effet	(+)	0	(++)	(+)
$\beta$ -OG	30,00	38,56	33,41	48,10
Effet	(+)	(+)	0	(++)
CHAPS	40,28	34,92	49,91	54,35
Effet	(++)	0	(++)	(+++)
NDSB-195	19,76	29,37	27,87	30,32
Effet	(- -)	0	(-)	(-)
NDSB-201	26,97	30,30	35,38	34,10
Effet	0	0	0	0
Diglycine	16,49	30,13	36,84	42,21
Effet	(- -)	0	(+)	(+)
Triglycine	34,67	31,47	30,78	38,04
Effet	(++)	0	0	(+)
Saccharose	27,63	32,74	31,91	32,82
Effet	0	0	0	0
L-Glu	34,49	28,89	25,23	46,09
Effet	(++)	(-)	(- -)	(++)
L-Arg	20,56	27,01	32,74	35,91
Effet	(- -)	(-)	0	(+)
Glycérol	31,46	31,07	34,44	29,09
Effet	(+)	0	0	(-)
TCEP	25,18	30,79	34,72	35,73
Effet	0	0	0	0

Additifs	Diglycine	Triglycine	Saccharose	L-Glu	L-Arg	Glycérol
Scores moyens	31,42	33,74	31,27	33,67	29,05	31,52
pH 6.0	16,49	34,67	27,63	34,49	20,56	31,46
Effet	(- -)	0	(-)	0	(- -)	0
pH 7.0	30,13	31,47	32,74	28,89	27,01	31,07
Effet	0	0	0	(-)	0	0
pH 8.0	36,84	30,78	31,91	25,23	32,74	34,44
Effet	(+)	0	0	(- -)	(+)	0
pH 9.0	42,21	38,04	32,82	46,09	35,91	29,09
Effet	(++)	(+)	0	(++)	(++)	0
200 NaCl	35,60	27,18	32,38	26,40	21,03	34,14
Effet	(+)	(-)	0	(- -)	(- -)	0
400 NaCl	18,26	37,15	25,02	43,16	16,46	28,82
Effet	(- -)	(+)	(-)	(++)	(- -)	0
600 NaCl	37,45	31,34	37,19	24,40	33,05	37,27
Effet	(+)	0	(+)	(- -)	(+)	(+)
200 HCOONa	24,65	44,36	26,94	42,31	32,98	30,16
Effet	(- -)	(++)	(-)	(++)	(+)	0
400 HCOONa	41,60	33,54	33,50	31,66	35,60	34,26
Effet	(++)	0	0	0	(++)	0
$\beta$ -OG	50,00	44,09	36,87	60,28	38,04	38,72
Effet	(+++)	(++)	(+)	(+++)	(++)	(++)
CHAPS	35,60	44,36	37,19	42,31	21,03	37,27
Effet	(+)	(++)	(+)	(++)	(- -)	(+)
NDSB-195	24,65	31,34	31,98	24,40	32,98	27,89
Effet	(- -)	0	0	(- -)	(+)	(-)
NDSB-201	34,43	29,39	18,06	26,12	33,77	25,65
Effet	0	(-)	(- -)	(- -)	(+)	(-)
TCEP	36,09	35,71	27,62	38,84	26,96	29,49
Effet	(+)	0	(-)	(+)	0	0

## Résultats

A partir de ces résultats, un second plan factoriel incomplet a été mis en place, se focalisant sur une gamme de pH et un nombre de sels et d'additifs moindres. Ainsi, la gamme de pH a été restreinte entre 8.5 et 9.5, avec le chlorure de sodium pour unique sel, potentiellement combiné au  $\beta$ -OG, à la diglycine et au L-glutamate. Les fractions de surnageant et de culot après sonication des échantillons ont également été déposées sur SDS-PAGE pour évaluer la solubilité d'OocK (**Figure 64**). Avec ce deuxième crible, les résultats les plus probants ont été obtenus à pH 9.0 et en présence de 400 mM de NaCl. L'ajout de 100 mM de diglycine et de L-glutamate semblait également améliorer la solubilité. La contribution la plus significative à l'amélioration de la solubilité a été apportée par 50 mM de  $\beta$ -OG puisque ce niveau de variable possède le plus haut score moyen et la présence concomitante d'autres composés modifie peu ce score. À l'inverse, comme attendu, les tampons ne contenant pas de sels (niveau 0 mM NaCl) ont obtenu un des plus faibles scores de solubilité et l'ajout de détergents et d'additifs à ces tampons a eu peu d'effet. De plus, contrairement aux observations du premier crible, OocK s'est avérée plus soluble en présence simultanée de 100 mM de diglycine et 400 mM de NaCl (**Tableau 8**).



**Figure 64 :** SDS-PAGE du second crible de solubilité de OocK. À chaque paire d'échantillons de surnageant (S) et de culot (C) est associé un tampon dont le numéro correspond à celui attribué dans le **Tableau 5**. Après migration des échantillons et révélation par coloration au bleu de Coomassie, l'intensité de chaque bande correspondant à OocK (délimitée par un rectangle rouge) a été calculée avec le logiciel Mesurim. L'aire des rectangles est identique pour chaque couple surnageant-culot afin de permettre une quantification relative des deux fractions. La première piste de chaque gel correspond au marqueur de taille (M) dont les bandes indiquant les masses 66 et 53 kDa sont indiquées. La protéine SUMO-OocK se situe entre ces deux bandes avec une masse moléculaire d'environ 62 kDa.

**Tableau 8 :** Influence du pH et des composants de tampon du second crible sur la solubilité d'Oock. Pour un niveau de variable donné (ex : 300 mM NaCl), un score de solubilité moyen est calculé puis comparé au score moyen des tampons contenant ce niveau en présence d'un autre (ex : 50 mM  $\beta$ -OG). L'influence sur la solubilité protéique est ensuite évaluée telle que : (-) = diminution entre 20 et 10 % par rapport au score moyen ; 0 = variation entre -10 et +10 % ; (+) = augmentation entre 10 et 20 % ; (++) = augmentation entre 20 et 50 % ; (+++) = augmentation de plus de 50 %.

pH	8.5	9.0	9.5
Scores moyens	37,41	38,04	36,98
0 mM NaCl	33,17	38,06	37,21
Effet	(-)	0	0
300 mM NaCl	38,87	33,20	32,39
Effet	0	(-)	(-)
400 mM NaCl	40,87	51,59	35,56
Effet	0	(++)	0
500 mM NaCl	36,08	36,08	40,58
Effet	0	0	0
25 mM $\beta$ -OG	34,05	37,79	34,80
Effet	0	0	0
50 mM $\beta$ -OG	45,69	45,52	44,39
Effet	(++)	(+)	(++)
50 mM diglycine	34,80	38,62	36,91
Effet	0	0	0
100 mM diglycine	43,38	42,98	36,99
Effet	(+)	(+)	0
50 mM L-Glu	38,87	37,24	34,20
Effet	0	0	0
100 mM L-Glu	35,24	43,46	40,80
Effet	0	(+)	(+)

NaCl	0	300 mM	400 mM	500 mM
Scores moyens	35,94	35,30	40,89	37,88
pH 8.5	33,17	38,87	40,87	36,08
Effet	0	(+)	0	0
pH 9.0	38,06	33,20	51,59	36,08
Effet	0	0	(++)	0
pH 9.5	37,21	32,39	35,56	40,58
Effet	0	0	(-)	0
25 mM $\beta$ -OG	37,21	33,37	33,73	37,79
Effet	0	0	(-)	0
50 mM $\beta$ -OG	39,44	43,38	49,80	44,39
Effet	0	(++)	(+++)	(+)
50 mM diglycine	37,34	34,36	36,91	37,79
Effet	0	0	0	0
100 mM diglycine	37,21	43,38	51,59	35,57
Effet	0	(++)	(+++)	0
50 mM L-Glu	36,68	38,87	34,20	37,79
Effet	0	(+)	(-)	0
100 mM L-Glu	36,22	35,32	51,59	44,39
Effet	0	0	(+++)	(+)

$\beta$ -OG	25 mM	50 mM
Scores moyens	35,10	45,36
0 mM NaCl	37,21	39,44
Effet	0	(-)
300 mM NaCl	33,37	43,38
Effet	0	0
400 mM NaCl	33,73	49,80
Effet	0	0
500 mM NaCl	37,79	44,39
Effet	0	0
pH 8.5	34,05	45,69
Effet	0	0
pH 9.0	37,79	45,52
Effet	0	0
pH 9.5	34,8	44,39
Effet	0	0
50 mM diglycine	36,08	39,44
Effet	0	(-)
100 mM diglycine	37,21	47,48
Effet	0	0
50 mM L-Glu	36,08	43,38
Effet	0	0
100 mM L-Glu	37,21	47,99
Effet	0	0

Diglycine	50 mM	100 mM
Scores moyens	36,75	40,66
0 mM NaCl	37,34	37,21
Effet	0	0
300 mM NaCl	34,36	43,38
Effet	0	0
400 mM NaCl	36,91	51,59
Effet	0	(++)
500 mM NaCl	37,79	35,57
Effet	0	0
pH 8.5	34,80	43,38
Effet	0	0
pH 9.0	38,62	42,98
Effet	0	0
pH 9.5	36,91	36,99
Effet	0	0
25 mM $\beta$ -OG	36,08	37,21
Effet	0	0
50 mM $\beta$ -OG	39,44	47,48
Effet	0	(+)
50 mM L-Glu	36,08	43,38
Effet	0	0
100 mM L-Glu	35,24	37,21
Effet	0	0

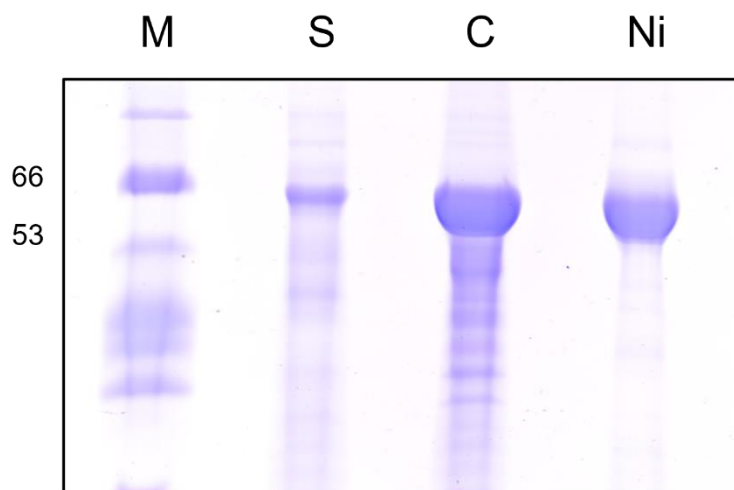
  

L-Glu	50 mM	100 mM
Scores moyens	37,28	40,75
0 mM NaCl	36,68	36,22
Effet	0	(-)
300 mM NaCl	38,87	35,32
Effet	0	(-)
400 mM NaCl	34,20	51,59
Effet	0	(++)
500 mM NaCl	37,79	44,39
Effet	0	0
pH 8.5	38,87	35,24
Effet	0	(-)
pH 9.0	37,24	43,46
Effet	0	0
pH 9.5	34,2	40,8
Effet	0	0
25 mM $\beta$ -OG	36,08	37,21
Effet	0	0
50 mM $\beta$ -OG	43,38	47,99
Effet	(+)	(+)
50 mM diglycine	36,08	35,24
Effet	0	(-)
100 mM diglycine	43,38	37,21
Effet	(+)	0

Sur la base de ces deux cribles, un tampon optimal de solubilisation de la BVMO Oock a été défini selon la composition suivante : 30 mM CAPSO pH 9.0, 400 mM NaCl, 50 mM  $\beta$ -OG, 100 mM diglycine, 100 mM L-glutamate.

### b. Purification des BVMO discrètes

Des tests de purification de OocK ont été ensuite réalisés afin de vérifier que le tampon optimal créé permettait bien d'améliorer la solubilité d'OocK. En effet, après lyse par sonication et filtration du surnageant de centrifugation, des quantités satisfaisantes de BVMO ont été adsorbées sur la résine Ni-sépharose puis éluées par un palier d'imidazole. Cependant, la majorité de la protéine précipitait de nouveau lors de l'étape limitante de dialyse sur la nuit. OocK ne semblait donc pas assez stable en solution pour permettre un protocole de purification sur deux jours malgré l'utilisation d'un tampon optimisé. Il a donc été décidé d'adopter un nouveau protocole limitant au maximum le temps de manipulation de la protéine en diminuant le nombre d'étapes de chromatographie. Ainsi, une unique chromatographie d'affinité nickel a été réalisée, à l'issue de laquelle la pureté de l'échantillon était déjà suffisante pour envisager des essais de cristallisation (**Figure 65**). L'étape de dialyse a été supprimée afin d'éviter un long temps d'incubation à 4°C qui provoquait inévitablement la précipitation de la protéine. Par conséquent, OocK a été purifiée en conservant son étiquette N-terminale His<sub>6</sub>-SUMO.

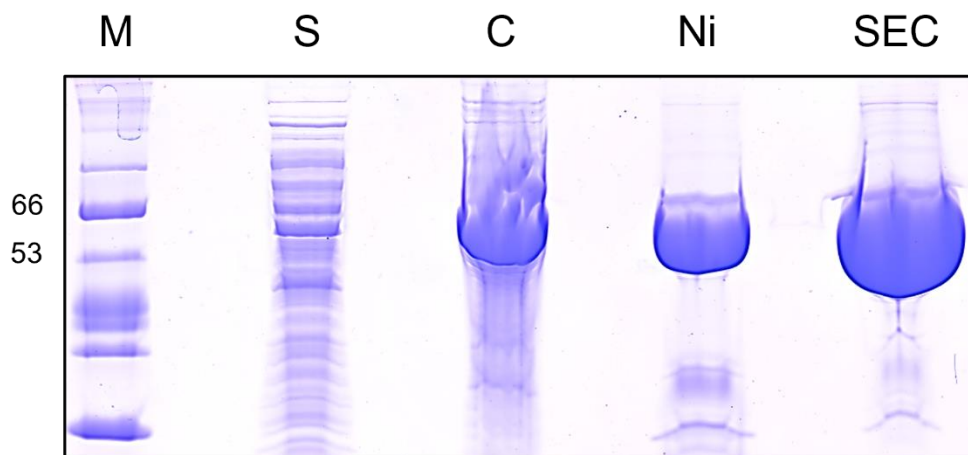


**Figure 65 :** SDS-PAGE bilan de purification de SUMO-OocK. La pureté de l'échantillon a été évaluée dans le surnageant (S) et le culot (C) de centrifugation après lyse des cellules et dans l'éluat de colonne de chromatographie d'affinité nickel (Ni). Le protocole mis en place a permis la purification de la BVMO avec un rendement de 7,8 mg par litre de culture. Les bandes du marqueur de taille (M) correspondant aux masses de 53 et 66 kDa sont indiquées. SUMO-OocK possède une masse moléculaire de 62 kDa.

Parallèlement à ces tests, la séquence codant la BVMO discrète SbnP de la PKS *trans*-AT de la sesbanimide nous a été envoyée par l'équipe du Pr. Jörn Piel, clonée dans le vecteur d'expression pET28b-SUMO. Les essais de purification ont rapidement montré que, tout comme son homologue OocK de la PKS de l'oocydine, SbnP était relativement instable en solution et précipitait lors d'incubations prolongées à 4°C. Nous avons donc également



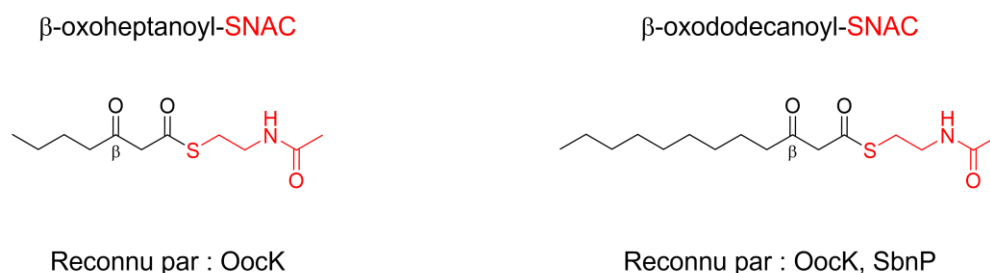
appliqué un protocole de purification plus court pour SbnP, sans clivage de l'étiquette N-terminale. Une deuxième étape de chromatographie par exclusion de taille a cependant été ajoutée pour affiner la pureté de l'échantillon (**Figure 66**).



**Figure 66 :** SDS-PAGE bilan de purification de SUMO-SbnP. La pureté de l'échantillon a été évaluée dans le surnageant (S) et le culot (C) de centrifugation après lyse des cellules et dans l'éluat des colonnes de chromatographie d'affinité nickel (Ni) et d'exclusion de taille (SEC). Le protocole mis en place a permis la purification de la BVMO, bien que des impuretés subsistent. Les rendements obtenus sont de 11,61 mg de BVMO par litre de culture. Les bandes du marqueur de taille (M) correspondant aux masses de 53 et 66 kDa sont indiquées. SUMO-SbnP possède une masse moléculaire de 63 kDa.

### c. Essais de cristallisation des BVMO discrètes

Les BVMO discrètes SbnP et Oock ont été soumises à des tests de cristallisation à l'aide de cribles commerciaux pour tenter à terme d'en résoudre la structure à haute résolution. L'importante instabilité de ces FMO ayant été constatée, nous avons essayé de rigidifier leurs structures en les co-cristallisant en présence de molécules biologiquement pertinentes. Les deux échantillons de protéines présentaient une couleur jaune en fin de purification, caractéristique de leur cofacteur flavinique, ce qui indiquait que celui-ci avait été co-purifié. Leur cofacteur NADP (identifié lors des tests d'activité *in vitro*<sup>308</sup>) a été ajouté à l'échantillon de cristallisation à 2 mM sous sa forme oxydée NADP<sup>+</sup> afin d'empêcher le transfert d'hydrure. De plus, deux analogues de substrats synthétisés par l'équipe du Pr. Jörn Piel, nous avaient été fournis afin de potentiellement observer leurs modes de fixation dans le site actif des BVMO. Les deux analogues sont des dérivés SNAC de chaînes carbonées plus ou moins longues (7 et 12 carbones à partir du carbonyle thioester), comportant une fonction  $\beta$ -cétone nécessaire à l'activité BVMO<sup>308</sup> (**Figure 67**). La reconnaissance et l'insertion d'oxygène avait été préalablement démontrées avec les deux analogues pour Oock<sup>308</sup> et avec l'analogue à plus longue chaîne pour SbnP (expérience réalisée par le Dr. Amy E. Fraley, données non publiées).



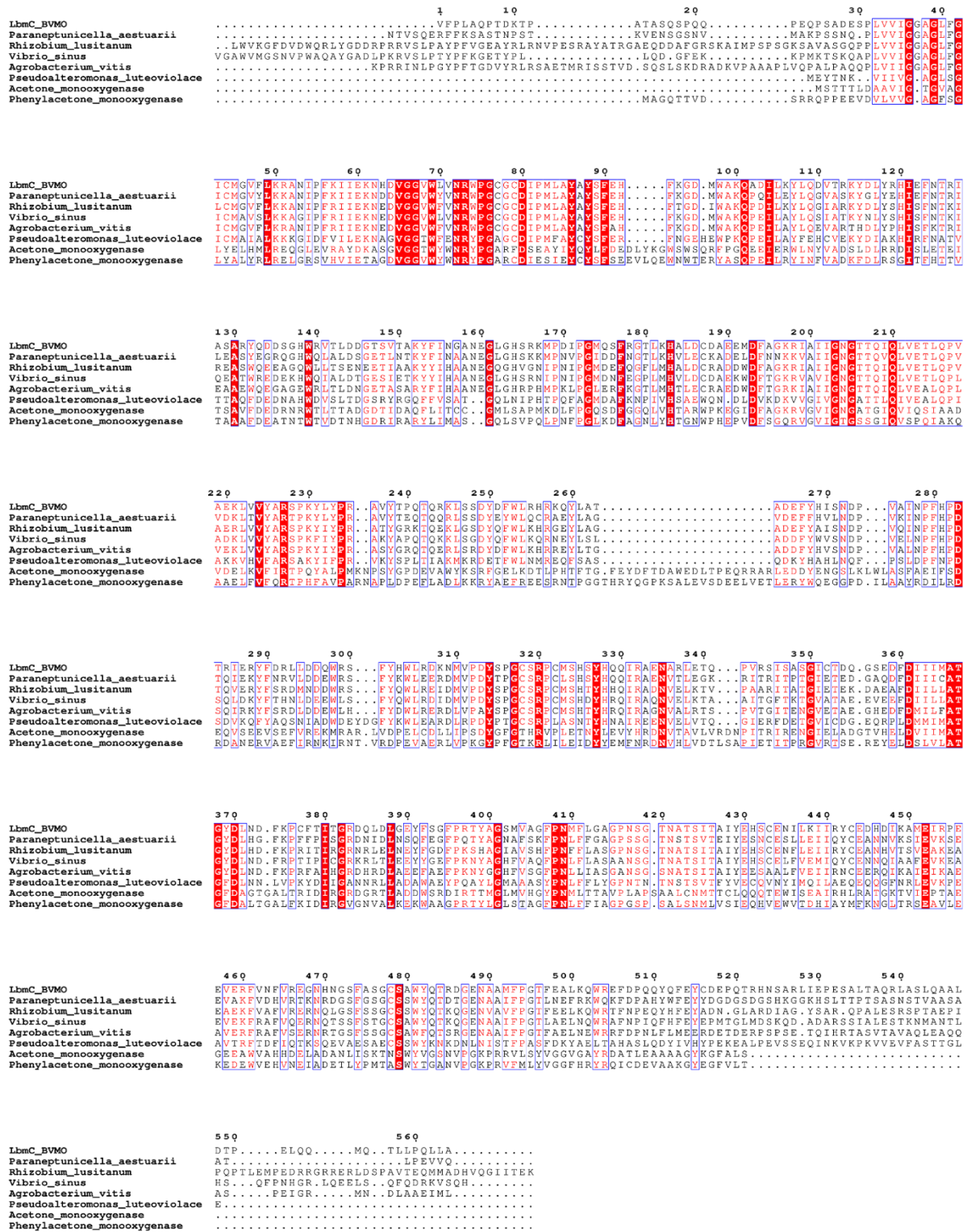
**Figure 67 :** Analogues de substrat de OocK et SbnP utilisés à 1 mM lors des essais de co-cristallisation. La fonction cétone reconnue par les BVMO est indiquée en position  $\beta$ , et la partie SNAC des molécules mimant l'extrémité du groupement Ppant est représentée en rouge.

À ce jour, aucun cristal de BVMO n'a encore été obtenu, en absence ou en présence de cofacteur et d'analogue de substrat. Il est fortement probable que l'instabilité de ces BVMO discrètes soit la principale cause de cette difficulté à cristalliser, couplée à une pureté de l'échantillon protéique qui nécessiterait d'être optimisée pour éliminer les contaminants observés sur gel d'acrylamide.

## 2. Le domaine LbmC-BVMO

### a. Analyses bioinformatiques

À la différence des enzymes discrètes OocK et SbnP, l'activité BVMO de la PKS *trans*-AT de la lobatamide est présente au sein d'une de ses sous-unités, portée par un domaine du module LbmC<sub>ox</sub>. Par conséquent, il a d'abord fallu identifier les bornes de ce domaine LbmC-BVMO afin de cloner sa séquence codante et espérer l'exprimer de manière stable en solution, sans omettre ou interrompre de structures secondaires. Pour ce faire, le BGC *lbm* codant la PKS de la lobatamide a été analysé à l'aide du serveur de détection et d'annotation de clusters biosynthétiques antiSMASH<sup>117</sup>. La séquence protéique comprise entre l'adaptateur KS-AT en amont et la cMT en aval du domaine BVMO a été extraite et utilisée comme point de départ d'une recherche de protéines homologues par BLASTP<sup>345</sup>. Un alignement de séquences des homologues les plus proches a été réalisé avec le serveur ClustalW<sup>346</sup> et mis en forme avec le programme ESPript<sup>347</sup>, en incluant également la séquence des BVMO acétone et phénylacétone monooxygénases (**Figure 68**).

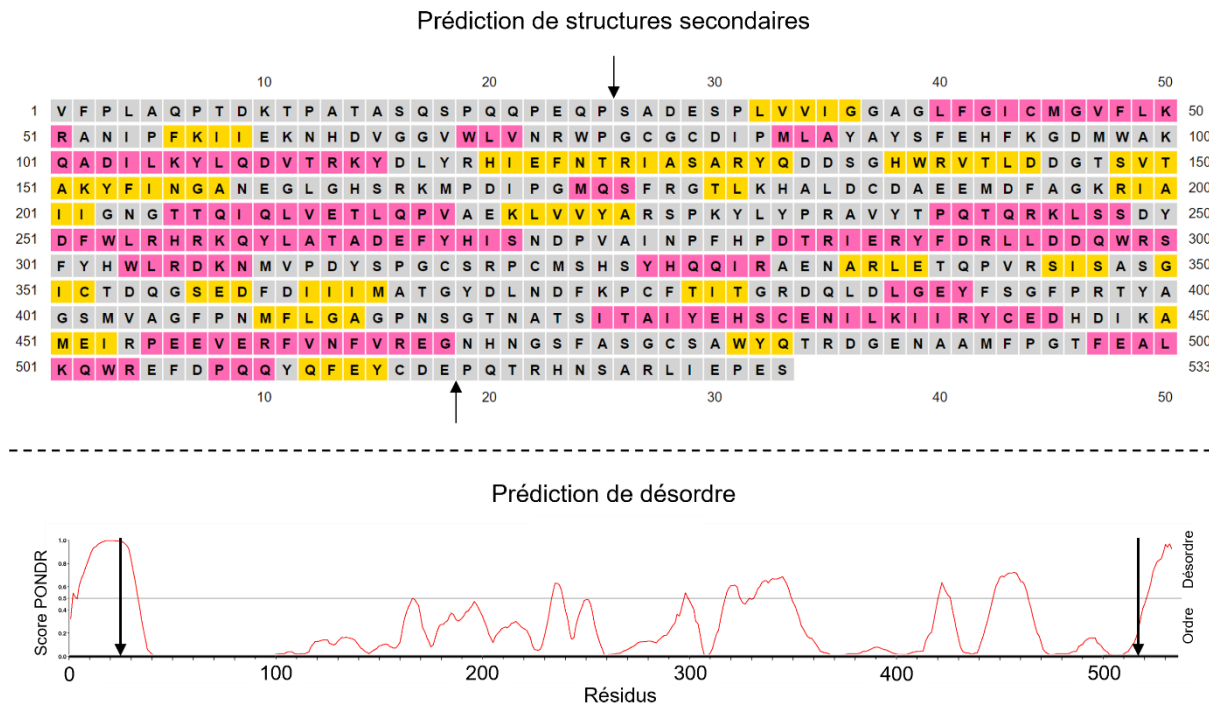


**Figure 68 :** Alignement de séquences de protéines homologues à LbmC-BVMO. La numérotation des positions alignées est basée sur la séquence de LbmC-BVMO. Les positions conservées entre au moins 50 % des séquences sont encadrées, et celles strictement identiques entre toutes les séquences sont surlignées en rouge.

Une région d'intérêt contenant le domaine LbmC-BVMO a ainsi été déterminée et soumise à une prédiction de structures secondaires et de désordre, à l'aide de serveurs

## Résultats

PSIPRED<sup>348</sup> et PONDR<sup>349</sup>, respectivement (**Figure 69**). Les informations obtenues nous ont alors permis de définir les bornes N- et C-terminales optimales du domaine BVMO, permettant de l'exprimer sous forme soluble et stable en solution.



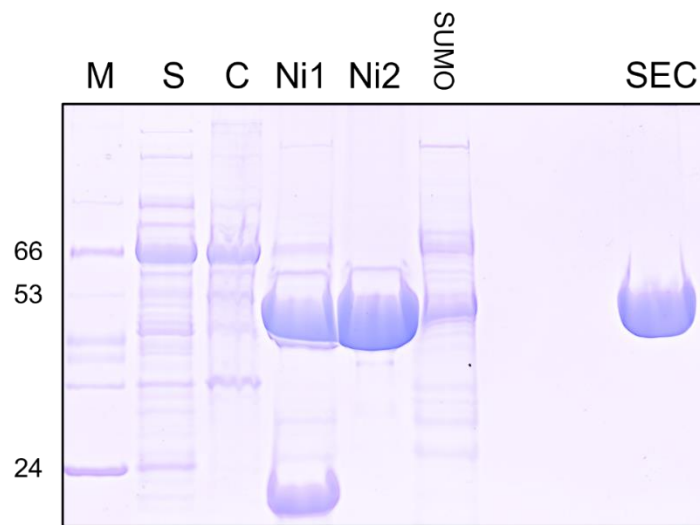
**Figure 69** : Prédictions de structures secondaires et de désordre du domaine LbmC-BVMO, calculées par PSIPRED et PONDR, respectivement. Les acides aminés indiqués en rose sont prédits au sein d'hélices  $\alpha$  et ceux en jaune sont prédits au sein de brins  $\beta$ . Quelques régions désordonnées semblent présentes au sein de la protéine, ainsi qu'à ses deux extrémités, ce qui coïncide avec l'absence de structures secondaires prédites en N- et C-terminal. Les flèches indiquent les bornes du domaine LbmC-BVMO finalement définies.

### b. Purification de LbmC-BVMO

La séquence codant LbmC-BVMO a été clonée en vecteur pBG102 selon les bornes du domaine définies précédemment, et la protéine a été purifiée selon le protocole détaillé en section IV. de la partie Matériels et Méthodes. Tout comme dans le cas des BVMO discrètes OocK et SbnP, l'échantillon après la première étape de chromatographie d'affinité présentait une couleur jaune intense, ce qui indiquait que l'enzyme avait été co-purifiée avec son cofacteur flavinique. Le maintien de ce cofacteur au sein de LbmC-BVMO, OocK et SbnP, couplé à leur activité d'oxydation Baeyer-Villiger, concorde avec le mécanisme des FMO de classe B<sup>303</sup>.

L'étiquette His<sub>6</sub>-SUMO a été clivée lors de l'étape de dialyse par ajout de protéase HRV 3C, puis, la protéine a été soumise à une seconde chromatographie d'affinité nickel et une étape finale de SEC. Contrairement à ses homologues discrets, aucune précipitation n'a entravé la purification de LbmC-BVMO, ce qui atteste de sa stabilité et par conséquent, de la

pertinence des extrémités du domaine sélectionnées. Un rendement de 26,2 mg de protéine par litre de culture a été obtenu et un SDS-PAGE a permis d'observer la pureté optimale de la protéine (**Figure 70**).



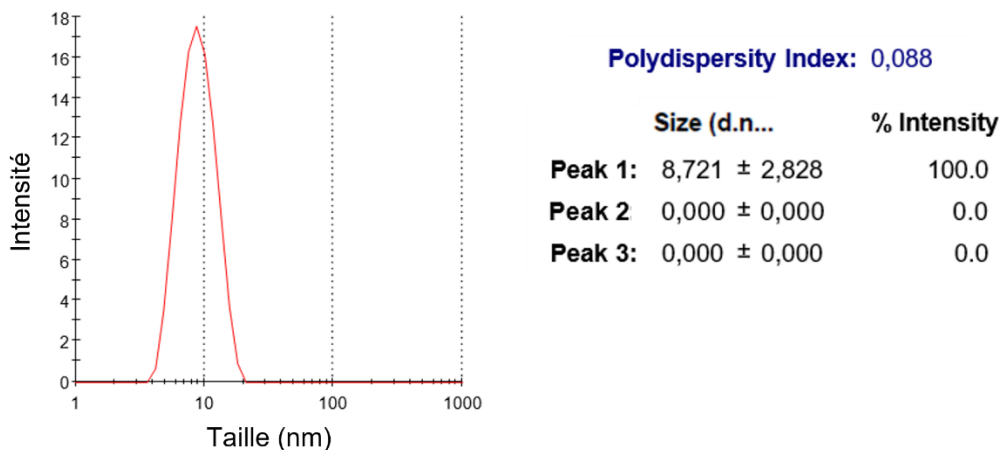
**Figure 70 :** SDS-PAGE de bilan de purification de LbmC-BVMO. La pureté de l'échantillon a été évaluée dans le surnageant (S) et le culot (C) de centrifugation après lyse des cellules et dans les différents éluats de chromatographie : Ni1 = première chromatographie d'affinité nickel et dialyse ; Ni2 = seconde chromatographie d'affinité nickel ; SUMO = étiquette His<sub>6</sub>-SUMO retenue sur chromatographie d'affinité nickel ; SEC = chromatographie d'exclusion de taille. Les bandes du marqueur de taille (M) correspondant aux masses de 24, 53 et 66 kDa sont indiquées. LbmC-BVMO et l'étiquette His<sub>6</sub>-SUMO possèdent des masses moléculaires respectives de 56 et 12 kDa.

### c. Homogénéité en solution

Avant de procéder à des tests de cristallisation, l'état d'agrégation de LbmC-BVMO a été évalué par une expérience de DLS. Les mesures de diffusion ont été réalisées à 6,5 mg/mL, ce qui correspond à la concentration protéique obtenue en fin de purification. Les données enregistrées ont révélé la présence d'une seule population de particules responsable de la totalité de l'intensité diffusée par l'échantillon ainsi qu'un faible indice de polydispersité de 8,8 % (**Figure 71**), ce qui atteste à la fois de la pureté de la protéine et de sa stabilité en solution.

Un diamètre hydrodynamique de  $87,21 \text{ \AA} \pm 28,28$  a été calculé. Cette taille est cohérente avec la masse moléculaire de 56 kDa de LbmC-BVMO et avec la nature monomérique des BVMO caractérisées.

## Résultats



**Figure 71:** Résultats de DLS de LbmC-BVMO. La présence d'un seul pic et le faible indice de polydispersité traduisent la pureté et l'homogénéité de la protéine. Le diamètre hydrodynamique est indiqué en nm.

### d. Essais de cristallisation de LbmC-BVMO

Le domaine LbmC-BVMO a été soumis à des essais de cristallisation en plaques 96 puits afin de cribler un maximum de solutions précipitantes pouvant potentiellement conduire à la formation de cristaux de protéines. Tout comme pour les BVMO discrètes OocK et SbnP, la pertinence des analogues de substrats  $\beta$ -oxoheptanoyl-SNAC et  $\beta$ -oxododecanoyl-SNAC (**Figure 67**) avait été préalablement démontrée par des tests d'activité *in vitro* réalisés par les Dr Reiko Ueoka et Roy A. Meoded<sup>307</sup> et le Dr Amy E. Fraley (données non publiées) de l'équipe du Pr. Piel. Des essais de co-cristallisation ont donc été réalisés avec ces dérivés SNAC ainsi que le donneur d'électron NADP sous forme oxydée pour empêcher la catalyse enzymatique d'avoir lieu.

À l'instar de ses homologues discrets, aucun cristal de LbmC-BVMO n'a pour l'instant été obtenu, en présence ou en absence des cofacteurs et analogues de substrats, malgré une stabilité bien meilleure que celle de OocK et SbnP. Néanmoins, de nombreuses constructions incluant le domaine LbmC-BVMO peuvent être envisagées. En effet, dans son contexte natif, ce domaine ne représente qu'une petite partie d'une protéine de haut poids moléculaire, la sous-unité LbmC. Par conséquent, la production et la purification de constructions alternatives incorporant les domaines N- et C-terminaux de la BVMO pourraient conduire à la cristallisation de la protéine. Cette stratégie s'est révélée payante dans le cas de la FMO LbmA-Ox dont la cristallisation a été permise en la produisant avec son domaine ACP C-terminal.

### III. L' $\alpha$ -hydroxylase discrète OocM : une autre classe de FMO au sein des modules de PKS *trans*-AT

La voie de biosynthèse des polycétides oocydines est notable par son abondance d'activités inhabituelles au sein des PKS. Les investigations menées par Hemmerling *et al.* sur

cette PKS *trans*-AT ont révélé l'origine biosynthétique de l'atome d'oxygène en position C14 du polycétide, intégré à un groupement tétrahydrofurane<sup>311</sup>. L' $\alpha$ -hydroxylase discrète OocM est responsable de l'insertion de cet oxygène par monooxygénation en position  $\alpha$  de l'intermédiaire de synthèse au niveau de cinquième module de la mégasynthase. Des tests d'activité *in vitro* ont également démontré que la réaction d'hydroxylation précédait la kétoréduction catalysée par le domaine KR du même module puisque qu'un analogue de substrat  $\beta$ -hydroxyl-SNAC n'a pas été reconnu par OocM, au contraire de son équivalent oxydé  $\beta$ -két-SNAC. Ces travaux menés par l'équipe du Pr. Jörn Piel fournissent les premières informations fonctionnelles sur cette  $\alpha$ -hydroxylase discrète et prédisent la présence d'enzymes homologues au sein de nombreuses autres PKS *trans*-AT<sup>311</sup>. Ainsi, nous avons entrepris d'étudier structurellement OocM afin d'élucider son mécanisme catalytique peu commun et de comprendre son mode de recrutement par ses domaines partenaires.

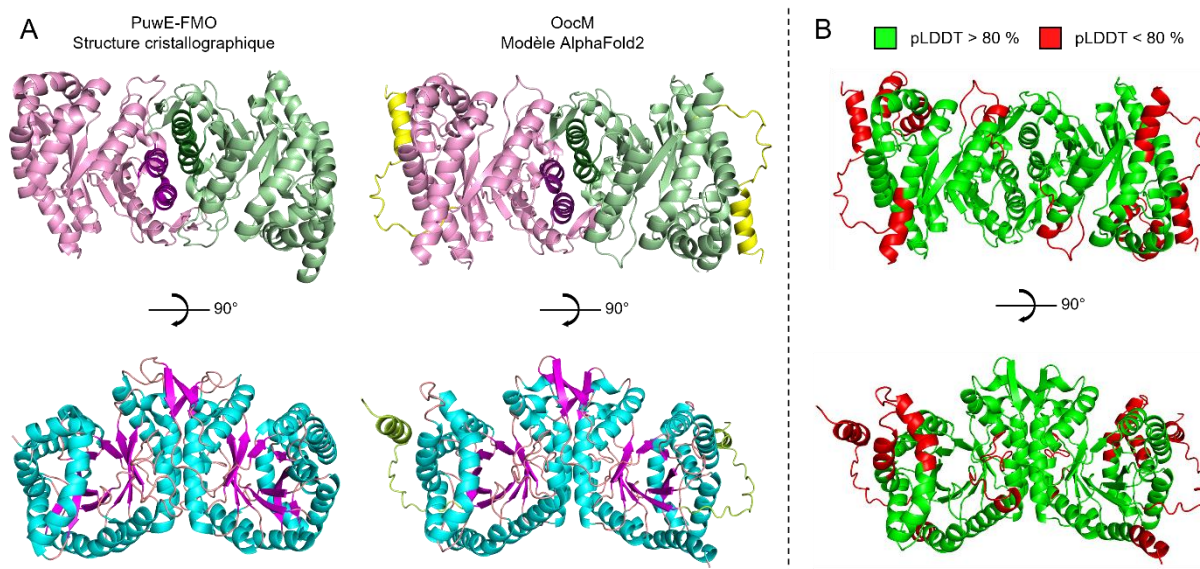
## 1. Analyses bioinformatiques

Une recherche de protéines homologues a été effectuée sur la PDB à partir de la séquence de OocM, et a révélé que les structures présentant l'identité de séquence la plus élevée sont également des FMO. La structure d'un domaine monooxygénase situé dans la sous-unité PuwE de l'hybride PKS-NRPS de la puwainaphycine (PDB : 6KET) se démarque particulièrement des autres structures homologues identifiées, par son identité de séquence de 29.8 % avec OocM. En se basant sur la règle de colinéarité, il a été proposé que ce domaine catalyse également l'hydroxylation de l'intermédiaire de synthèse en position  $\alpha$ <sup>350</sup>, ce qui expliquerait sa forte homologie à OocM. Les autres FMO identifiées présentent une identité de séquence moindre, inférieure à 20 %, et incluent une monooxygénase de PKS de type II (PDB : 7BIP, 18.0 % d'identité de séquence), plusieurs structures de type luciférase (PDB : 3RAO, 1LUC, 1BSL, 1BRL, 3FGC, 1XKJ, 6FRI, 14.3–17.1 % d'identité de séquence) et une déshydrogénase flavine-dépendante (PDB : 5LXE, 15.8 % d'identité de séquence).

Toutes les monooxygénases identifiées présentent une organisation tertiaire et quaternaire similaire. Ces protéines sont homodimériques (exceptées les hétérodimères 1LUC et 1BRL) et adoptent un repliement en tonneau  $(\beta/\alpha)_8$ -TIM. De plus, leur catalyse enzymatique requiert la présence d'un cofacteur flavinique FMN<sup>351,352</sup> pour activer l'oxygène moléculaire. Un modèle d'OocM a alors été généré à l'aide d'AlphaFold2 afin de voir si la structure proposée était en accord avec les caractéristiques de ses homologues (**Figure 72**). Le modèle homodimérique calculé suggère en effet un repliement de chaque monomère en tonneau  $(\beta/\alpha)_8$ -TIM, avec une large interface inter-monomère stabilisée notamment par la formation d'un fagot central de quatre hélices  $\alpha$  et d'interactions avec des boucles adjacentes, à l'image de la structure de PuwE-FMO. La différence majeure entre cette structure et le modèle d'OocM

## Résultats

réside en N-terminal de chaque monomère. À cette extrémité, la structure prédite d'OocM présente une hélice  $\alpha$  supplémentaire de onze résidus, liée au tonneau TIM par une boucle de 17 résidus (**Figure 72**). Cet élément pourrait avoir un rôle central dans l'interaction d'OocM avec ses partenaires puisqu'elle est absente de la structure cristallographique de PuwE-FMO, qui se trouve nativement liée covalamment à un domaine aminotransférase en N-terminal et à un domaine C en C-terminal <sup>350</sup>. L'indice de confiance (pLDDT) calculé par AlphaFold2 pour chaque position du modèle concorde avec les caractéristiques structurales communes et divergentes entre OocM et PuwE-FMO. En effet, les éléments de structure participant à la formation du tonneau TIM et de l'interface de dimérisation sont modélisés avec une confiance importante (pLDDT > 80 %), tandis que l'hélice  $\alpha$  N-terminale ainsi que les deux autres hélices  $\alpha$  et une boucle qui la stabilisent possèdent un pLDDT inférieur à 80 %. De façon similaire, l'extrémité de l'hélice  $\alpha$  C-terminale et deux boucles  $\beta/\alpha$  probablement flexibles sont modélisées avec une confiance faible (pLDDT < 80 %) (**Figure 72**).



**Figure 72 :** Modèle de l' $\alpha$ -hydroxylase OocM obtenu avec AlphaFold2 <sup>336</sup> et structure cristallographique de l' $\alpha$ -hydroxylase PuwE-FMO (PDB : 6KET). Les deux protéines présentent une forte homologie de structure, reflétée par un r.m.s.d de 1.083, obtenu par superposition structurale. A) Les deux hydroxylases sont colorées par monomère (en haut) et le fagot de quatre hélices  $\alpha$  qui stabilise l'interface est indiqué en couleurs plus foncées. La coloration par structures secondaires (en bas) permet d'observer clairement le repliement de chaque monomère en tonneau  $(\beta/\alpha)_8$ -TIM. La différence la plus notable entre ces deux structures est la présence d'une hélice  $\alpha$  et d'une boucle flexible, représentées en jaune, en N-terminal de chaque monomère d'OocM. B) Représentation du modèle structural d'OocM en fonction de son indice de confiance. Les zones prédites avec une confiance importante (pLDDT > 80%) sont colorées en vert, et celles modélisées avec une confiance plus faible (pLDDT < 80 %) sont colorées en rouge.

Ainsi, la structure tridimensionnelle prédite de OocM et la nature potentielle de son cofacteur flavinique favorisent l'hypothèse d'une FMO de classe C <sup>303</sup>, à l'inverse de LbmA-Ox, FMO de classe B, et des différentes BVMO étudiées lors de cette thèse vraisemblablement

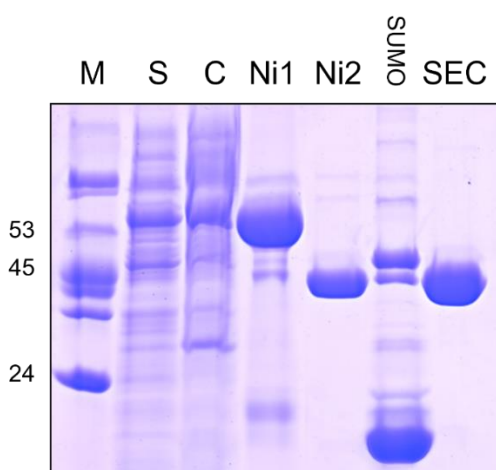


aussi de classe B. En effet, leurs activités de formation d'oxime et d'oxydation Baeyer-Villiger respectives requièrent la présence des cofacteurs FAD et NADPH, et la structure de LbmA-Ox a révélé un repliement en deux domaines de type Rossmann, de même que les prédictions de structures des BVMO. OocM serait donc potentiellement la première monooxygénase modulaire discrète de classe C caractérisée structuralement.

## 2. Purification de OocM

OocM a été exprimée à partir du vecteur pET28b-SUMO et purifiée selon le protocole détaillé en section IV. de la partie Matériels et Méthodes. L'échantillon après la première étape de chromatographie d'affinité était incolore, ce qui indiquait que le cofacteur flavinique de l'enzyme n'avait pas été co-purifié. Cette observation apporte encore un peu plus de crédit à l'hypothèse d'une FMO de classe C, puisque ces monooxygénases ne conservent pas le FMN dans leur site actif au cours du cycle catalytique. En effet, elles font partie de systèmes à deux composants dont le deuxième est une flavine réductase NAD(P)H-dépendante qui fournit le FMN réduit à la monooxygénase.

L'étiquette His<sub>6</sub>-SUMO a été clivée lors de l'étape de dialyse, en présence de la protéase ULP1 qui reconnaît spécifiquement la protéine SUMO, puisque le plasmide pET28b-SUMO ne contient pas de site de clivage protéique entre l'étiquette et la MCS. Après une seconde chromatographie d'affinité nickel et une étape finale de SEC, la pureté de la protéine a été évaluée par SDS-PAGE (**Figure 73**) avant d'être stockée à -80 °C. Un rendement de 10,3 mg d' $\alpha$ -hydroxylase par litre de culture a été obtenu, avec une pureté optimale de la protéine.



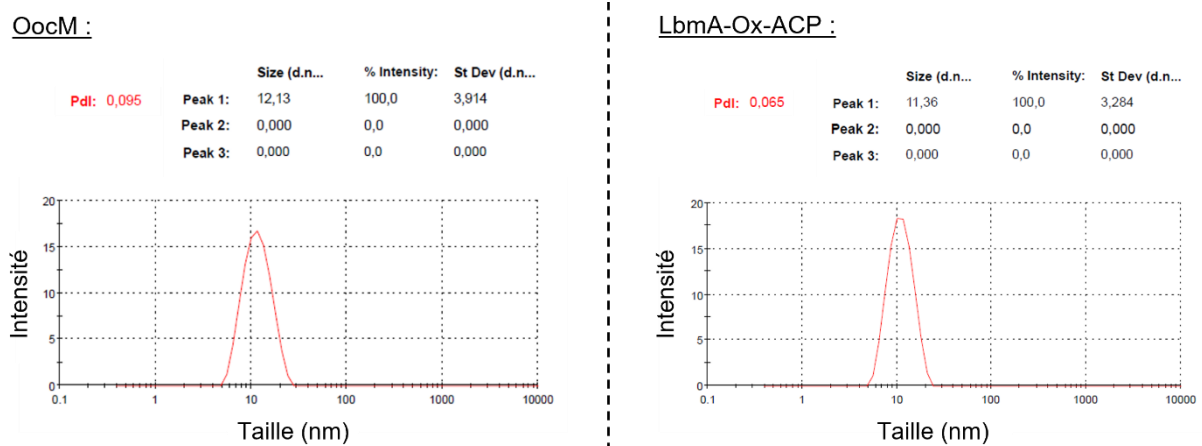
**Figure 73 :** SDS-PAGE bilan de purification de OocM. La pureté de l'échantillon est évaluée dans le surnageant (S) et le culot (C) de centrifugation après lyse des cellules et dans les différents éluats de chromatographie : Ni1 = première chromatographie d'affinité nickel ; Ni2 = seconde chromatographie d'affinité nickel ; SUMO = étiquette His<sub>6</sub>-SUMO retenue sur chromatographie d'affinité nickel ; SEC = chromatographie d'exclusion de taille. Les bandes du marqueur de taille (M) correspondant aux masses de 24, 53 et 66 kDa sont indiquées. OocM et l'étiquette SUMO possèdent des masses moléculaires respectives de 43 et 12 kDa.

### 3. Caractérisation de OocM en solution : stabilité et oligomérisation

OocM a été soumise à une expérience de DLS préalablement aux essais de cristallisation afin de s'assurer de la stabilité de la protéine et de l'homogénéité de l'échantillon. L'expérience a été réalisée à 6,5 mg/mL de protéine, concentration de l' $\alpha$ -hydroxylase obtenue en fin de purification, après concentration.

La mesure de DLS a révélé la présence d'une unique population de particules ainsi qu'un indice de polydispersité de 9,5 % de l'échantillon, amplement suffisant pour envisager la cristallisation de la protéine. En effet, il est généralement considéré que cet indice doit être inférieur à 20 % pour permettre une caractérisation structurale.

Un diamètre hydrodynamique de  $121 \text{ \AA} \pm 39$  a été calculé pour OocM. Malgré un écart type important, ce diamètre est incompatible avec une protéine monomérique de 43 kDa (**Figure 74**). À titre de comparaison, une expérience de DLS menée sur la protéine LbmA-Ox-ACP à une concentration de 10 mg/mL a révélé un diamètre hydrodynamique similaire de  $114 \text{ \AA} \pm 33$  (**Figure 74**), bien que le didomaine possède une masse moléculaire de 72 kDa, soit 60 % de plus qu'OocM. Ces résultats permettent donc de conclure qu'OocM est homodimérique en solution, ce qui concorde avec la nature dimérique des protéines homologues identifiées sur la PDB (voir la section IV.1.).

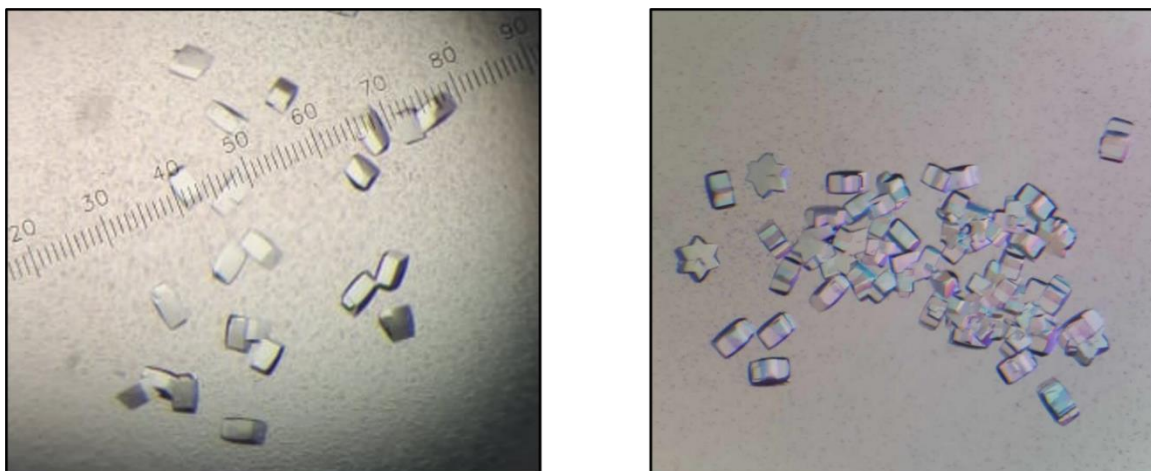


**Figure 74 :** Résultats de DLS de OocM (à gauche) et de LbmA-Ox-ACP (à droite) permettant la comparaison des paramètres calculés. La détection d'un seul pic responsable de 100 % de l'intensité diffusée atteste de la pureté de l'échantillon. De plus, l'indice de polydispersité (Pdl) est inférieur à 10 %, ce qui indique une bonne homogénéité de l'échantillon. Les deux constructions présentent un diamètre hydrodynamique de taille comparable, bien que le didomaine monomérique LbmA-Ox-ACP soit 60 % plus massif qu'OocM. Cela témoigne de l'homodimérisation d'OocM en solution.

### 4. Essais de cristallisation

Des cribles commerciaux ont été utilisés pour identifier des conditions de cristallisation

de OocM prometteuses. Dans un premier temps, les essais ont été réalisés avec 5 mg/mL d'hydroxylase, en absence de cofacteur et de substrat afin de tester la propension de la protéine à cristalliser seule. En l'espace d'une journée, deux formes cristallines sont apparues dans plusieurs solutions précipitantes. La première forme a été obtenue à l'aide du crible Wizard Classic (Rigaku Reagents) en présence de 1 M ammonium phosphate dibasique, 200 mM NaCl et 100 mM imidazole/HCl pH 8.0 (**Figure 75**). La deuxième forme cristalline est caractérisée par la croissance d'étoiles à six branches obtenues avec le crible Wizard Precipitant Synergy (Rigaku Reagents) en présence de 2 M ammonium sulfate, 5 % (v/v) PEG400, 50 mM magnésium sulfate et 100 mM Tris base/HCl pH 8.5 (**Figure 75**).



**Figure 75 :** Formes cristallines d'OocM obtenues par diffusion de vapeur. Les parallélépipèdes (à gauche) sont apparus en présence de 1 M ammonium phosphate dibasique, 200 mM NaCl et 100 mM imidazole/HCl pH 8.0. Quant aux étoiles à six branches (à droite), elles sont apparues en présence de 2 M ammonium sulfate, 5 % (v/v) PEG400, 50 mM magnésium sulfate et 100 mM Tris base/HCl pH 8.5.

Après affinement des conditions de cristallisation, plusieurs cristaux ont été collectés et congelés en flux d'azote après cryoprotection en présence de glycérol ou d'éthylène glycol. Des tests de diffraction des rayons X ont été ensuite réalisés avec ces cristaux sur la ligne de lumière PROXIMA-2A du synchrotron SOLEIL. Les données enregistrées ont révélé que la diffraction générée par ces cristaux ne dépassait pas les 6 Å de résolution, faisant donc obstacle à l'enregistrement d'un jeu de données permettant la résolution de la structure d'OocM.

La co-cristallisation d'OocM avec ses cofacteurs FMN et NADPH pourrait constituer une approche permettant de stabiliser davantage une conformation de la protéine et ainsi augmenter l'ordre dans le cristal pour en améliorer la diffraction. De la même manière, la co-cristallisation en présence d'un analogue de substrat, de FMN et de NADP oxydé pourrait rigidifier la protéine tout en interdisant la catalyse sans avoir recours à un mutant inactif.



---

*Discussion et  
perspectives*

---



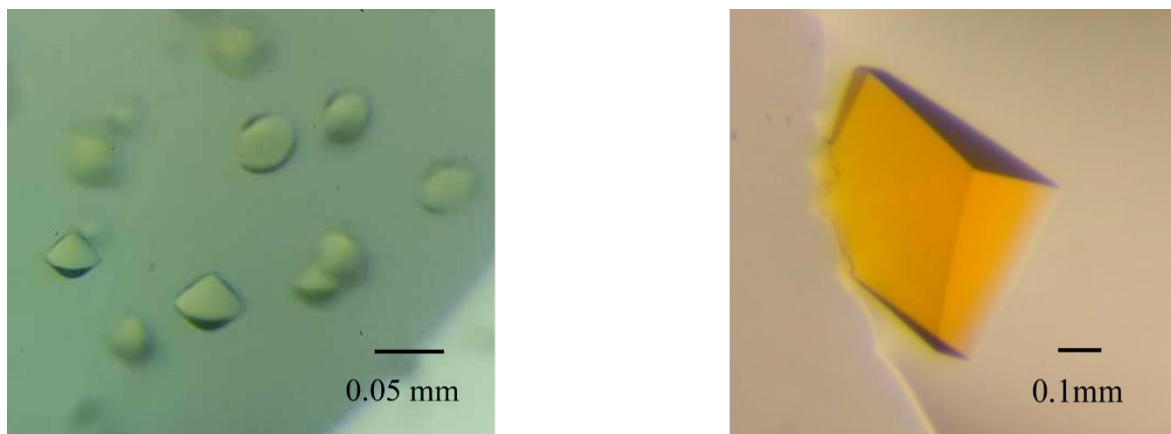
## Relations structure-fonction d'un domaine monooxygénase générateur d'oxime

Avant les travaux de cette thèse, la présence d'activités de formation d'oxime au sein de voies de biosynthèse de polycétides et NRP était déjà connue de longue date et les mécanismes proposés suggéraient une *N*-hydroxylation catalysée par une CYP<sup>353,354</sup> ou une FMO<sup>312</sup>. Cependant, les réactions décrites impliquées systématiquement le recrutement d'enzymes discrètes, soit de manière post-PKS, c'est-à-dire après synthèse et libération du polycétide de sa PKS<sup>355</sup>, soit au cours de la synthèse du métabolite<sup>356</sup>. La découverte d'une telle activité portée par un domaine intégré à un module au sein de la PKS-NRPS de la lobatamide A a donc naturellement attiré notre attention et nous a conduit à investiguer le mécanisme catalytique de cette enzyme sans précédent. Pour ce faire, l'activité de la FMO LbmA-Ox a été reconstituée *in vitro* et sa spécificité de substrat a été caractérisée, prouvant son rôle de formation d'oxime par oxydation d'une amine terminale. Une approche combinant la cristallographie aux rayons X et des tests d'activité *in vitro* de mutants couplés à la spectrométrie de masse a été ensuite mise place pour élucider les bases structurales de cette monooxygénation.

Le domaine seul LbmA-Ox a montré une très faible propension à la cristallisation, ce qui nous a conduit à modifier la construction protéique pour y inclure le domaine ACP en aval de la FMO. Cette stratégie a conduit à l'obtention de cristaux de forme bipyramide hexagonale, à partir desquels un jeu de données de diffraction a été enregistré. Un phasage par remplacement moléculaire à l'aide d'un modèle calculé par AlphaFold2<sup>336</sup> a ensuite permis la résolution de la structure du didomaine.

L'ajout du domaine ACP à la protéine a engendré une importante flexibilité dont les conséquences sont doubles. Premièrement, la présence du linker interdomaine a permis à la protéine d'adopter différentes conformations, augmentant de ce fait le nombre potentiel d'interactions protéine-protéine responsables de l'agencement du réseau cristallin. C'est donc grâce à la flexibilité de ce linker que deux conformations du didomaine ont pu être observées dans l'unité asymétrique, chacun des positionnements de l'ACP par rapport à la FMO participant de manière différente à l'empilement cristallin. Deuxièmement, la flexibilité apportée par le linker contribue inévitablement à l'augmentation de l'agitation thermique de la protéine. En effet, l'une des conformations cristallographiques de LbmA-Ox-ACP est stabilisée par des interactions protéine-protéine médiées principalement par son domaine Ox. Par conséquent, le domaine ACP est plus bien plus mobile, et entraîne une agitation importante du didomaine. Au niveau des données, cela s'est traduit par une densité électronique mal définie de cette conformation, surtout au niveau de l'ACP, et par une faible stabilisation des cofacteurs de la

FMO. Cela explique ainsi que les données de diffraction n'aient pu être traitées que jusqu'à 3.07 Å de résolution, et que seuls des cristaux de petite taille aient été obtenus, avec une couleur jaune relativement pâle comparée à celle de cristaux de flavoprotéines dont la flavine est étroitement liée <sup>357</sup> (**Figure 76**).



**Figure 76 :** Cristaux de flavoprotéines. Les cristaux de LbmA-Ox-ACP (à gauche) sont de petite taille (~ 50  $\mu\text{m}$ ), présentent une couleur jaune pâle qui atteste de la fixation partielle du cofacteur flavinique, et ont permis de traiter un jeu de données de diffraction jusqu'à 3.07 Å de résolution. Au contraire, le cristal d'une flavodoxine de *Desulfovibrio gigas* (à droite) mesure près de 800  $\mu\text{m}$  dans sa plus grande dimension, possède une couleur jaune intense due à la fixation étroite de son cofacteur flavinique, et a permis l'enregistrement de données de diffraction jusqu'à 1.3 Å de résolution <sup>357</sup>.

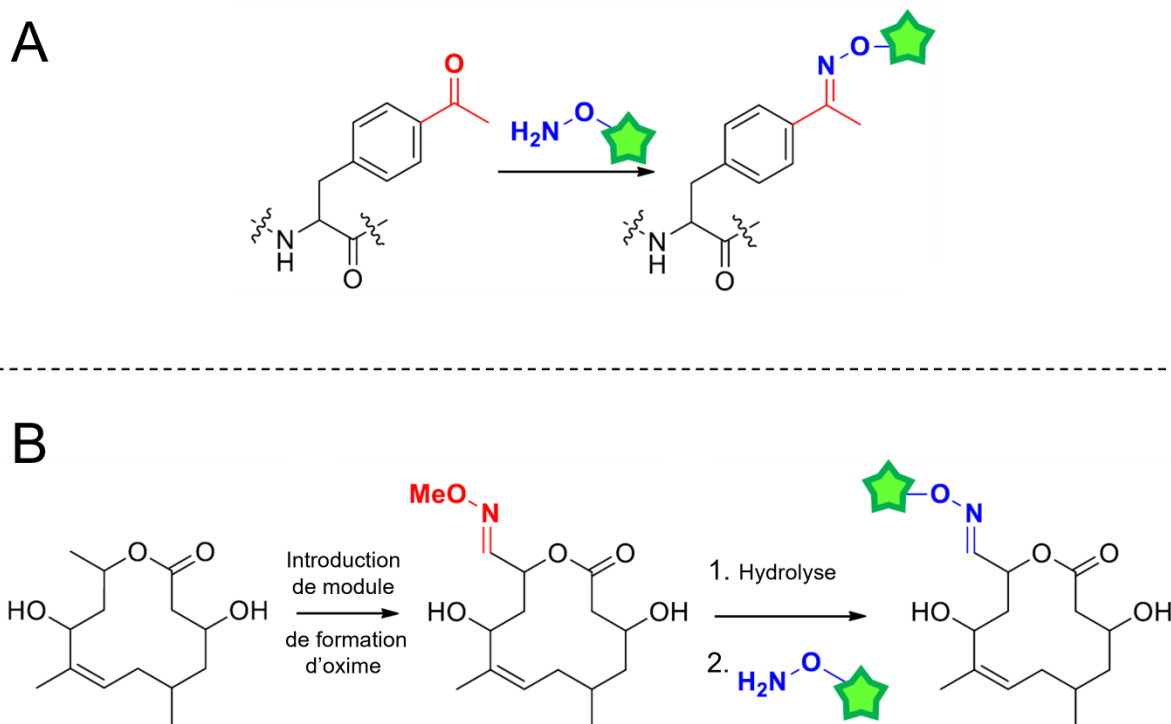
L'étude de la structure de LbmA-Ox nous a permis d'identifier les acides aminés potentiellement impliqués dans la fixation des cofacteurs et du substrat, ainsi que dans la catalyse. Leurs rôles ont pu être confirmés ou infirmés par le biais de mutagenèses dirigées couplées à des tests d'activités et des analyses par HPLC-MS. Sur la base de ces résultats, nous avons proposé deux mécanismes enzymatiques plausibles, se déroulant par déshydrogénation d'un intermédiaire monohydroxylé ou déshydratation d'un intermédiaire dihydroxylé. Toutefois, les données de mutagenèse font pencher la balance en faveur du second mécanisme, suggérant ainsi que LbmA-Ox est une enzyme bifonctionnelle *N*-hydroxylase/déshydratase.

Ces résultats étendent le répertoire connu de réactions chimiques catalysés par les PKS modulaires, fournissant un outil supplémentaire aux approches d'ingénierie génétique. L'insertion rationnelle d'oxime dans d'autres voies de biosynthèse par transposition de sa machinerie enzymatique pourrait ainsi permettre de modifier les propriétés biologiques des analogues générés. En effet, des études de relations structure-activité menées sur l'althiomycine ont démontré l'implication de sa fonction oxime dans son activité antibactérienne <sup>356,358</sup>.

Une seconde application potentielle prometteuse pourrait être l'insertion d'oxime dans



d'autres polycétides en tant qu'étiquette aldéhyde masquée. La bioconjugaison de protéines à des molécules d'intérêt (molécules bioactives, agents de contraste, polymères,...) est traditionnellement accomplie par introduction d'acides aminés kéto-fonctionnalisés<sup>359</sup> ou d'un groupement aldéhyde génétiquement encodé<sup>360</sup> (**Figure 77**). La protéine est ensuite liée covalamment à la molécule d'intérêt préalablement fonctionnalisée par un groupement amine. Dans le cas des polycétides, la fusion N-terminale de la machinerie de formation de méthyloxime aux PKS permettrait des applications de conjugaison du même type. En effet, l'hydrolyse réversible d'une oxime entraîne la formation d'un aldéhyde, pouvant ensuite être conjugué à une molécule d'intérêt, ce qui permettrait de fonctionnaliser les polycétides ne possédant pas de carbonyles réactifs (**Figure 77**).



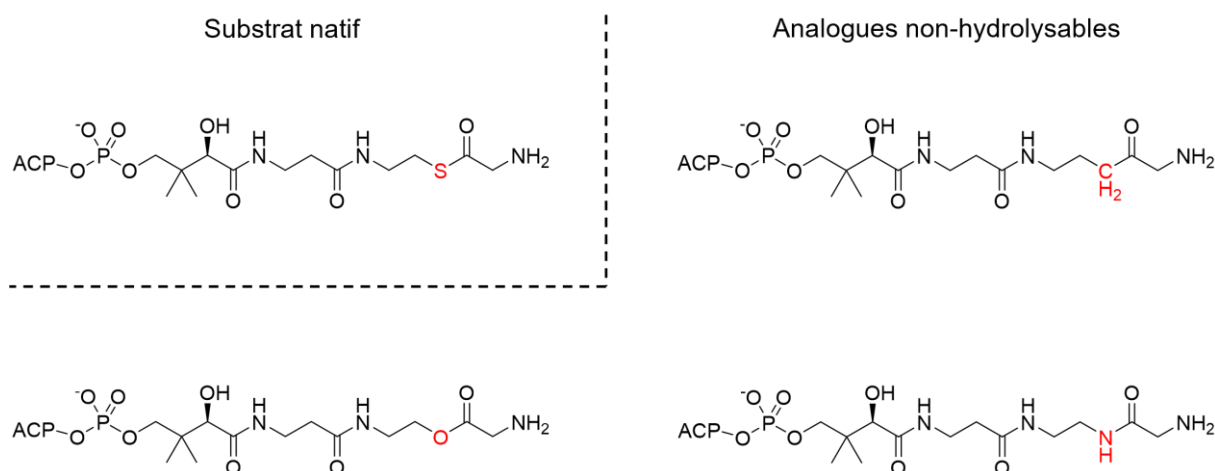
**Figure 77 :** Conjugaison site-spécifique via une étiquette oxime. A) Méthode de fonctionnalisation de protéines par introduction d'acides aminés non naturels kéto-fonctionnalisés. B) Méthode proposée de fonctionnalisation de polycétides par introduction d'une étiquette aldéhyde masquée sous forme d'oxime (figure réalisée par Jörn Piel).

L'étude de spécificité de substrat réalisée sur LbmA-Ox a révélé sa capacité à accepter le glycyI-SNAC ainsi que l'alanyl-SNAC. En parallèle à ces expériences, des travaux ont été menés sur un domaine générateur d'oxime homologue, OxiB-Ox, présent dans la synthase hybride de l'oximidine par l'équipe du Pr. Joleen Masschelein et ont démontré que ce domaine acceptait également le thréonyl-SNAC en plus des deux cités précédemment<sup>361</sup>. Toutefois, ces réactions n'ont été permises que dans un contexte *in vitro* dans lequel LbmA-Ox et OxiB-Ox ont été extraites de leurs PKS respectives. L'analyse d'extraits organiques de la bactérie *Gyvuella sunshinyii* produisant la lobatamide n'a pas permis d'identifier de dérivés contenant

une alanine, ce qui indique que dans le contexte natif de l'enzyme, le domaine d'adénylation en aval de LbmA-Ox présente probablement une forte spécificité pour la glycine. Le domaine KS<sup>0</sup> en N-terminal de la FMO pourrait également jouer un rôle dans l'acheminement sélectif d'unité d'initiation glycine. Des études visant à introduire une fonction oxime sur un acide aminé différent de la glycine dans d'autres voies de biosynthèse devraient donc tenir compte de ces spécificités de substrats importantes des domaines en amont et les modifier en conséquence.

Des expériences de *docking in silico* de LbmA-Ox avec son ACP C-terminal ont montré que la longueur du linker entre ces deux domaines ne permet pas à l'ACP de former un complexe productif avec la FMO. Le site de formation de l'oxime a alors été investigué par mutagenèse du PCP N-terminal de LbmA et des deux ACP suivants, désignés ACP<sub>1</sub> et ACP<sub>2</sub>. Le produit de synthèse monooxygéné n'a pas été détecté lorsque l'ACP<sub>1</sub> ou les deux ACP étaient mutés, éliminant l'hypothèse d'une formation d'oxime au niveau du domaine PCP. En revanche, le produit de synthèse a été détecté lorsque seul l'ACP<sub>2</sub> était muté, prouvant que la réaction de monooxygénation a lieu au niveau de l'ACP<sub>1</sub>, celui du didomaine LbmA-Ox-ACP. Les résultats de *docking* et de mutagenèse suggèrent ensemble qu'un seul mécanisme d'interaction entre Ox et ACP<sub>1</sub> peut permettre la catalyse. Selon cette hypothèse, la catalyse du domaine FMO de la sous-unité LbmA survient sur l'ACP<sub>1</sub> de la deuxième sous-unité LbmA appairée dans le contexte homodimérique des PKS. Une telle coopération entre sous-unités identiques a déjà été observée au sein de systèmes *cis*-AT<sup>362</sup> mais serait une première dans le cas de systèmes *trans*-AT. En accord avec cette hypothèse, la réitération des expériences de *docking* entre les domaines Ox et ACP<sub>1</sub> en absence de contraintes de longueur de linker a conduit au calcul de complexes productifs.

Afin d'apporter des preuves expérimentales de ce mécanisme, la structure quaternaire de LbmA-Ox et LbmA-ACP<sub>1</sub> doit être étudiée. Dans un premier temps, des tests d'interactions par titration calorimétrique isotherme ou SEC-MALS pourraient permettre de s'assurer que la FMO et l'ACP forment un complexe. Ces expériences devraient être réalisées en présence de l'ACP sous forme *apo*, ainsi que sous forme *holo* et acylé afin d'évaluer la contribution du groupement Ppant et du substrat à l'interaction interdomaine. L'*holo*-ACP peut être généré par expression dans la souche *E. coli* BAP1, dérivée de la souche BL21 (DE3), qui contient le gène *sfp* codant la PPTase de *Bacillus subtilis*, sous contrôle du promoteur T7. L'incubation de l'analogue de substrat glycyI-SNAC en présence de DTNB a révélé une hydrolyse de son thioester en quelques minutes, empêchant d'envisager son utilisation pour les caractérisations d'interactions et de structures. Des analogues non-hydrolysables devraient donc être employés, dont l'atome de soufre de la fonction thioester, serait remplacé par un atome d'oxygène, une fonction amide ou méthylène (**Figure 78**)<sup>363</sup>.



**Figure 78 :** Substrat natif de LbmA-Ox lié à l'*holo*-ACP et analogues non-hydrolysables dont le thioester est remplacé par une cétone, un ester, ou un amide.

Le comportement mono- ou dimérique du didomaine LbmA-Ox-ACP pourrait ensuite être caractérisé en solution par diffusion des rayons X aux petits angles (SAXS). Cette technique particulièrement adaptée à l'étude de protéines flexibles permet d'évaluer la masse moléculaire, les dimensions, et donc l'état oligomérique d'une protéine en solution. De plus, elle permet de déterminer l'enveloppe globale de la protéine, dans laquelle la structure à haute résolution des différents domaines peut ensuite être replacée. Le complexe potentiellement formé entre deux didomaines Ox-ACP ou entre un domaine Ox et un domaine ACP<sub>1</sub> pourrait ainsi être caractérisé. Il est toutefois probable que la dimérisation de LbmA soit nécessaire à la proximité spatiale du domaine Ox d'une sous-unité et de l'ACP<sub>1</sub> de la sous-unité appairée. La caractérisation en solution d'une construction incluant les deux KS dimériques adjacentes au didomaine Ox-ACP (KS<sup>0</sup><sub>1</sub>-Ox-ACP<sub>1</sub>-KS<sup>0</sup><sub>2</sub>) permettrait alors de vérifier cette hypothèse et de fournir de précieuses informations sur la communication interdomaine au sein des PKS.

### **Etudes préliminaires d'activités BVMO modulaires : une cristallisation entravée par leur instabilité**

Les activités BVMO identifiées au sein des systèmes *trans*-AT sont remarquables par leur diversité d'architectures modulaires, de régiospécificité et de structures polycétidiques oxydées générées. Les esters formés par les BVMO discrètes constituent le site d'hydrolyse de la chaîne polycétidique pour engendrer un produit final plus court, tandis que le domaine BVMO de LbmC catalyse la formation d'un ester interne conservé dans la structure du polycétide final. Ces enzymes élargissent donc l'éventail d'activités et de groupements synthétisés par les PKS modulaires. De plus, elles représentent un des seuls exemples d'activité de PKS responsable de l'insertion d'un atome autre que le carbone dans le squelette

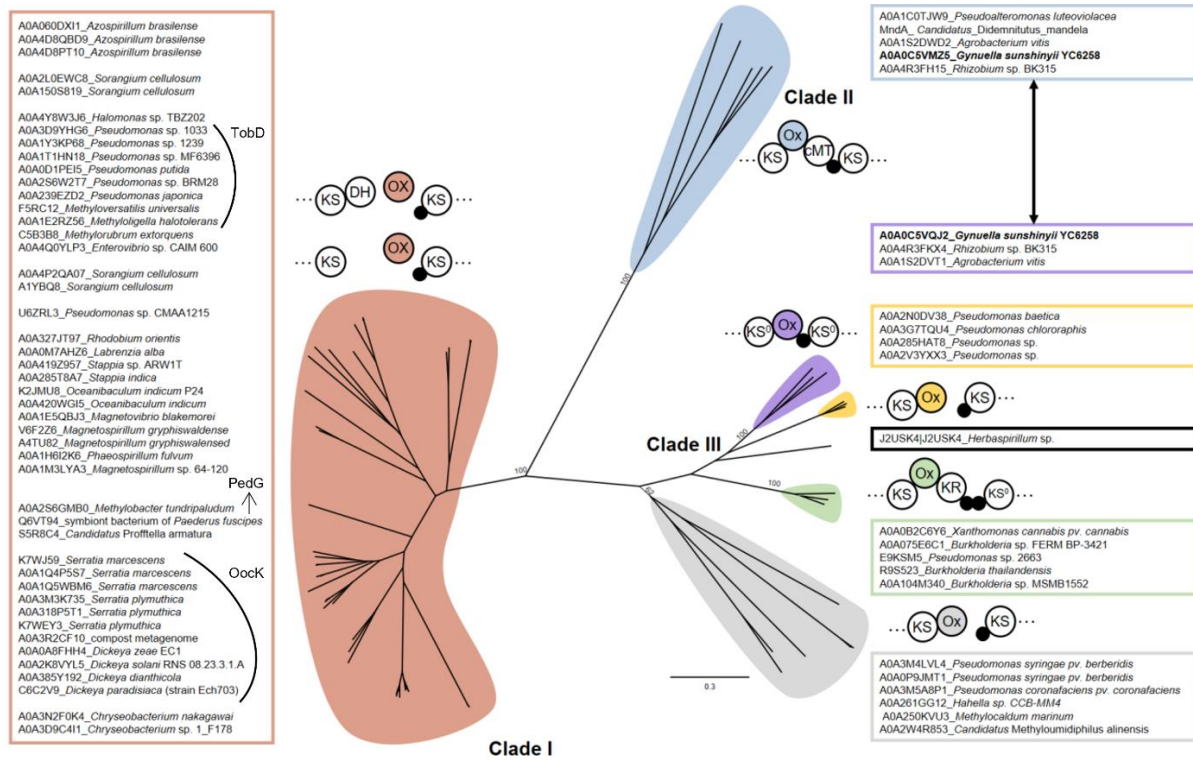
polycétidique.

Afin d'élucider les bases structurales et fonctionnelles de ces activités BVMO ainsi que les caractéristiques spécifiques à chacune, des études préliminaires de production et purification des BVMO discrètes OocK et SbnP et du domaine LbmC-BVMO ont été réalisées afin de mettre en place des essais de cristallisation. OocK et SbnP ont montré une forte propension à la précipitation avec une perte quasiment totale des échantillons lors des étapes de dialyse. Un plan factoriel incomplet a alors été réalisé pour identifier le pH, les sels, les additifs et leurs concentrations permettant d'améliorer la solubilité d'OocK. La composition d'un tampon de lyse optimale a ainsi été établie et a permis d'obtenir une fraction soluble plus importante de la BVMO discrète dans le surnageant de centrifugation après sonication des cellules. Néanmoins, la dialyse de la protéine demeurait l'étape critique du procédé de purification, avec une précipitation toujours importante de OocK et SbnP. Pour contourner ce problème, le temps de purification a été réduit en supprimant la dialyse et l'étape suivante de chromatographie d'affinité nickel. Le domaine LbmC-BVMO quant à lui s'est révélé beaucoup plus stable en solution ce qui a permis sa purification dans des rendements supérieurs aux BVMO discrètes. Les trois monooxygénases ont ainsi été purifiées et la couleur jaune des échantillons attestait de la présence de leur cofacteur flavinique.

Les BVMO ont été ensuite soumises à des essais de cristallisation à l'aide de cribles commerciaux de 96 conditions. Dans le but de stabiliser les BVMO discrètes, celles-ci ont été mises en présence de leur cofacteur NADP oxydé et d'un analogue de substrat SNAC. Des tests de co-cristallisation du domaine LbmC-BVMO ont également été initiés en incorporant les mêmes cofacteurs et substrats aux gouttes de cristallisation. À ce jour, aucun cristal de BVMO n'a été obtenu, et la proportion importante de gouttes présentant un précipité protéique dans le cas des BVMO discrètes témoigne une fois de plus de leur instabilité, malgré l'utilisation de concentrations protéiques variées.

La cristallisation des BVMO discrètes OocK et SbnP semble donc difficilement réalisable, en dépit de l'optimisation de leur tampon de solubilisation. Cependant, de nombreux homologues de ces enzymes ont été identifiés bioinformatiquement dans d'autres systèmes PKS *trans*-AT. En effet, la recherche de gènes codant une FMO au sein de PKS sans produit de synthèse associé, ou une FMO homologue à OocK, a permis à Ueoka *et al.*<sup>307</sup> de mettre en lumière 69 PKS *trans*-AT dont 46 possèdent une BVMO discrète phylogénétiquement proche de OocK. Parmi elles, on retrouve les enzymes PedG et TobD des voies de biosynthèse de la pédérine et du toberol, mais également de nombreuses BVMO au sein de PKS non caractérisées (**Figure 79**). Cette diversité d'organisme et de voies de biosynthèse représente autant d'occasions potentielles de pouvoir identifier une BVMO discrète plus stable et adaptée à des études structurales. En effet, ces enzymes peuvent avoir un comportement différent en solution d'une PKS à l'autre bien que leurs modules et activités enzymatiques soient

identiques. C'est ce qui avait d'ailleurs conduit l'équipe du Pr. Piel à se détourner des BVMO discrètes PedG et TobD dont la purification avait échoué, et à s'intéresser davantage à OocK<sup>308</sup> puis SbnP dont les faibles rendements de purification permettaient tout de même la mise en place de tests d'activité *in vitro*.



**Figure 79 :** Phylogramme de nombreuses FMO associées à des voies de biosynthèse PKS. Le clade I majoritaire regroupe les homologues de OocK. Il contient les BVMO discrètes PedG et TobD, ainsi que des FMO associées à des PKS non caractérisées. Il est à noter que les clades II et III contiennent le domaine LbmC-BVMO et le domaine LbmA-Ox, respectivement (figure reproduite de Ueoka *et al.*, 2020<sup>307</sup>).

Concernant le domaine LbmC-BVMO, les expériences de DLS ont montré que la protéine était stable en solution et tout à fait approprié pour des essais de cristallisation, avec un indice de polydispersité de l'échantillon inférieur à 10 %. Afin de favoriser sa cristallisation, la stratégie qui avait été adoptée pour le domaine LbmA-Ox pourrait dans ce cas s'avérer également bénéfique. En effet, tout comme le domaine générateur d'oxime, LbmC-BVMO a été extraite de sa sous-unité et par conséquent n'est plus liée en N- et C-terminal aux domaines KS et cMT, respectivement. Différentes constructions protéiques pourraient donc être envisagées en incluant ces domaines, afin d'apporter une certaine flexibilité à la protéine et établir de potentiels interactions protéine-protéine pour former le cristal.

## Etude d'une $\alpha$ -hydroxylase discrète : une FMO de classe C portant une activité modulaire

La PKS *trans*-AT de l'oocydine comprend de nombreuses activités enzymatiques non-canoniques. En plus de la FMO discrète OocK qui catalyse une réaction de type BVMO, une seconde FMO discrète, OocM, est responsable de l' $\alpha$ -hydroxylation de la chaîne polycétidique liée au module scindé OocLMN<sup>311</sup>. Des analyses bioinformatiques par homologie de séquences et prédiction de structure ont révélé un repliement de cette protéine en tonneau ( $\beta/\alpha$ )<sub>8</sub>-TIM, caractéristique des FMO de classe C FMN-dépendantes. OocM se distingue donc nettement des BVMO et FMO génératrice d'oxime étudiées dans cette thèse, ce qui explique notamment pourquoi elle avait échappé à la recherche d'homologues de OocK<sup>307</sup> (**Figure 79**). L'hydroxylase a été ensuite produite et purifiée dans de bons rendements et avec une pureté optimale. Les mesures de DLS ont permis d'attester de la stabilité de la protéine en solution, au contraire des autres FMO discrètes d'intérêt, et le rayon hydrodynamique calculé indique une organisation homodimérique, en accord avec les structures de protéines homologues identifiées sur la PDB. Des essais de cristallisation ont alors été réalisés afin d'identifier des solutions de précipitants prometteuses. Les tests initiaux ont été menés avec la protéine seule, sans ajout de cofacteur ou analogue de substrat. Plusieurs conditions de cristallisation ont alors conduit à l'obtention de deux formes cristallines distinctes. Ces conditions ont été affinées et les cristaux formés ont été collectés et soumis à des expériences de diffraction des rayons X. La résolution maximale de 6 Å n'a pas permis d'enregistrer de jeu de données pouvant conduire à l'élucidation de la structure d'OocM. La cause la plus probable de cette faible qualité de diffraction est probablement l'évolution rapide des cristaux, apparaissant en moins d'une journée, et se maclant presque aussi rapidement.

La réalisation des cribles de cristallisation à 4 °C plutôt qu'à 20 °C pourrait permettre de ralentir ce phénomène et donc d'améliorer la qualité des données collectées. Par ailleurs, la reconstitution de l' $\alpha$ -hydroxylation *in vitro* et les analyses bioinformatiques ont montré que OocM reconnaît un analogue de substrat  $\beta$ -keto-SNAC<sup>311</sup> et utilise vraisemblablement le FMN pour activer l'oxygène moléculaire. Ces deux molécules pourraient donc être utilisées lors d'expériences de co-cristallisation afin de stabiliser une conformation de la protéine et établir les résidus impliqués dans leur fixation et dans la catalyse enzymatique.

De manière surprenante, Hemmerling *et al.*, ont démontré que pour que l'hydroxylation ait lieu, OocM devait impérativement se trouver en présence de la KS<sup>05</sup>, premier domaine de la sous-unité en aval<sup>311</sup>. Cette KS non-élongatrice pourrait ainsi jouer un rôle de plateforme intermédiaire entre OocM et l'ACP C-terminal de la sous-unité en amont. Des tests d'interactions entre OocM, la KS<sup>05</sup> et l'ACP permettraient de révéler si un complexe se forme

entre ces trois protéines, et si tel est le cas, la résolution de la structure du complexe apporterait de précieuses informations sur le recrutement d'OocM et plus généralement, sur le recrutement de protéines discrètes au sein de modules scindés.





---

# *Annexes*

---



## Supporting Information

### **Modular Oxime Formation by a *trans*-AT Polyketide Synthase**

*H. A. Minas, R. M. M. François, F. Hemmerling, A. E. Fraley, C. L. Dieterich, S. H. Rüdiger, R. A. Meoded, S. Collin, K. J. Weissman\*, A. Gruez\*, J. Piel\**

## SUPPORTING INFORMATION

## Table of Contents

Experimental Procedures.....	2
General.....	2
Phylogenetic analysis of lobatamide KS sequences .....	2
Construction of expression plasmids and strains .....	3
Heterologous gene expression and protein purification.....	4
Synthesis of test substrates and standards.....	5
<i>In vitro</i> enzyme activity assay .....	7
<i>In vivo</i> assay.....	7
FAD analysis.....	7
UHPLC-HRMS analysis.....	7
Kinetic and <sup>15</sup> N NMR analysis.....	8
Crystallization of LbmA-Ox-ACP, X-ray diffraction data collection and structure refinement .....	8
Supplementary tables .....	10
Supplementary figures.....	18
NMR spectra of synthesized compounds .....	83

## Experimental Procedures

## General

UHPLC-HRMS and MS/MS measurements were performed on a Thermo Scientific Q Exactive mass spectrometer coupled to a Dionex Ultimate 3000 UHPLC system.

NMR spectra were recorded on Bruker BBO 400 MHz S1 and Bruker BBO 500 MHz S2 instruments without cryoprobes for verification of organic syntheses and for kinetic analysis of the enzymatic assays. In addition, a Bruker Avance III spectrometer equipped with a cryoprobe at 600 MHz and 125 MHz for <sup>1</sup>H NMR, and 150 MHz for <sup>13</sup>C NMR at 298K, was used for in situ reaction monitoring. Chemical shifts are given in parts per million (ppm) and were referenced to the solvent peaks at  $\delta_{\text{H}}$  2.50 and  $\delta_{\text{C}}$  39.51 for DMSO-*d*<sub>6</sub> and  $\delta_{\text{H}}$  7.26 and  $\delta_{\text{C}}$  77.16 for chloroform-*d*.<sup>[1]</sup> Multiplicities are given as follows: s - singlet, d - doublet, t - triplet, q - quartet, m - multiplet. The obtained data were processed and analyzed with Bruker Topspin 3.5 and 4.0 software, as well as Mestrelab Research S.L. MestReNova 14.2.0 software.

Chemicals were purchased from Sigma-Aldrich (subsidiary of Merck KGaA), if not stated otherwise. Chemicals and solvents were purchased from commercial suppliers and were used without further purification. For silica gel chromatography, distilled technical grade solvents and silica gel SilicaFlash® P60 (Silicycle) were used. Thin layer chromatography (TLC) was performed using aluminum sheets "TLC Silica gel 60 F254" from Merck Millipore® and analyzed with UV-light, ninhydrin or by permanganate staining. Enzymes were purchased from New England Biolabs or Thermo Fisher. Protino Ni-NTA agarose and DNA purification kits were purchased from Macherey-Nagel. The PD MiniTrap™ from GE Healthcare was used for desalting of proteins, and Amicon® Ultra centrifugal filter from Merck Millipore were used for ultracentrifugation.®

## Phylogenetic analysis of lobatamide KS sequences

The amino acid sequences of the KSs from the *lbm* cluster were aligned to 1121 KS sequences from an in-house collection of annotated *trans*-AT PKS clusters with assigned biosynthetic models using the MUSCLE algorithm with default settings.<sup>[2]</sup> KS<sub>3</sub> and KS<sub>5</sub> from the erythromycin *cis*-AT PKS were used as outgroups. A phylogenetic tree was built using FastTree (version 2.1.10 +SSE3 +OpenMP, 16 threads, default settings)<sup>[3]</sup> and bootstrap analysis was performed with 1000 pseudo-replicate sequences. The phylogenetic tree was visualized using FigTree (version 1.4.3).<sup>[4]</sup>

## SUPPORTING INFORMATION

## Construction of expression plasmids and strains

The sequences encoding LbmA-Ox, LbmA-Ox-ACP and LbmA-MT were amplified from *G. sunshinyii* YC6258 liquid culture using the primer pairs LbmA-Ox-F/R, LbmA-Ox-ACP-F/R and LbmA-MT-F/R (Table S1). The fragments were gel purified, digested with *SpeI* and *NotI*, *BamHI* and *HindIII*, or *NdeI* and *SacI* and cloned into a derivative of pET28b, called pET28b-SUMO\_Ser, yielding plasmids pET28b-SUMO-*lbmA*\_Ox, pET28b-SUMO-*lbmA*\_Ox-ACP and pET28b-SUMO-*lbmA*\_MT. The sequence encoding the LbmA bimodule was amplified from *G. sunshinyii* YC6258 as two parts using the primer pairs LbmA-M3.2-F/R and LbmA-M3.3-F/R. The pET28b-SUMO\_Ser backbone was amplified using the primer pair LbmA-M3.1-F/R with overhangs to the LbmA bimodule. The fragments were gel purified and the LbmA-bimodule was assembled into pET28b-SUMO by Gibson assembly<sup>[6]</sup> to yield pET28b-SUMO-*lbmA*\_begLob. The sequences encoding LbmA-Ox-ACP mutants were amplified from the pET28b-SUMO-*lbmA*\_Ox-ACP plasmid using a single primer for each mutation (Table S3). The LbmA trimodule mutants were amplified from the pET28b-SUMO-*lbmA*-KS<sub>3</sub> plasmid using a single primer for each mutation (Table S3). The plasmid template was digested by incubation at 37 °C for one hour in the presence of *DpnI*. *E. coli* DH5α was transformed with the final constructs, and the plasmids re-isolated and introduced into *E. coli* DE3 Tuner (NEB) for expression of N-terminally His<sub>6</sub>- and SUMO-tagged LbmA-Ox wild type and mutants and LbmA-Ox-ACP, N-terminally and N-terminally His<sub>6</sub>- and SUMO-tagged LbmA-MT, or into *E. coli* BAP1<sup>[6]</sup> for expression of the LbmA-bimodule. As a negative control, *E. coli* BAP1 was also transformed with empty pET28b-SUMO\_Ser.

For clarity, in the following the amino acid sequence for the LbmA trimodule (Protein ID: AJQ95678.1) is given in the following diagram, with domain boundaries indicated as highlighted. Positions mutated in the carrier proteins are marked in bold pink and positions mutated in the Ox-domain are marked in bold red.

1	MKLSFNELLH	LKSTFPDTIS	FWKQTLAPL	PILNLPVDRP	RLTALDKHKT	
51	DTVVSVVSAH	TAQAVVNFAD	AQDTPATVF	LTLFIVFLYR	HTDQQDILFG	
101	LLMDMTRDSD	RAESTKQLVP	MRATAVGTQP	FGDLLHHVRH	TLSTMSAHS	
151	CTFENVCQVL	EVDADVPLED	PLFSVGFAPQ	DIKKNDEWQ	LNWKDGYELF	
201	VQITERRGNF	TVVITYDIGL	FEASTINYLQ	NQYVNLIDTV	MDDPGQLPEM	
251	YSLLSRQEEQ	TILDDWNATA	AEYPQEQCFH	ELLAEQVKKD	PDKSAVFHEG	
301	NHLSYQDLEH	RSTLLARYLQ	STGVGPGSLV	GLCVERSLDM	IVGLIGILKS	
351	GGAYVPPGPD	YPEDRLRYMM	QDSGVQLVLT	QDRLISVISA	WSGEAGSIIA	
401	LDRDWQRIEQ	AAQTDFFLPR	QVTPDDLAYL	MYTSGTTGRP	KGVMIPHRAL	
451	TNFLVSMRSRE	PGLQAYDRLL	AVTTYCFDIA	GLELYLPLLQ	GGECHICSEE	
501	NSRDAQQLLQ	QIRTVKPTVM	QATPATWNML	FHAGWKNEER	VKILCGGEAL	
551	TDSLKRRFAD	HDCEVWNMYG	PTETTIWSTL	HRLEAGSNVS	TIIGRPIANT	
601	QVYIVDKNIQ	LLPVGIADEL	CIAGHGLAKG	YWNLPDPTAE	KFRANPFSGQ	
651	TRMYRTGDLA	RWRSDGTIEY	LGRMDHQVKI	RGFRIELGEI	ETRISHPPEI	
701	LESVAVVKEY	GEERQLVCYY	TTIDAAPIAS	QQLREYLQHS	LPDYMVPGLF	
751	INLQAMPLTP	NGKTRKTLT	ERPLELTRAQ	TEQSPESALE	TQLWTLWKDL	
801	LHIENIGIND	GFFDVGGSV	SAAVLVQHIS	QNFDCDMTLA	TLFSHPTIRQ	PCP
851	LAHYISSADS	ESVSRENIGR	GRTVADVSES	PIAIVGMAGK	FPGA <del>DL</del> DEF	
901	WLNISSGRNC	VTTIPESRWD	WKSHFGDPLS	AQGKTTIMWG	GFIDGVAAFD	
951	PLFFGIAPRE	AELMDPQQLR	LLTYTWLAL	HAAITPATLA	NQDTGVFMAA	
1001	SAGEYRDCTS	LDPEYGYRM	TSEMVMIPS	RISYLLNLRG	PSELCETTCS	
1051	SSLVAIHRAM	QSIHTGECKQ	AIVGAVNLLL	SPAGFSSYEA	MGLLTSHKQV	KS <sup>0</sup> <sub>1</sub>
1101	KSFQADASGY	IRAEGVVAI	LKPLSSARAD	GDRIYAVLKG	SAVFHGGGRM	
1151	SLTAPNGAGM	EAAMIRAYVE	AGIDPRSVAY	IEAHGIASAV	GDSIEIEALK	
1201	SAYRNAAERC	QSDLENTAQC	HLSCCLKPTMG	HGEVVS <del>G</del> MAA	LFKTVYALRH	
1251	QVIPGLPGFS	GLSEHISLTD	TAFVITADNH	PWDPLTDRHG	TTLPRRASIN	
1301	SYGFGASAHL	VLEEYSSEPE	FTASSANVAE	I I ILSAKNRT	QLQAMAENLC	trans-AT
1351	AFVARQPRLC	LGDLAYTLQV	GREAMRELA	IVAADLNELQ	DSLEQVIGNN	docking
1401	ISTVLGTGVT	AEPIDQEALE	IEVRNASEQG	DFQALARAWI	NGADISWMDL	
1451	HQHKRRRKIP	LPTYPFQKQA	YWATPYRQEE	MKPMTDQTRT	KKRICIIGAG	
1501	PAGLVMAKSL	LEEGHEPVIY	ETESVLGGIW	NIKADKTAGV	<del>YNS</del> TRFQNSA	
1551	DTSFFSDFPA	DTTDGFFLGV	DQVRAYLQAY	ASRFDIHQYI	HYSKI IAVT	
1601	EHGDQWKVDI	GEGDQQQTRY	FDGVAMCHGR	<del>Y</del> KHPFIPTIP	GLDQFQGEVL	
1651	HSGQYYDNRI	FAGKRVLVIG	NGVSGMDIAE	EASHVASAVF	WSM <del>RSL</del> RLVL	Ox
1701	<del>PM</del> VGYLPND	FISPANLLIS	KDNSIIMERL	KNSMPEYYEC	YQKSGLFPSL	Mutated positions in red bold letters
1751	EDFRANPFVH	INDGVIQ <del>R</del> VA	EGAIQTHVED	IERFTGRGCI	FSASGTHIEN	
1801	IDMVVLTGTY	DNSQSF <del>D</del> YVK	QFSMRDDFAM	GLFYRQNP <del>S</del> L	VNTYGLQNVG	

## SUPPORTING INFORMATION

1851	TGTLPYLEM	VARWYAQIIS	GNYTLDAAEEL	NHRAGEGEIV	VAPLANVIMG	
1901	LKLGLLPDPK	TEFQAFWRCL	NYPSFPPMYR	LRGPHADPQA	QSVLSRSVQR	
1951	SLIQQGEHDS	QLQTVKXRL	AGLGEEVMOA	LLARQEISQE	EYLOAQRCE	
2001	NAIVLSWDTQ	VIRPVKKASS	AVSQQASVPA	DRLAEEAFQO	RITELMSQTL	
2051	KLDVQGITAD	RHLSYDGFSS	VTLTAFSRKI	TDEYNIRLQP	FVFLEYTTLK	
2101	ALTDFLYRKW	SEQQPARTPV	NAVREDKQQL	TTHSVSTALP	TTASHEVAII	ACP <sub>1</sub>
2151	GIGGRLPGAEL	SLTEFWQNL	TNKSSISTVP	ADRSWQACD	GDPQDNKTD	
2201	CHYGAFIANV	GYFDPLHFN	SPREAPQLDP	QQRLLLEATW	ETLENAGYAK	
2251	NALYGKKGVC	FIGVERQDYA	ARARALGYPV	EGYLNTGN	SHAMLVNRVSHF	
2301	FGWQGPAMAL	DTACSSSFSA	IQTALKSLWT	GDAELALAGG	VNLVLPD	KS <sub>2</sub>
2351	ICNRKLGFT	GEPQVRPFDR	DANGHFFSDA	LGLLLKPLA	AAERDRDHIY	
2401	GVIKGLSVRH	GGQGMVTV	NPASHQEV	IQEALQANLQ	PSDIDYIEAQ	
2451	SGNTLTDIAE	LKAYHQVFGD	RDDPMPIGTI	KGHMGHFAGA	SGVIAVIKSL	
2501	LSFTNNQLVK	VENCQHLNWD	ADDEPFRCHI	LQQHQNLPR	RDRGRRIGVH	
2551	NFGFGVGTGH	LILEEHRQRD	QHLAEAPASE	YLIVFSARKA	EQLVAMAQRL	
2601	LSYLQNKYR	YHGFETISLR	DLAYTLQGR	EAMTQRLALL	VKNHAELLQK	trans-AT docking
2651	LQQIADGAAH	LNHCWRGNAE	EGCDGLTTQA	LRQANLSVLA	ESWCCGMTLD	oMT
2701	WEVLYHGQKP	AKVVLPTFPF	LTTTCYWI	DAVDPEPPHTDDK	PESAMPEAF	
2751	KGMIPTLNGT	GTMTRQLLSC	SEAFVDYAGQ	CDGEVLDLGC	AYGAAALSAL	
2801	ARGATVLAVD	MDARHLEILT	RQVNGEARQR	ISTQQGTVPE	LDLPGNRFGA	
2851	IHAGRVLHFL	SPQALQLT	LTMYRWLKP	GGKIFLTC	TPYFPQWAANVSD	
2901	YETRKQTGDL	WPGYIADLKH	YFASHADISG	YGDASAHADA	LQGMTEINLL	
2951	DPDILRRECQ	NAGFTVEQAE	FETLAVSVNG	ARPKGGVEHV	SLIAVKPVTA	
3001	TAEELTDATDE	SPAAELQLER	VLHSITEMLA	EELDLQDQEI	SANHKFFELG	
3051	LNSAIGVSWM	SRINERYGLL	VPAARLFAYP	TLQELARYLL	GEISKQPSDV	ACP <sub>2</sub>
3101	TNSTRQSSPS	RNESAASSVS	SFSVSSQQTQ	IATIVGLSCR	FPGSDNAEEFW	
3151	NTIQNKQDCI	DEVPPERW	NMADYFHADPEQ	PDKTDCRYMG	VLQDADCFDP	
3201	MFFNITPLEA	RHMDPQQLF	LQAAWSCLED	AGYASEVISG	SRCGVFAGYG	
3251	NYGQNMESG	LNAQRFLGGS	PSILAARIAY	FLNLQGPCLT	IDTACSASLV	KS <sub>3</sub>
3301	AVAEACNSLV	LGDCDMALAG	GISVLTGPDV	HIMATKAGML	STDGRCYTLD	
3351	QRANGFV	PSEGIGVVALKRL	SDAQKDNDII	YGLVQGWGVN	QDGRSNGITA	
3401	PNQQAQVRLQ	KHVDYKFGIN	PEHIQYVELH	GTGTRLGDPV	EVDALTQTFQ	
3451	SYTAKQQYCA	IGSVKSNVGH	ALAAAGIAGL	IKMLMAMQHR	QLPPTLHVED	
3501	VNEHIVLNQS	PFYINTESRD	WETDGVDRRR	AAISSFGYSG	TNAHVVEEY	
3551	VSPVPVAPAQ	LQPLLIVLSA	RTPERLRVMA	QQLHDFIGHH	PDLDLTNLAY	trans-AT docking
3601	TLQVGRQAMA	ARSGIMAESR	EDILEKLQAV	SAGAKTSGSG	WLGQSLQHEP	
3651	VLSLFSDDDD	MDKTITAWMH	KGKYAKLLSL	WVSGFDIDWQ	QLYQPASTAE	
3701	PSASPRRISL	PTYPFARERF	W			

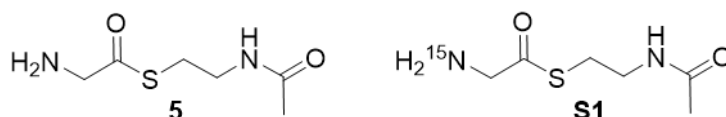
## Heterologous gene expression and protein purification

LB medium containing 50 µg/mL kanamycin was inoculated with the respective *E. coli* expression strain and incubated overnight at 37 °C. TB medium containing 0.4% glycerol and 50 µg/mL kanamycin was inoculated (1:100 with starter culture) and cultured at 37 °C until an OD<sub>600</sub> of 0.6–1.0. After cooling at 4 °C for 15 min, 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce gene expression. The culture was grown at 16 °C for 16–20 h and then harvested by centrifugation. Cell pellets were either frozen and stored at -80 °C or directly processed at 4 °C by resuspension in lysis buffer (50 mM sodium phosphate buffer, 300 mM NaCl, 10% [v/v] glycerol, pH 8.0). The Sonicator Q700 (QSonica, Newton, USA) was used to disrupt the cells and the lysate was centrifuged to remove cell debris. The supernatant was incubated with Ni-NTA agarose for 30 min at 4 °C and subsequently transferred onto a fretted column. The resin was washed once with 3 mL lysis buffer, once with 3 mL washing buffer (lysis buffer with 40 mM imidazole) and then eluted with elution buffer (lysis buffer with 250 mM imidazole). The proteins were desalted into lysis buffer and subsequently concentrated to 5–10 mg/mL by ultracentrifugation. For crystallization assays of LbmA-Ox-ACP, the His<sub>6</sub>-SUMO tag was cleaved by incubation of the protein in the presence of the SUMO protease ULP1 and subjected to a second round of purification with Ni-NTA agarose resin. The flowthrough was further purified by size exclusion chromatography (SEC) on a HiLoad™ 16/600 Superdex™ 200 Prep Grade column washed with SEC buffer (20 mM Tris pH 8.0, 200 mM NaCl, 5% [v/v] glycerol) and coupled to an ÄKTA Purifier FPLC system (GE Healthcare). The protein was finally concentrated to 20 mg/mL and stored at -80°C. The presence of protein in the elution fractions was assessed by SDS-PAGE.

## SUPPORTING INFORMATION

## Synthesis of test substrates and standards

*S*-(2-Acetamidoethyl) 2-aminoethanethioate; Gly-SNAC (**5**) and *S*-(2-acetamidoethyl) 2-(amino-<sup>15</sup>N)ethanethioate; <sup>15</sup>N-Gly-SNAC (**S1**)

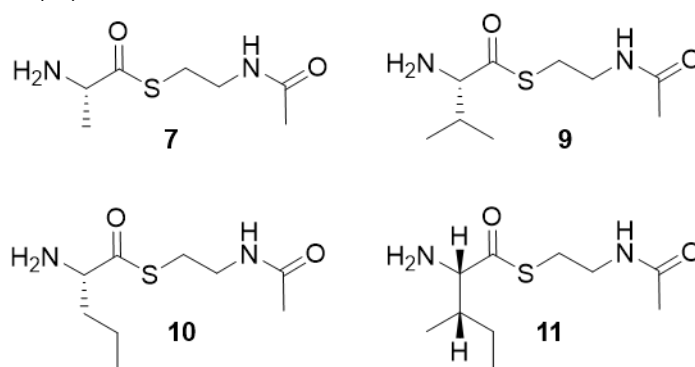


Coupling of *N*-(*tert*-butoxycarbonyl)glycine and *N*-acetylcysteamine was performed according to a previously published procedure to obtain *N*-(*tert*-butoxycarbonyl)glycine-SNAC.<sup>[7]</sup> At 0 °C, *N*-(*tert*-butoxycarbonyl)glycine (206.7 mg, 1.0 eq., 1.18 mmol) was dissolved in 5 mL dry dichloromethane (DCM), followed by the addition of a catalytic amount of 4-dimethylaminopyridine. Then, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (226.2 mg, 1.3 eq., 1.45 mmol) was added and the mixture was stirred for a few minutes. Finally, *N*-acetylcysteamine (132  $\mu$ L, 1.1 eq., 1.24 mmol) was added and the reaction was stirred overnight at room temperature. The reaction was quenched by addition of sat. aq. NH<sub>4</sub>Cl (15 mL) and extracted twice with DCM. The combined organic phases were washed with water and dried over MgSO<sub>4</sub>, and the organic phase was evaporated. For deprotection, 115 mg *N*-(*tert*-butoxycarbonyl)glycine-SNAC was dissolved in 4 mL 4 N dioxane-HCl on ice and stirred for 60 min. The solution was dried with compressed air and purified by flash chromatography using silica gel (5:1:1:1 butanol:water:acetic acid:acetone to 4:1:1:1 butanol:water:acetic acid:acetone).

**5** (70.7 mg, 0.40 mmol) was isolated as a white solid with a yield of 34%. HRMS (*m/z*): [M+H]<sup>+</sup> calcd. for C<sub>6</sub>H<sub>12</sub>N<sub>2</sub>O<sub>2</sub>S = 177.0692; obs. = 177.0693. <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  8.42 (s, br, 3H, NH<sub>2</sub>), 8.15 (t, *J* = 5.60 Hz, 1H, NH), 4.06 (s, 2H, C $\alpha$ H<sub>2</sub>), 3.22 (q, *J* = 6.50 Hz, 2H, CH<sub>2</sub>NH), 3.03 (t, *J* = 6.80 Hz, 2H, SCH<sub>2</sub>), 1.80 (s, 3H, COCH<sub>3</sub>). <sup>13</sup>C NMR (500 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  193.6 (C $\alpha$ C=O), 169.4 (NHC=O), 47.1 (C $\alpha$ ), 37.9 (CH<sub>2</sub>NH), 28.1 (SCH<sub>2</sub>), 22.6 (NHCH<sub>3</sub>).

**S1** was synthesized as described for **5**, starting with *N*-(*tert*-butoxycarbonyl)glycine-<sup>15</sup>N. **S1** (77.4 mg, 0.44 mmol) was isolated with impurities (*N*-acetylcysteamine) as a white solid with a yield of 18%. HRMS (*m/z*): [M+H]<sup>+</sup> calcd. for C<sub>6</sub>H<sub>12</sub>NO<sub>2</sub>S<sup>15</sup>N = 178.0663; obs. = 178.0659. <sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  8.04 (t, *J* = 5.40 Hz, 1H, NH), 7.68 (t, *J* = 6.20 Hz, 1H, NH<sub>2</sub>), 7.45 (t, *J* = 6.20 Hz, 1H, NH<sub>2</sub>), 3.81 (d, *J* = 6.20 Hz, 2H, C $\alpha$ H<sub>2</sub>), 3.14 (m, 2H, CH<sub>2</sub>NH), 2.87 (t, *J* = 6.60 Hz, 2H, SCH<sub>2</sub>), 1.78 (s, 3H, COCH<sub>3</sub>). <sup>13</sup>C NMR (400 MHz, DMSO):  $\delta$  199.4 (C $\alpha$ C=O), 169.3 (NHC=O), 50.1 (C $\alpha$ ), 38.7 (CH<sub>2</sub>NH), 27.9 (SCH<sub>2</sub>), 23.1 (NHCH<sub>3</sub>).

*S*-(2-Acetamidoethyl) 2-aminopropanethioate; Ala-SNAC (**7**), *S*-(2-acetamidoethyl) 2-amino-3-methylbutanethioate; Val-SNAC (**9**), *S*-(2-acetamidoethyl) 2-aminopentanethioate; Nva-SNAC (**10**), and *S*-(2-acetamidoethyl) 2-amino-3-methylpentanethioate; Ile-SNAC (**11**)



Synthesis of **7** and **9–11** was performed according to previously published procedures<sup>[8]</sup> with slight modifications.<sup>[9]</sup> At room temperature, Boc-*N*-L-amino acid (1.06 mmol, 1.0 eq.) was dissolved in 4–5 mL dry DCM. After the addition of diisopropylethylamine (740  $\mu$ L, 4.24 mmol, 4.0 eq.) and PyBOP (1.1 g, 2.12 mmol, 2.0 eq.), *N*-acetylcysteamine (0.12 mL, 1.16 mmol, 1.1 eq.) was added and the reaction was stirred overnight. The solvent was removed *in vacuo* and the crude mixture taken up in 25 mL of ethyl acetate. The organic layer was washed with 5% KHSO<sub>4</sub>, 5% NaHCO<sub>3</sub> and brine before being dried over MgSO<sub>4</sub>. The solvent was evaporated, and the resulting oil directly subjected to deprotection. At 0 °C, the residue was dissolved in 1 mL DCM and 1 mL trifluoroacetic acid (TFA) was added dropwise. The reaction was allowed to warm to room temperature and stirred overnight. Then, the TFA was removed *in vacuo* and the residue purified by silica gel chromatography (5:1:1:1 butanol/water/acetic acid/acetone to 4:1:1:1 butanol/water/acetic acid/acetone) to give the resulting SNAC conjugates as yellow or colorless oils.

**7** (141 mg, 0.64 mmol) was prepared following the general procedure stated above and obtained as a colorless oil with a yield of 59%. HRMS (*m/z*): [M+H]<sup>+</sup> calcd. for C<sub>7</sub>H<sub>14</sub>N<sub>2</sub>O<sub>2</sub>S = 191.0849; obs. = 191.0845. <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>):  $\delta$  8.44 (s, br, 3H, NH<sub>2</sub>), 8.10 (t, *J* = 5.50 Hz, 1H, NH), 4.31 (m, 1H, C $\alpha$ H), 3.23 (m, 2H, CH<sub>2</sub>NH), 3.03 (m, 2H, SCH<sub>2</sub>), 1.79 (s, 3H, COCH<sub>3</sub>), 1.44 (d, *J* = 7.20 Hz,

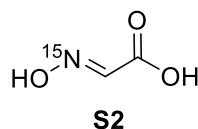
## SUPPORTING INFORMATION

3H, C<sub>β</sub>H<sub>3</sub>). <sup>13</sup>C NMR (500 MHz, DMSO-*d*<sub>6</sub>): δ 197.3 (C<sub>α</sub>C=O), 169.4 (NHC=O), 54.5 (C<sub>α</sub>), 37.8 (CH<sub>2</sub>NH), 28.3 (SCH<sub>2</sub>), 22.6 (NHCH<sub>3</sub>), 17.0 (C<sub>β</sub>).

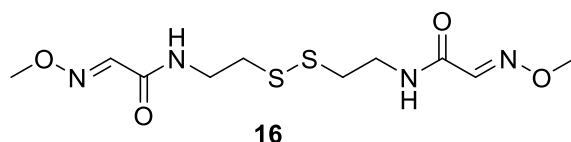
**9** (124 mg, 0.572) was prepared following the general procedure stated above and obtained as a colorless oil with a yield of 54%. HRMS (*m/z*): [M+H]<sup>+</sup> calcd. for C<sub>9</sub>H<sub>18</sub>N<sub>2</sub>O<sub>2</sub>S = 219.1162; obs. = 219.1155. <sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>): δ 8.44 (s, br, 2H, NH<sub>2</sub>), 8.09 (t, *J* = 5.40 Hz, 1H, NH), 4.15 (m, 1H, C<sub>α</sub>H), 3.23 (m, 2H, CH<sub>2</sub>NH), 3.06 (m, 2H, SCH<sub>2</sub>), 2.18 (m, 1H, C<sub>β</sub>H), 1.79 (s, 3H, COCH<sub>3</sub>), 0.97 (dd, *J* = 6.95, 14.20 Hz, 6H, C<sub>β</sub>(CH<sub>3</sub>)<sub>2</sub>). <sup>13</sup>C NMR (400 MHz, DMSO-*d*<sub>6</sub>): δ 196.2 (C<sub>α</sub>C=O), 169.4 (NHC=O), 63.5 (C<sub>α</sub>), 37.8 (CH<sub>2</sub>NH), 30.1 (C<sub>β</sub>), 28.4 (SCH<sub>2</sub>), 18.1, 17.3 (C<sub>γ</sub>).

**10** (141 mg, 0.64 mmol) was prepared following the general procedure stated above and obtained as a colorless oil with a yield of 61%. HRMS (*m/z*): [M+H]<sup>+</sup> calcd. for C<sub>9</sub>H<sub>18</sub>N<sub>2</sub>O<sub>2</sub>S = 219.1162, obs. = 219.1156. <sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>): δ 8.48 (s, br, 3H, NH<sub>2</sub>), 8.09 (t, *J* = 5.60 Hz, 1H, NH), 4.25 (t, *J* = 6.30 Hz, 1H, C<sub>α</sub>H), 3.22 (m, 2H, CH<sub>2</sub>NH), 3.05 (m, 2H, SCH<sub>2</sub>), 1.79 (s, 3H, COCH<sub>3</sub>), 1.77 (m, 2H, C<sub>β</sub>H<sub>2</sub>), 1.36 (m, 2H, C<sub>γ</sub>H<sub>2</sub>), 0.98 (t, *J* = 7.30 Hz, 6H, C<sub>δ</sub>H<sub>3</sub>). <sup>13</sup>C NMR (400 MHz, DMSO-*d*<sub>6</sub>): δ 196.8 (C<sub>α</sub>C=O), 169.4 (NHC=O), 58.0 (C<sub>α</sub>), 37.6 (CH<sub>2</sub>NH), 33.0 (C<sub>β</sub>), 27.9 (SCH<sub>2</sub>), 17.2 (C<sub>γ</sub>), 13.3 (C<sub>δ</sub>).

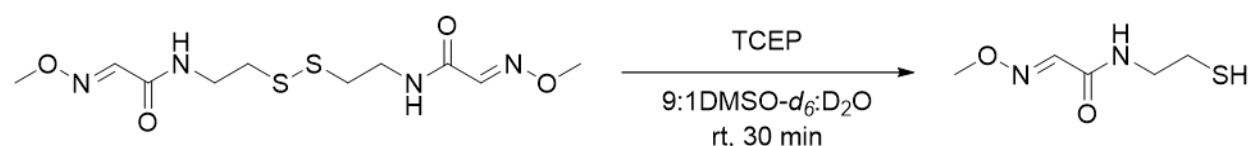
**11** (128 mg, 0.55 mmol) was prepared following the general procedure stated above and obtained as a colorless oil with a yield of 52%. HRMS (*m/z*): [M+H]<sup>+</sup> calcd. for C<sub>10</sub>H<sub>20</sub>N<sub>2</sub>O<sub>2</sub>S = 233.1318; obs. = 233.1318. <sup>1</sup>H NMR (500 MHz, DMSO-*d*<sub>6</sub>): δ 8.48 (s, br, 2H, NH<sub>2</sub>), 8.10 (s, br, 1H, NH), 4.19 (m, 1H, C<sub>α</sub>H), 3.24 (m, 2H, CH<sub>2</sub>NH), 3.07 (m, 2H, SCH<sub>2</sub>), 1.91 (m, 1H, C<sub>β</sub>H), 1.80 (s, 3H, COCH<sub>3</sub>), 1.47 (m, 1H, C<sub>γ</sub>H), 1.27 (m, 1H, C<sub>γ</sub>H), 0.95 (d, *J* = 6.91 Hz, 3H, C<sub>β</sub>CH<sub>3</sub>), 0.90 (t, *J* = 4.25 Hz, 3H, C<sub>δ</sub>H<sub>3</sub>). <sup>13</sup>C NMR (500 MHz, DMSO-*d*<sub>6</sub>): δ 195.9 (C<sub>α</sub>C=O), 169.4 (NHC=O), 62.7 (C<sub>α</sub>), 37.8 (CH<sub>2</sub>NH), 36.5 (C<sub>β</sub>), 28.5 (SCH<sub>2</sub>), 24.3 (C<sub>γ</sub>), 14.3 (C<sub>β</sub>CH<sub>3</sub>), 11.5 (C<sub>δ</sub>).

<sup>15</sup>N-Hydroxyiminoacetic acid (**S2**)

Synthesis of <sup>15</sup>N-hydroxyiminoacetic acid (**S2**) was performed according to a previously published procedure using <sup>15</sup>N-hydroxylamine.<sup>[10]</sup> HRMS (*m/z*): [M-H]<sup>-</sup> calcd. for C<sub>2</sub>H<sub>2</sub><sup>15</sup>NO<sub>3</sub> = 89.0011; obs. = 88.9991.

*(2E,2'E)-N,N'*-(Disulfanediy)bis(ethane-2,1-diyl)bis(2-(methoxyimino)acetamide) (**16**)

Methoxyimino acetic acid (25 mg, 1.0 eq., 0.243 mmol) was dissolved in dry DCM. At room temperature, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (56 mg, 1.5 eq., 0.364 mmol) and a catalytic amount of 4-dimethylaminopyridine were added. After stirring for 10 min at room temperature, the reaction mixture was cooled to 0 °C and cysteamine dihydrochloride (74 mg, 2.0 eq., 0.485 mmol) was added, and the reaction allowed to warm up to room temperature and stirred overnight. The reaction was quenched by addition of sat. aq. NH<sub>4</sub>Cl (15 mL) and extracted twice with DCM. The combined organic phases were washed with water and dried over MgSO<sub>4</sub>. The organic phase was evaporated and subjected to purification by preparative HPLC (water + 0.1% TFA, acetonitrile + 0.1% TFA, 20% — 98% 5 — 35 min, rt 22 min). Compound **16** was isolated as a white solid (30 mg, 0.093 mmol, 38%). HRMS (ESI) *m/z* 323.0842 ([M+H]<sup>+</sup>, calcd. for C<sub>10</sub>H<sub>19</sub>N<sub>4</sub>O<sub>4</sub>S<sub>2</sub> 323.0842, found 323.0829). <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ 7.41 (s, 2H, NCHCO), 3.98 (s, 6H, NOCH<sub>3</sub>), 3.68 (q, *J* = 6.45 Hz, 4H, CH<sub>2</sub>NH), 2.87 (t, *J* = 6.38 Hz, 4H, SCH<sub>2</sub>). <sup>13</sup>C NMR (400 MHz, CDCl<sub>3</sub>): δ 162.8 (C=O), 143.3 (CHNO), 63.8 (NOCH<sub>3</sub>), 38.7 (CH<sub>2</sub>NH), 38.3 (SCH<sub>2</sub>). Signals annotated by HSQC and HMBC. <sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>:D<sub>2</sub>O, 9:1): δ 7.50 (s, 2H, NCHCO), 3.71 (s, 6H, NOCH<sub>3</sub>), 3.42 (m, 4H, CH<sub>2</sub>NH), 2.81 (t, *J* = 6.85 Hz, 4H, SCH<sub>2</sub>).

*(E)-N*-(2-Mercaptoethyl)-2-(methoxyimino)acetamide (**15**)

The free thiol compound **15** was generated *in situ* and the reaction monitored by NMR according to a previously described procedure.<sup>[11]</sup> *(2E,2'E)-N,N'*-(Disulfanediy)bis(ethane-2,1-diyl)bis(2-(methoxyimino)acetamide) (**16**) (2 mg, 1.0 eq., 0.0062 mmol) was dissolved in



## SUPPORTING INFORMATION

600  $\mu\text{L}$  of a 9:1 mixture of  $\text{DMSO-}d_6$  and  $\text{D}_2\text{O}$ . Then, TCEP (2.3 mg, 1.5 eq., 0.0093 mmol) was added and the reaction progression monitored by NMR at room temperature. After 5 min, most of the starting material **16** was reduced. After 30 min, the starting material had been quantitatively converted to free thiol compound (*E*)-*N*-(2-mercaptoethyl)-2-(methoxyimino)acetamide (**17**).

HRMS (ESI)  $m/z$  163.0534 ( $[\text{M}+\text{H}]^+$ , calcd. for  $\text{C}_5\text{H}_{11}\text{N}_2\text{O}_2\text{S}$  = 163.0536, obs. = 163.0534).  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO-}d_6\text{:D}_2\text{O}$ , 9:1):  $\delta$  7.51 (s, 2H, NCHCO), 3.89 (s, 6H, NOCH<sub>3</sub>), 3.28 (m, 4H, CH<sub>2</sub>NH), 2.54 (t,  $J$  = 6.85 Hz, 4H, SCH<sub>2</sub>). Signals assigned based on literature<sup>[11]</sup> and dimer **16**.

### **In vitro** enzyme activity assay

The assay was set up in a volume of 100  $\mu\text{L}$ , containing purified LbmA-Ox (2–5  $\mu\text{M}$ ), substrate (2–5 mM), 1 mM NADPH, 2.5% [v/v] glycerol, and 50 mM sodium phosphate buffer pH 6. Negative controls included boiled LbmA-Ox. The reactions were incubated for 45 min at 25 °C. Absorption at 340 nm was measured on an Infinite M200 Pro (Plex) Tecan spectrometer at 2 scans per second. Reactions were extracted with ethyl acetate, dried under compressed air flow or *in vacuo*, and resuspended in methanol. The samples were centrifuged at maximum speed for 10 min and the supernatant was subjected to UHPLC-HRMS analysis. For assays in  $\text{H}_2^{18}\text{O}$ , assay buffer was lyophilized and stored at -20 °C. For the assay, the lyophilized pellet was resuspended in  $\text{H}_2^{18}\text{O}$ . For assays under  $^{18}\text{O}_2$ , the buffer was degassed under  $\text{N}_2$  flow and  $^{18}\text{O}_2$  was added. The remaining assay components were added under oxygen-free conditions. Extraction and analysis were performed as described above. For one-pot LbmA-Ox/LbmA-MT assays, purified LbmA-MT (2–5  $\mu\text{M}$ ), 0.5 mM *S*-adenosyl methionine (SAM), 150 mM NaCl, and 5 mM  $\text{MgCl}_2$  were added to the assay mixture. For cascaded LbmA-Ox/LbmA-MT assays, extracts from an LbmA-Ox assay were resuspended in water and used as substrate. The assay was set up in a volume of 100  $\mu\text{L}$ , containing purified LbmA-MT (2–5  $\mu\text{M}$ ), 2.5% [v/v] glycerol, 50 mM phosphate buffer pH 6, 0.5 mM SAM, 150 mM NaCl, and 5 mM  $\text{MgCl}_2$ . After incubation at 25 °C for 45 min, the reactions were extracted with ethyl acetate and subjected to UHPLC-HRMS analysis. To check for hydrolyzed products, the assay was maintained at 25 °C for an additional 48 h.

Enzyme activity assays with LbmA-Ox mutants were carried out in quadruplicate with 2 mg/ml purified protein (30  $\mu\text{M}$ ), 5 mM glycine SNAC, 1 mM NADPH, 1 mM FAD, 2.5% glycerol and 50 mM sodium phosphate buffer pH 6. Negative controls included no protein. The reactions were incubated for 60 min at 30 °C and worked up as described above.

### **In vivo** assay

The LbmA bimodule was expressed as described above using TB medium containing 4% glycerol. For assays using labelled methionine, the medium was supplemented with 5 mM L-methionine-(*methyl*- $^{13}\text{C}$ ). The cell lysate after sonication was incubated with 0.2 M cysteamine hydrochloride for 3 hours at 30 °C with occasional shaking, and 400  $\mu\text{L}$  were extracted three times with 400  $\mu\text{L}$  ethyl acetate. The organic phases were combined, the solvent was evaporated, and the residue was resuspended in 100  $\mu\text{L}$  methanol. The samples were centrifuged at 20,000  $\times$  g for 15 min and the supernatant was subjected to UHPLC-HRMS analysis. After checking that **14** is present in the sample, it was treated with 1 mM tris(2-carboxyethyl)phosphine (TCEP, stock 1 M in 9:1  $\text{DMSO}/\text{H}_2\text{O}$ ), mixed briefly and incubated for 20 min. The sample was reanalyzed by UHPLC-HRMS.

### FAD analysis

The LbmA-Ox protein was boiled for 10 min at 95 °C and centrifuged for 10 min at maximum speed. The supernatant was diluted 1:10 with 50 mM phosphate buffer pH 6 and then subjected to UHPLC-HRESI-MS and the mass spectrum and retention time were compared with an FAD standard. Likewise, UV-Vis spectra were recorded using a Nanodrop spectrophotometer for the supernatant (and the FAD standard), revealing local maxima at 374 (375) and 450 (451) nm. The UV-Vis measurement was repeated with native purified protein, resulting in slightly shifted local maxima at 377 and 443–445 nm. Therefore, the absorbance at 443 nm was used to quantify the amount of bound oxidized FAD for all LbmA-Ox mutants, with an extinction coefficient of  $\epsilon = 11,300 \text{ M}^{-1} \text{ cm}^{-1}$ .

### UHPLC-HRMS analysis

A solvent gradient (A =  $\text{H}_2\text{O}$  + 0.1% formic acid and B = acetonitrile + 0.1% formic acid) was used to inject 10  $\mu\text{L}$  sample on a Kinetex 2.6  $\mu\text{m}$  C18 100 Å, LC Column (150  $\times$  4.6 mm; Phenomenex). The following gradient of solvent B was employed: 2% for 0–2 min, 2–60% for 2–7 min, 60–95% for 7–10 min, 95% for 10–12 min at a flow rate of 1.0 mL/min and the column oven set to 27 °C. MS measurements were conducted in positive ionization mode (or negative ionization mode, for  $^{15}\text{N}$ -hydroxyiminoacetic acid) in a mass range of 50–250  $m/z$ . The spray voltage was set to 3.5 kV and the capillary temperature to 300 °C. MS/MS data were acquired in a data-dependent fashion with the parent ion scan at a resolution of 70,000 and the MS/MS scan at a resolution of 17,500. The 10 most abundant peaks of each parent ion scan were subjected to CID fragmentation with a normalized collision energies of 15, 20 and 25, and the dynamic exclusion time was set to 10 sec. MS/MS scans were conducted with an AGC target of  $3 \times 10^6$  or a maximum injection time of 150 ms. Acquired MS data were analyzed with the Thermo Xcalibur 2.2. software. For the collection of MS and MS data for **13**, a targeted selected ion monitoring (tSIM) using an inclusion list of 163.0536 and a selection window of 5 ppm was applied.

## SUPPORTING INFORMATION

To carry out FAD analysis, the same UHPLC column and settings were used with the following alterations: The gradient was solvent B at 5% for 0–2 min, 5–50% for 2–4 min, 50–95% for 4–10 min, 95% for 10–13 min, 95–5% for 13–13.1 min, 5% for 13.1–15.1 min with the column oven set to 30 °C. MS measurements were conducted in a mass range of 400–850  $m/z$  with the capillary temperature set to 250 °C.

**Kinetic and  $^{15}\text{N}$  NMR analysis**

NMR experiments were performed on a Bruker AVNEO 700 MHz spectrometer equipped with a CP-QCI H&F-N/C/P-D 05 Z probe or on a Bruker AVNEO 500 MHz spectrometer equipped with a CP-TCI H-C/N-D 05 Z probe using the topspin 4.0.7 software for data acquisition and processing. NMR data were recorded at 298 K in standard 5 mm NMR tubes. To follow the enzymatic reaction, a series of 1D  $^1\text{H}$  experiments with a 3-9-19 watergate sequence<sup>[12]</sup> was applied. A total of 128 transients was averaged with the number of datapoints for the FID set to 32768. Protons which are scalar coupled to  $^{15}\text{N}$  are selectively detected with isotope edited experiments. Isotope edited experiments result in clean spectra with only the signals of interest from the  $^{15}\text{N}$  labeled substrate or product observable.

To observe protons which are scalar coupled to  $^{15}\text{N}$ , an isotope edited 1D experiment with watergate<sup>[13]</sup> for water suppression was performed. The experimental parameters were set to observe one of the following moieties: *E*- or *Z*-oxime, amide, imine, or amide protons coupled to  $^{15}\text{N}$ . The INEPT delays for the isotope filters were set according to the coupling constant  $J_{\text{H-N}}$ . The carrier offset for  $^{15}\text{N}$  was set to the reported values for  $^{15}\text{N}$  in the corresponding functional group. A total of 128 scans was averaged for the kinetic experiments and 5120 transients were averaged at the end of the reactions with the size of the FID set to 2794 complex data points. NMR samples containing the enzyme, NADPH and the substrate in assay buffer were prepared immediately before data acquisition.

For quantification, sodium trimethylsilylpropanesulfonate (DSS) was added as an internal standard. The first time point was acquired approximately 5 min after sample mixing and experiments were repeatedly acquired up to 10 hours after sample mixing. For each time point, a 1D and an isotope edited 1D experiment were performed. The total experiment time for both experiments is about 20 min. At the end of the reaction, an isotope edited 1D experiments and 2D  $^{15}\text{N}$ - $^1\text{H}$  HSQC experiments employing watergate<sup>[13]</sup> water suppression were performed to identify possible reactions products.

Chemical shifts and scalar coupling constants (values for  $^{15}\text{N}$  are from Witanowski<sup>[14]</sup>):

Functional group	$\delta(^{15}\text{N})$ / ppm	$J_{\text{H-N}}$ / Hz	$\delta(^1\text{H})$ / ppm
<i>E</i> -oxime	350	2.7	7–8
<i>Z</i> -oxime	350	13.9	7–8
amine	30	1	4–5
amide	120	90	7–8
imine	340	4	10
hydroxylamine	120	1	3

NMR samples containing the enzyme, NADPH and the substrate in assay buffer were prepared immediately before data acquisition. For quantification, sodium trimethylsilylpropanesulfonate (DSS) was added as an internal standard. The first time point was acquired approximately 5 min after sample mixing and experiments were repeatedly acquired up to 10 hours after sample mixing. For each time point, a 1D and an isotope edited 1D experiment were performed. The total experiment time for both experiments is about 20 min. At the end of the reaction, an isotope edited 1D experiments and 2D  $^{15}\text{N}$ - $^1\text{H}$  HSQC experiments employing watergate<sup>[13]</sup> water suppression were performed to identify possible reactions products.

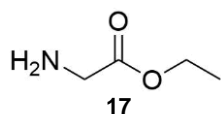
**Crystallization of LbMA-Ox-ACP, X-ray diffraction data collection and structure refinement**

The purified LbMA-Ox-ACP was incubated with 2 mM ethyl glycinate (**17**) and 2 mM  $\text{NADP}^+$  prior to crystallization. Initial crystallization hits were obtained using the Wizard Precipitant Synergy kit (Molecular Dimensions). Crystals were grown by the vapor diffusion method using a 1:1 ratio of 10 mg/mL LbMA-Ox-ACP with the well solution containing 1.8 M potassium phosphate dibasic/sodium phosphate monobasic pH 6.0 and 2% [v/v] PEG 400. Crystals were cryoprotected with a solution of the well buffer containing 30% glycerol. The crystals were then flash-frozen. X-ray diffraction data were collected on the beamline Proxima-2A at the SOLEIL synchrotron. The data set was indexed and integrated with XDS<sup>[15]</sup> and scaled using pointless and aimless.<sup>[15]</sup> The LbMA-Ox-ACP structure was solved by molecular replacement using Molrep<sup>[15c, 16]</sup>, with a computed AlphaFold2<sup>[17]</sup> structure of LbMA-Ox as a search model. The initial backbones of the ACP were then built using Buccaneer<sup>[15c, 18]</sup>. The overall structure was manually built and refined through iterative processing using Coot<sup>[19]</sup> and Phenix<sup>[20]</sup>. Statistics are reported in Table S1.

During the first refinement cycles, the electron density of chain B remained weaker than chain A, while B factors for the overall chain B increased. The crystal structure reveals that the two domains of chain A are responsible for the crystal packing. Surprisingly, chain B is located in the solvent columns of the crystals, belongs to the P6<sub>5</sub>22 space group, and is less stabilized within the lattice. Accordingly, the structure was refined with an occupation of 0.5 for chain B atoms, which is consistent with an equilibrium between the presence and absence of chain B in the crystals and their poor diffraction quality.

## SUPPORTING INFORMATION

*Ethyl 2-aminoacetate; Ethyl glycinate; EGE (PDB ID: GEE) (17)*



## SUPPORTING INFORMATION

## Supplementary tables

Table S1. Data collection and refinement statistics for LbmA-Ox-ACP.

Data collection	LbmA-Ox-ACP
Space group	P 6 <sub>2</sub> 22
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	206.75, 206.75, 178.40
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00, 90.00, 120.00
Wavelength	0.980104
Resolution (Å)	47.84–3.07 (3.18–3.07)
<i>R</i> <sub>merge</sub> (%)	13.7 (67.4)
<i>R</i> <sub>pim</sub> (%)	6.9 (34.1)
<i>I</i> / $\sigma$ <i>I</i>	9.7 (2.3)
Number of total observations	239189 (24078)
Number of unique reflections	42439 (4264)
Completeness (%)	99.5 (96.6)
Redundancy	5.6 (5.6)
Wilson <i>B</i> -factors (Å <sup>2</sup> )	46.867
CC <sub>1/2</sub>	0.994 (0.718)
<b>Refinement</b>	
Resolution (Å)	47.84–3.07
No. reflections	42381
<i>R</i> , <i>R</i> <sub>free</sub>	0.2383, 0.2915
No. atoms	9778
Protein	9409
FAD/NADP+	202
Other ligands/ions	35
Water	132
Average B-factors, all atoms (Å <sup>2</sup> )	50.0
R.m.s deviations	
Bond lengths (Å)	0.013
Bond angles (°)	1.38
<b>PDB ID</b>	<b>8ODW</b>

## SUPPORTING INFORMATION

**Table S2.** Accession numbers for PKS pathways shown in Figure 1.

Natural product pathway	Organism	Identifier
Lobatamide	<i>Gyvuella sunshinyii</i>	WP_132721172 for LbmA
Oximidine III	<i>Pseudomonas baetica</i>	PKA71712.1–PKA71714.1
Oximidine III	<i>Chromobacterium phragmitis</i>	WP_168194823.1, WP_114060631.1, WP_168194822.1
Necroxime	<i>Burkholderia</i> sp. B8	QIE07362.1–QIE07364.1 for NecC–E
Apicularen A	<i>Chondromyces apiculatus</i>	WP_081865114.1

## SUPPORTING INFORMATION

**Table S3.** Primers used in this work.

Primer Name	Primer Sequence
LbmA-Ox-F	CGGACTAGTCCGGAAATCATCATTCTC
LbmA-Ox-R	TTGCGCCGCTTTCAGATACGTTGC
LbmA-Ox-ACP-F	TTCCGGATCCATGACAGACCAGACCAGGACGAAA
LbmA-Ox-ACP-R	TTCCAAGCTTTCAAACAGCTGAGGACGCCTTTTTGA
LbmA-MT-F	CGGACTAGTCCGAGCGCGATGCCAGAGGCATTTC
LbmA-MT-R	TTGCGCCGCTTTGACAGCTCTGCCGCTGGGGATTTC
LbmA-M3.1-F	ACCTGCTTGGCGAAATATCGCGGCCGCACTCGAGCACCACCACCACCACC
LbmA-M3.1-R	TCGTTGAAAGAAAGTTTCATCTAGTGATCCACCAATCTGTTCTCTGTGAGCCTCAATAATATCGTTATCCTCC
LbmA-M3.2-F	CAGATTGGTGGATCCACTAGATGAAACTTTCTTTCAACGAATTGTTACACTTAAAGAGCAGCTTTC
LbmA-M3.2-R	TTTCATTTCTCTGTCTGTACGGGGTGGCCAGTAGGCTTTCTGATCAA
LbmA-M3.3-F	AGCCTACTGGGCCACCCCGTACAGACAAGAGGAAATGAAACCAATGACAGACC
LbmA-M3.3-R	GTGGTCTCGAGTGCAGGCCGCGATATTTCCGAAGCAGGTACCGGGCAAG
LbmA-Ox-N1548D	TACAACAGCACCCGGTTTCAGGATCCGCCGATACCAGCTTTTTTC
LbmA-Ox-N1548A	TACAACAGCACCCGGTTTCAGGCATCCGCCGATACCAGCTTTTTTC
LbmA-Ox-N1709D	ATGGTTGGTTATCTGCCAGATGATTTTATTTCCACCGGCTAATCT
LbmA-Ox-N1709A	ATGGTTGGTTATCTGCCAGCAGATTTTATTTCCACCGGCTAATCT
LbmA-Ox-T1851V	TATGGACTGCAAAATGTCGGCGTGACCGGAACCCCTCCGTAC
LbmA-Ox-Q1547E	GTATACAACAGCACCCGGTTTGAAAATTCGCCGATACCAGCTTTT
LbmA-Ox-Q1547A	GTATACAACAGCACCCGGTTTGCGAATTCGCCGATACCAGCTTTT
LbmA-Ox-Y1541F	GACAAAACCGCCGGGGTATTCAACAGCACCCGGTTTCAGAATTC
LbmA-Ox-T1544V	GCCGGGGTATACAACAGCGTGCGGTTTCAGAATTCGCCGAT
LbmA-Ox-R1702A	CTGCGACTGGTATTGCCGGCGATGGTTGGTTATCTGCCAAACG
LbmA-Ox-D1677N	GGTAACGGCGTCAGCGGTATGAACATCGCCGAAGAAGCCAGTCAT
LbmA-Ox-D1677A	GGTAACGGCGTCAGCGGTATGGCAATCGCCGAAGAAGCCAGTCAT
LbmA-Ox-E1680Q	GTCAGCGGTATGGATATCGCCCAGGAAGCCAGTCATGTCGCCAG
LbmA-Ox-E1680A	GTCAGCGGTATGGATATCGCCGAGAAGCCAGTCATGTCGCCAG
LbmA-Ox-E1521A	GGACATGAGCCGGTCATCTATGCAACAGAATCAGTCCTGGGTGG
LbmA-Ox-K1632A	CATGTGCATGGCCGTTACGCGCATCCGTTTCATTCCAATTTCC
LbmA-Ox-R1694A	GCCGTGTTCTGGTCCATGGCGTCCCTGCGACTGGTATTG
LbmA-Ox-R1697A	TTCTGGTCCATGCGATCCCTGGCGCTGGTATTGCCGCGCATGG
PCP-S819A	GATGTTGGCGGGACAGCGGTTTCCGCAGCGG
ACP1-S2070A	GTGACTATGGTTTCAGCGCGGTCACACTGACCGC
ACP2-S3053A	GTTTTTTGAACTTGGGCTGAATGCGGCCATCGGCGTATC

## SUPPORTING INFORMATION

**Table S4.** Data for UV-Vis measurements with different LbmA-Ox mutant strains. M1–M6 give the values for absorbance at 443 nm,  $C_{\text{FAD}}$  is the concentration of FAD in mol/l (calculated with an extinction coefficient of  $\epsilon = 11.300 \text{ M}^{-1} \text{ cm}^{-1}$  and a path length of  $b = 1 \text{ mm}$ ).  $C_{\text{P}}$  is the protein concentration in mol/l (calculated with a MW = 73.99 kg/mol). Inc. indicates the incorporation of FAD into the protein in %.

Strain	M1	M2	M3	M4	M5	M6	Mean	$C_{\text{FAD}}$	$C_{\text{P}}$	Inc.
WT	0.944	0.929	0.965	0.941			0.943	0.834	1.527	54.6
R1694A	0.677	0.682	0.681				0.681	0.603	0.866	69.6
E1521A	0.018	0.02	0.007				0.018	0.016	0.747	2.1
E1680A	0.599	0.616	0.609	0.593			0.604	0.535	0.559	95.6
E1680Q	0.42	0.423	0.42				0.420	0.372	0.522	71.2
K1632A	0.949	1	0.964	0.966			0.965	0.854	1.148	74.4
N1709A	1.018	0.994	0.985	0.988			0.991	0.877	1.584	55.4
N1709D	0.418	0.418	0.426				0.418	0.370	1.556	23.8
Q1547A	0.645	0.655	0.646				0.646	0.572	0.684	83.6
Q1547E	0.694	0.704	0.71	0.695			0.700	0.619	0.660	93.8
N1548A	0.74	0.734	0.759	0.764	0.783	0.774	0.759	0.672	0.754	89.1
N1548D	0.7	0.667	0.679				0.679	0.601	1.870	32.1
D1677A	0.166	0.164	0.162				0.164	0.145	0.754	19.3
D1677N	0.444	0.457	0.446	0.455			0.451	0.399	1.954	20.4
T1544V	0.444	0.446	0.432				0.444	0.393	0.711	55.2
R1702A	0.294	0.301	0.297				0.297	0.263	0.887	29.6
Y1541F	0.795	0.821	0.761	0.784			0.790	0.699	0.733	95.3
R1697A	0.516	0.496	0.51				0.510	0.451	0.655	68.9
T1851V	1.173	1.188	1.168	1.148			1.171	1.036	1.690	61.3

## SUPPORTING INFORMATION

**Table S5.** Peak areas calculated with Xcalibur Quan software for UHPLC-HRMS measurements of different LbmA-Ox mutant strains that were used to estimate substrate conversions. The respective extracted ion chromatograms (EICs) are shown in Figures S38–61. The mean conversion in % was normalized to 100% conversion by the wildtype. Sample sets are separated by horizontal black lines and \* indicates outliers that were not considered for the calculations.

Sample name	Peak area substrate	Peak area product	Mean conv.
WT1_1	237803141.59	20042564.77	
WT1_2	205222689.75	14859454.67	
WT1_3	172127547.52	15643808.19	
WT1_4	203039438.17	15525650.89	100±9
R1694A_1	182621407.94	8568843.30	
R1694A_2	173169092.94	7653315.45	
R1694A_3	156866747.72	6880402.85	
R1694A_4	160722705.34	6193102.46	54±4
E1521A_1*	137868098.89	3915753.18	
E1521A_2	151923554.88	2570345.14	
E1521A_3	139177123.46	3051980.27	
E1521A_4	135341028.41	2701823.04	27±3
E1680Q_1	133348294.78	7067138.71	
E1680Q_2	121372511.97	6071673.64	
E1680Q_3	112488818.39	5973165.29	
E1680Q_4	103015331.77	6479367.87	68±7
T1851V_1	171243489.81	19122458.16	
T1851V_2	161453819.07	16620600.05	
T1851V_3	182185603.23	18829121.26	
T1851V_4	167399998.88	15999013.75	128±8



## SUPPORTING INFORMATION

Table S5. Continued.

Sample name	Peak area substrate	Peak area product	Mean conv.
WT6_1	94910004.31	17152302.28	
WT6_2	99864511.17	18533501.82	
WT6_3	99575494.74	18920125.11	
WT6_4	97077992.65	17051394.89	100±3
K1632A_1	87922758.56	33515288.40	
K1632A_2	80699024.25	31247322.10	
K1632A_3	86578261.57	33823368.55	
K1632A_4	76175256.26	30639345.39	213±5
N1709D_1	74968731.01	3004050.98	
N1709D_2	76723954.91	2665365.96	
N1709D_3	69818638.49	2886194.64	
N1709D_4	66892377.50	2613925.02	21±2
Q1547E_1	71513652.43	59243.08	
Q1547E_2	71286012.99	47796.35	
Q1547E_3	68521996.21	73086.86	
Q1547E_4	80716567.63	74069.92	0.5±0.1
WT10_1	113793819.71	10139673.45	
WT10_2	138427444.41	12090710.93	
WT10_3	99598996.98	13525128.76	
WT10_4	123332758.41	15862350.28	100±20
N1548D_1	85542406.21	1645094.59	
N1548D_2	87251697.77	1667099.64	
N1548D_3	76964085.26	1477508.54	
N1548D_4	72498225.03	1438887.79	18±0.3
D1677N_1	56411861.24	118034.06	
D1677N_2	63607994.26	66166.26	
D1677N_3	67535699.10	78441.63	
D1677N_4	62497377.07	71632.57	1±0.4
N1709A_1	27642406.28	680474.63	
N1709A_2	52194363.24	639510.64	
N1709A_3	28289977.11	569112.17	
N1709A_4	28679164.12	488045.66	17±5

## SUPPORTING INFORMATION

Table S5. Continued.

Sample name	Peak area substrate	Peak area product	Mean conv.
WT15_1	127191637.21	13392131.44	
WT15_2	111521078.64	10752650.39	
WT15_3	114926281.29	11795851.36	
WT15_4	138540133.44	12706610.69	100±5
Q1547A_1	126990383.02	5958.89	
Q1547A_2	87934664.40	0	
Q1547A_3	89319833.84	0	
Q1547A_4	103857407.21	0	0.0±0.02
N1548A_1	83161222.03	20360.93	
N1548A_2	74550713.98	16812.65	
N1548A_3	59064588.33	15095.68	
N1548A_4	46911573.87	15328.58	0.3±0.04
R1702A_1*	82059656.84	2578287.65	
R1702A_2	70752926.35	1497681.90	
R1702A_3	80815670.70	1533428.23	
R1702A_4	63057937.71	1531479.01	24±3
WT19_1	121771079.41	11348916.75	
WT19_2	111585917.54	9270376.10	
WT19_3	116076266.42	10346847.74	
WT19_4	116888628.79	10330323.45	100±4
Y1541F_1	106431736.32	1499661.92	
Y1541F_2	82262205.45	1301919.39	
Y1541F_3	77706953.14	1287009.80	
Y1541F_4	90102361.10	1447414.23	18±1
D1677A_1	72880767.83	147931.03	
D1677A_2	62796350.64	111711.69	
D1677A_3	67467753.61	114161.79	
D1677A_4	63000487.72	79426.96	2±0.4
R1697A_1	78019179.29	6638175.21	
R1697A_2	68017098.77	6753623.20	
R1697A_3	61048977.48	6232355.32	
R1697A_4*	41162243.98	6879560.93	128±10
E1680A_1	47060268.74	6491742.44	
E1680A_2	38512099.76	4724112.44	
E1680A_3	45065546.83	5486373.71	
E1680A_4	37335672.08	5031204.23	146±9

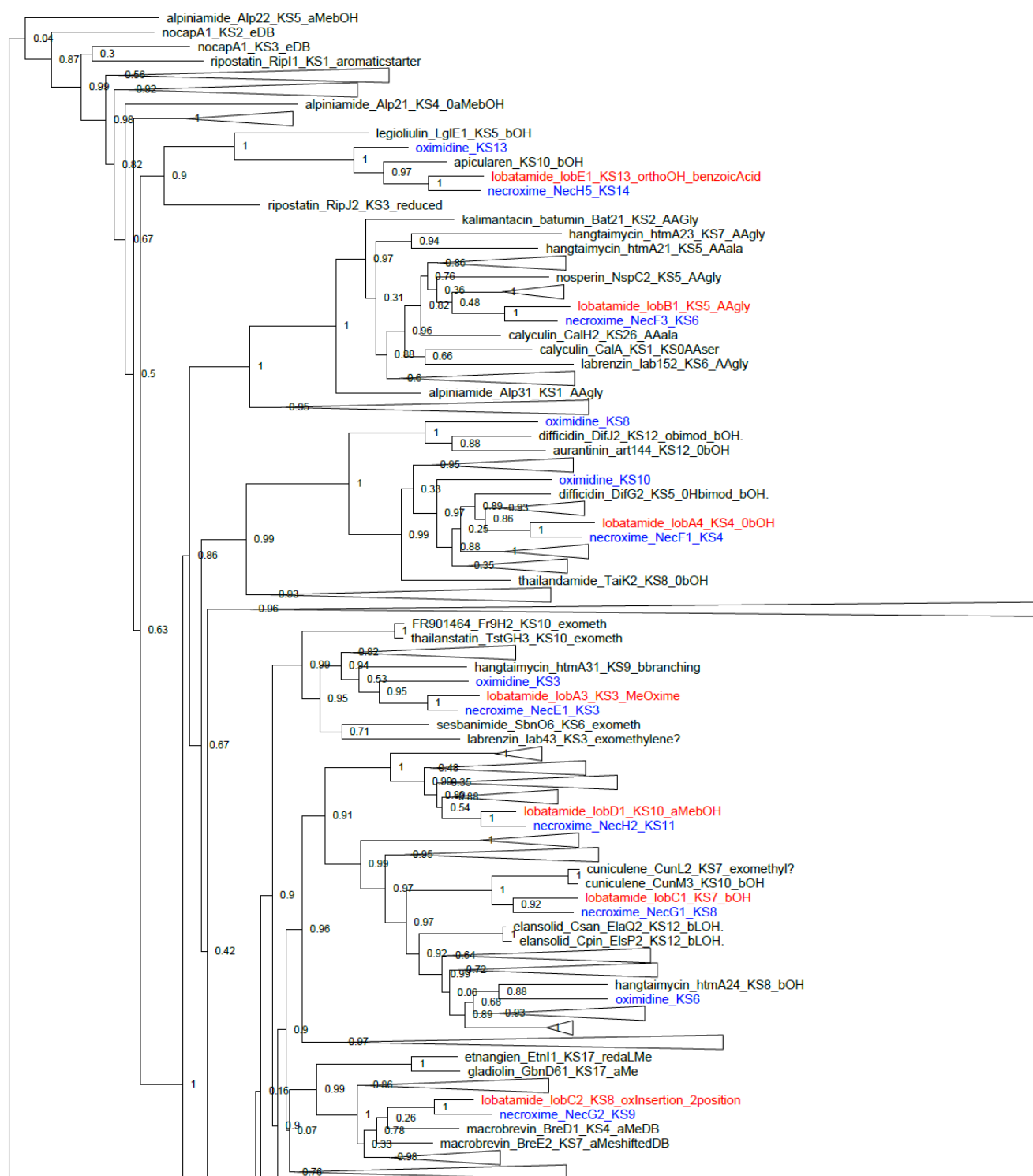
## SUPPORTING INFORMATION

Table S5. Continued.

Sample name	Peak area substrate	Peak area product	Mean conv.
WT20_1	228325862	27647446	
WT20_2	207052133	25980046	
WT20_3	222299720	26325736	
WT20_4	202243699	26146992	100±3
T1544V_1	196104704	16654351	
T1544V_2	195245271	15740143	
T1544V_3	189323594	15882150	
T1544V_4	191295735	18687132	69±5

## SUPPORTING INFORMATION

## Supplementary figures



**Figure S1.** Maximum likelihood phylogram of 1134 KS domains from 88 trans-AT PKS clusters using two KSs from the cis-AT PKS for erythromycin as outgroup (bottom). KS sequences are named as compound\_protein\_number of KS\_assigned biochemical transformation. Ketide clades for necroxime, oximidine, and lobatamide are expanded. Lobatamide KSs are marked in red, necroxime and oximidine KSs in blue. Nodes are labelled with their numerical bootstrap values. The figure is continued on the following pages.

## SUPPORTING INFORMATION

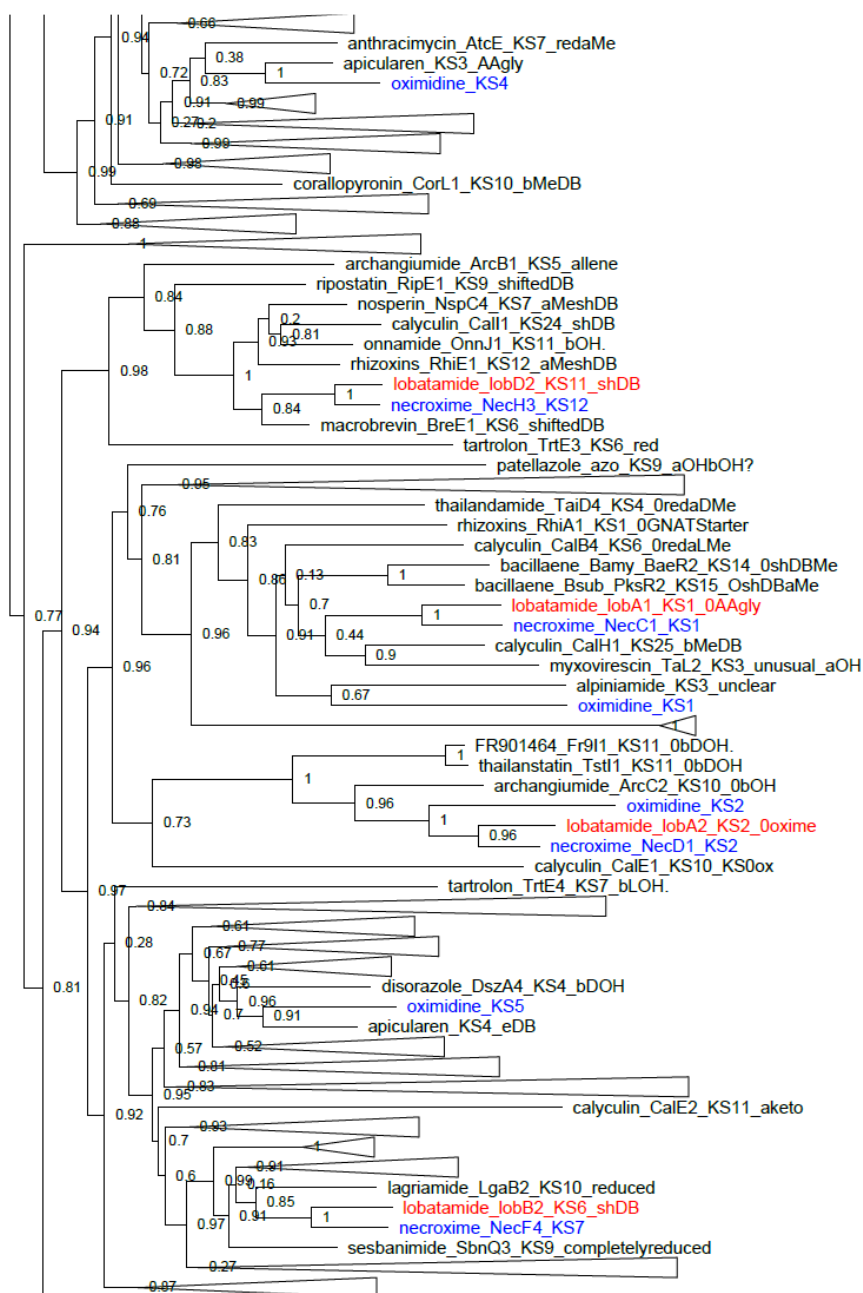


Figure S1. Continued.

## SUPPORTING INFORMATION

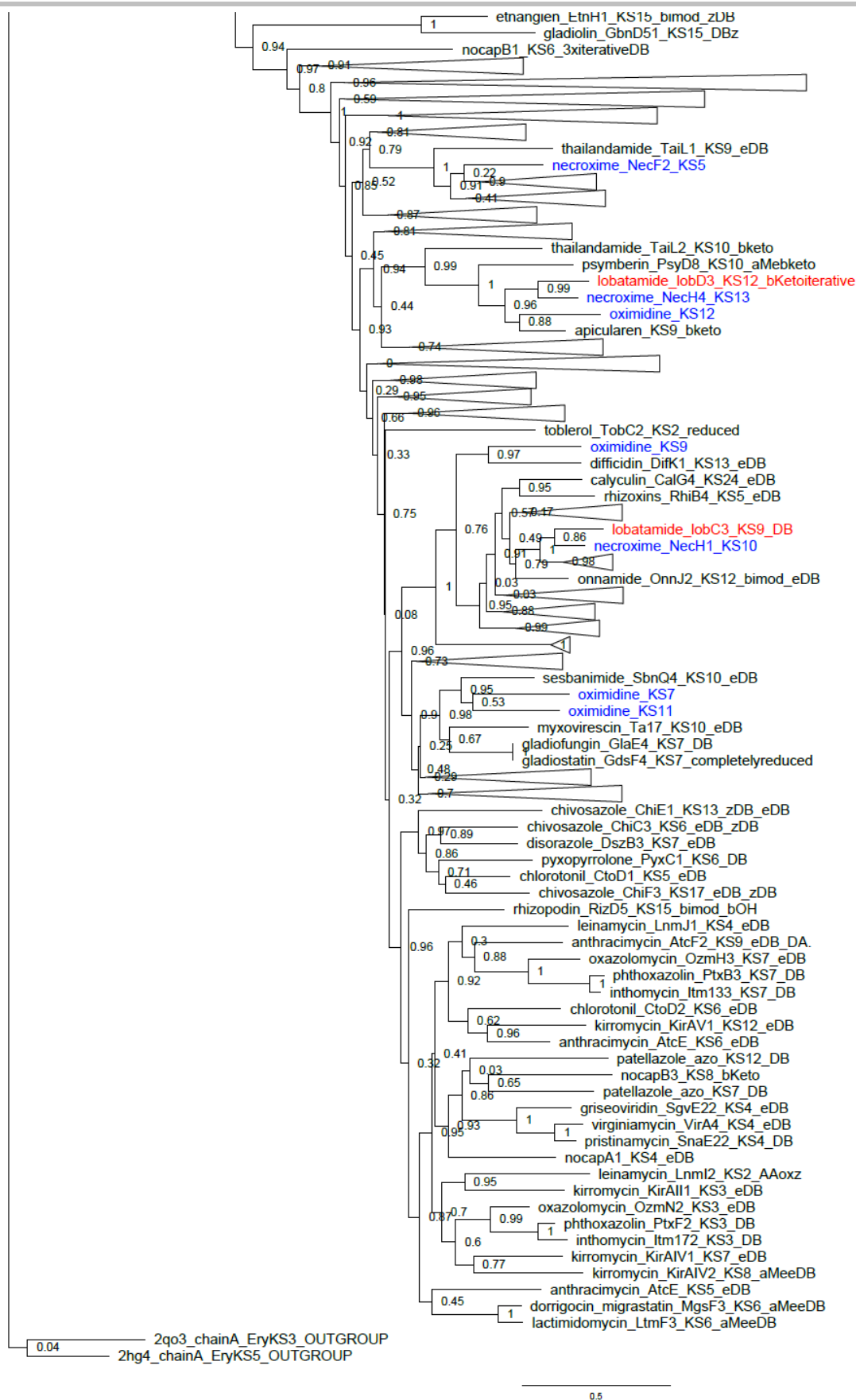
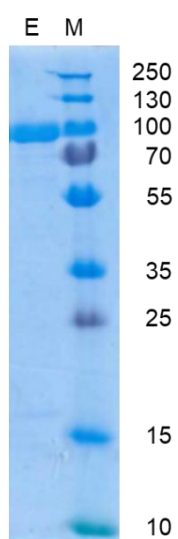


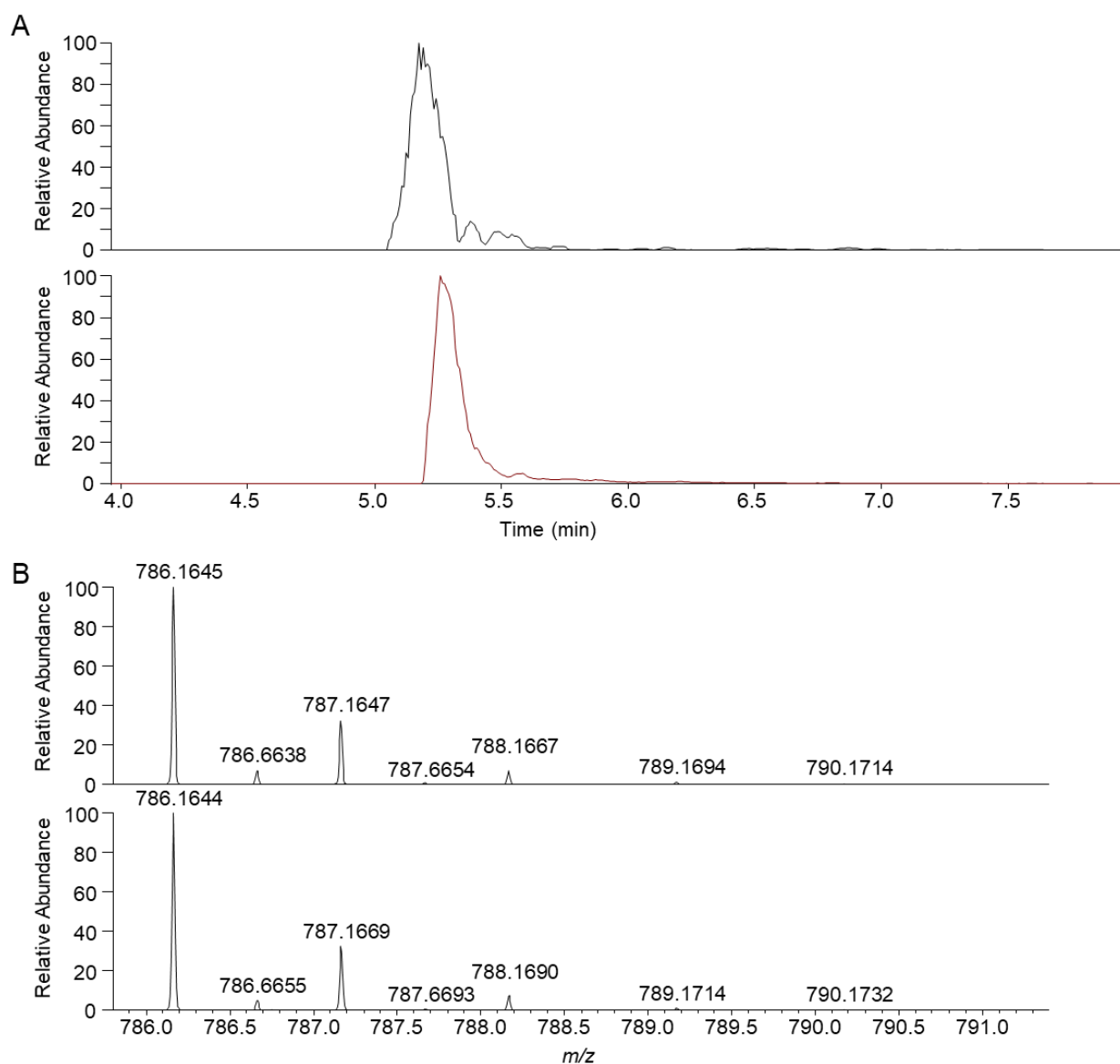
Figure S1. Continued.

## SUPPORTING INFORMATION



**Figure S2.** 12% SDS-PAGE gel of His<sub>6</sub>-SUMO-tagged LbmA-Ox post Ni-NTA affinity purification. Expected molecular weight: 91.3 kDa. E: elution fraction, M: protein ladder.

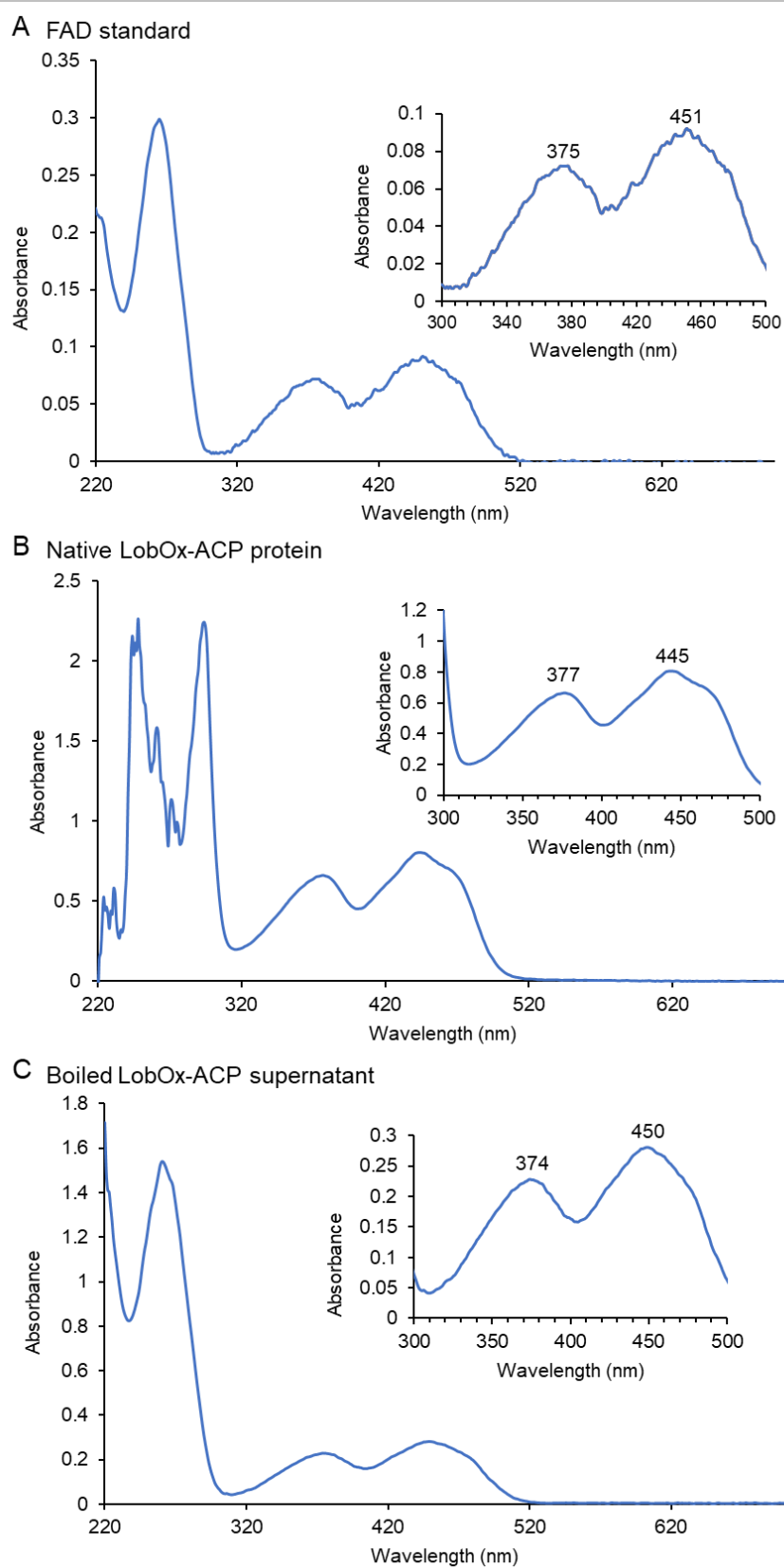
## SUPPORTING INFORMATION



**Figure S3.** UHPLC-HRMS data for the FAD analysis. (A) EIC of the FAD standard (calcd. for  $C_{27}H_{34}P_2N_9O_{15}$   $[M+H]^+ = 786.1645$   $m/z$ , upper, intensity =  $4.46 \times 10^6$ ) and of FAD extracted from purified LobOx-ACP (lower, intensity =  $7.27 \times 10^5$ ). (B) Mass spectrum of the FAD standard (upper) and of FAD extracted from purified LobOx-ACP (lower) at a retention time of 5.22 min.



## SUPPORTING INFORMATION



**Figure S4.** UV-Vis analysis of the flavin cofactor. (A) UV-Vis spectrum of the FAD standard with absorbance maxima at 375 and 451 nm. (B) UV-Vis spectrum of purified, native LbMA-Ox protein with absorbance maxima at 377 and 445 nm. (C) UV-Vis spectrum of boiled LbMA-Ox protein supernatant with absorbance maxima at 374 and 450 nm.

## SUPPORTING INFORMATION

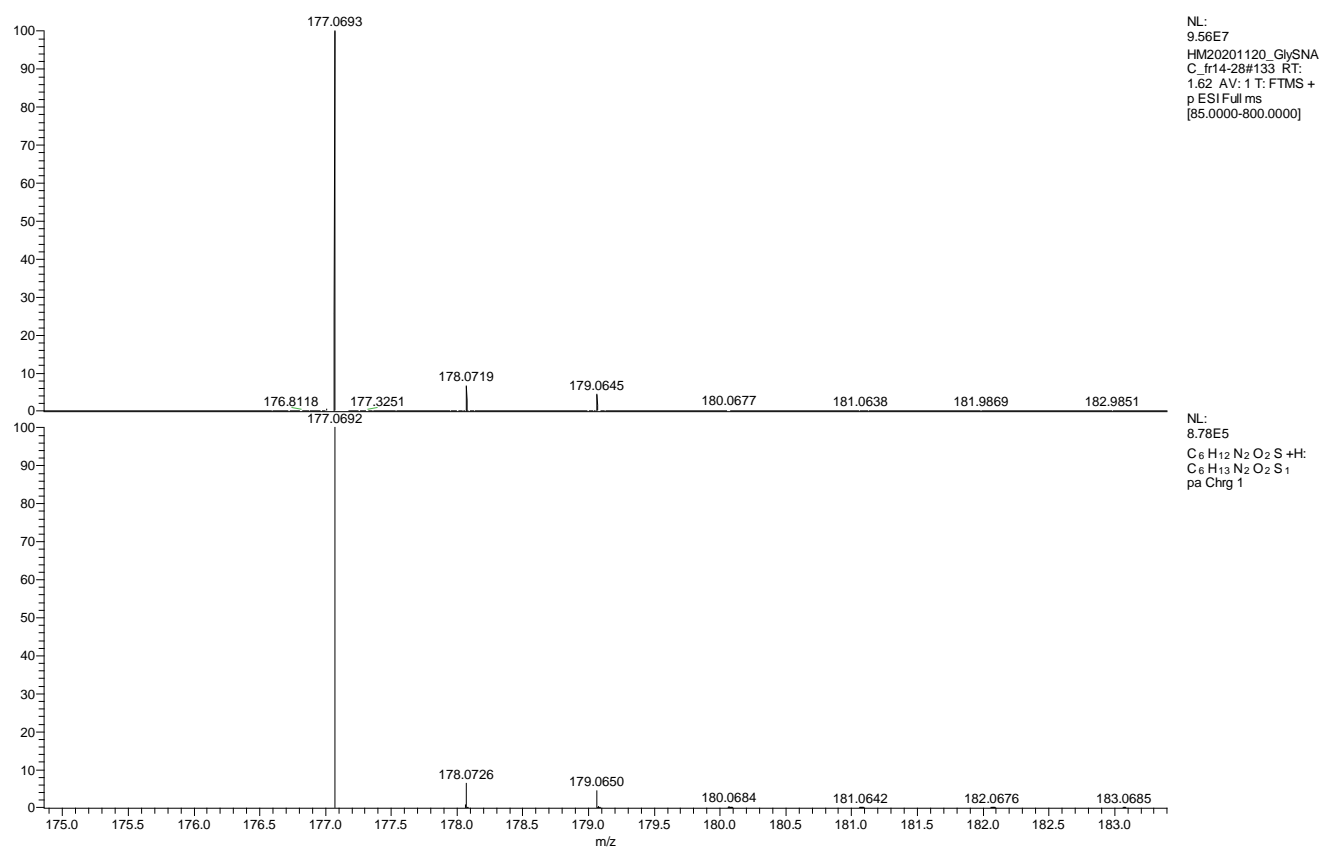
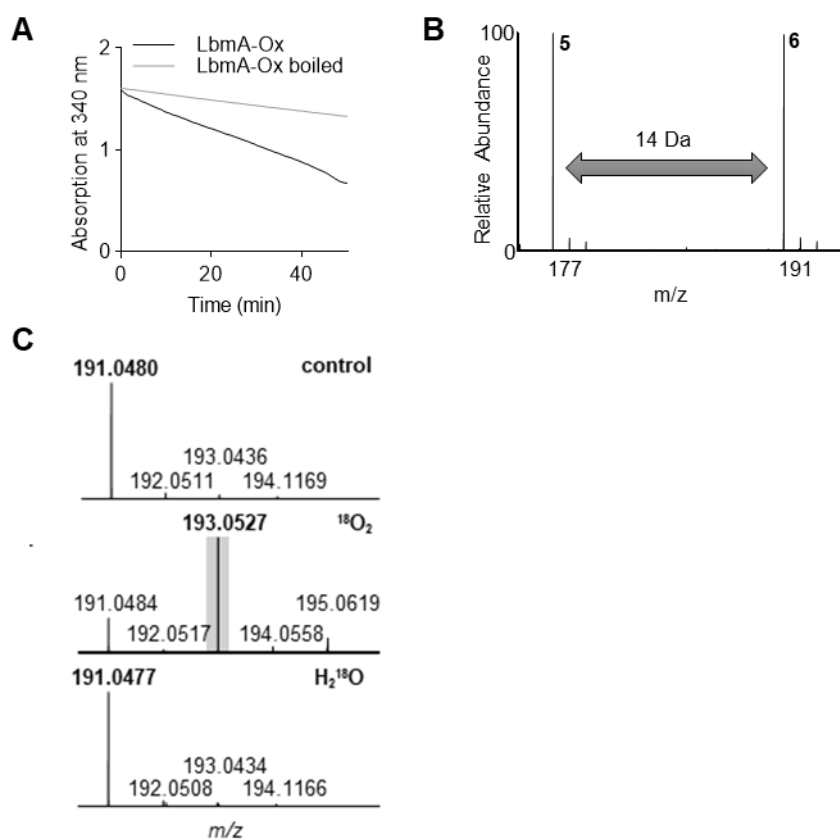


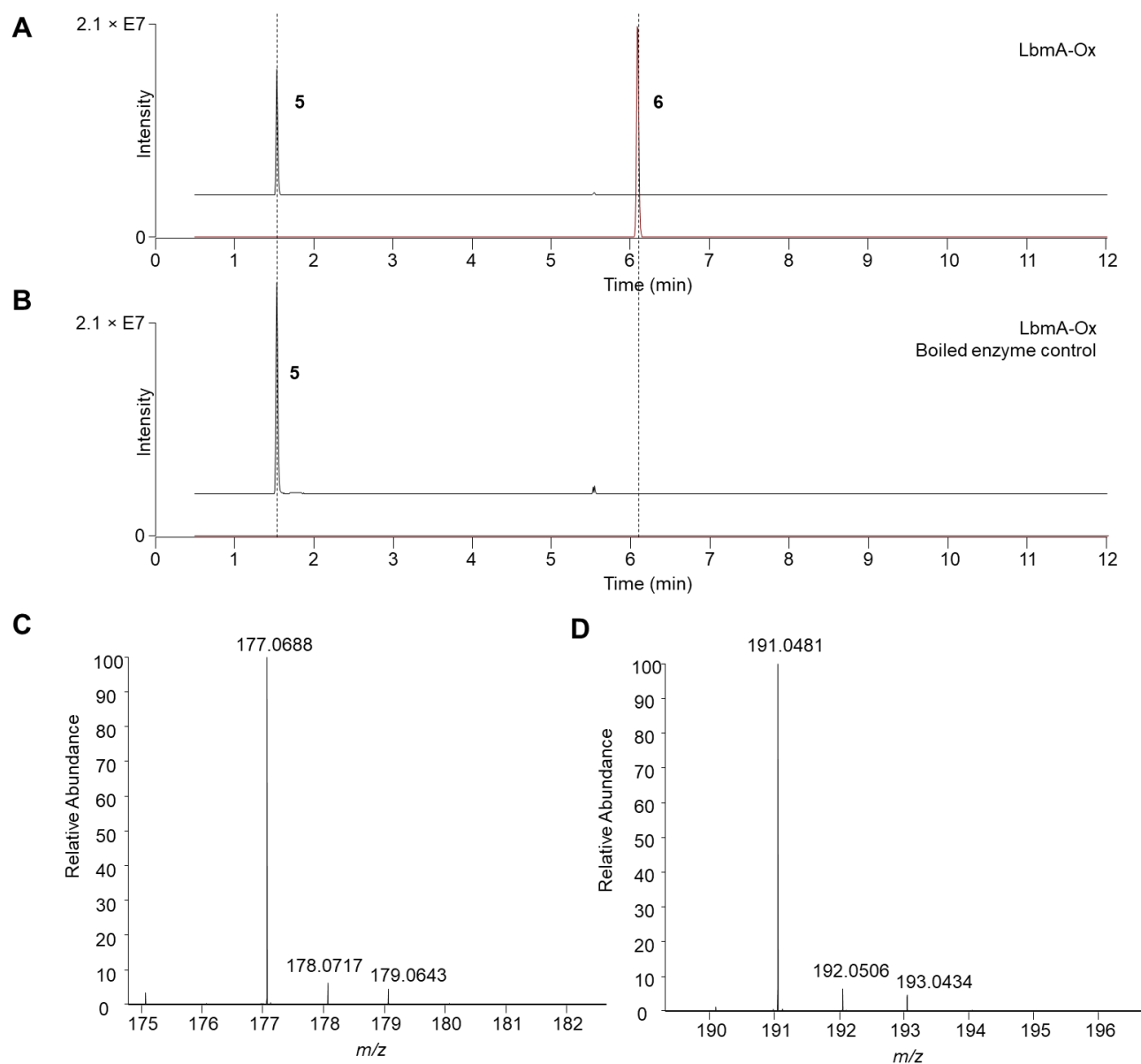
Figure S5. Top: mass spectrum for Gly-SNAC (5). Bottom: simulated mass spectrum.

## SUPPORTING INFORMATION



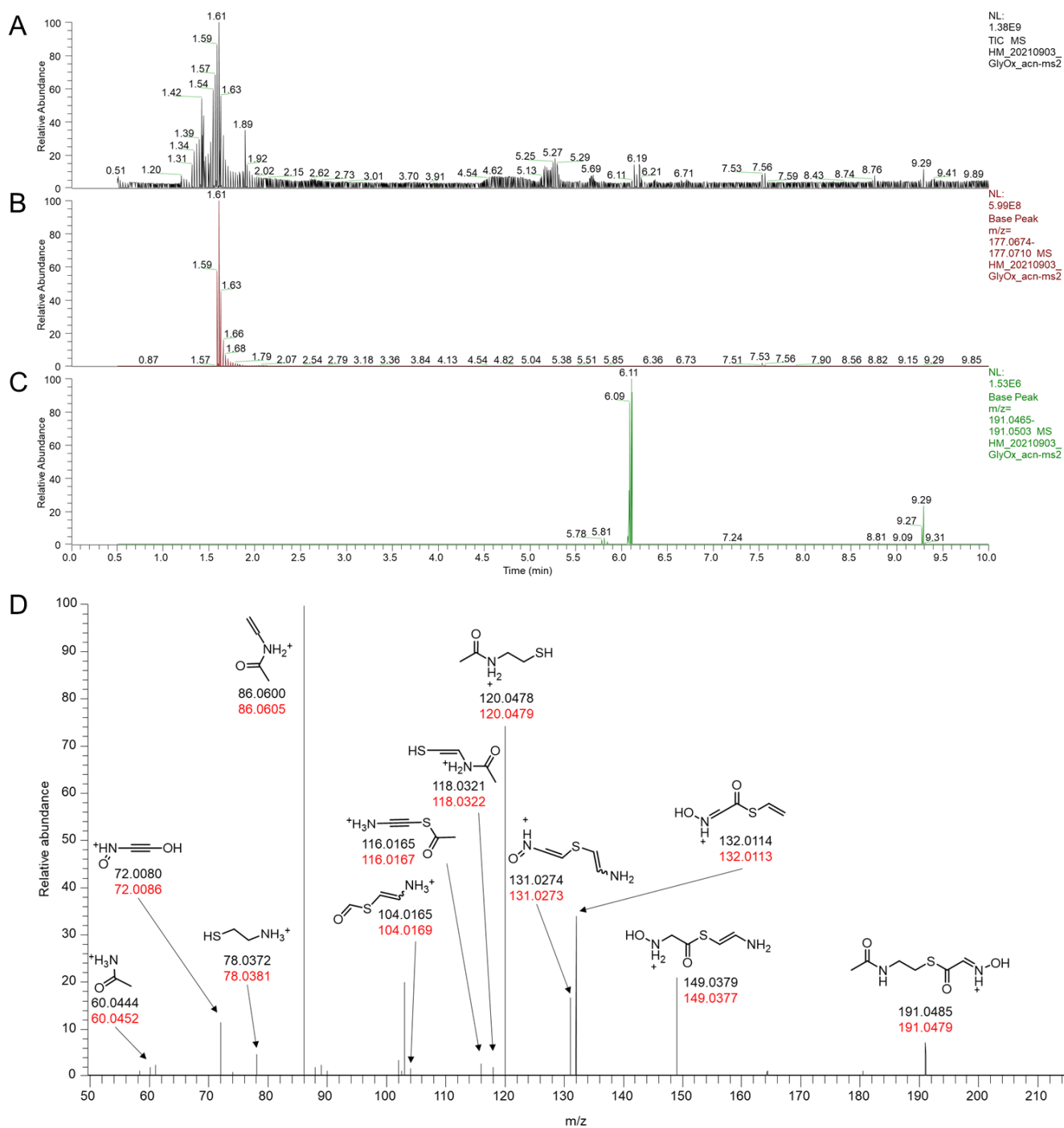
**Figure S6.** Additional experimental data for LbmA-Ox *in vitro* assays using **5** as test substrate. (A) Absorption measurement at 340 nm to monitor NADPH conversion during the assay. (B) Overlay of the MS spectra of **5** and **6** indicate a difference of 14 Da between substrate and product. Intensity for **5** is  $7.85 \times 10^7$ , and intensity for **6** is  $6.24 \times 10^5$ . (C) D Mass spectra of the putative product peak (**6**) and its natural isotope distribution for an assay under normal conditions (control), under <sup>18</sup>O<sub>2</sub> enriched atmosphere (<sup>18</sup>O<sub>2</sub>), or in an H<sub>2</sub><sup>18</sup>O background (H<sub>2</sub><sup>18</sup>O). The +2 Da shift corresponding to incorporation of <sup>18</sup>O is marked with a grey box.

## SUPPORTING INFORMATION



**Figure S7.** Extracted ion chromatogram (EIC) for *in vitro* LbmA-Ox assay mixtures using Gly-SNAC (**5**) as substrate. (A) EIC for substrate **5** (black) and product **6** (red). (B) EIC for substrate **5** (black) and product **6** (red) for the control assay using boiled enzyme. (C) Mass spectrum of **5** (calcd. for  $[M+H]^+ = 177.0692$ ) at 1.54 min, with an intensity of  $7.85 \times 10^7$  (D) Mass spectrum of **6** (calcd. for  $[M+H]^+ = 191.0484$ ) at 6.09 min, with an intensity of  $6.24 \times 10^6$ .

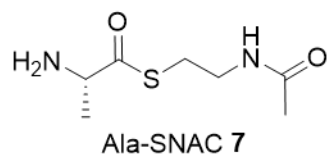
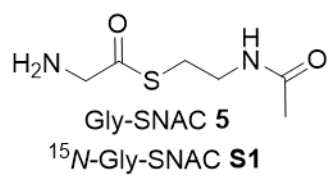
## SUPPORTING INFORMATION



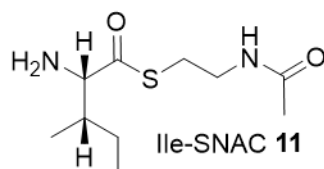
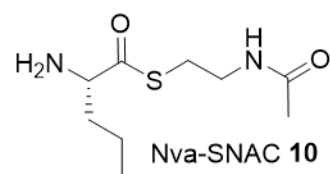
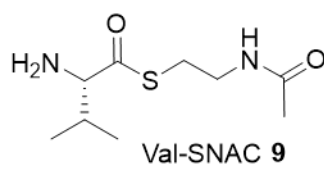
**Figure S8.** UHPLC-HRMS data for the *in vitro* LbmA-Ox assay reaction with Gly-SNAC (**5**) as substrate. (A) Total ion chromatogram. (B) Extracted ion chromatogram for substrate **5**. (C) Extracted ion chromatogram for product **6**. (D) Filtered MS/MS spectrum of mass 191.0480 at 6.09 min. Black numbers indicate calculated  $m/z$  and red numbers indicate measured  $m/z$ . Possible fragment ions shown were deduced from predicted structures using the competitive fragmentation modeling for metabolite identification (CFM-ID 3.0) tool.<sup>[21]</sup>

## SUPPORTING INFORMATION

Accepted by LbmA-Ox



Not accepted by LbmA-Ox



Free amino acids:  
Gly, Ala, Val, Ile

**Figure S9.** Test substrates for conversion with LbmA-Ox. Analytical data are shown in Figures 2B–C and Figures S10–13.

## SUPPORTING INFORMATION

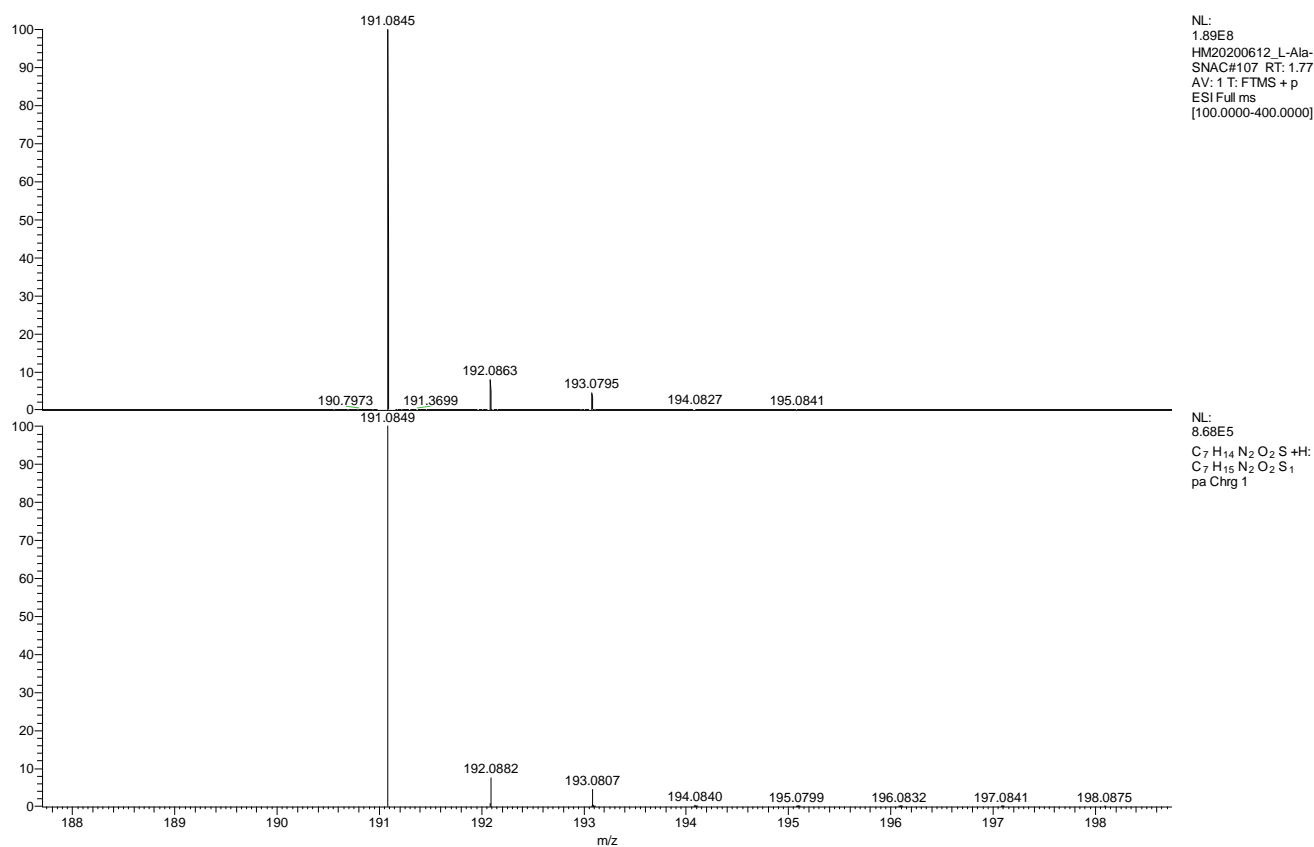
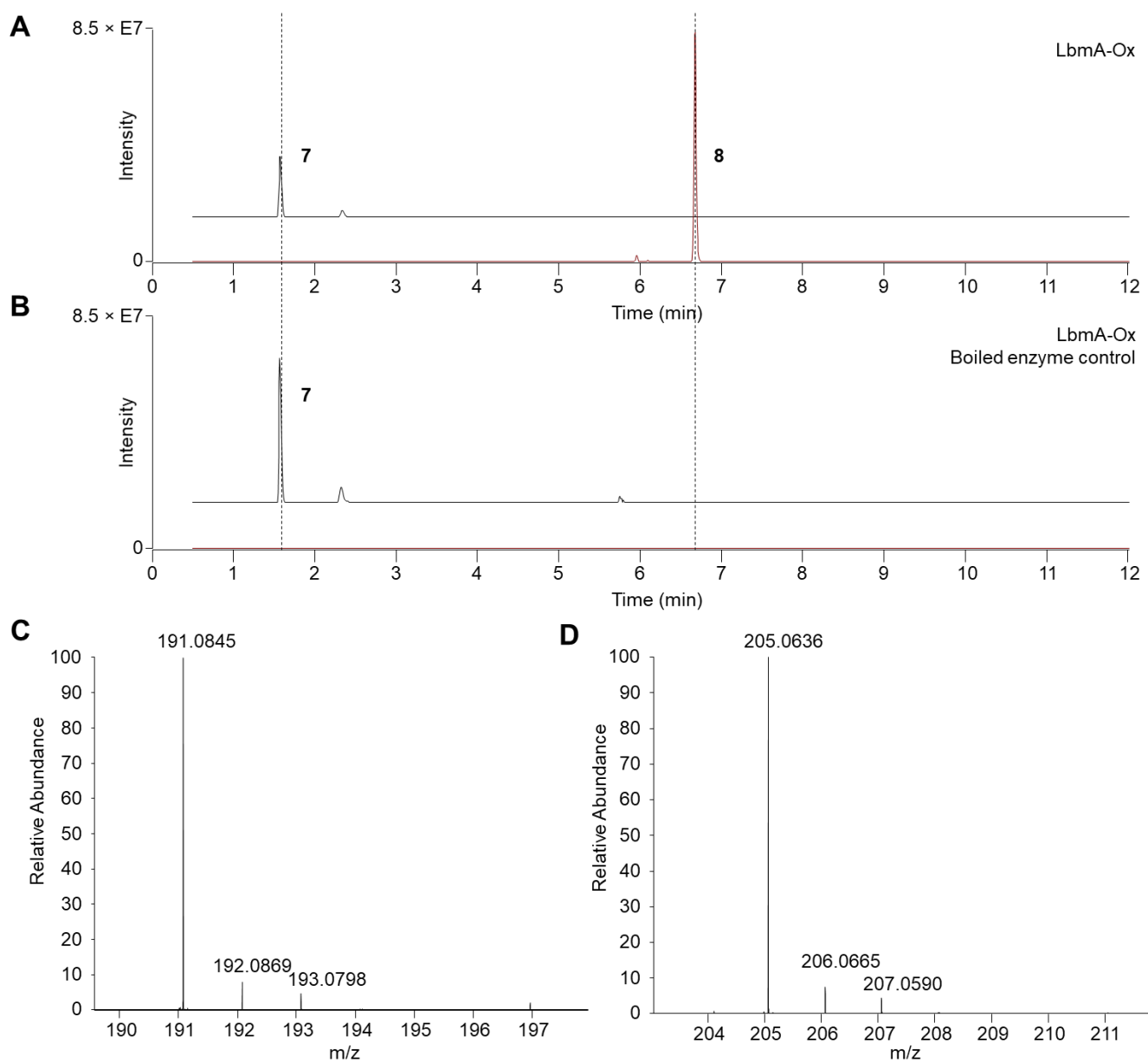


Figure S10. Top: Mass spectrum for Ala-SNAC (7). Bottom: Simulated mass spectrum.

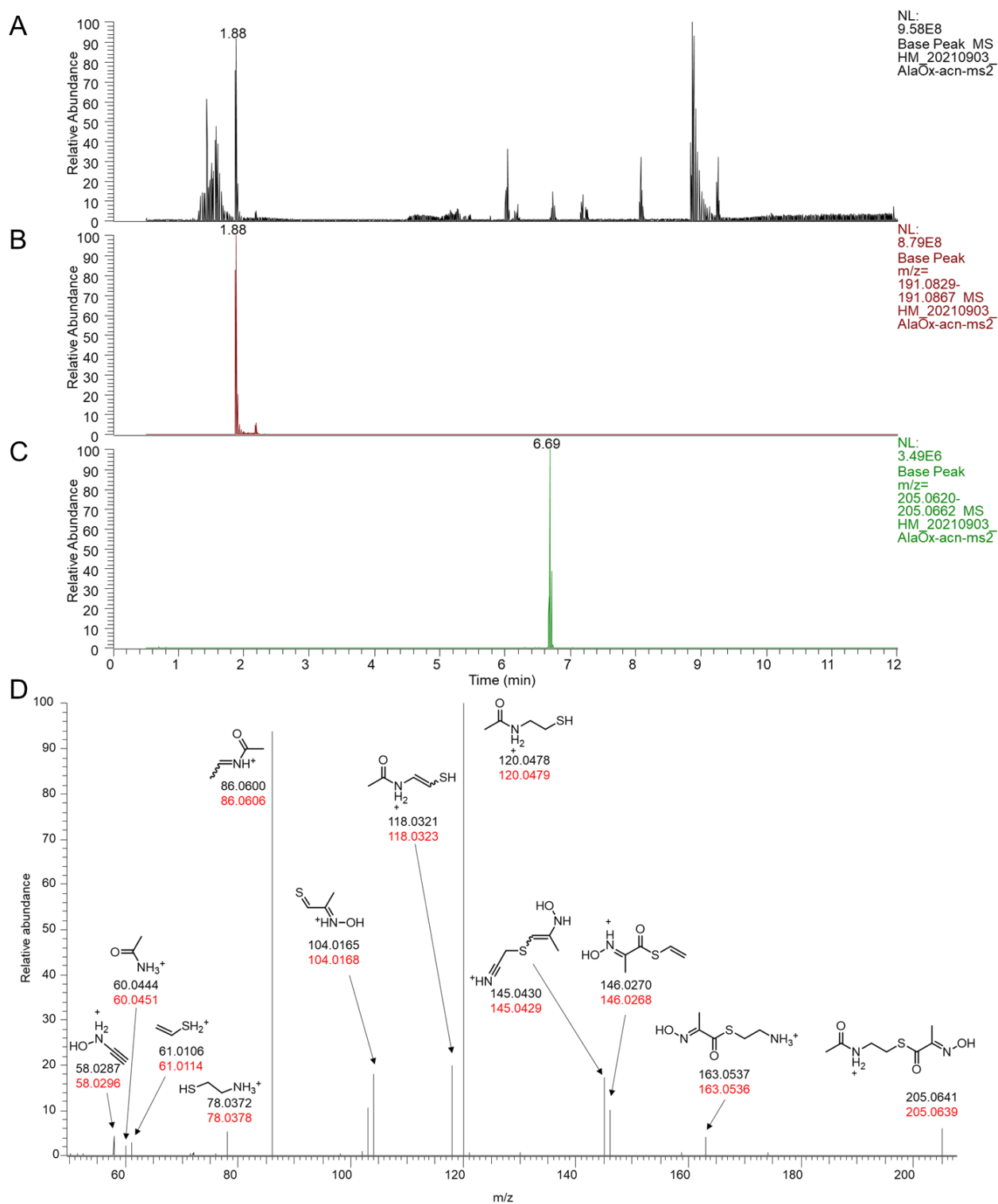
## SUPPORTING INFORMATION



**Figure S11.** Extracted ion chromatograms (EIC) and mass spectra for the *in vitro* LbmA-Ox assay mixtures using Ala-SNAC (**7**) as substrate. (A) EIC for substrate **7** (black) and product **8** (red). (B) EIC for substrate **7** (black) and product **8** (red) for the control assay using boiled enzyme. (C) Mass spectrum of **7** (calcd. for  $[M+H]^+ = 191.0848$ ) at 1.58 min. (D) Mass spectrum of **8** (calcd. for  $[M+H]^+ = 205.0641$ ) at 6.68 min.

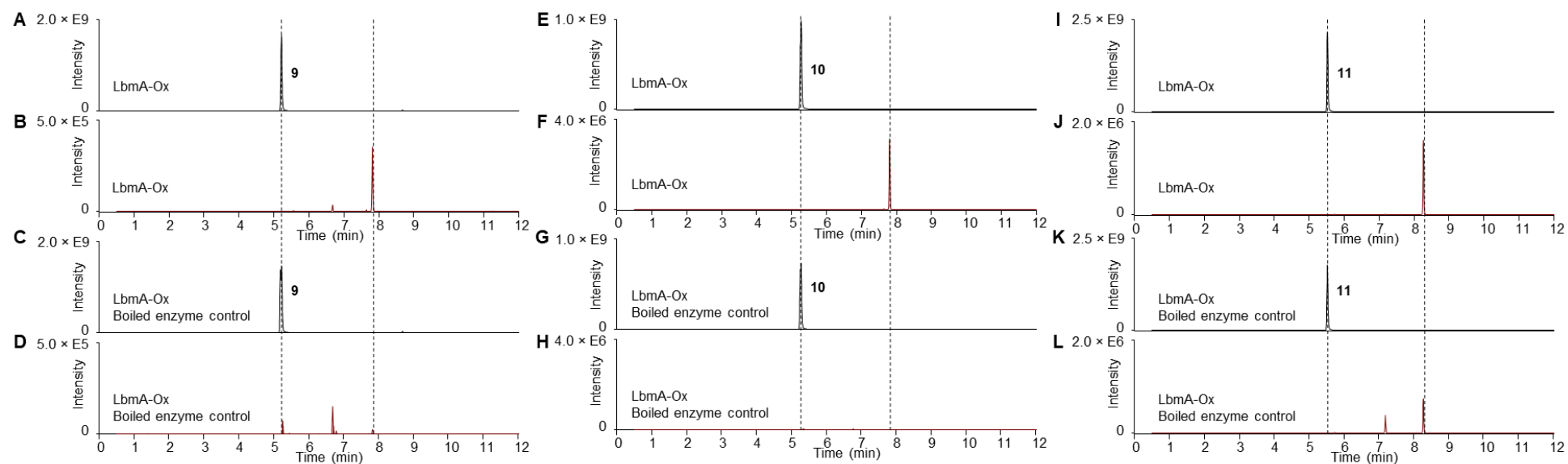


## SUPPORTING INFORMATION



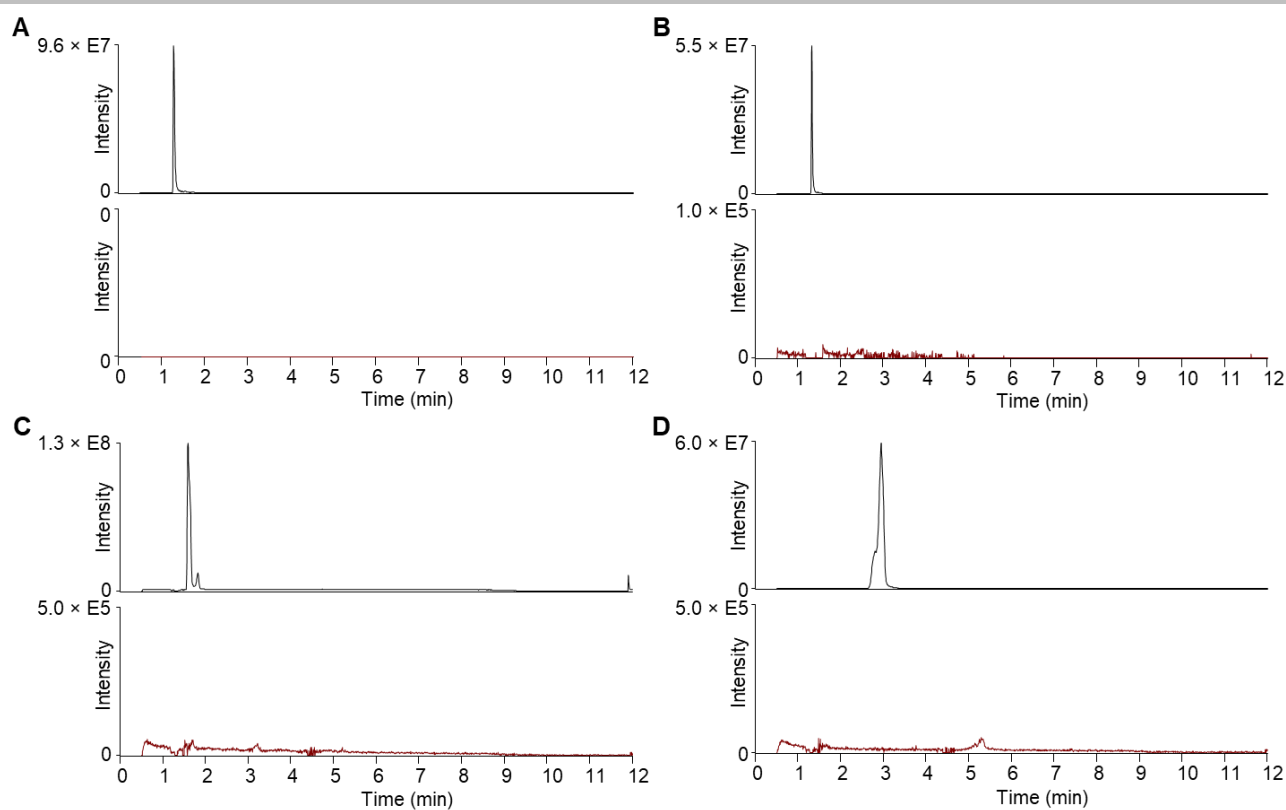
**Figure S12.** HPLC-HRMS data for the *in vitro* LbMA-Ox assay reaction with Ala-SNAC (**7**) as substrate. (A) Total ion chromatogram. (B) Extracted ion chromatogram for substrate **7**. (C) Extracted ion chromatogram for product **8**. (D) Filtered MS/MS spectrum of mass 205.1097 at 6.67 min. Black numbers indicate calculated  $m/z$  and red numbers indicate measured  $m/z$ . Possible fragment ions shown were deduced from predicted structures using the competitive fragmentation modeling for metabolite identification (CFM-ID 3.0) tool.<sup>[21]</sup>

## SUPPORTING INFORMATION



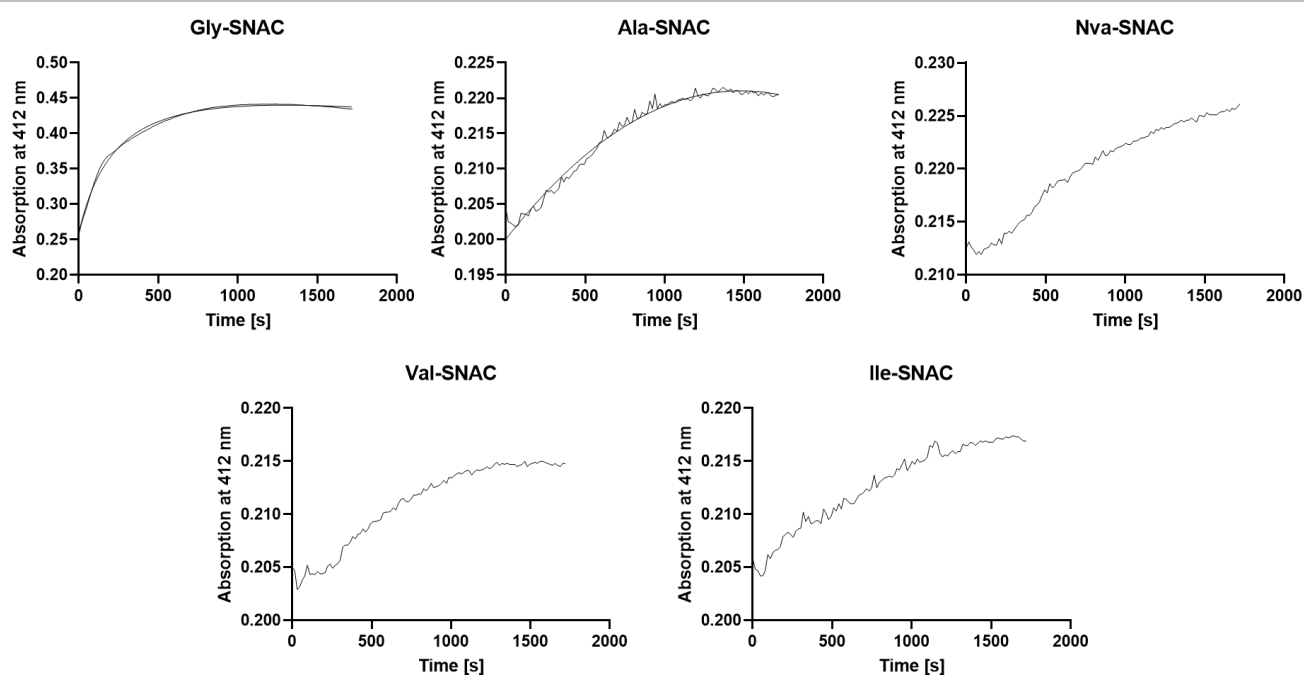
**Figure S13.** Extracted ion chromatograms (EICs) and mass spectra for assay mixtures using Val-SNAC (**9**), Nva-SNAC (**10**) or Ile-SNAC (**11**) as substrate. (A) EIC for **9**. (B) EIC for a putative product of **9**. (C) EIC for **9** in the control assay. (D) EIC for a putative product of **9** in the control assay. (E) EIC for **10**. (F) EIC for a putative product of **10**. (G) EIC for **10** in the control assay. (H) EIC for a putative product of **10** in the control assay. (I) EIC for **11**. (J) EIC for a putative product of **11**. (K) EIC for **11** in the control assay. (L) EIC for a putative product of **11** in the control assay.

## SUPPORTING INFORMATION



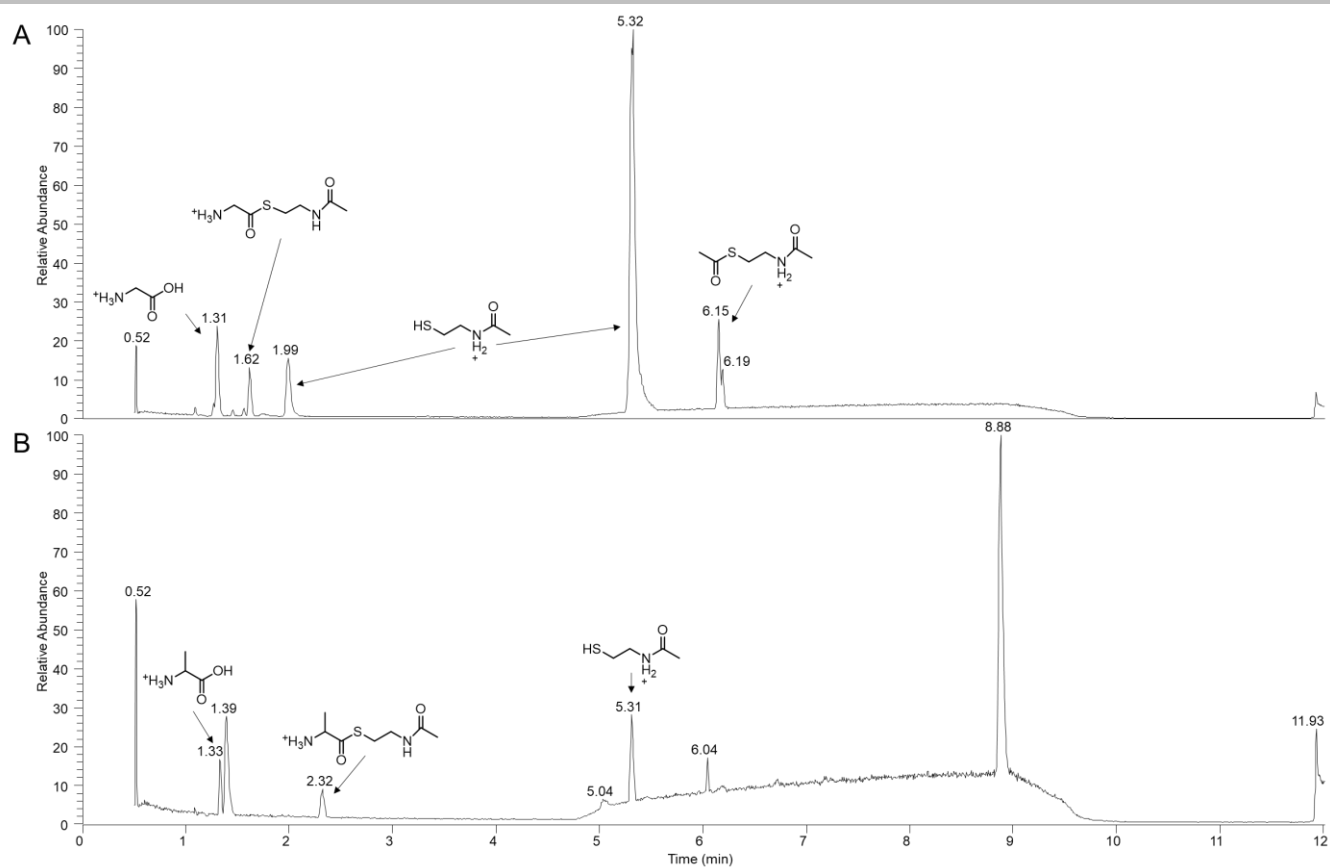
**Figure S14.** Extracted ion chromatograms (EICs) for *in vitro* LbmA-Ox assay mixtures using free amino acids as substrate. (A) Upper trace: EIC for glycine as substrate (calcd. for  $[M+H]^+ = 76.0393$ ). Lower trace: EIC for the product (calcd. for  $[M+H]^+ = 90.0186$ ). (B) Upper trace: EIC for alanine as substrate (calcd. for  $[M+H]^+ = 90.0550$ ). Lower trace: EIC for the product (calcd. for  $[M+H]^+ = 104.0342$ ). (C) Upper trace: EIC for valine as substrate (calcd. for  $[M+H]^+ = 118.0863$ ). Lower trace: EIC for the product (calcd. for  $[M+H]^+ = 132.0655$ ). (D) Upper trace: EIC for isoleucine as substrate (calcd. for  $[M+H]^+ = 132.1019$ ). Lower trace: EIC for the product (calcd. for  $[M+H]^+ = 146.0812$ ).

## SUPPORTING INFORMATION



**Figure S15.** Ellman's assay of the various SNAC substrates. The data represent absorption measurements at 412 nm over 2.5 h for the indicated substrates.

## SUPPORTING INFORMATION



**Figure S16.** UHPLC-HRMS data for *in vitro* LbmA-Ox assay reactions using Gly-SNAC (**5**) or Ala-SNAC (**7**) show that the products **6** and **8** hydrolyze during the isolation and purification process. (A) Total ion chromatogram for an assay with **5** as substrate. (B) Total ion chromatogram for an assay with **7** as substrate. The main peaks are assigned with putative hydrolysis products.

## SUPPORTING INFORMATION

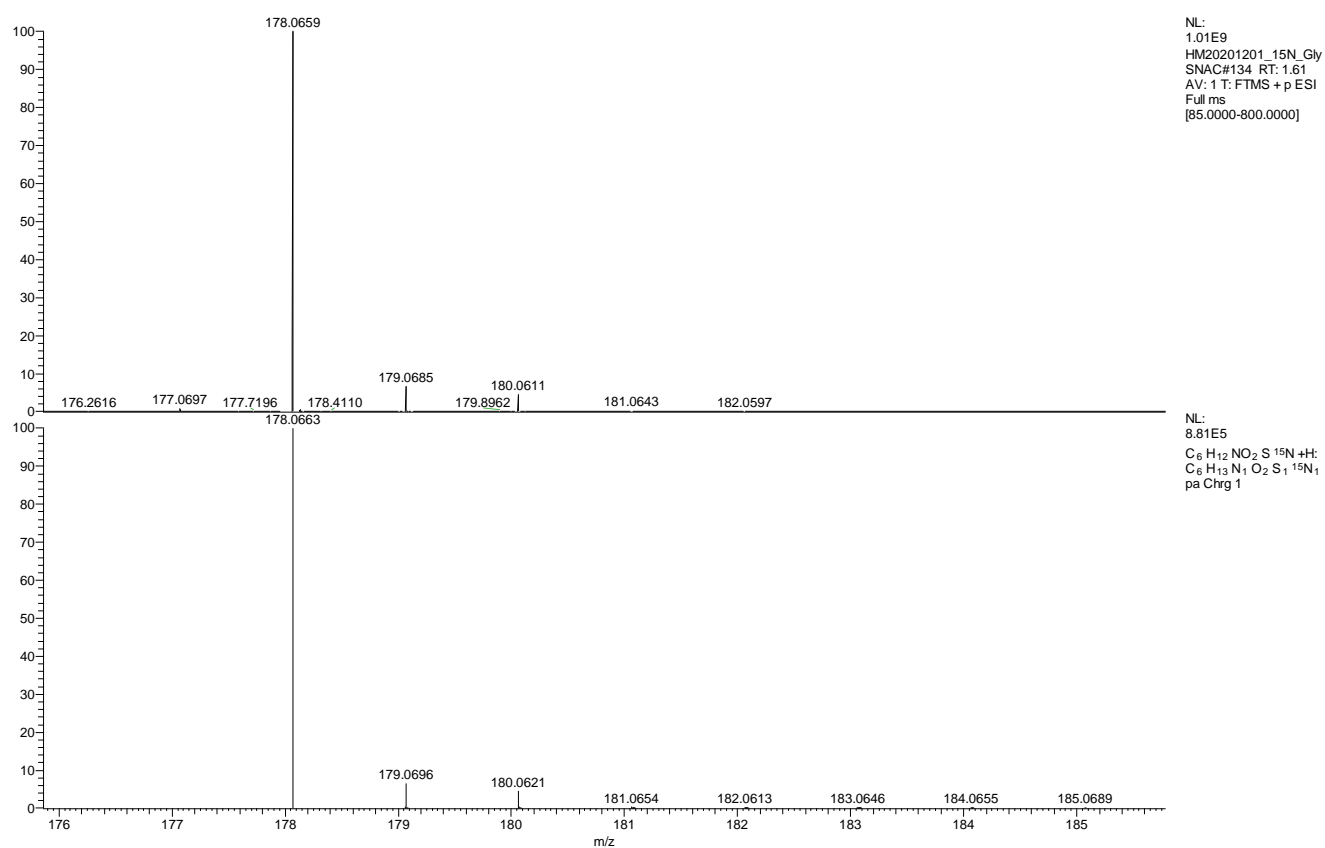
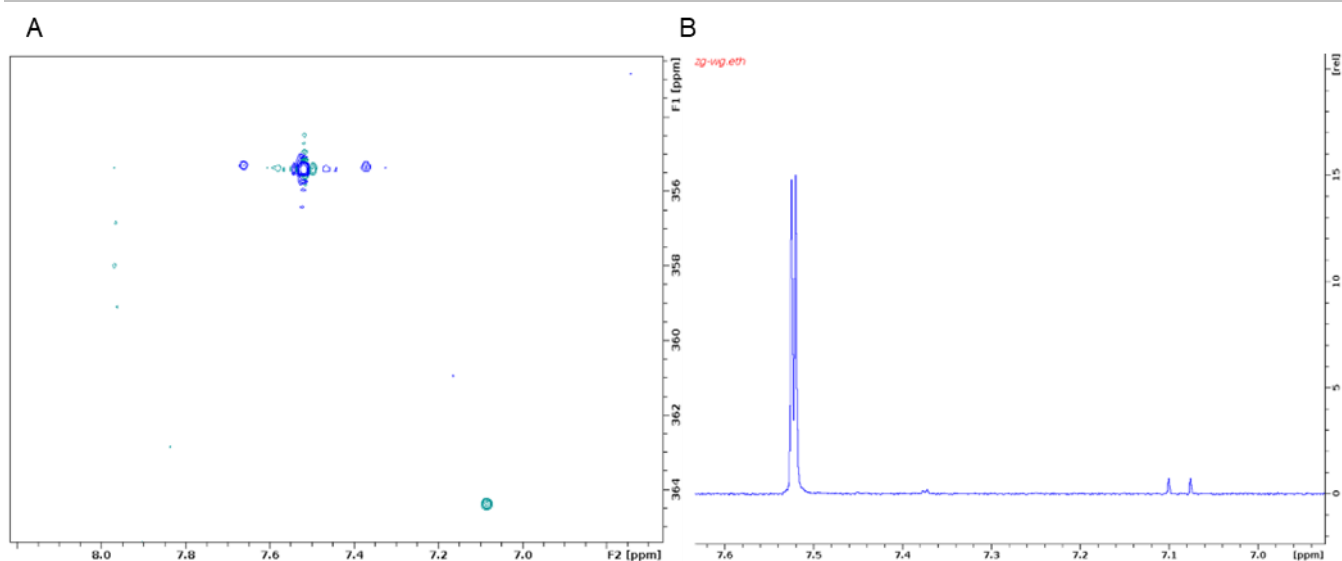


Figure S17. Top: mass spectrum for  $^{15}\text{N}$ -Gly-SNAC (S1). Bottom: simulated mass spectrum.

## SUPPORTING INFORMATION



**Figure S18.** NMR spectra of  $^{15}\text{N}$ -hydroxyiminoacetic acid (**S2**) in 10%  $\text{D}_2\text{O}$ . A)  $^{15}\text{N}$   $^1\text{H}$ -HSQC and B)  $^1\text{H}$ -NMR spectra of **S2** with two characteristic signals for the Z-oxime and E-oxime at 7.52 ppm and 7.09 ppm, respectively. The INEPT delay for (A) was set to 87 ms for optimal sensitivity for the Z-oxime.

## SUPPORTING INFORMATION

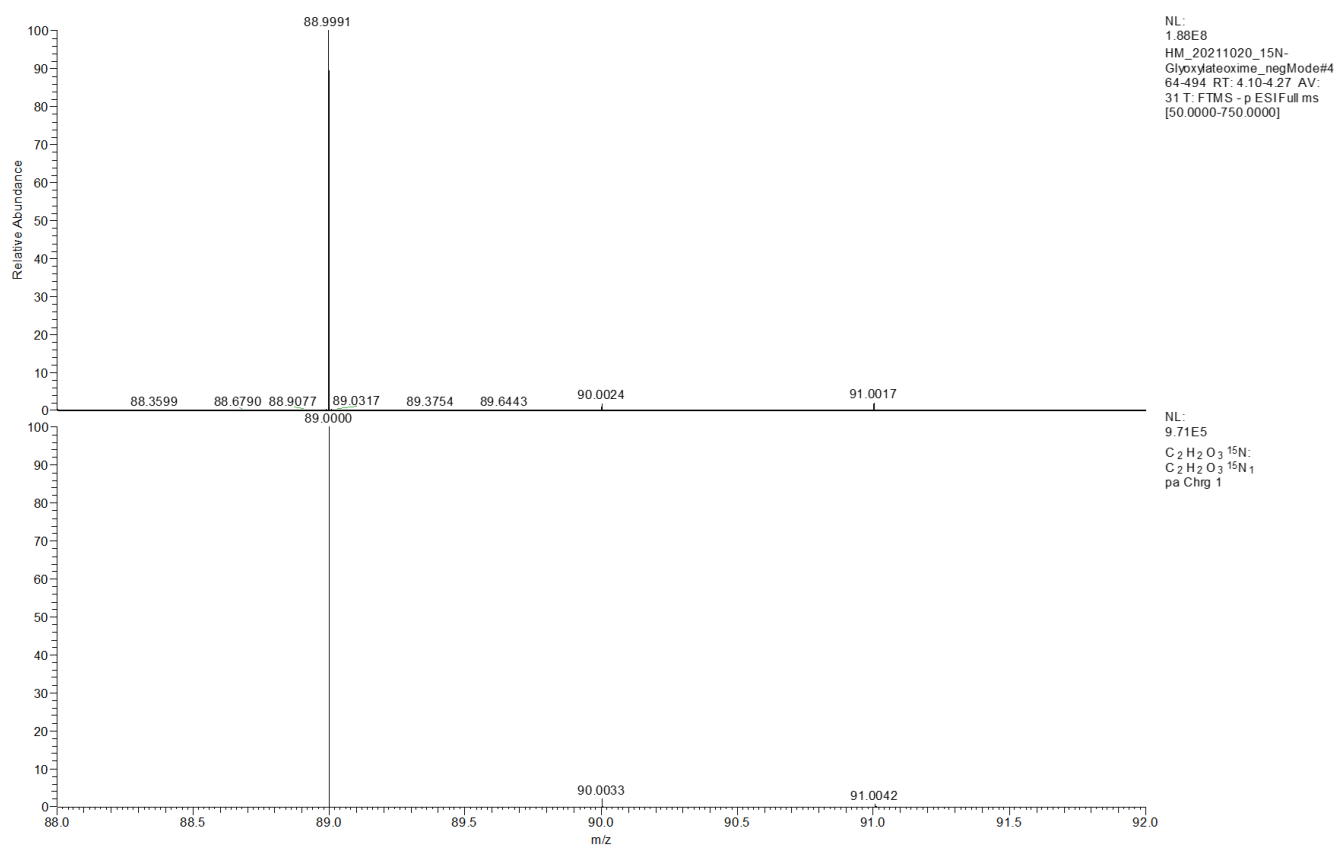
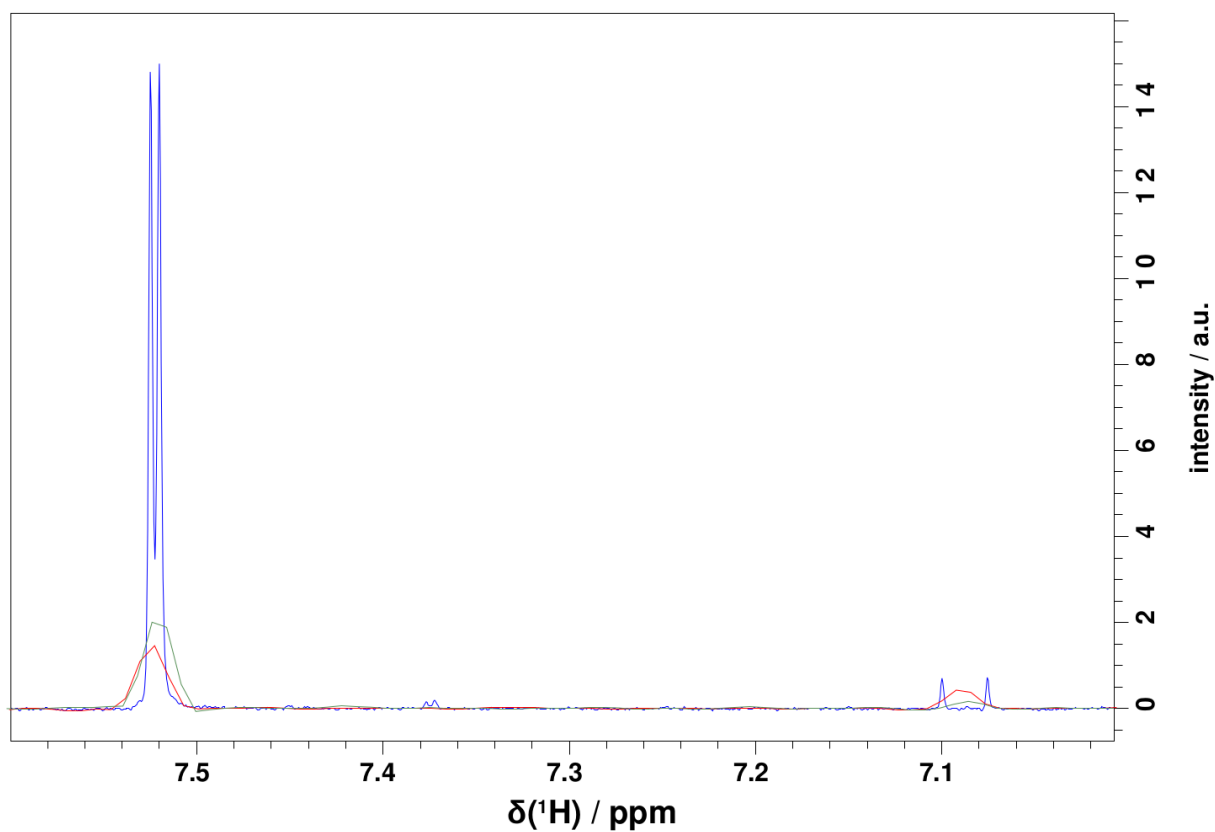


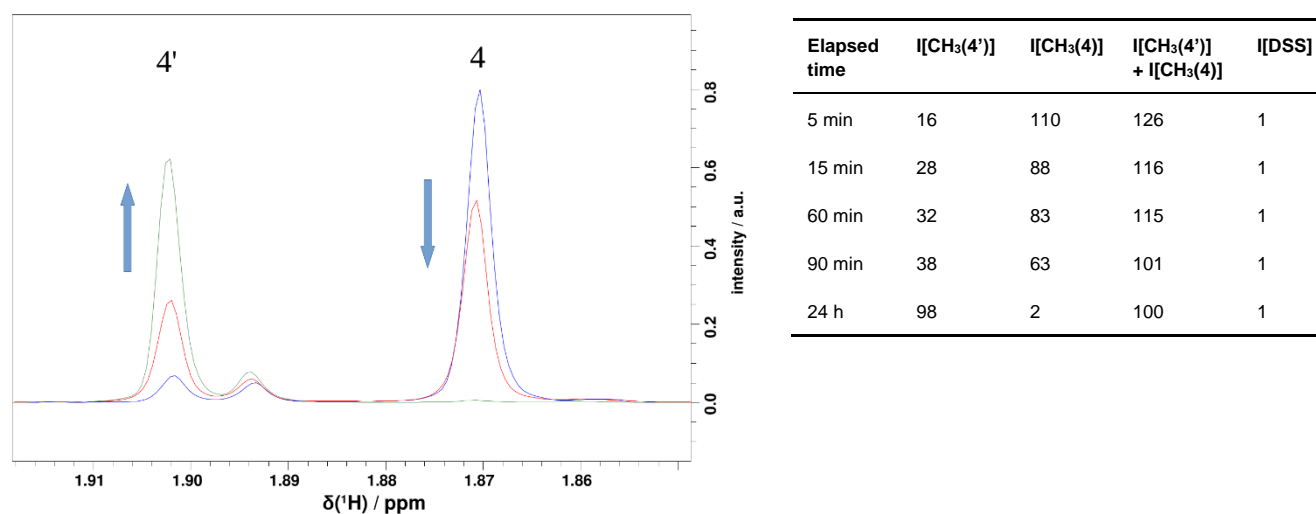
Figure S19. Top: mass spectrum for <sup>15</sup>N-hydroxyiminoacetic acid (S2). Bottom: simulated mass spectrum





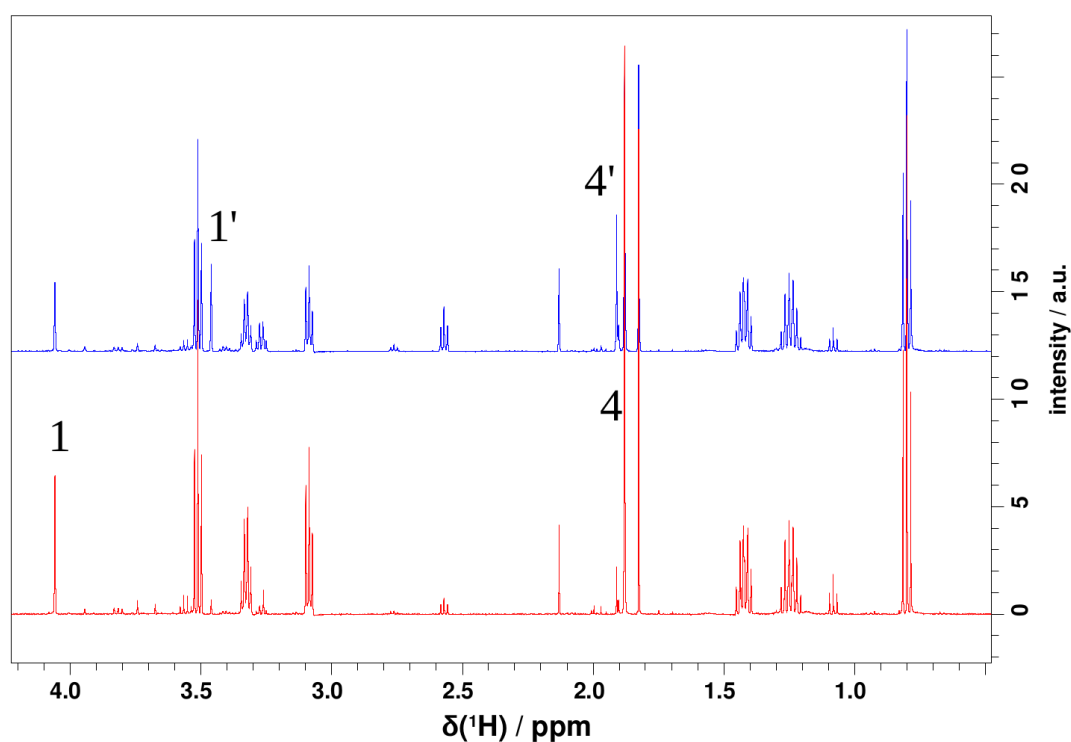
**Figure S20.**  $^1\text{H}$ - $^{15}\text{N}$  selected NMR experiments for  $^{15}\text{N}$ -hydroxyiminoacetic acid (**S2**) as a standard for hydrolyzed oxime-**S1**. The  $^1\text{H}$  NMR spectrum is overlaid with isotope-edited spectra with settings for *Z*-oxime (green, 2.88 Hz, 350 ppm) and *E*-oxime (red, 13.9 Hz, 350 ppm). In this sample, the concentration of the *E*-oxime (doublet at 7.09 ppm,  $^2J_{\text{HN}} = 14.7$  Hz) is about 10% of the *Z*-oxime (doublet at 7.52 ppm,  $^2J_{\text{HN}} = 2.82$  Hz). Note that the signal loss due to  $T_2$ -relaxation in the  $^1\text{H}$ - $^{15}\text{N}$  selected NMR experiment is more pronounced for the experiment with the INEPT delay set to 87 ms for  $^2J_{\text{HN}} = 2.82$  Hz (*Z*-oxime).

## SUPPORTING INFORMATION



**Figure S21.** Kinetic NMR experiments with LbmA-Ox and <sup>15</sup>N-Gly-SNAC (**S1**) as substrate. Left: The <sup>15</sup>N edited spectrum for three different time points showing the interconversion of **S1** to a putative degradation product. (blue: start of the experiment, red: after 90 min, green: after 24 h). Right: Quantification of the signal intensities corresponding to **S1** and the degradation product by comparison to the internal standard sodium trimethylsilylpropanesulfonate (DSS) show a net decrease of substrate signal. The amide signal at about 8 ppm (not shown), which is putatively formed by Beckmann rearrangement of the oxime, has a relative intensity of less than 1 compared to DSS for all time points. The degradation product is most likely the hydrolyzed thioester resulting in glycine.

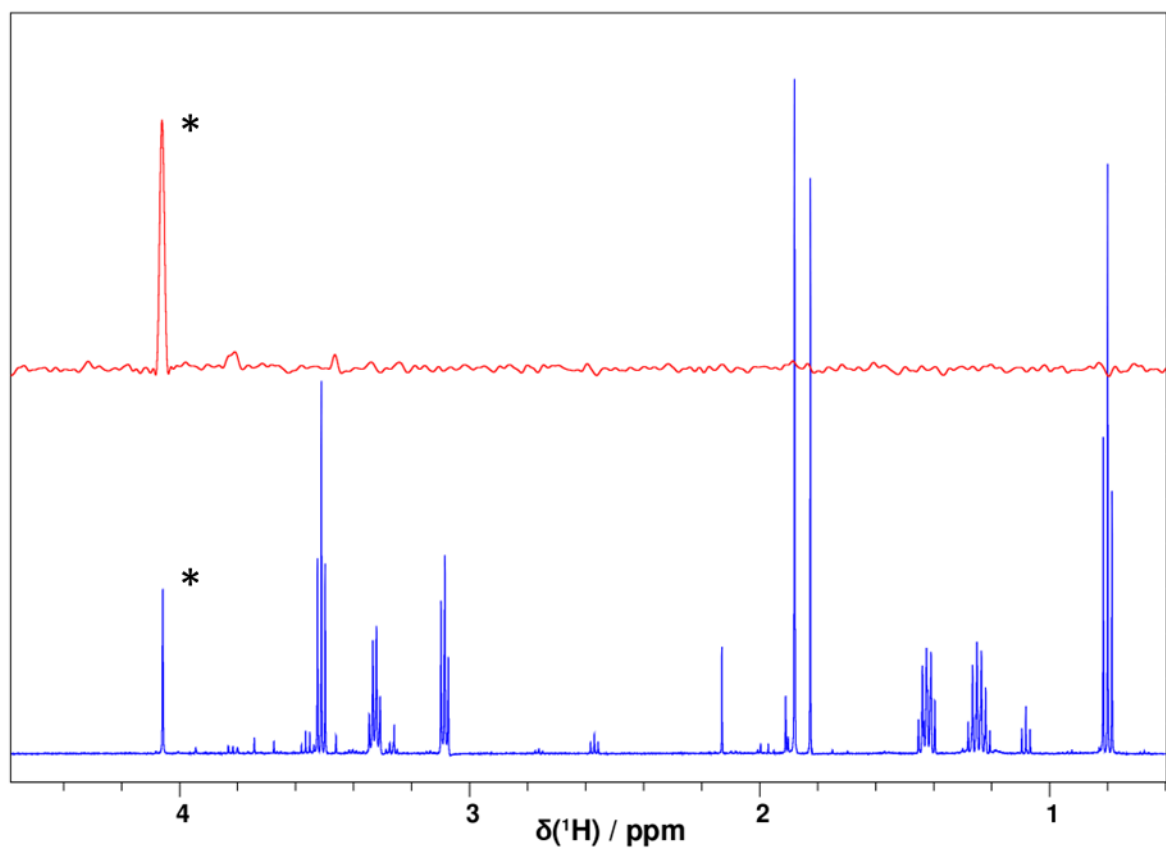
## SUPPORTING INFORMATION



Elapsed time	I[CH <sub>3</sub> (4')]	I[CH <sub>3</sub> (4)]	I[CH <sub>3</sub> (4')] + I[CH <sub>3</sub> (4)]	I[8.35 ppm]
5 min	0.93	8.06	8.99	1
18 h	3.25	5.52	8.77	1

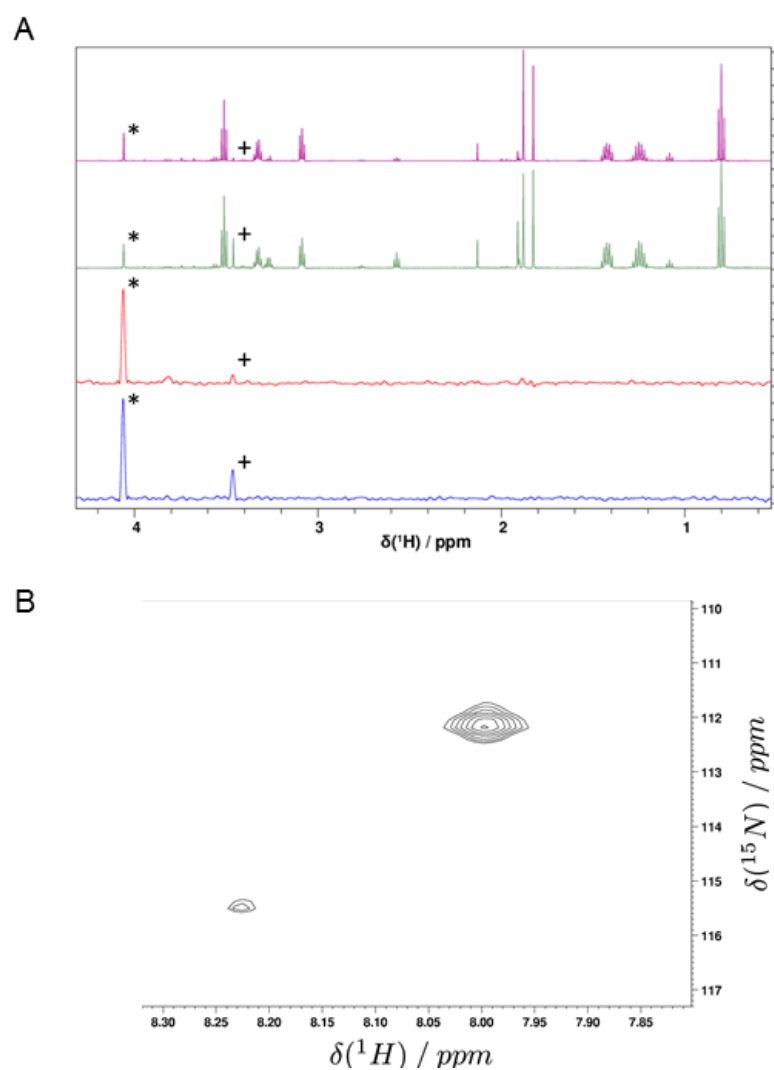
**Figure S22.** Kinetic NMR control experiments with <sup>15</sup>N-Gly-SNAC (**S1**) as substrate in the absence of LbmA-Ox. Top: The <sup>1</sup>H NMR spectrum for two different time points showing the interconversion of **S1** to a putative degradation product (red: start of the experiment, blue: after 18 h). Bottom: Quantification of the signal intensities corresponding to **S1** and the degradation product by comparison to the internal standard at 8.35 ppm shows no significant net decrease of substrate signal. The degradation product is most likely the hydrolyzed thioester resulting in glycine with the H $\alpha$  signal (1') at 3.46 ppm.

## SUPPORTING INFORMATION



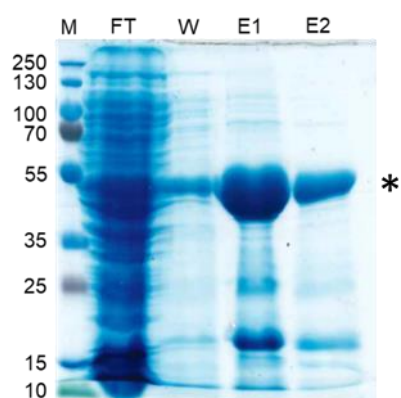
**Figure S23.** The upper trace in red shows the  $^{15}\text{N}$  edited spectrum of  $^{15}\text{N}$ -Gly-SNAC (**S1**), the lower trace in blue shows the 1D spectrum of the same sample. The  $\text{H}_\alpha$  signal is observable at 4.06 ppm and marked with an asterisk. The  $\text{H}_\alpha$  proton is weakly coupled to  $^{15}\text{N}$  with a coupling constant of about 1 Hz.

## SUPPORTING INFORMATION



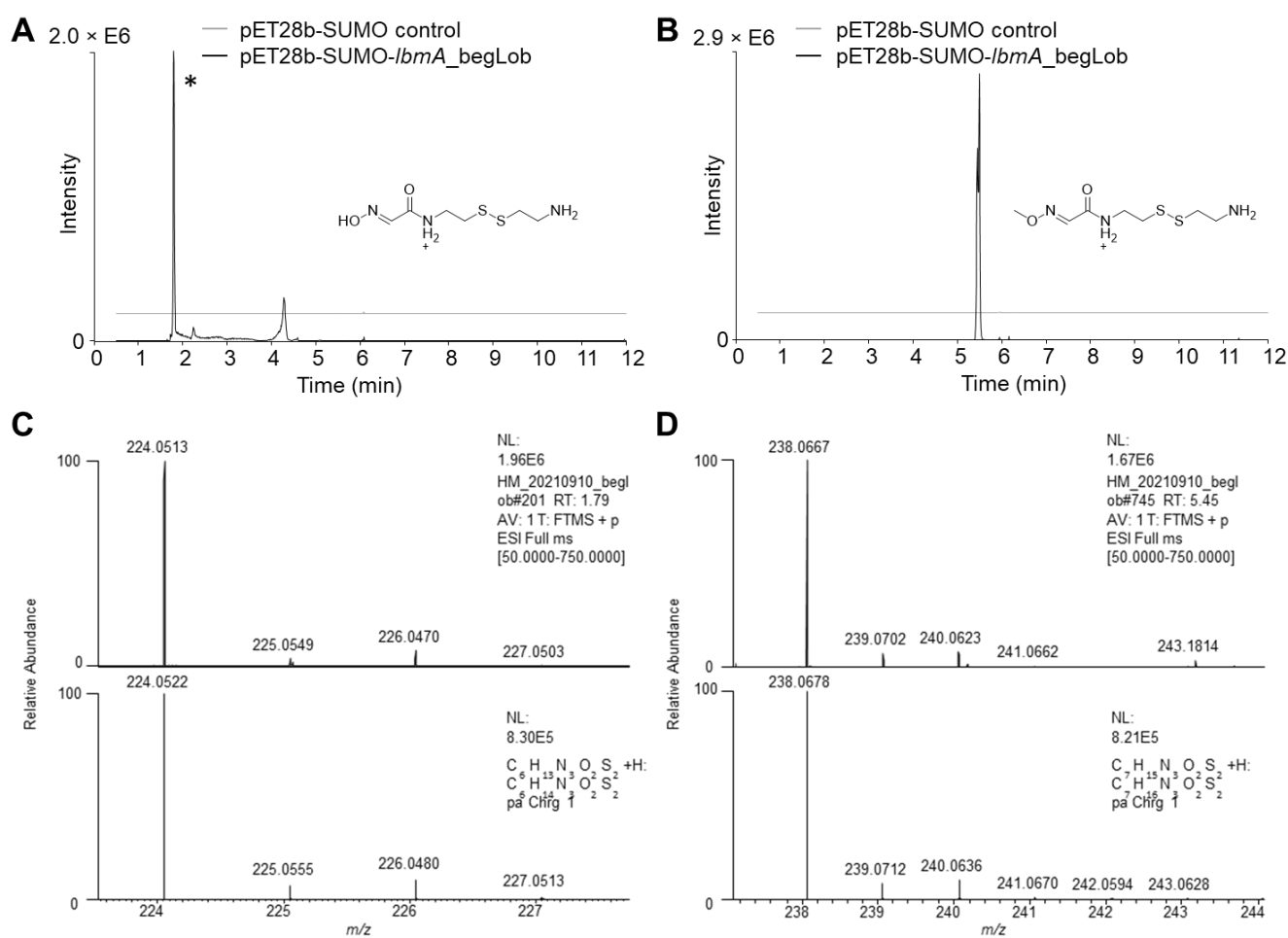
**Figure S24.** NMR data of assay mixtures with LbmA-Ox and  $^{15}\text{N}$ -Gly-SNAC (**S1**) as substrate. (A) The lower trace in red shows the  $^{15}\text{N}$  edited spectrum of **S1** at time 0 of an assay using LbmA-Ox, observed directly in an NMR tube. The  $^{15}\text{N}$  edited spectrum after 6 h is shown in blue. The upper trace in pink shows the 1D spectrum of the same sample at time 0 and in green after 6 h. The  $\text{H}_\alpha$  signal is observable at 4.06 ppm and marked with asterisks. An additional signal, marked with a '+' appears at 3.46 ppm indicating degradation of **S1** over time. (B) The  $^{15}\text{N}/^1\text{H}$ -HSQC of the reactions mixture show signals with the characteristic chemical shift of an amide group at 8.0 / 112.2 ppm. For the preparation of the NMR sample, 20 ml of reaction mixture containing **S1**, NADPH and LbmA-OX were incubated for 6 h, lyophilized and dissolved in  $\text{D}_2\text{O}$ .

## SUPPORTING INFORMATION



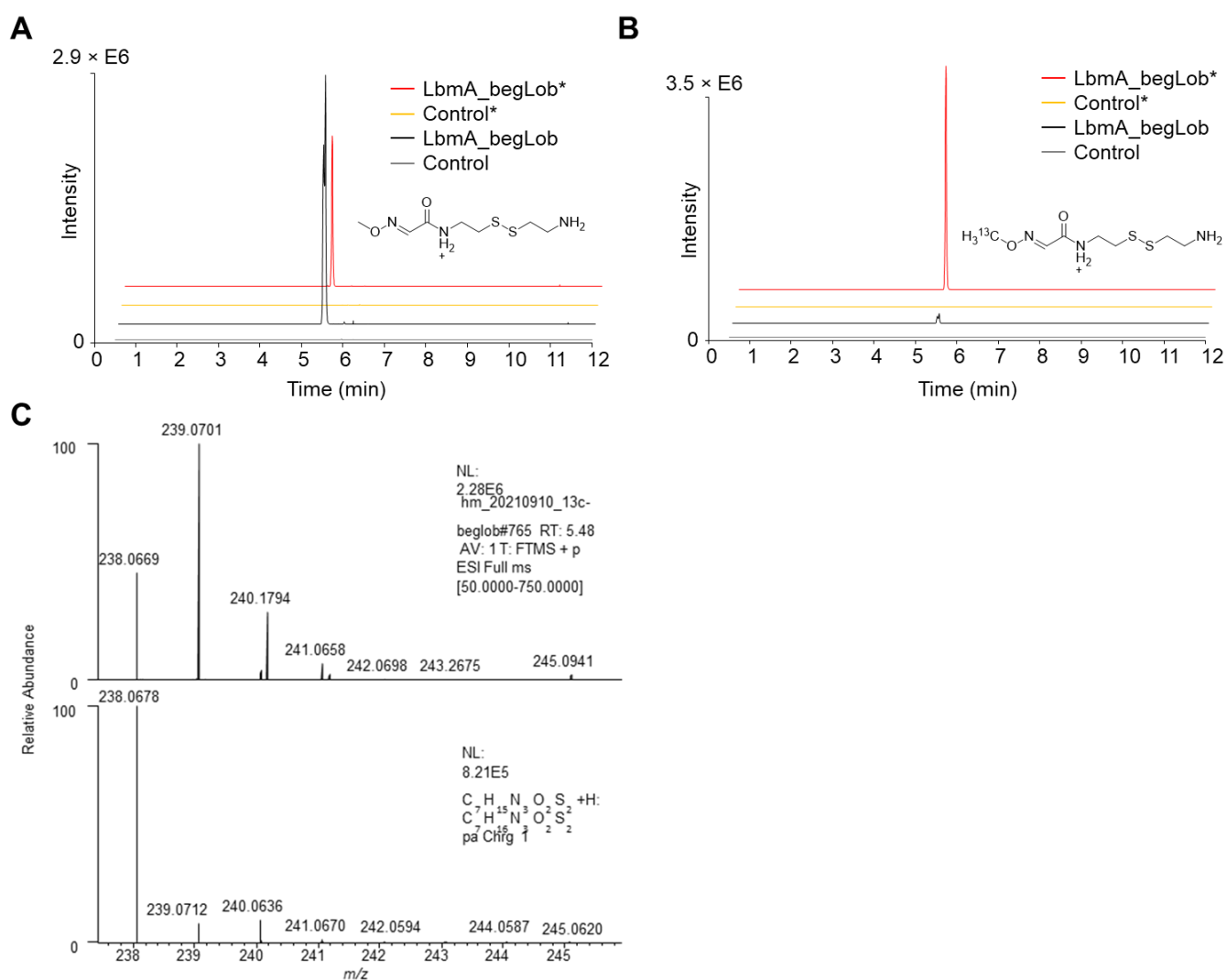
**Figure S25.** 12% SDS-PAGE gel of His<sub>6</sub>-SUMO-tagged LbmA-MT following Ni-NTA affinity purification. Expected molecular weight, marked with an asterisk: 40.4 kDa. M: protein ladder, FT: flowthrough, W: wash, E1: elution 1, E2: elution 2.

## SUPPORTING INFORMATION



**Figure S26.** UHPLC-HRMS data of the *in vivo* assay with the *LbaA* bimodule, encoded on pET28b-SUMO-*lbaA\_begLob*. (A) Extracted ion chromatograms (EICs) of **13** (calcd. for  $[M+H]^+ = 224.0522$ , marked with an asterisk) for extracts of *E. coli* BAP1 expressing pET28b-SUMO-*lbaA\_begLob* (black) or the pET28b-SUMO plasmid as a negative control (grey). (B) EIC of **14** (calcd. for  $[M+H]^+ = 238.0678$ ) for extracts of *E. coli* BAP1 expressing pET28b-SUMO-*lbaA\_begLob* (black) or the pET28b-SUMO plasmid as a negative control (grey). (C) Top: Mass spectrum for **13** at  $t_r = 1.79$  min, bottom: simulated mass spectrum for **13**. (D) Top: Mass spectrum for **14** at  $t_r = 5.45$  min, bottom: simulated mass spectrum for **14**.

## SUPPORTING INFORMATION



**Figure S27.** UHPLC-HRMS data of the *in vivo* assay using medium enriched with  $^{13}C$ -methyl-methionine. Shown are extracted ion chromatograms (EICs) for extracts of *E. coli* BAP1 expressing pET28b-SUMO-*lbmA\_begLob* in medium enriched with  $^{13}C$ -methyl-methionine (red, \*) or non-labelled medium (black) or the pET28b-SUMO plasmid as a negative control in medium enriched with  $^{13}C$ -methyl-methionine (yellow, \*) or in non-labelled medium (grey). (A) EICs of **14** (calcd. for  $[M+H]^+ = 238.0678$ ). (B) EICs of **14** incorporating  $^{13}C$  (calcd. for  $[M+H]^+ = 239.0712$ ). (C) Top: mass spectrum for  $^{13}C$ -labelled **14**, bottom: simulated mass spectrum for **14**.



## SUPPORTING INFORMATION

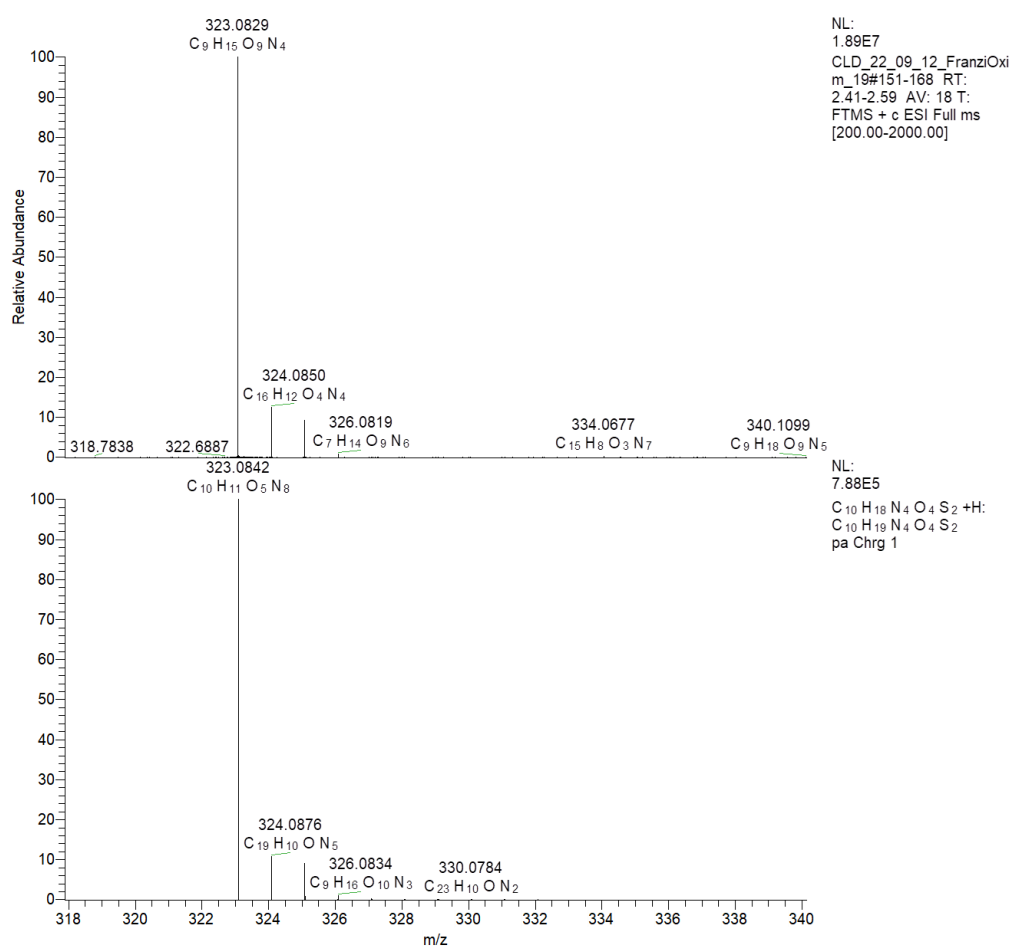
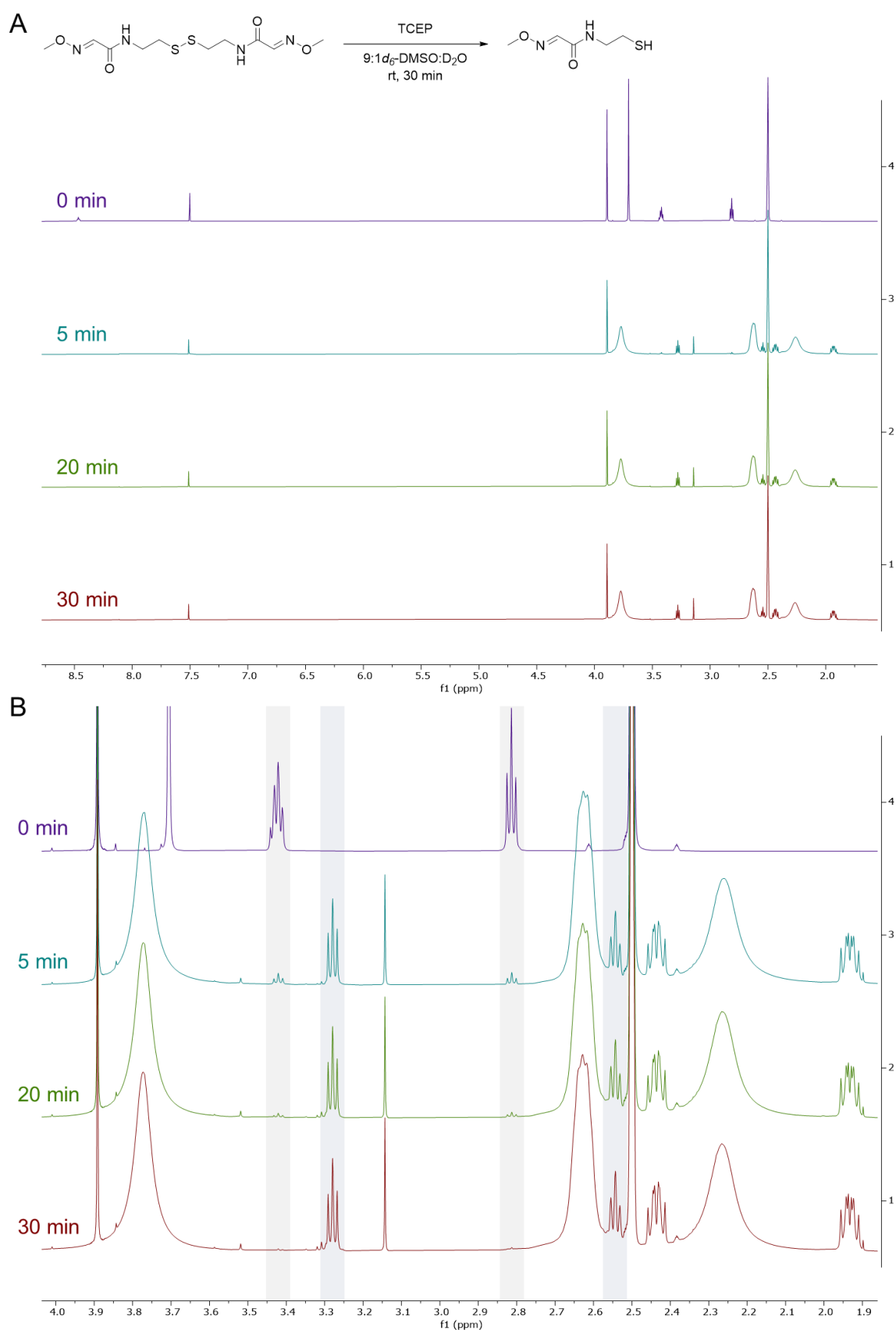


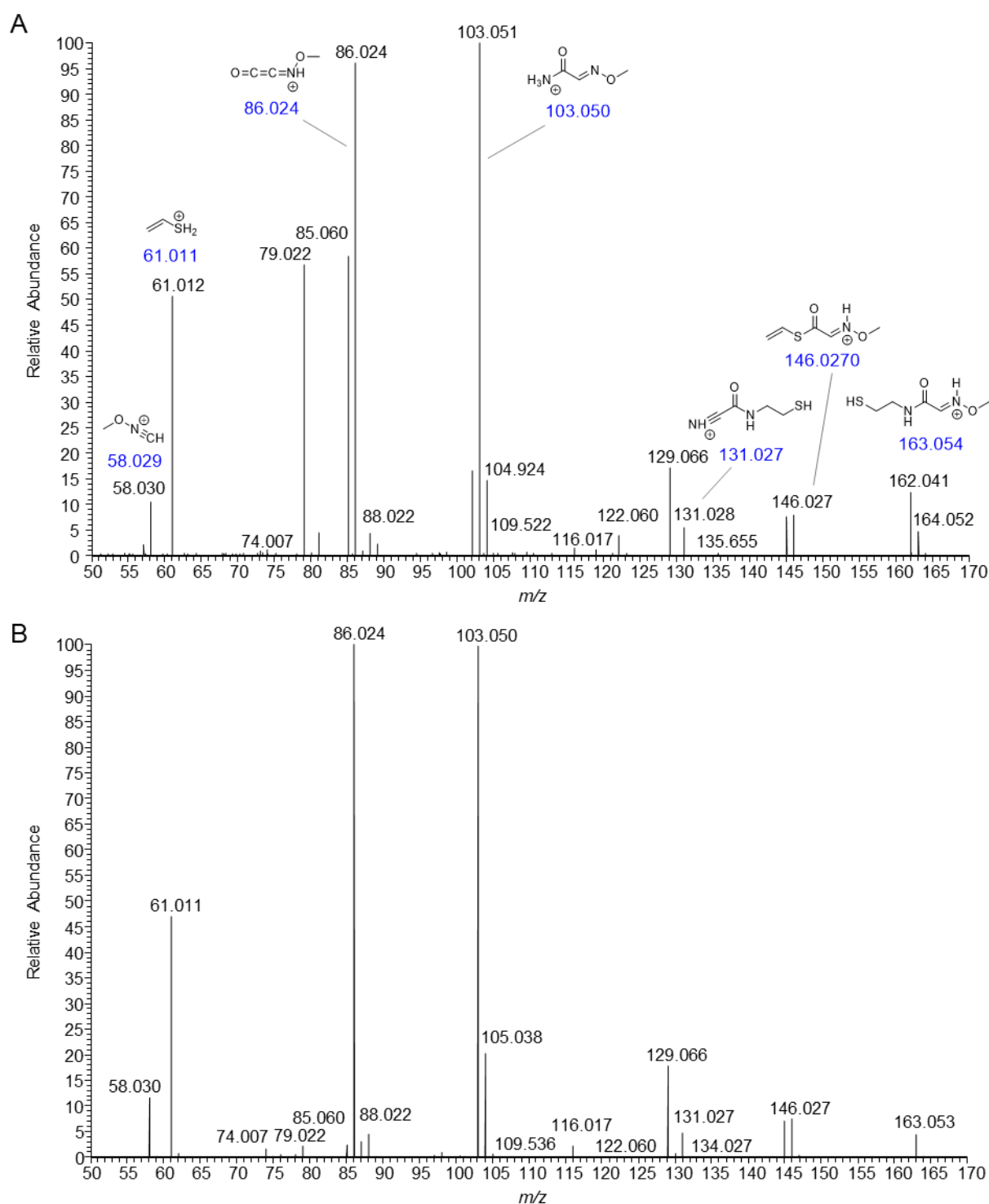
Figure S28. Top: mass spectrum for (*E*)-*N*-(2-mercaptoethyl)-2-(methoxyimino)acetamide (**16**). Bottom: simulated mass spectrum.

## SUPPORTING INFORMATION



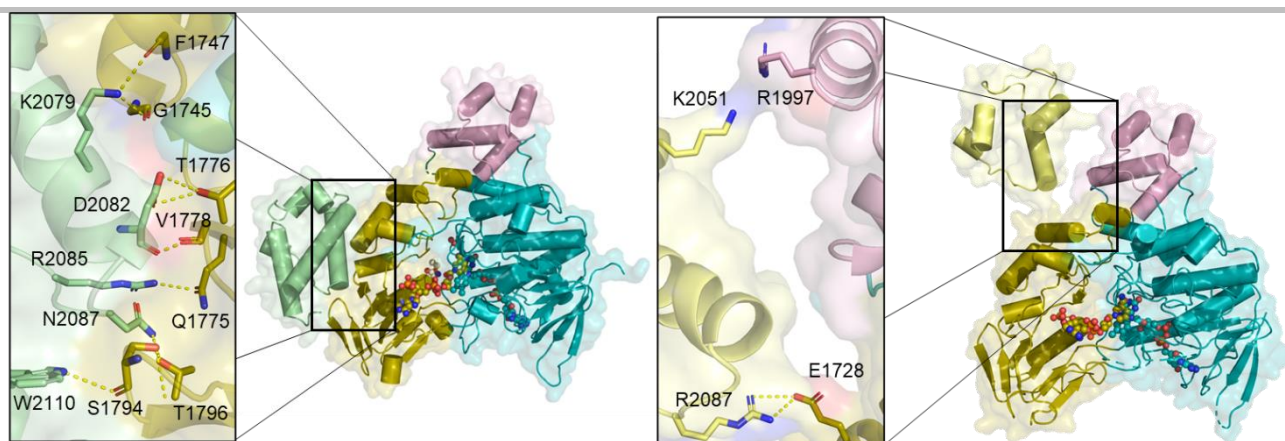
**Figure S29.** *In situ* formation of (*E*)-*N*-(2-mercaptoethyl)-2-(methoxyimino)acetamide (**15**) through reduction of (2*E*,2'*E*)-*N,N*-(disulfanediybis(ethane-2,1-diyl))bis(2-(methoxyimino)acetamide) (**16**) with tris-(2-carboxyethyl)-phosphine (TCEP). (A) <sup>1</sup>H NMR spectra of the reaction mixture in 9:1 DMSO-*d*<sub>6</sub>:D<sub>2</sub>O at 400 MHz from 0–30 min. (B) Enlargement of (A) from 1.9 to 4.0 ppm. The regions with diagnostic signal changes are highlighted by grey boxes.

## SUPPORTING INFORMATION



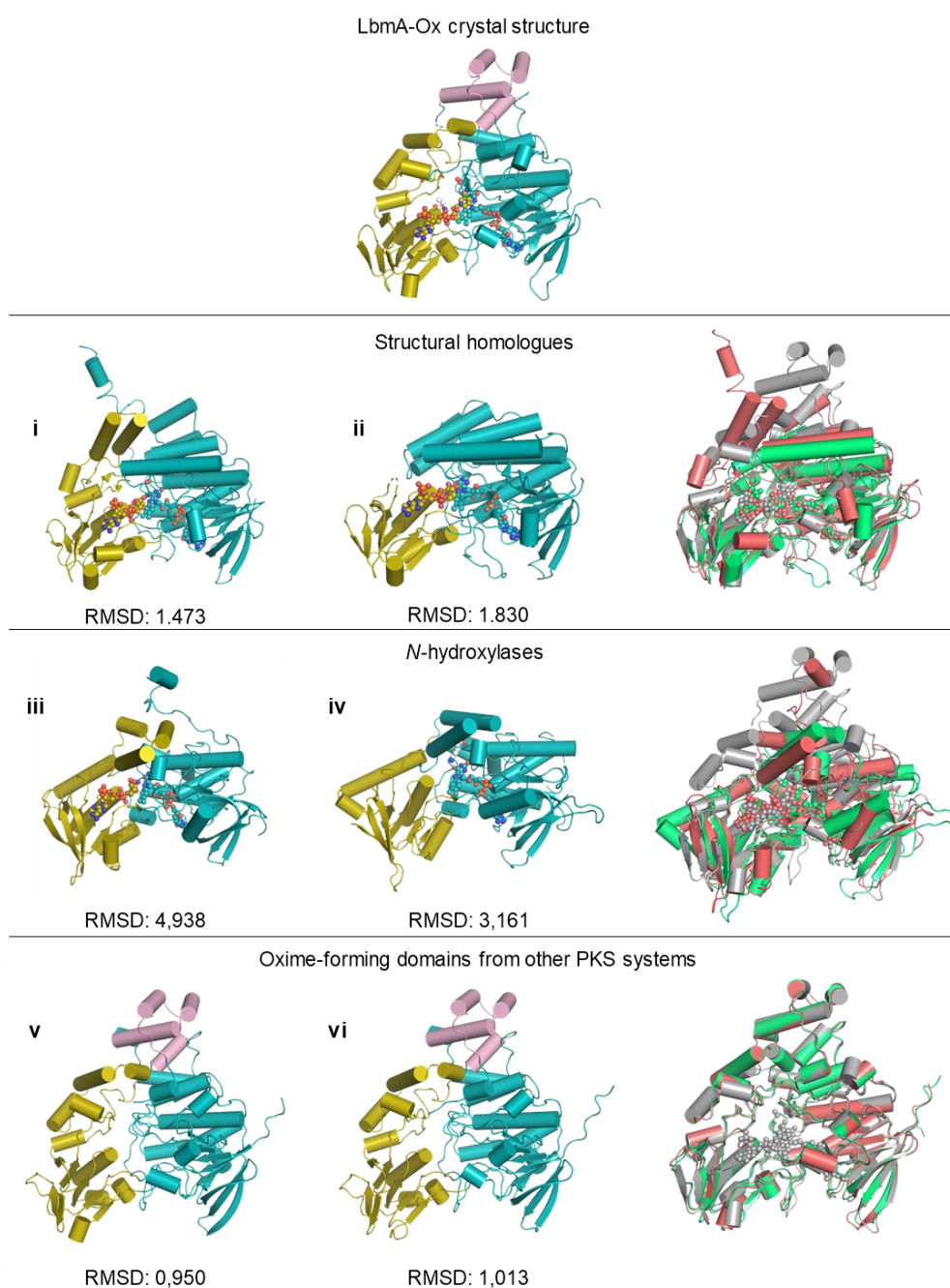
**Figure S30.** MS/MS analyses for (*E*)-*N*-(2-mercaptoethyl)-2-(methoxyimino)acetamide (**15**, calcd. for  $C_5H_{11}N_2O_2S$ ,  $[M+H]^+ = 163.053$ ). (A) *E. coli* lysate expressing the Lob starter module, was treated with cysteamine and TCEP, extracted with EtOAc and an MS/MS spectrum was recorded over a retention time range of 6.71–6.77 min. (B) MS/MS spectrum of synthetic **15** standard, generated *in situ* by reduction of dimer **16** with TCEP, at a retention time of 6.74 min.

## SUPPORTING INFORMATION



**Figure S31.** Conformations of the two domains of LbmA-Ox-ACP in the crystal, with magnified views of the Ox/ACP interfaces. The residues involved in the stabilization of conformation A (left) and conformation B (right) are shown as sticks. The monooxygenase domains are colored according to Figure 5A, and the ACPs as in Figure 5B. The dashed yellow lines represent the electrostatic interactions contributing to the domain-domain interfaces.

## SUPPORTING INFORMATION



**Figure S32.** Structures and models of LbmA-Ox homologues. The closest structural homologues **i** (ancestral flavin-containing monooxygenase (PDB ID: 6SE3)) and **ii** (pyrrolizidine alkaloid *N*-oxygenase (PDB ID: 5NMX)) were found using the DALI server, the *N*-hydroxylases **iii** (putrescine *N*-hydroxylase (PDB ID: 7US3)) and **iv** (ornithine hydroxylase (PDB ID: 6X0K)) were identified in the Protein Data Bank, and the models of the *N*-hydroxylating domains **v** (necroxime oxime-forming domain (*Burkholderia* sp. B8)) and **vi** (oximidine oxime-forming domain (*Chromobacterium phragmitis*)) were calculated with AlphaFold2<sup>17</sup>. Structural superpositions with LbmA-Ox (grey) and the respective RMSDs are displayed for each group, with **i**, **iii** and **v** in red and **ii**, **iv** and **vi** in green. The protein domains are colored according to Figure 5A.

## SUPPORTING INFORMATION

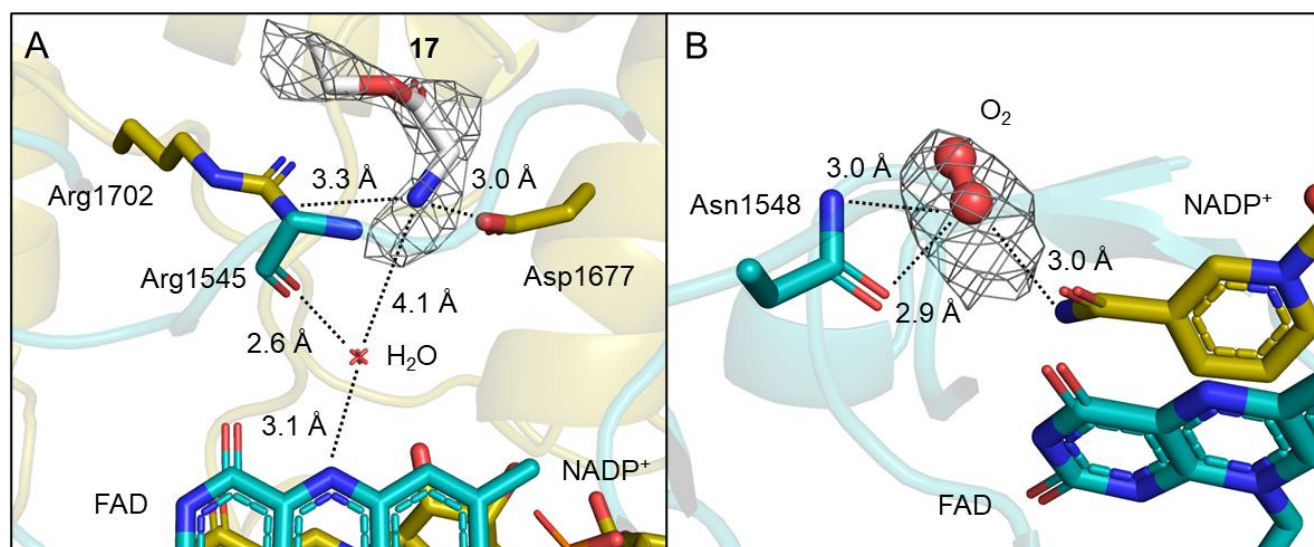
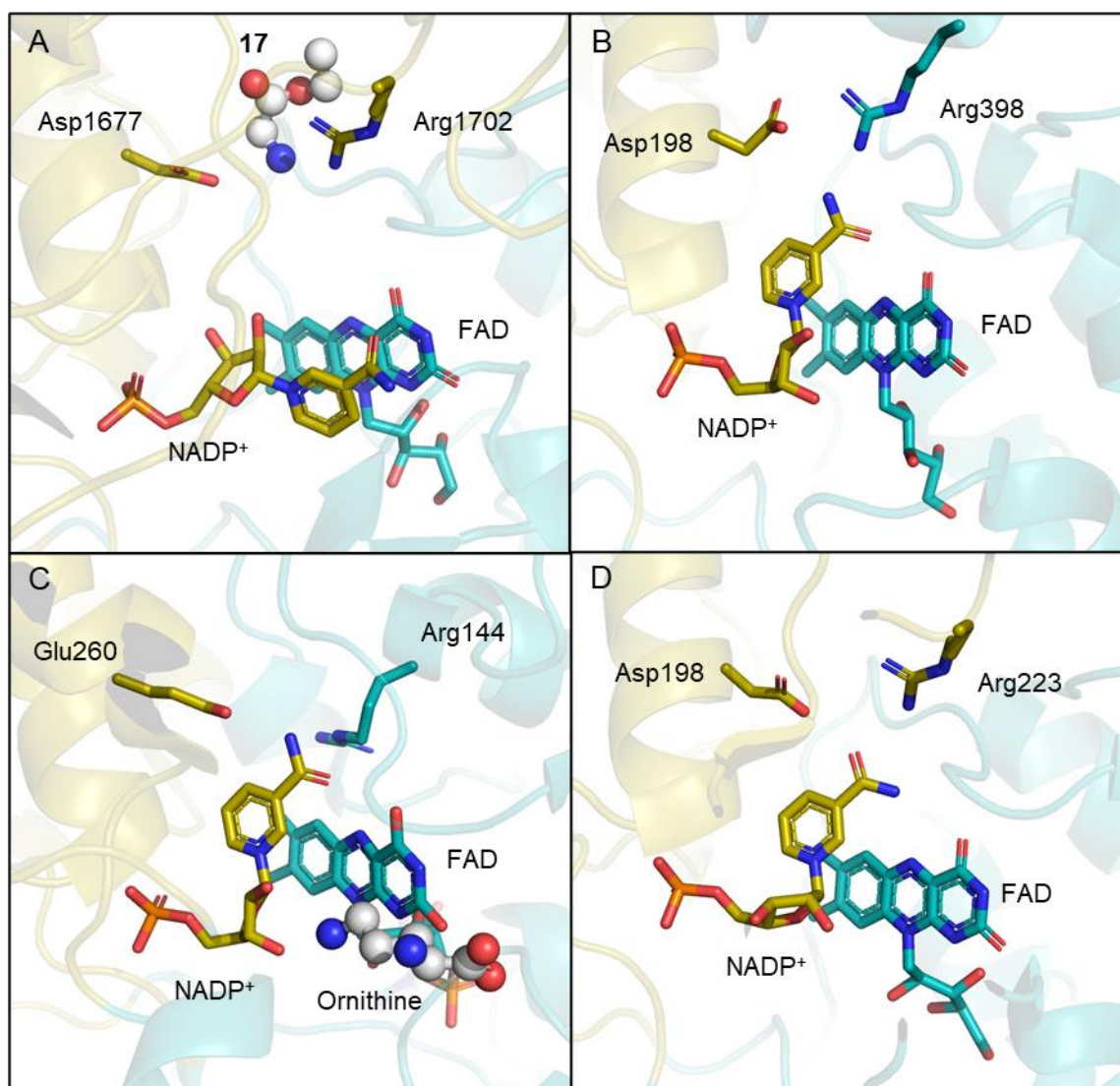


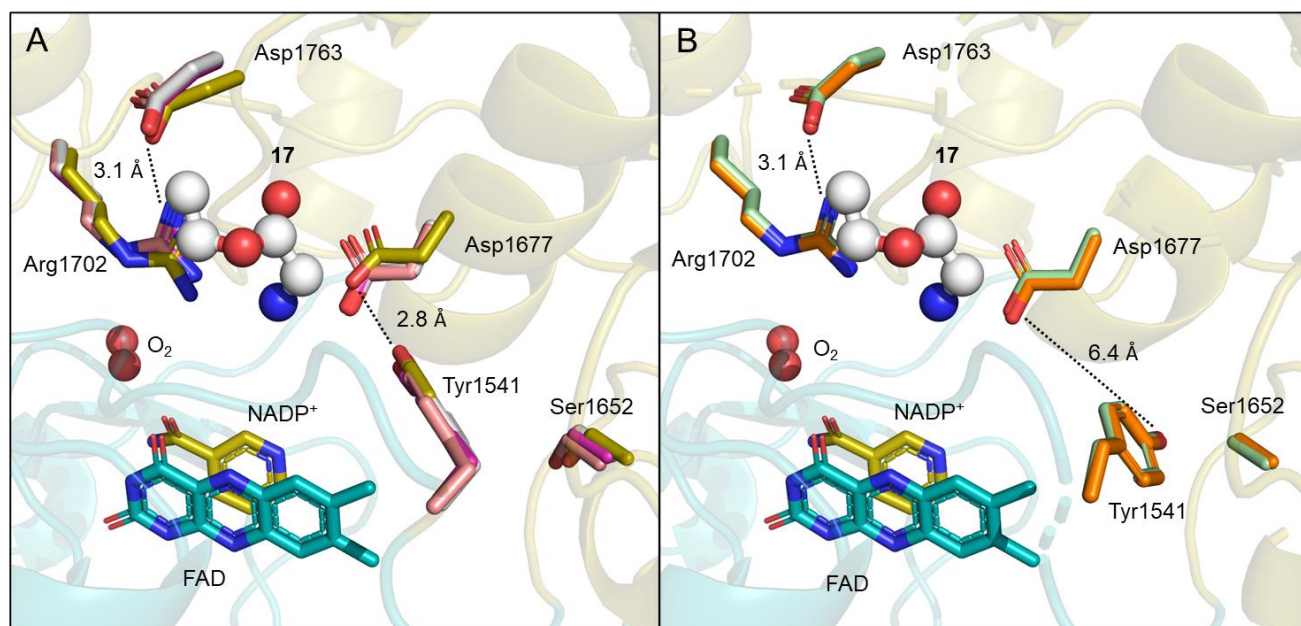
Figure S33. Ethyl glycinate (17) (A) and molecular dioxygen (B) in the LbMA-Ox active site, with the omit maps contoured at 3  $\sigma$ .

## SUPPORTING INFORMATION



**Figure S34.** FAD-NADP<sup>+</sup> conformations captured in various *N*-monoxygenase crystal structures. The NADP<sup>+</sup> nicotinamide moiety stacks against the FAD isoalloxazine in LbmA-Ox (A), while it is located above the flavin N5 atom in (B) pyrrolizidine alkaloid *N*-oxygenase (PDB ID: 5NMX), (C) ornithine hydroxylase (PDB ID: 6X0J) and (D) an ancestral flavin-containing monooxygenase (PDB ID: 6SE3). The structures and cofactor conformers are colored as in Figure 5A. The side chains of Asp1677 and Arg1702 in LbmA-Ox and its homologues are shown as sticks, and the substrates as spheres. 17 = ethyl glycinate.

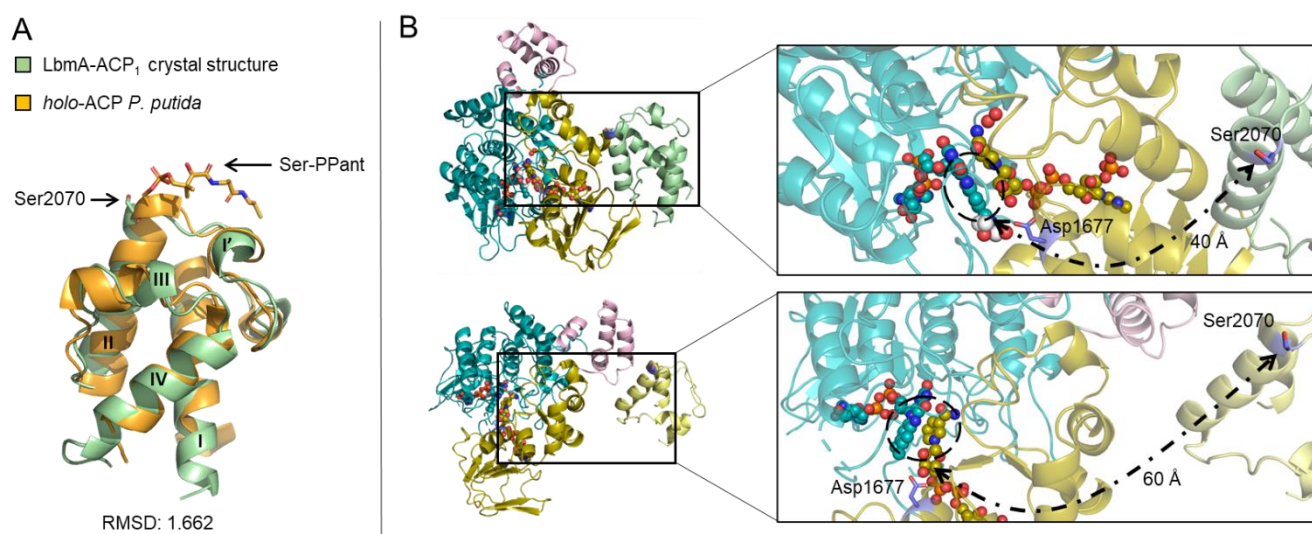
## SUPPORTING INFORMATION



**Figure S35.** Positioning of the residues involved in catalysis of the dehydration reaction in LbmA-Ox in models of its homologues from various PKS systems. (A) In the Ox domains lacking a downstream DH, the Tyr1541 side chain points toward Asp1677 as in the LbmA-Ox crystal structure, whereas (B) in the Ox domains followed by a DH, Tyr1541 is oriented toward Ala1652 (which is a serine in DH-lacking systems). The substrate analogue, dioxygen, FAD and NADP<sup>+</sup> of the LbmA-Ox structure are shown for comparison and colored as in Figure 5A. For the sake of clarity, only the isoalloxazine and nicotinamide moieties of FAD and NADP<sup>+</sup> are represented. **17** = ethyl glycinate. In (A), Ox domains without a downstream DH: LbmA-Ox crystal structure (gold), oximidine-Ox (*Pseudomonas baetica*) (salmon), necroxime-Ox (*Rhizobium* sp. BK315) (magenta). An AlphaFold model of LbmA-Ox is also shown (grey) to demonstrate that the side chain orientation prediction for the other Ox domains is accurate. In (B), Ox domains with a downstream DH: oximidine-Ox (*Chromobacterium phragmitis*) (orange), necroxime-Ox (*Burkholderia* sp. B8) (light green).

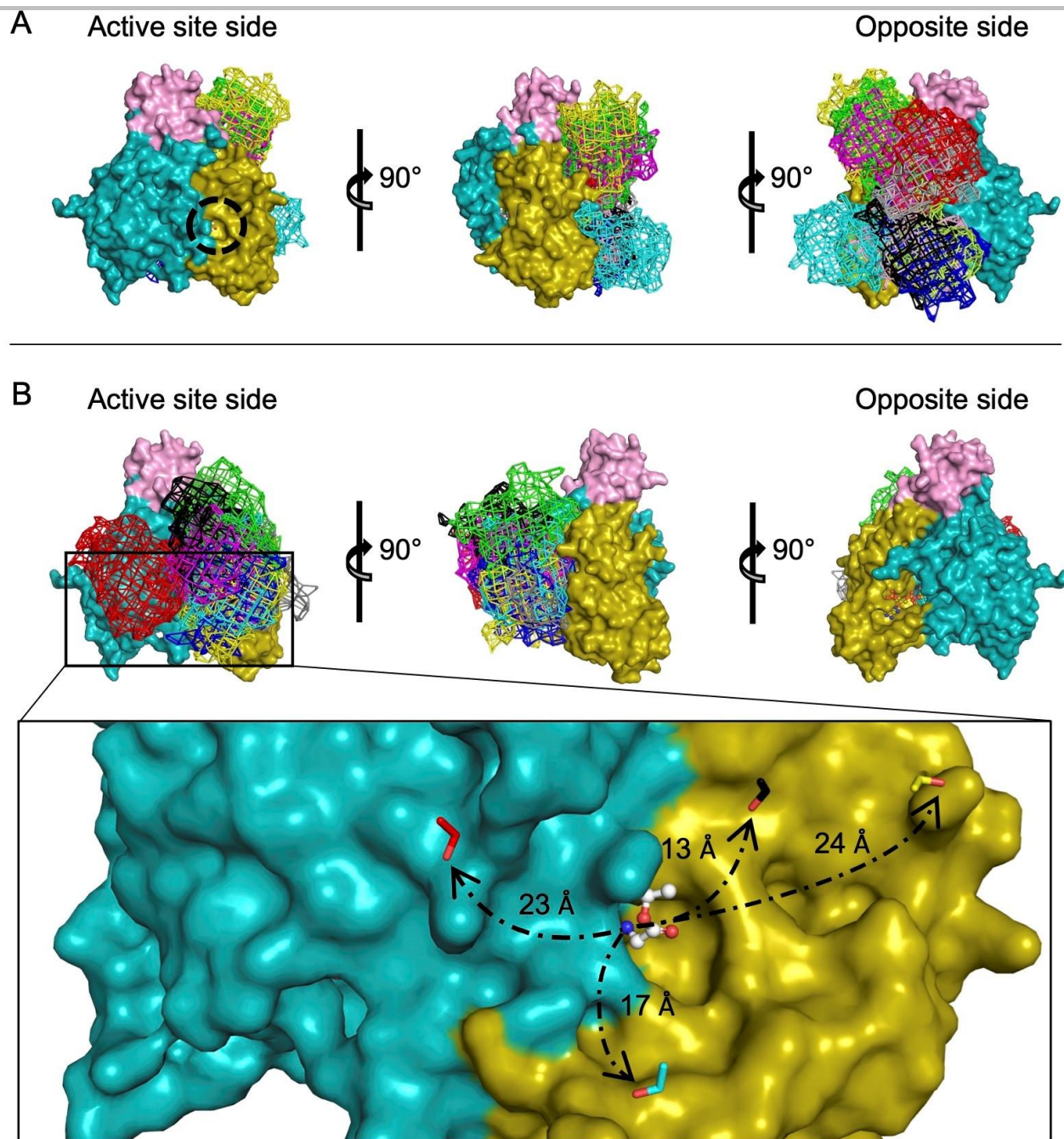


## SUPPORTING INFORMATION



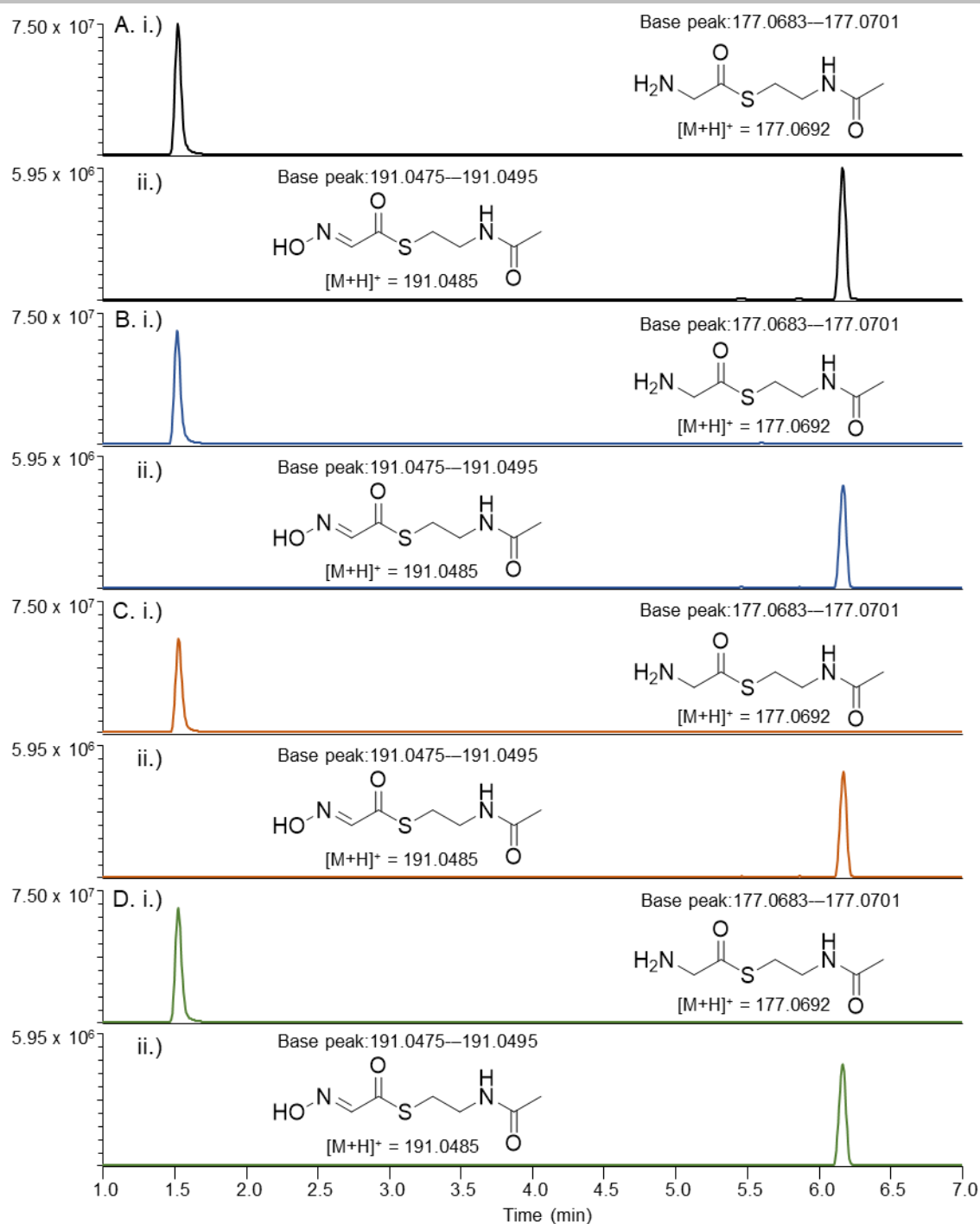
**Figure S36.** Identification of the substrate-bound serine of LbmA-ACP<sub>1</sub> and position in the Ox-ACP didomain context. A) Superposition of the ACP<sub>1</sub> crystal structure of chain A (light green) and the *holo*-ACP crystal structure from *Pseudomonas putida* (PDB ID: 7PDI) (orange). The  $\alpha$ -helices of ACP<sub>1</sub> are numbered I, II, III and IV and the pseudo-helix between helices  $\alpha$ I and  $\alpha$ II is marked as helix I'. Despite a low percentage sequence identity (12.6%), the two ACP structures are highly similar (RMSD = 1.662), allowing the identification of Ser2070 at the N-terminus of helix  $\alpha$ II as the most likely residue to carry the phosphopantetheine-bound substrate. B) Positions of Ser2070 relative to the FAD/NADP<sup>+</sup> interface. Ser2070 is located 40 Å and 60 Å away respectively from Asp1677 in chain A (top) and chain B (bottom), distances that cannot be spanned by the 25 Å long native substrate. Nonetheless, the ACP conformations are necessary for stabilizing the crystal lattice via interactions with didomains of other asymmetric units, potentially revealing why the isolated Ox domain was recalcitrant to crystallization. The monooxygenase domains are colored according to Figure 5A, the ACPs are colored as in Figure 5B, Ser2070 and Asp1677 are shown as lavender sticks, and the substrate channel entrance is indicated by a dashed circle.

## SUPPORTING INFORMATION



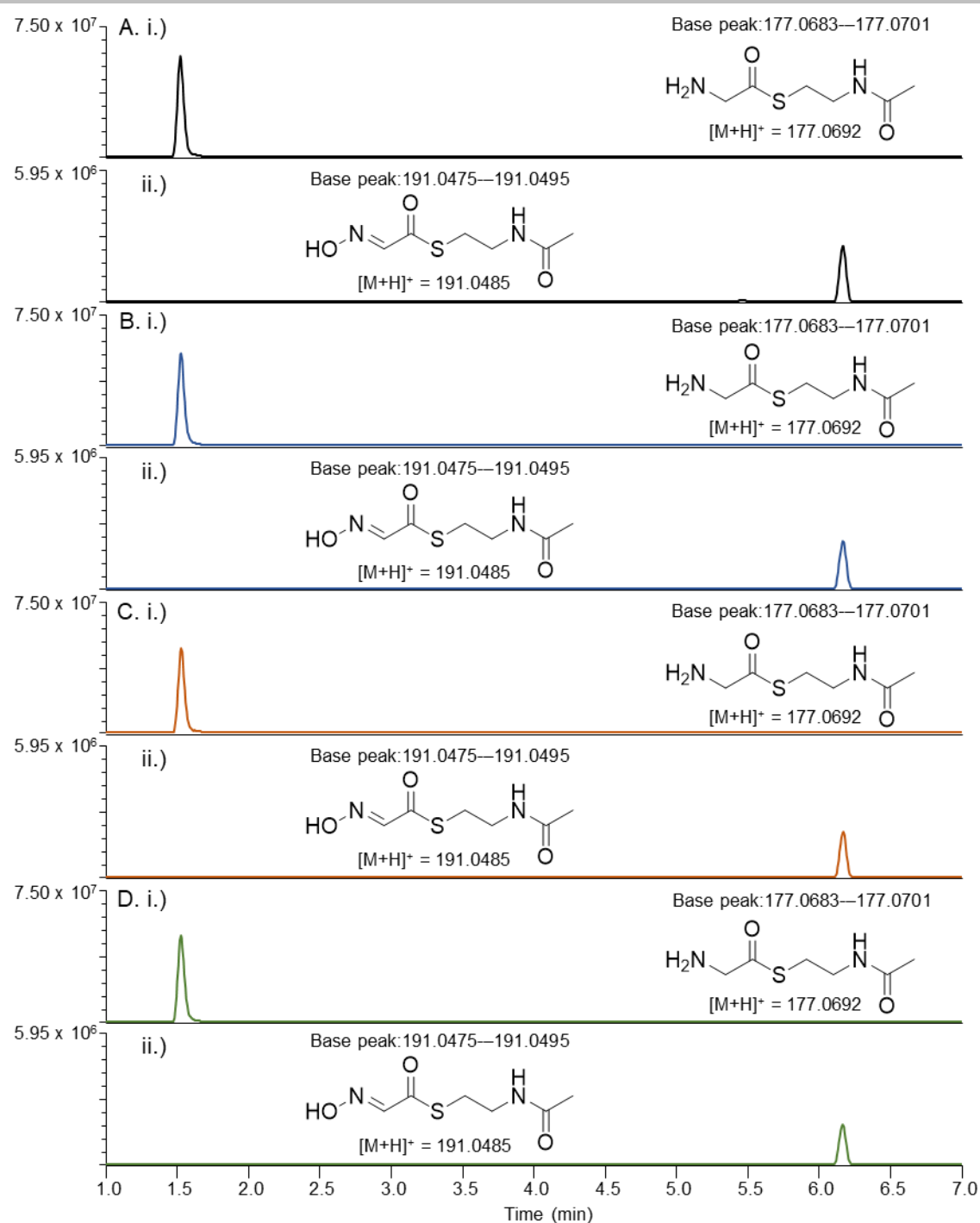
**Figure S37.** Docking models of LbmA-Ox and LbmA-ACP<sub>1</sub> provided by the HDock server.<sup>[23]</sup> These models are characterized by docking and confidence scores. Protein-protein complexes deposited in the PDB display a docking score around -200 or better, and a confidence score of 0.7 or more describes a high binding likelihood of the two molecules. A) A 55 Å maximum length restraint was set between the LbmA-Ox C-terminus and the LbmA-ACP<sub>1</sub> N-terminus. The ten most favorable solutions based on the docking and confidence scores are displayed. In all of these docking solutions, the ACP is located on the side of the monooxygenase opposite the entrance of the substrate channel, arguing that the observed ACP conformations are not directly relevant to the oximation reaction. The docking and confidence scores of these models range from -116.4 to -187.7 and from 0.34 to 0.68, respectively. B) The docking models were calculated without any linker length restraint. Among the ten most favorable solutions, eight place the ACP on the active site side of the monooxygenase and are shown here. Notably, taking into account the 25 Å length of the native substrate in linear form, four of these docking models place Ser2070 of ACP<sub>1</sub> close enough to the terminal amine of the substrate analogue to allow for a productive interaction. The docking and confidence scores of these models range from -196.2 to -256.0 and from 0.72 to 0.89, respectively, consistent with native protein-protein complex scores. The substrate-carrying Ser2070 of the four ACPs and their distances relative to the terminal amine of **17** are shown in magnified view. LbmA-Ox is colored according to Figure 5A, the ACPs are depicted in mesh representation using a different color for each solution, and the substrate channel entrance is indicated by a dashed circle.

## SUPPORTING INFORMATION



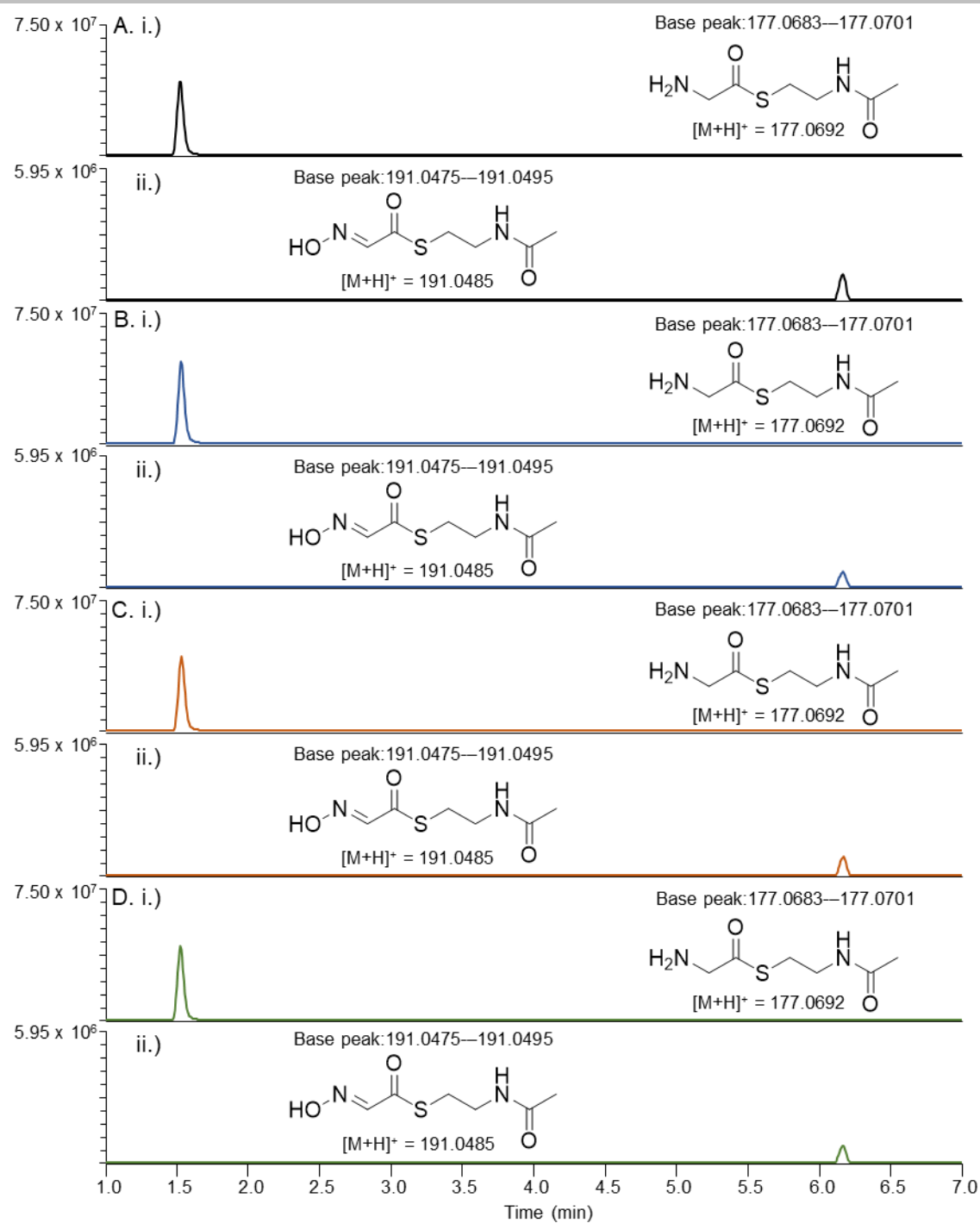
**Figure S38.** Base peak analyses for enzyme activity assays with LbmA-Ox wildtype (WT1), with the intensities scaled to the maximum value of assay set 1. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



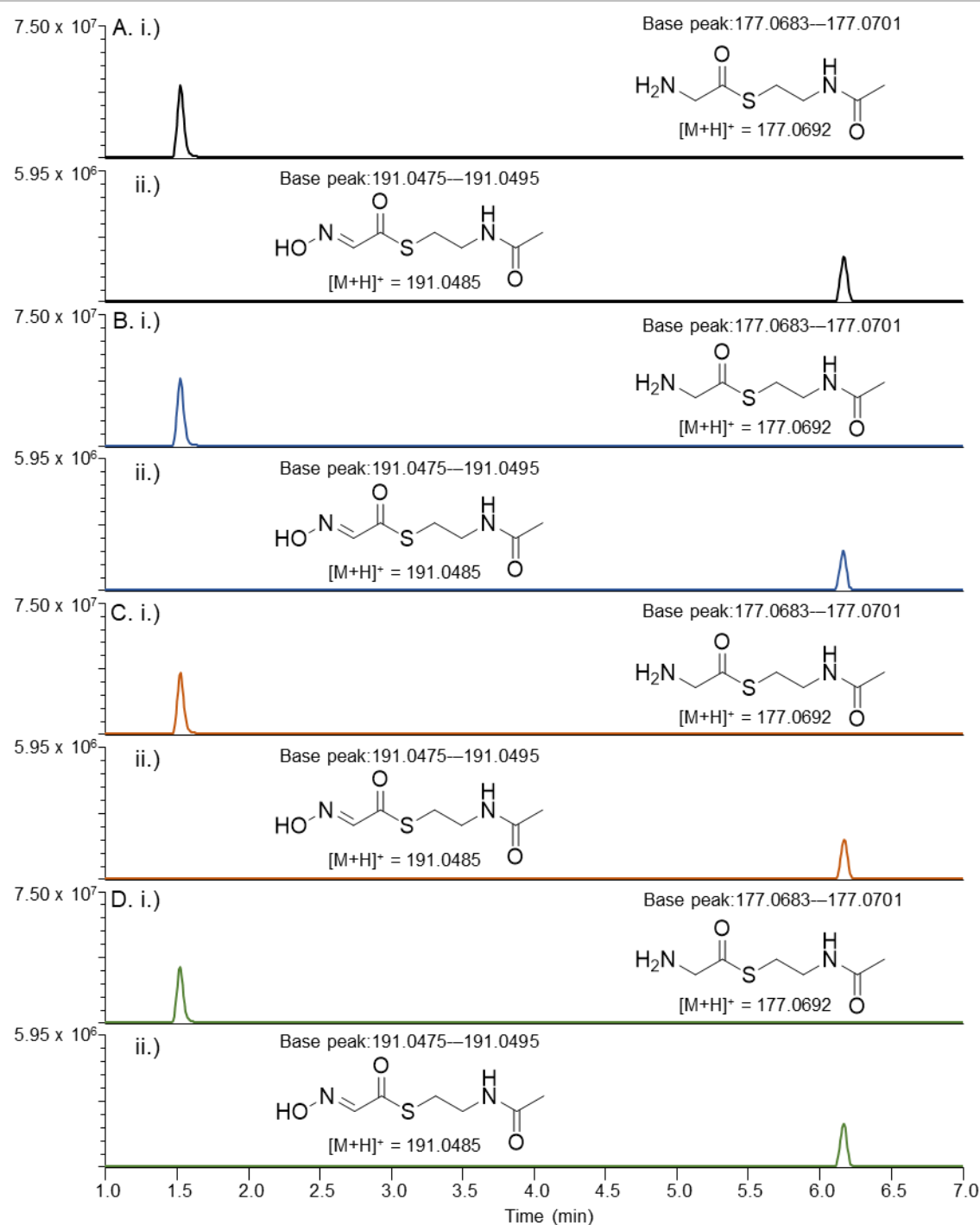
**Figure S39.** Base peak analyses for enzyme activity assays with Lbma-Ox R1694A, with the intensities scaled to the maximum value of assay set 1. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



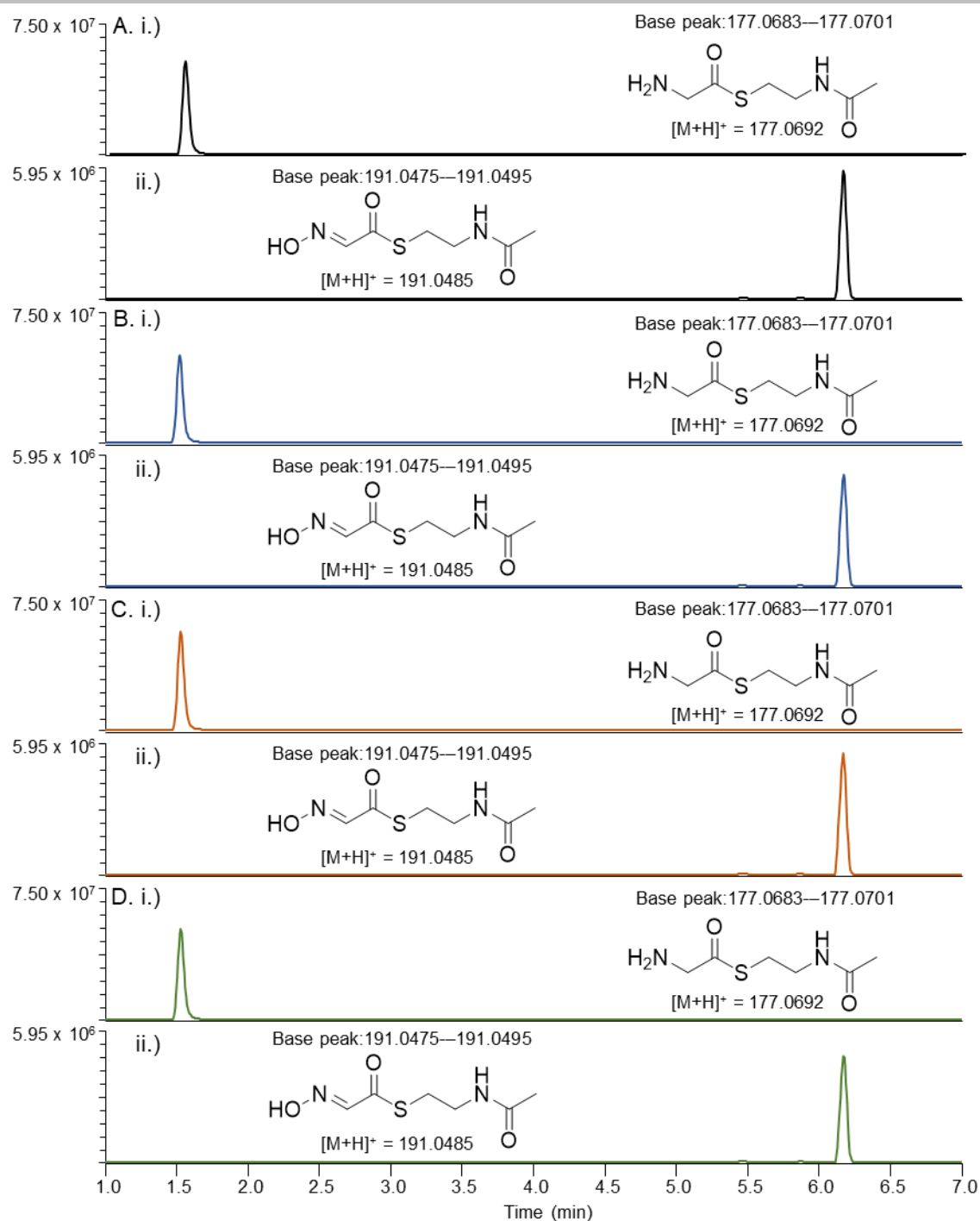
**Figure S40.** Base peak analyses for enzyme activity assays with LbmA-Ox E1521A, with the intensities scaled to the maximum value of assay set 1. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



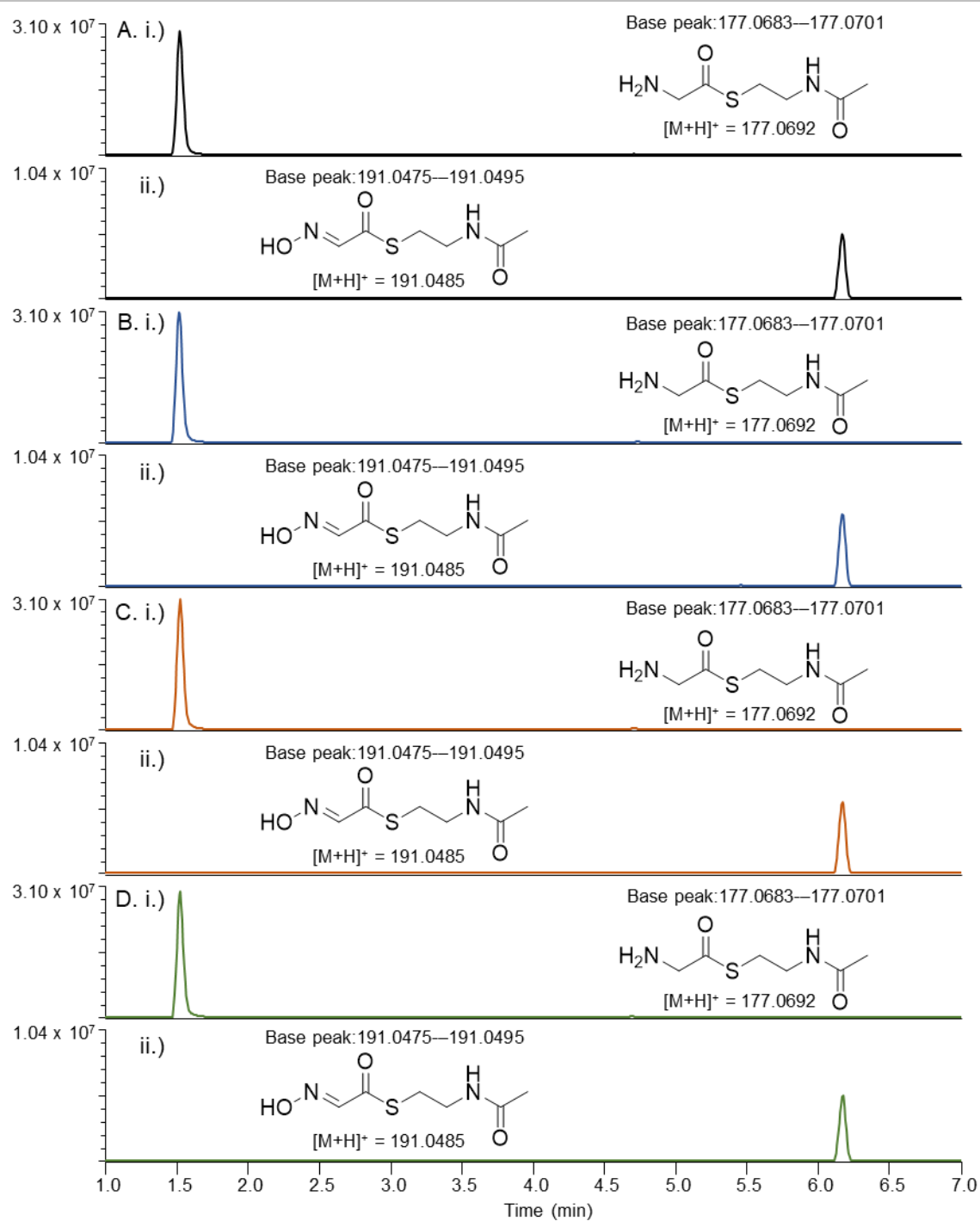
**Figure S41.** Base peak analyses for enzyme activity assays with LbmA-Ox E1680Q, with the intensities scaled to the maximum value of assay set 1. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



**Figure S42.** Base peak analyses for enzyme activity assays with LbmA-Ox T1851V, with the intensities scaled to the maximum value of assay set 1. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

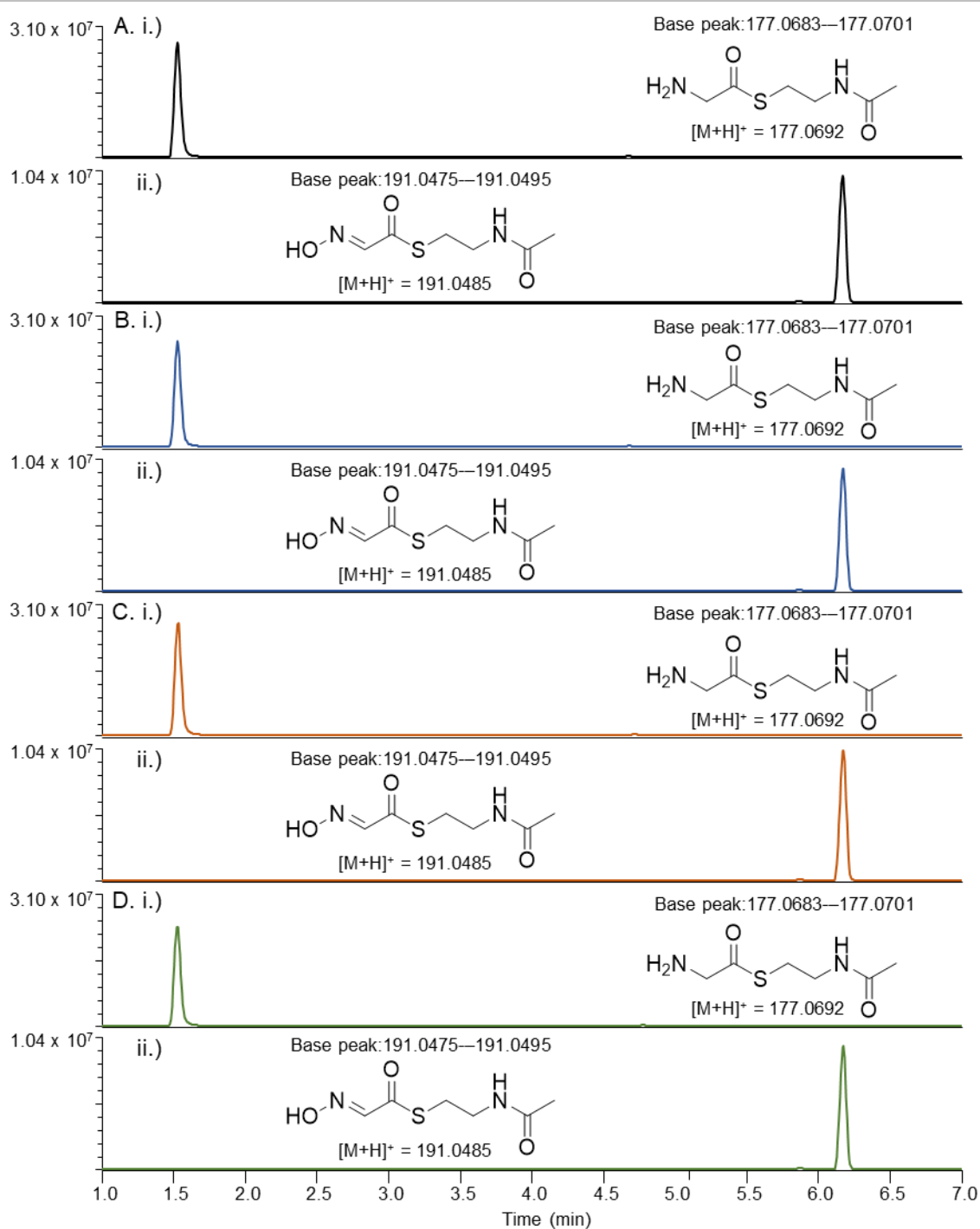
## SUPPORTING INFORMATION



**Figure S43.** Base peak analyses for enzyme activity assays with LbmA-Ox wildtype (WT6), with the intensities scaled to the maximum value of assay set 2. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate 6 and ii) GlySNAC oxime product 7.

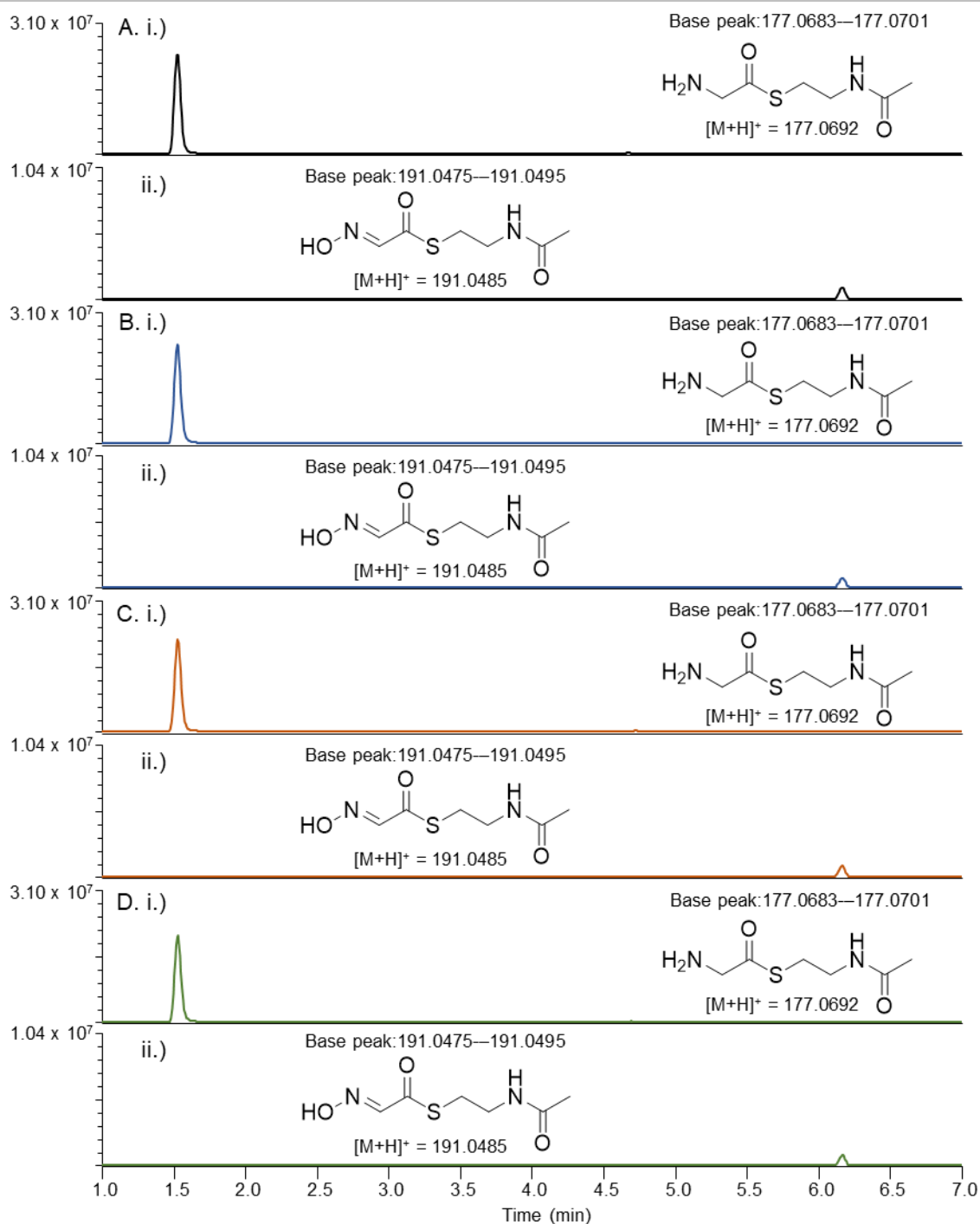


## SUPPORTING INFORMATION



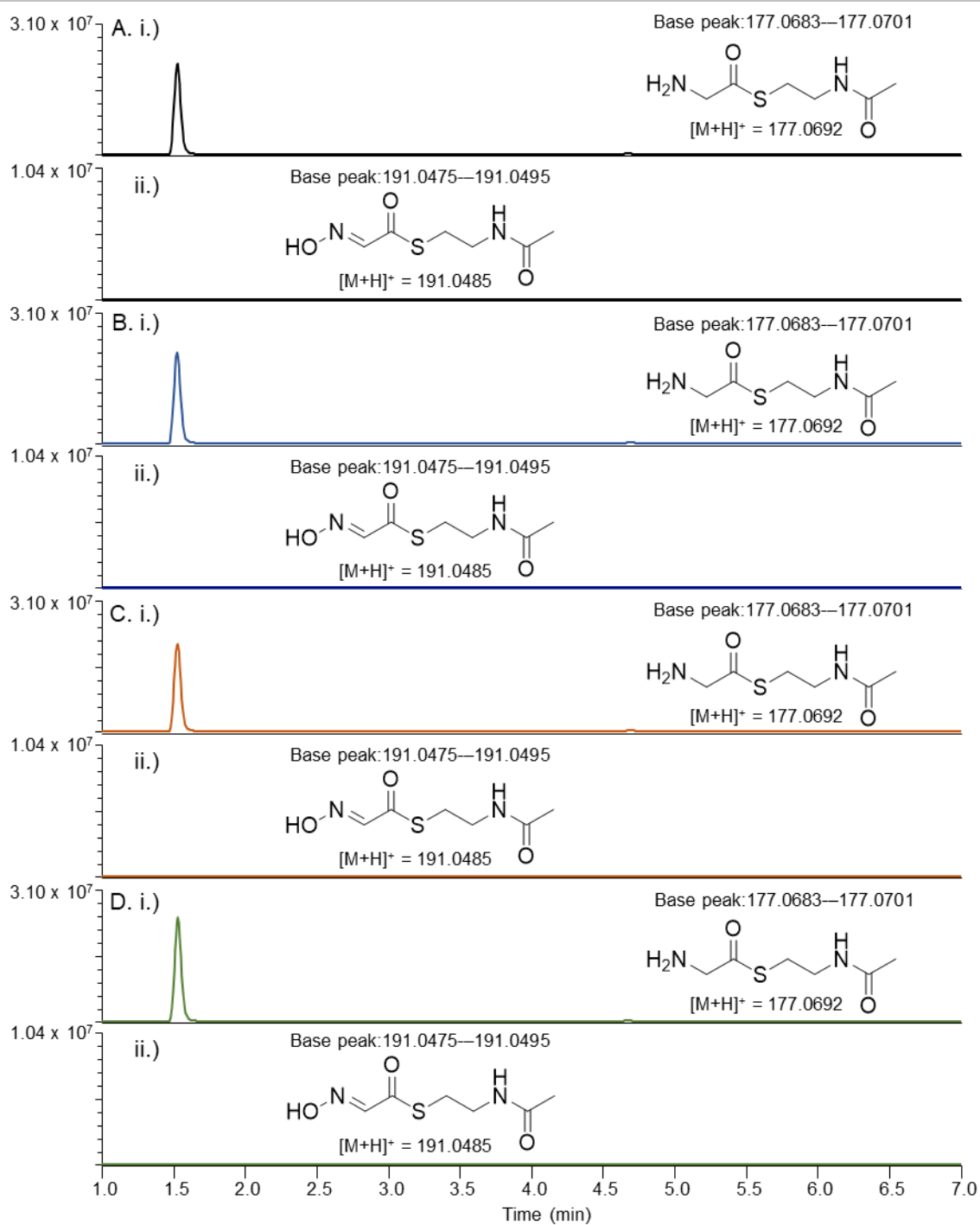
**Figure S44.** Base peak analyses for enzyme activity assays with LbmA-Ox K1632A, with the intensities scaled to the maximum value of assay set 2. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



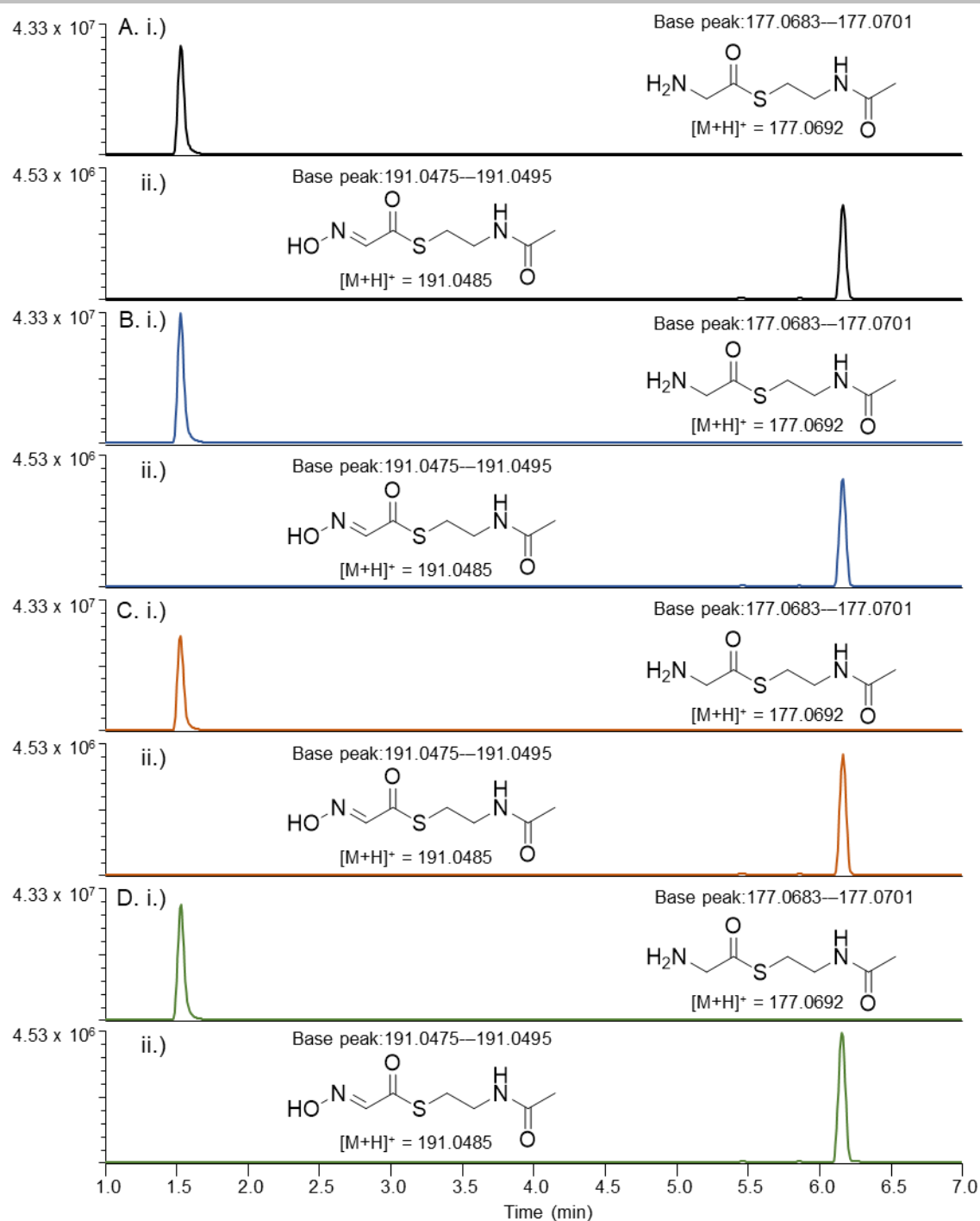
**Figure S45.** Base peak analyses for enzyme activity assays with LbmA-Ox N1709D, with the intensities scaled to the maximum value of assay set 2. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



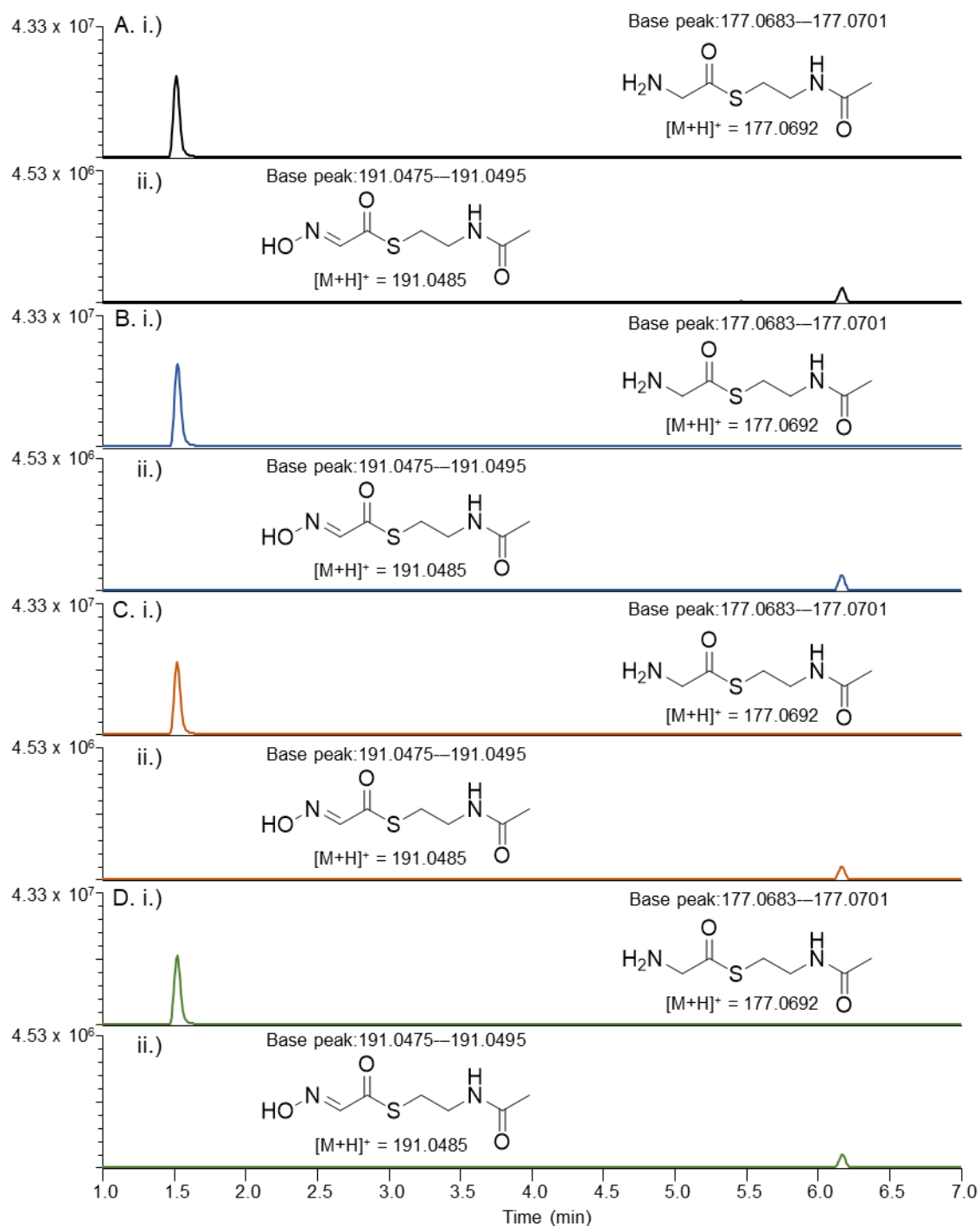
**Figure S46.** Base peak analyses for enzyme activity assays with LbmA-Ox Q1547E, with the intensities scaled to the maximum value of assay set 2. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



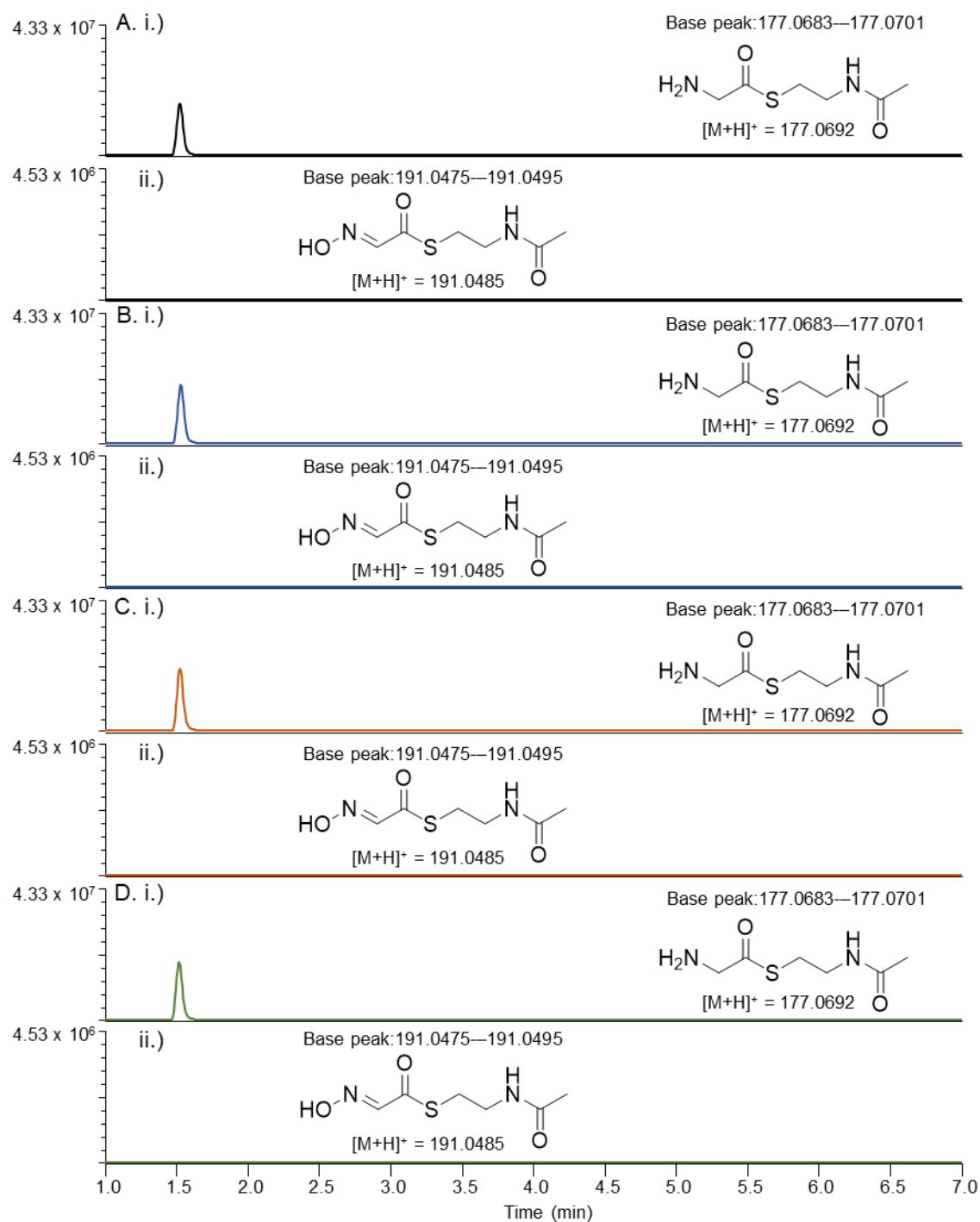
**Figure S47.** Base peak analyses for enzyme activity assays with Lbma-Ox wildtype (WT10), with the intensities scaled to the maximum value of assay set 3. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



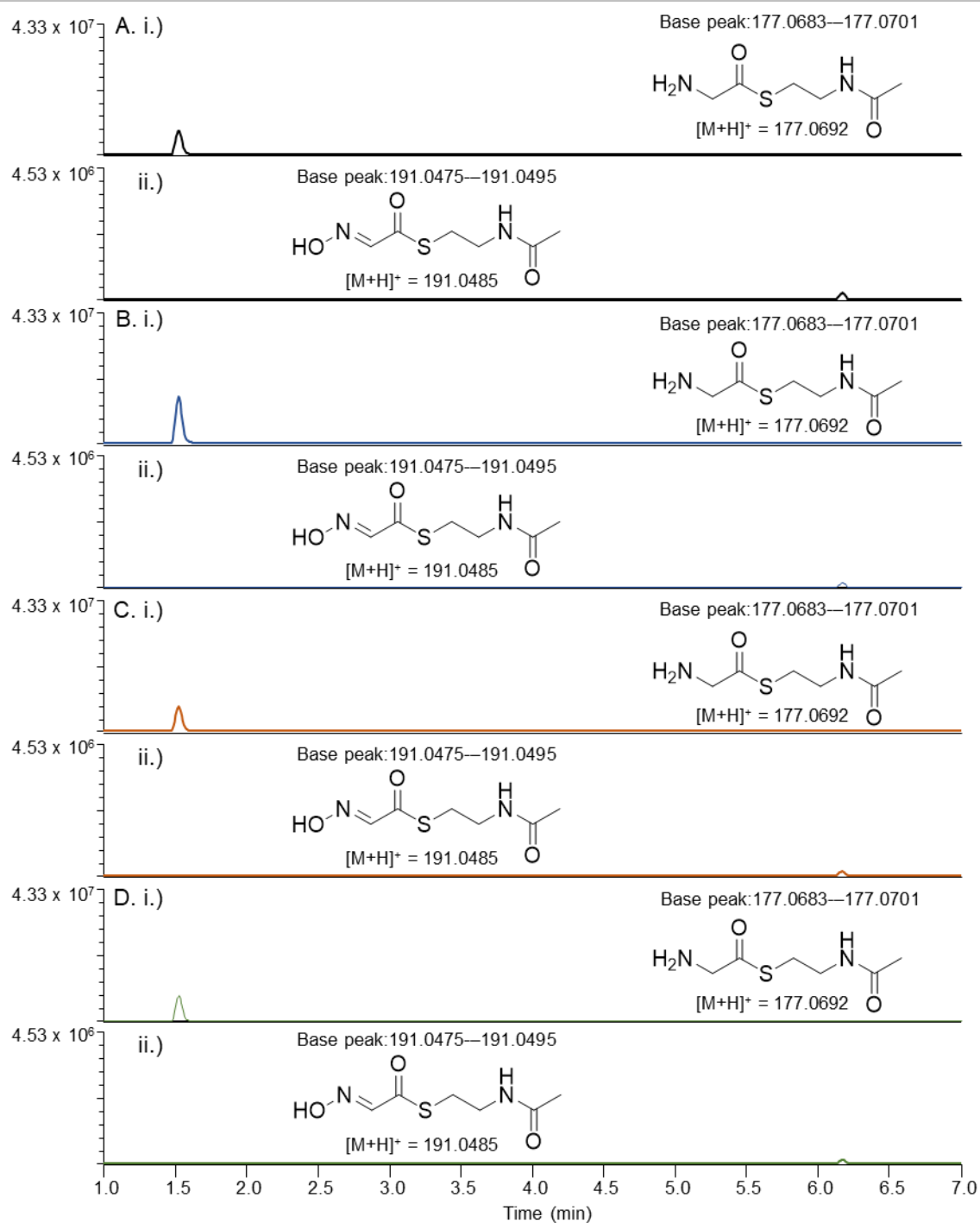
**Figure S48.** Base peak analyses for enzyme activity assays with LbmA-Ox N1548D, with the intensities scaled to the maximum value of assay set 3. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



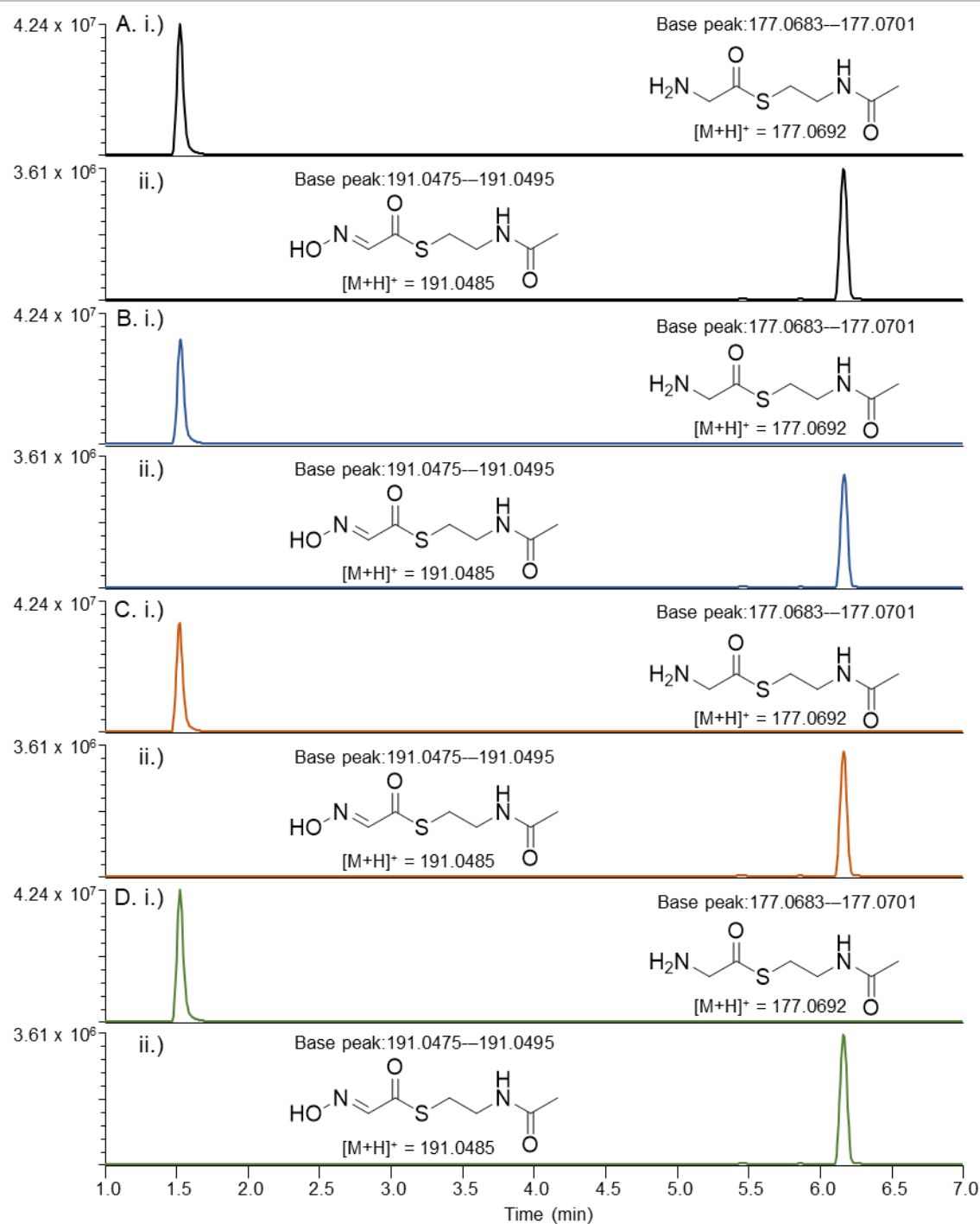
**Figure S49.** Base peak analyses for enzyme activity assays with LbmA-Ox D1677N, with the intensities scaled to the maximum value of assay set 3. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



**Figure S50.** Base peak analyses for enzyme activity assays with LbmA-Ox N1709A, with the intensities scaled to the maximum value of assay set 3. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

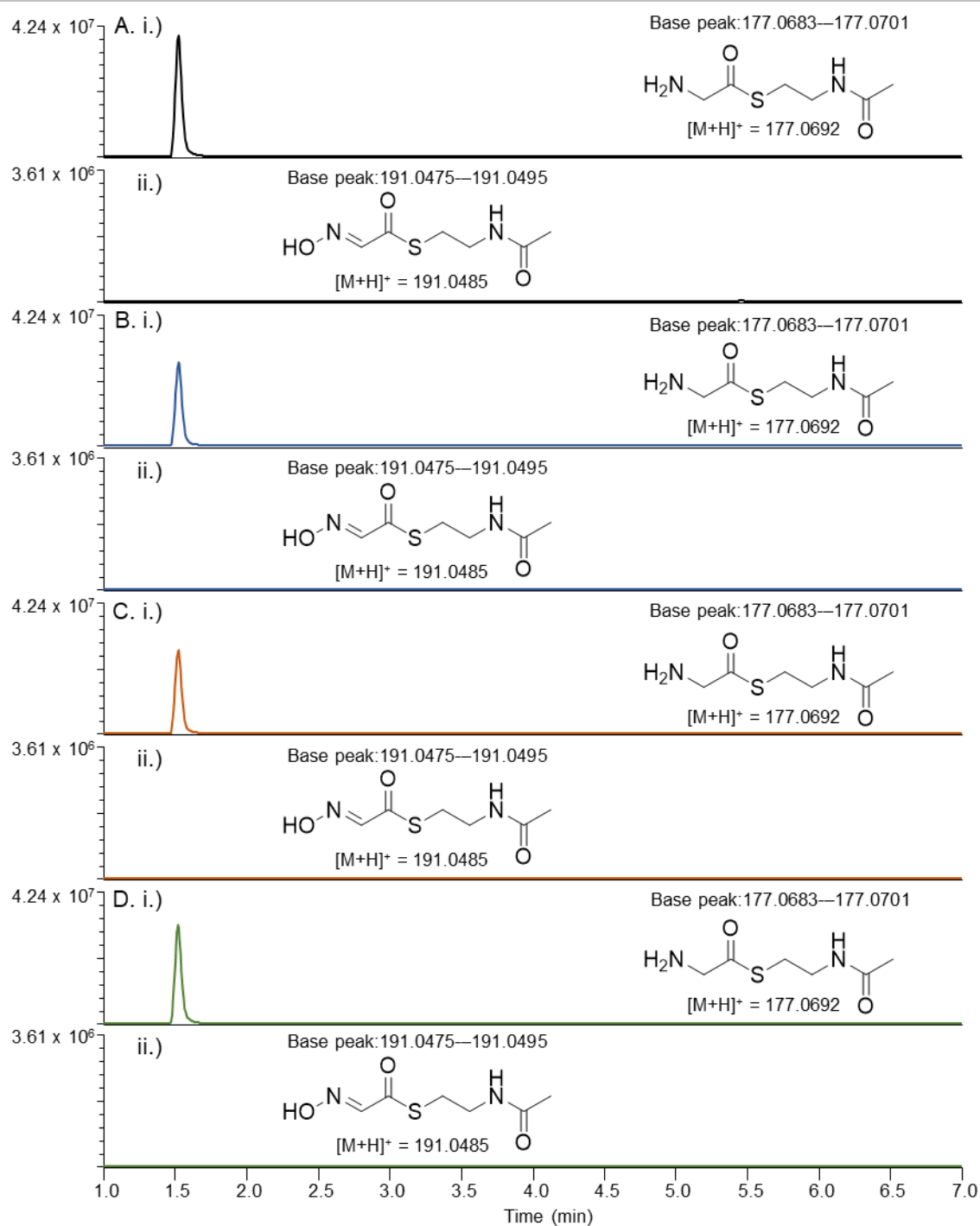
## SUPPORTING INFORMATION



**Figure S51.** Base peak analyses for enzyme activity assays with LbmA-Ox wildtype (WT15), with the intensities scaled to the maximum value of assay set 4. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

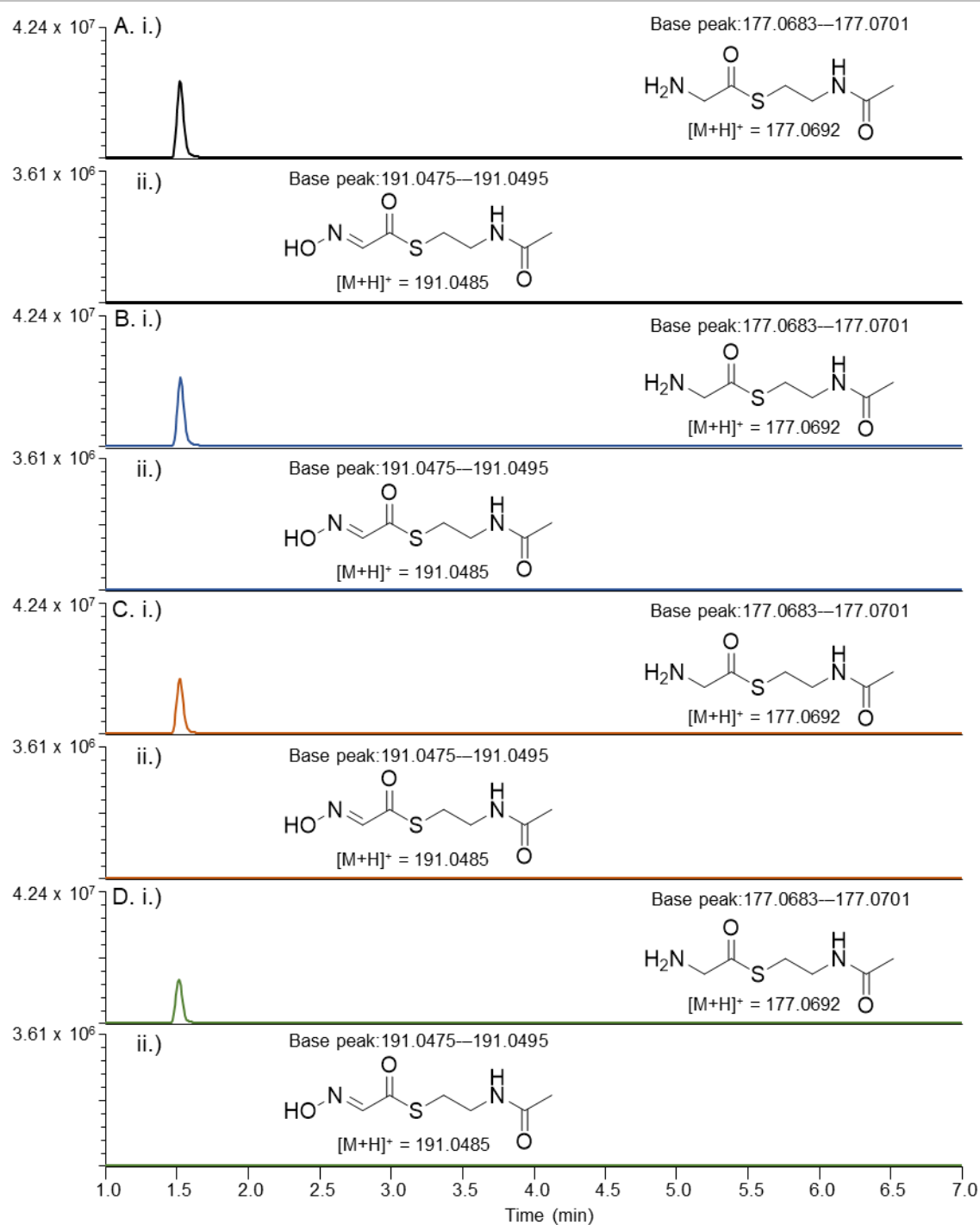


## SUPPORTING INFORMATION



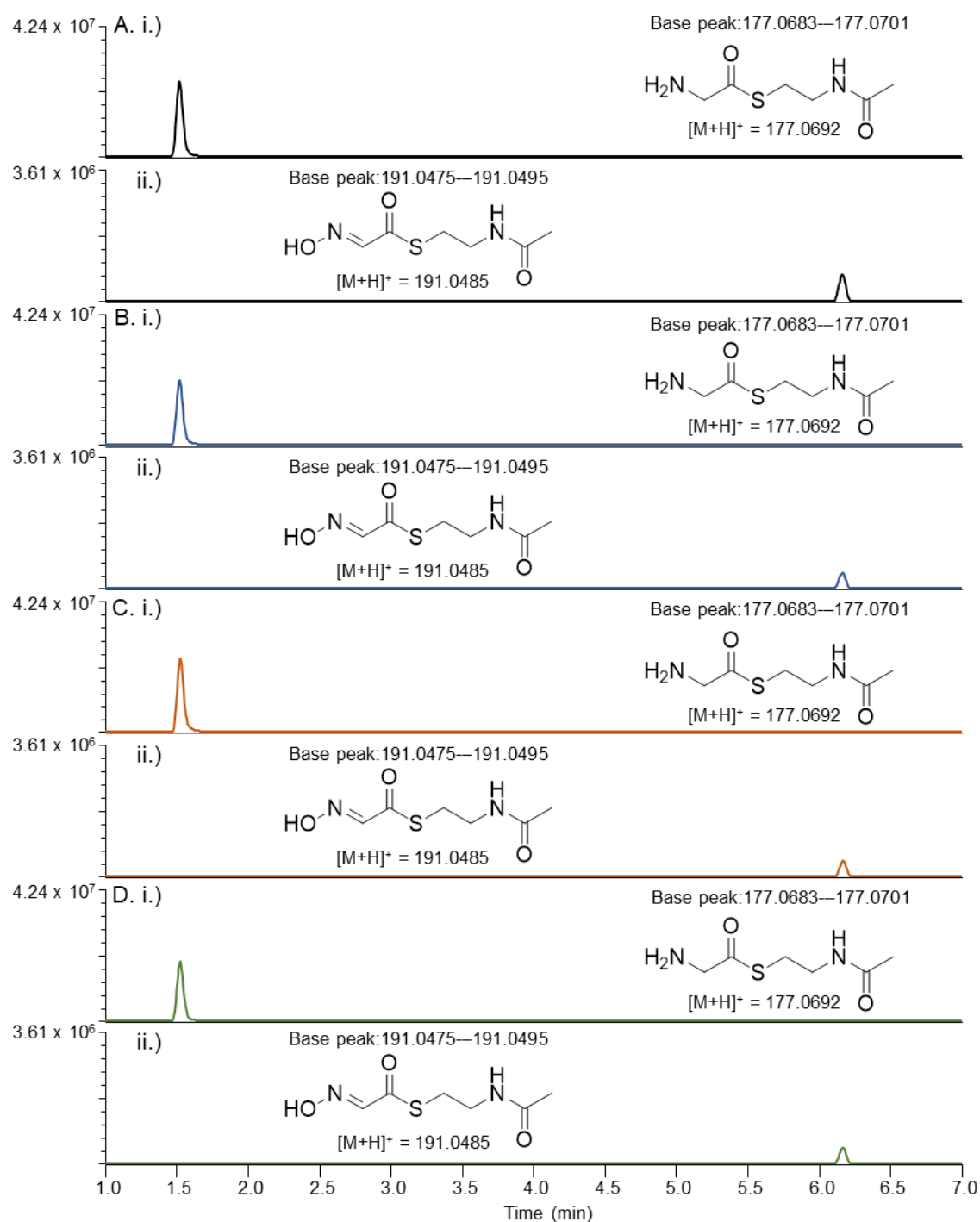
**Figure S52.** Base peak analyses for enzyme activity assays with LbMA-Ox Q1547A, with the intensities scaled to the maximum value of assay set 4. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



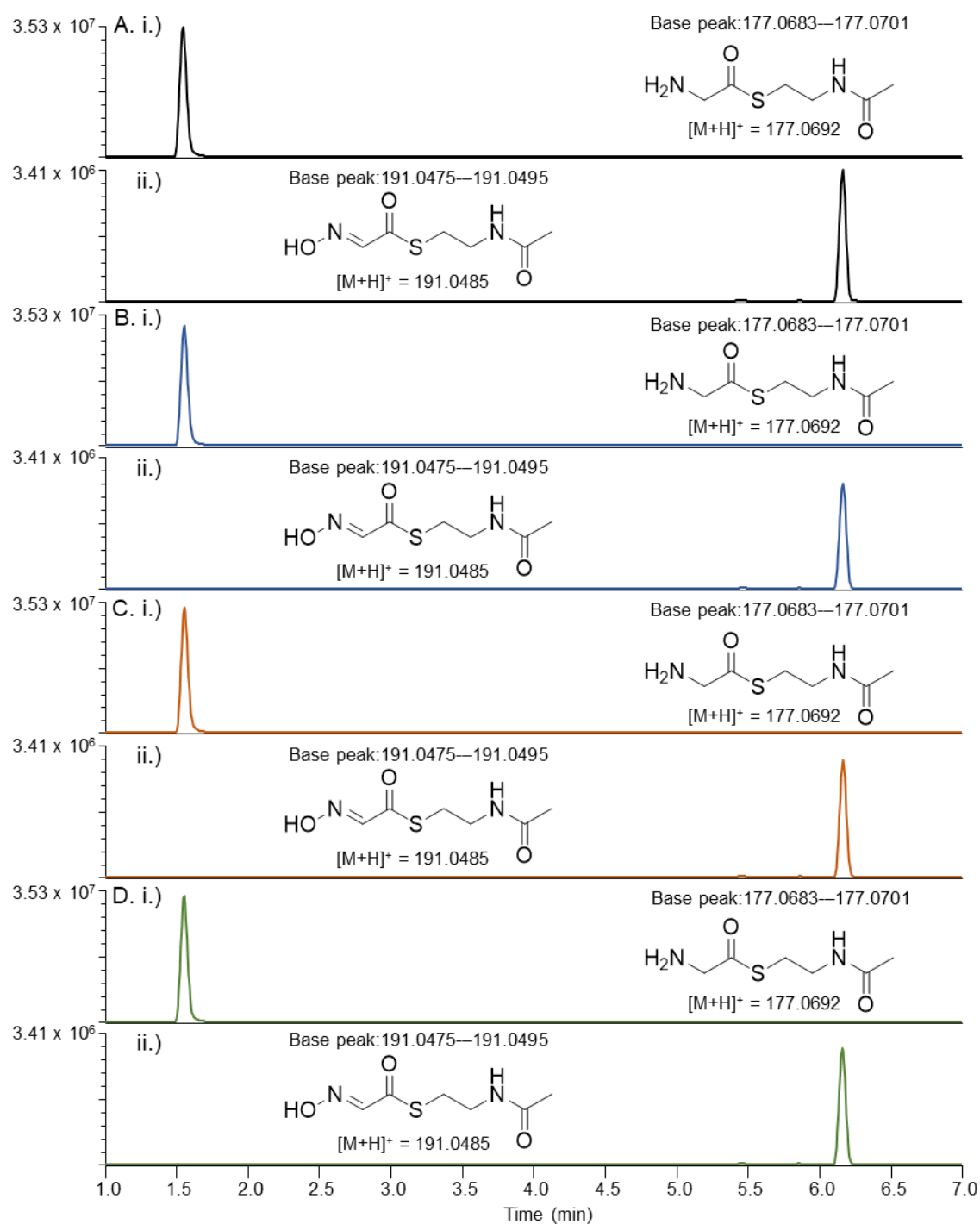
**Figure S53.** Base peak analyses for enzyme activity assays with LbmA-Ox N1548A, with the intensities scaled to the maximum value of assay set 4. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



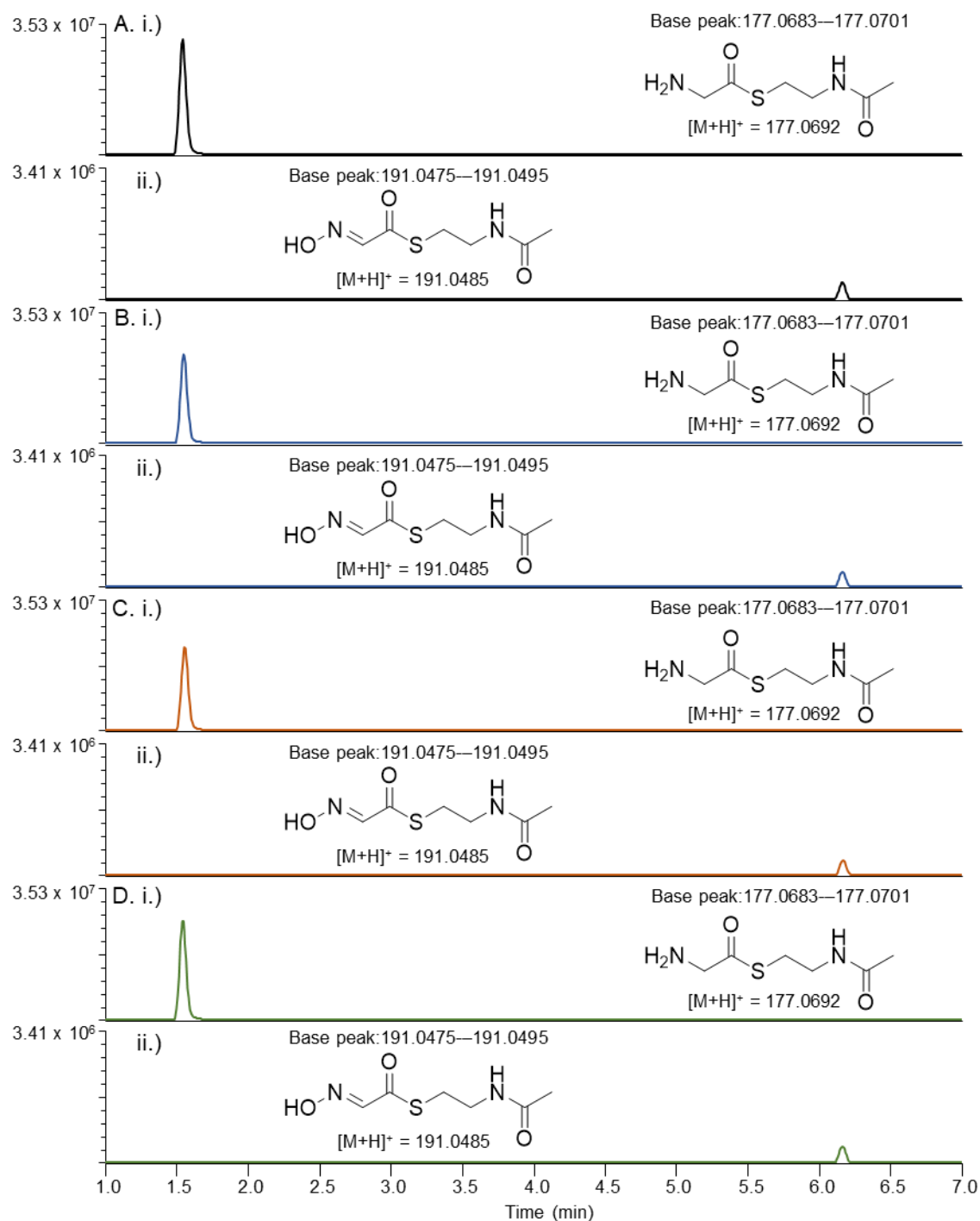
**Figure S54.** Base peak analyses for enzyme activity assays with LbmA-Ox R1702A, with the intensities scaled to the maximum value of assay set 4. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



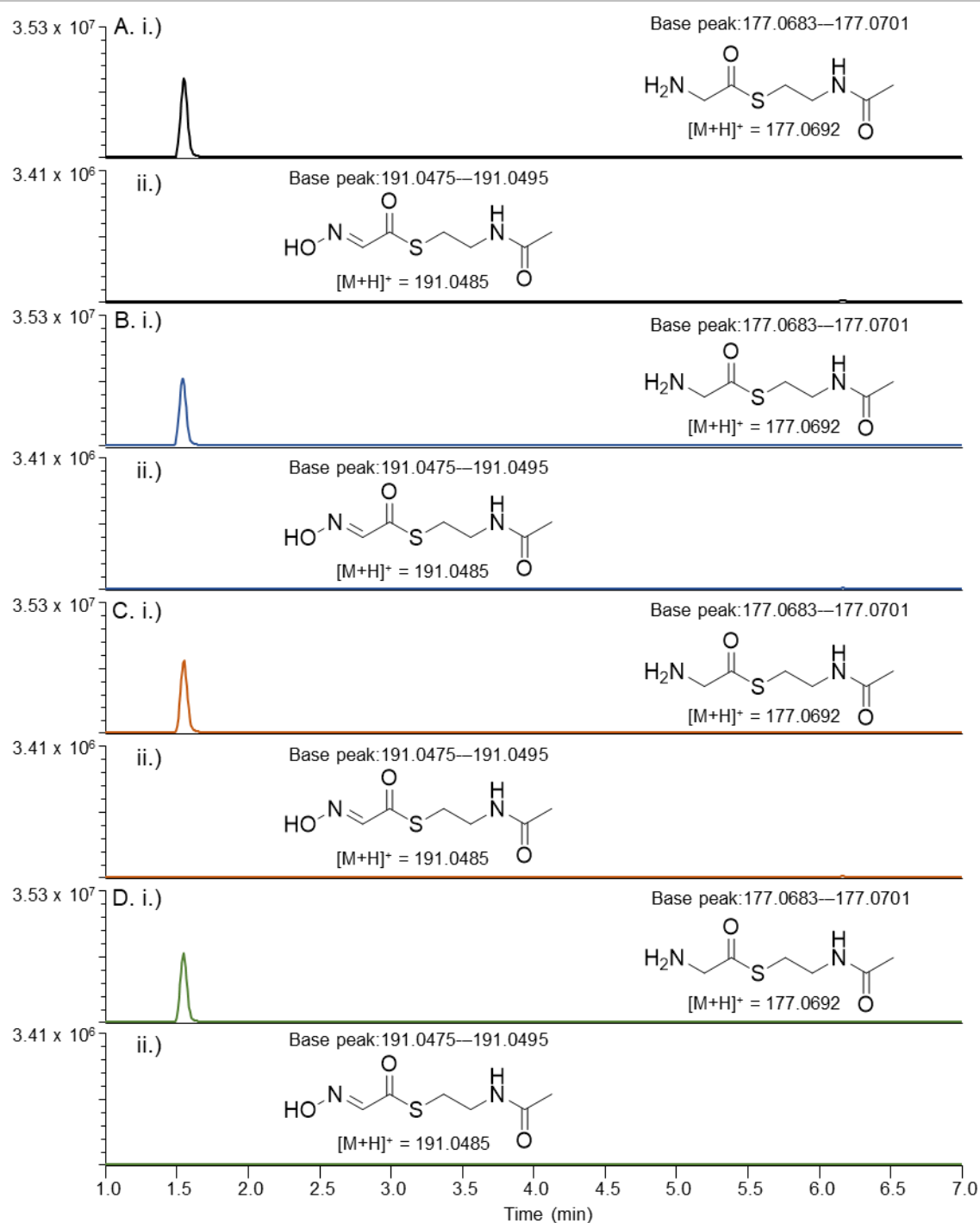
**Figure S55.** Base peak analyses for enzyme activity assays with Lbma-Ox wildtype (WT19), with the intensities scaled to the maximum value of assay set 5. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



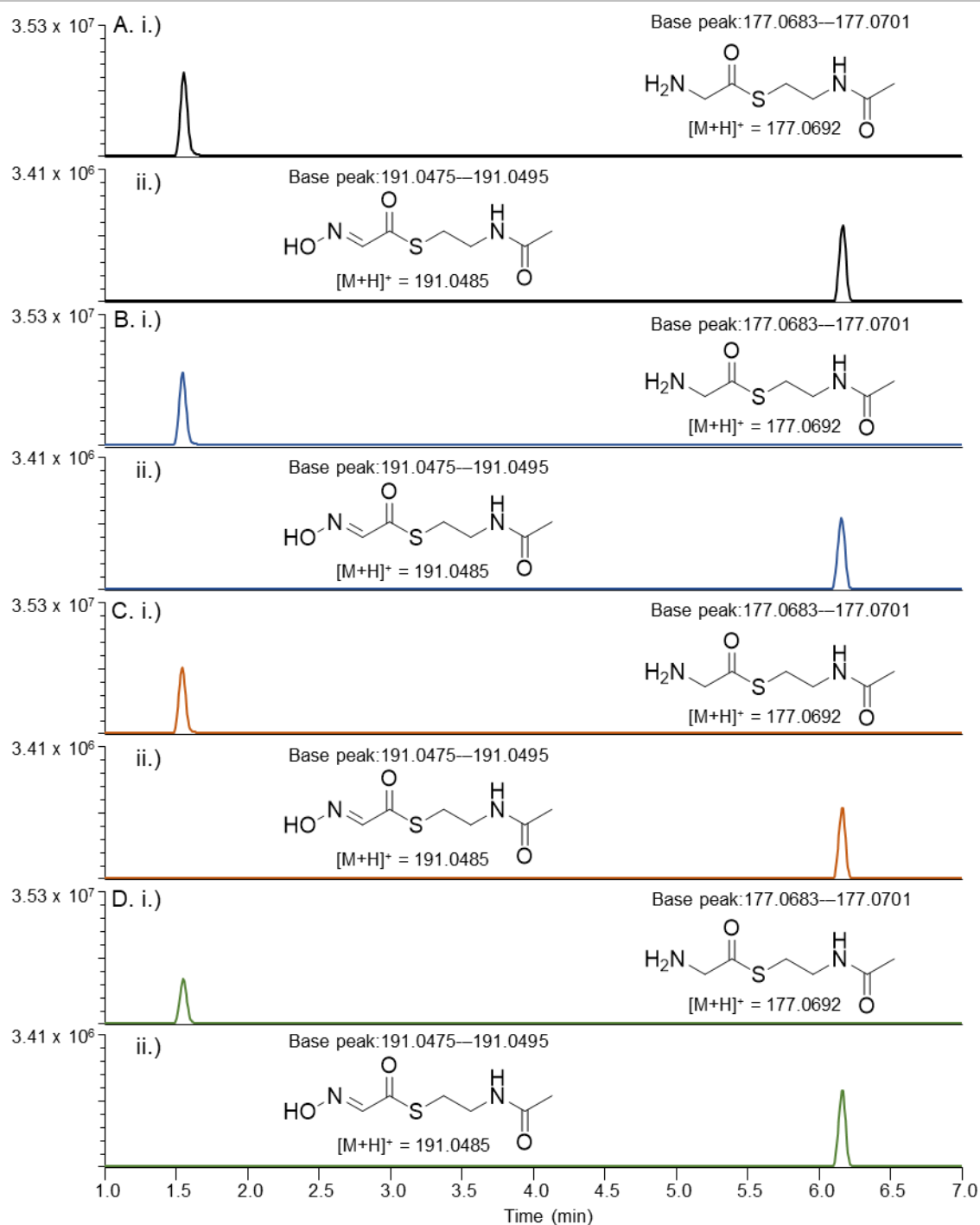
**Figure S56.** Base peak analyses for enzyme activity assays with LbmA-Ox Y1541F, with the intensities scaled to the maximum value of assay set 5. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



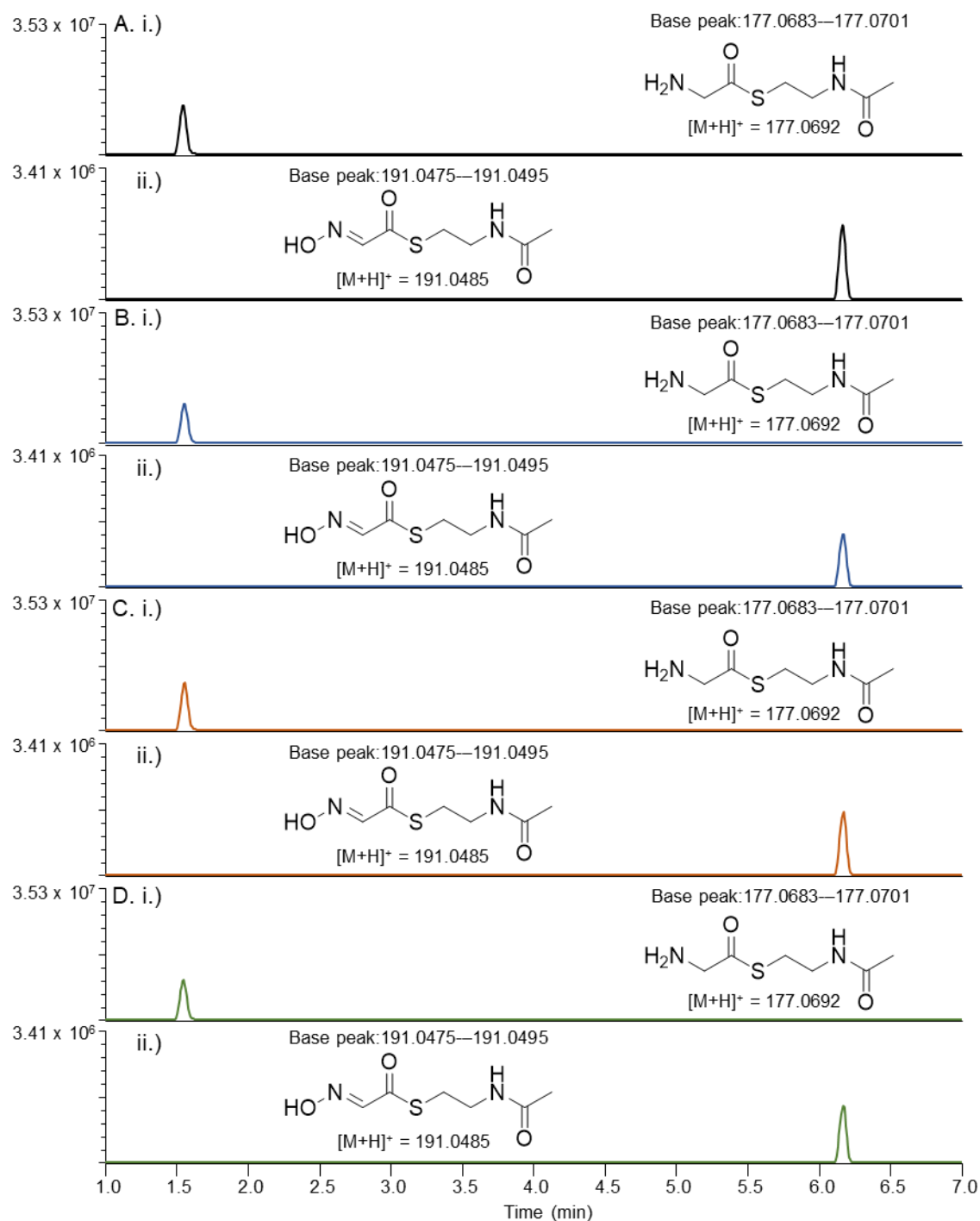
**Figure S57.** Base peak analyses for enzyme activity assays with LbMA-Ox D1677A, with the intensities scaled to the maximum value of assay set 5. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



**Figure S58.** Base peak analyses for enzyme activity assays with LbMA-Ox R1697A, with the intensities scaled to the maximum value of assay set 5. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

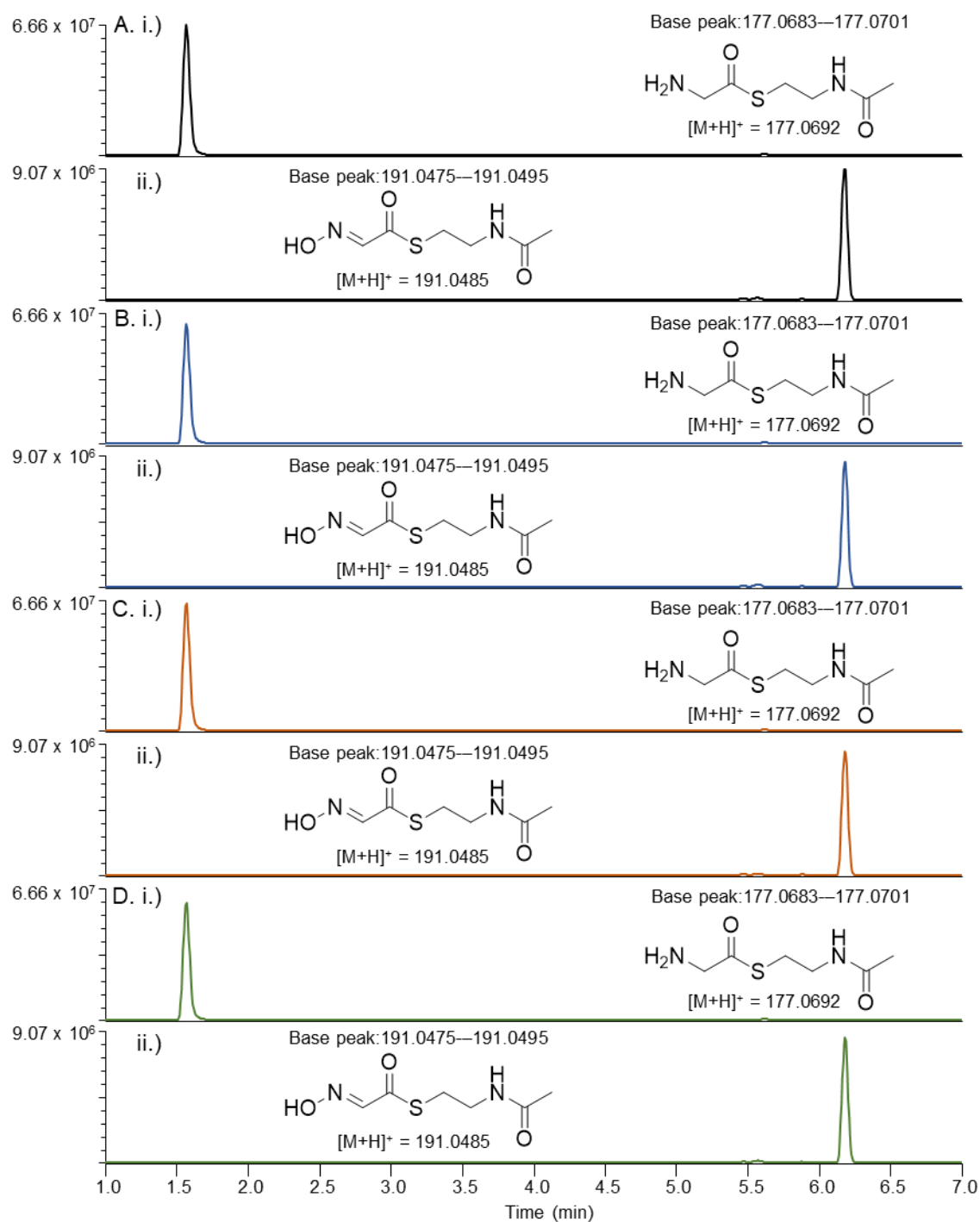
## SUPPORTING INFORMATION



**Figure S59.** Base peak analyses for enzyme activity assays with LbmA-Ox E1680A, with the intensities scaled to the maximum value of assay set 5. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

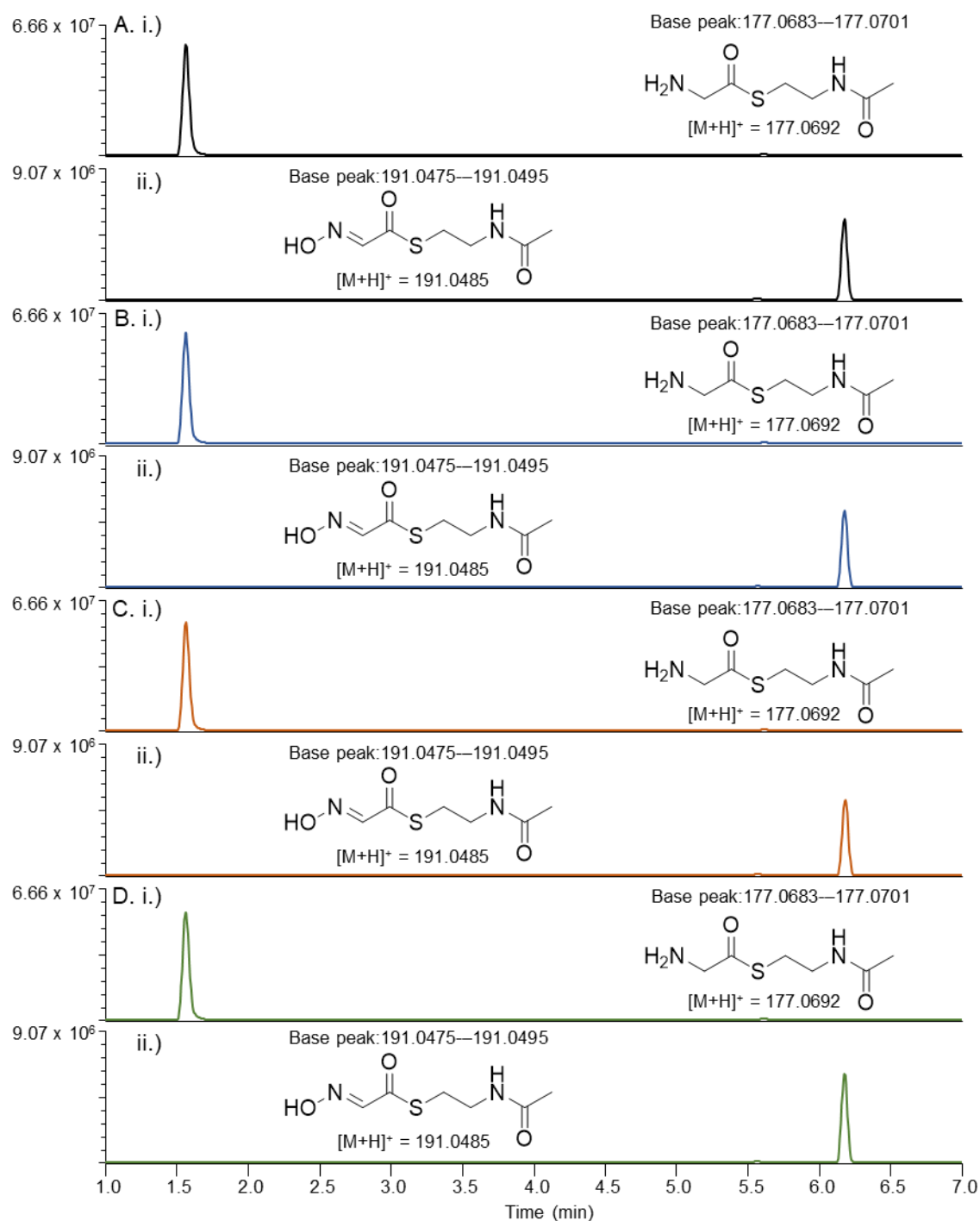


## SUPPORTING INFORMATION



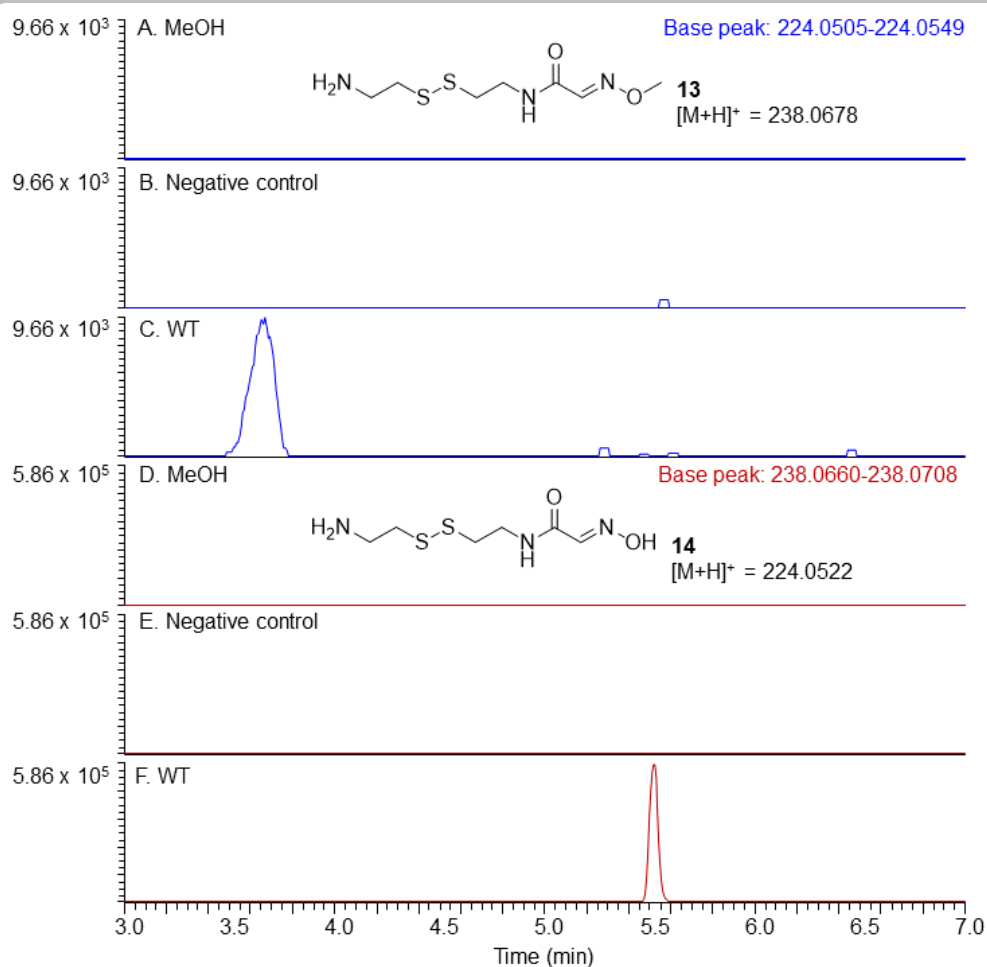
**Figure S60.** Base peak analyses for enzyme activity assays with Lbma-Ox wildtype (WT20), with the intensities scaled to the maximum value of assay set 6. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



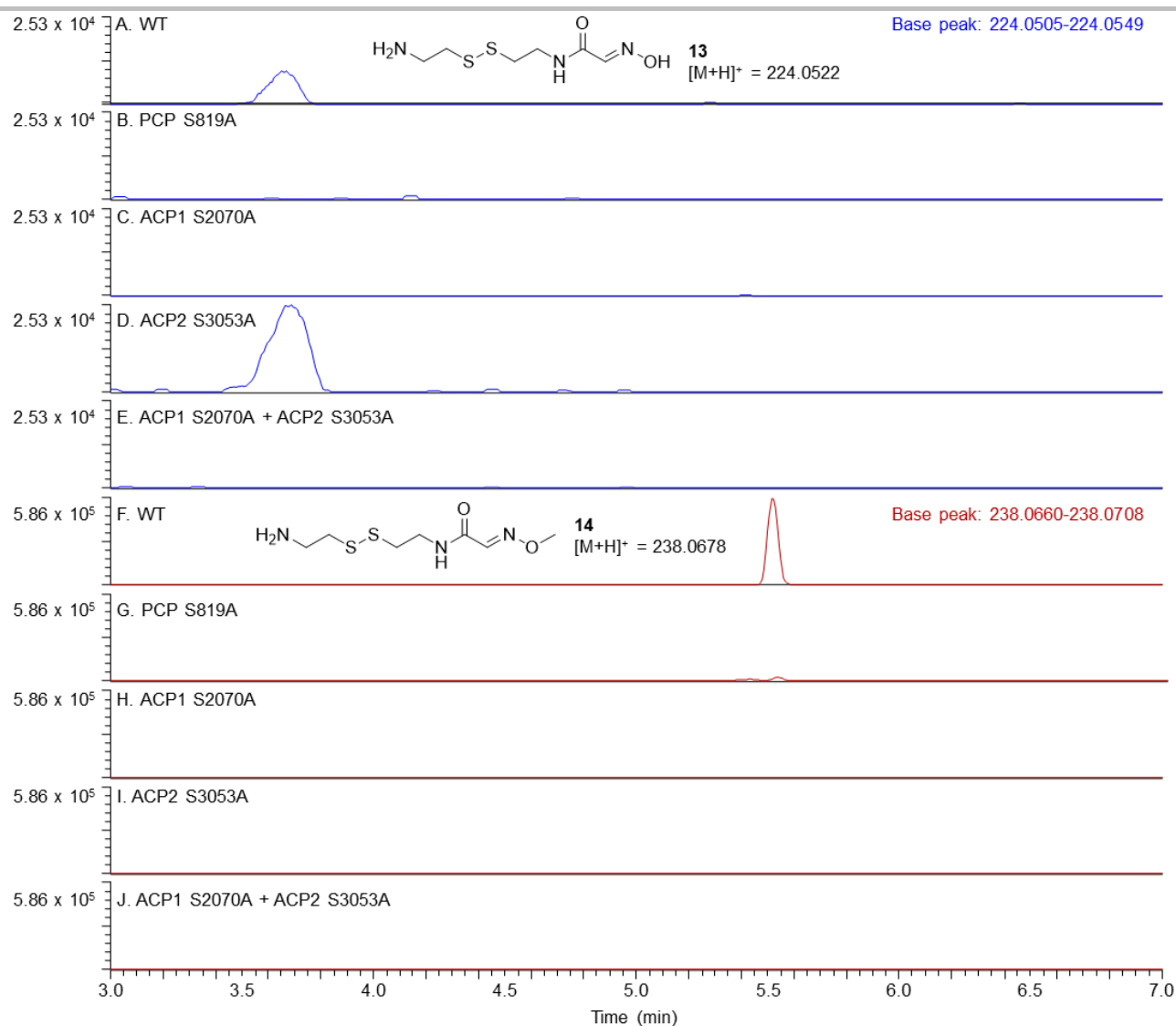
**Figure S61.** Base peak analyses for enzyme activity assays with LbmA-Ox T1544V, with the intensities scaled to the maximum value of assay set 6. A–D) Shown are extracted ion chromatograms (EICs) of UHPLC-MS data of quadruplicate analyses for i) GlySNAC substrate **6** and ii) GlySNAC oxime product **7**.

## SUPPORTING INFORMATION



**Figure S62.** Base peak analyses for *in vivo* assays with the wild type (WT) LbmA trimodule, with the intensities scaled to the maximum value of the WT assay. Shown are extracted ion chromatograms (EICs) of UHPLC-MS data for production of **13** (A-C) and **14** (D-F), resulting from offloading of the intermediates from the carrier proteins. The negative control consisted of WT *E. coli* BAP1.

## SUPPORTING INFORMATION



**Figure S63.** Base peak analyses for *in vivo* assays with the wild type (WT) and point mutants of the active site serine in the carrier protein domains of the LbmA trimodule. Intensities are scaled to the maximum value for each assay set. Shown are extracted ion chromatograms (EICs) of UHPLC-MS data for production of **13** (A-E) and **14** (F-J), resulting from offloading of the intermediates from the carrier proteins.

## SUPPORTING INFORMATION

## NMR spectra of synthesized compounds

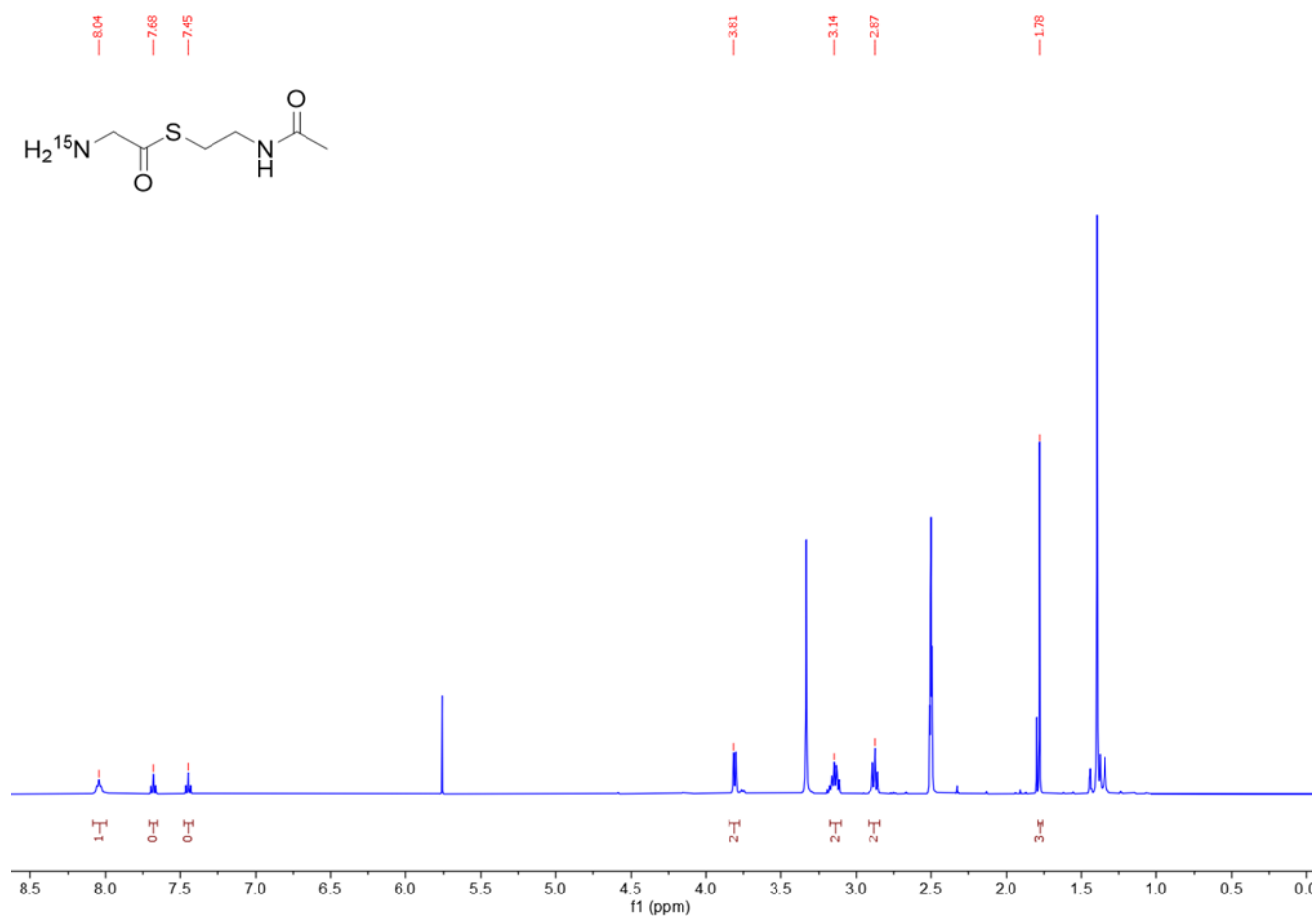
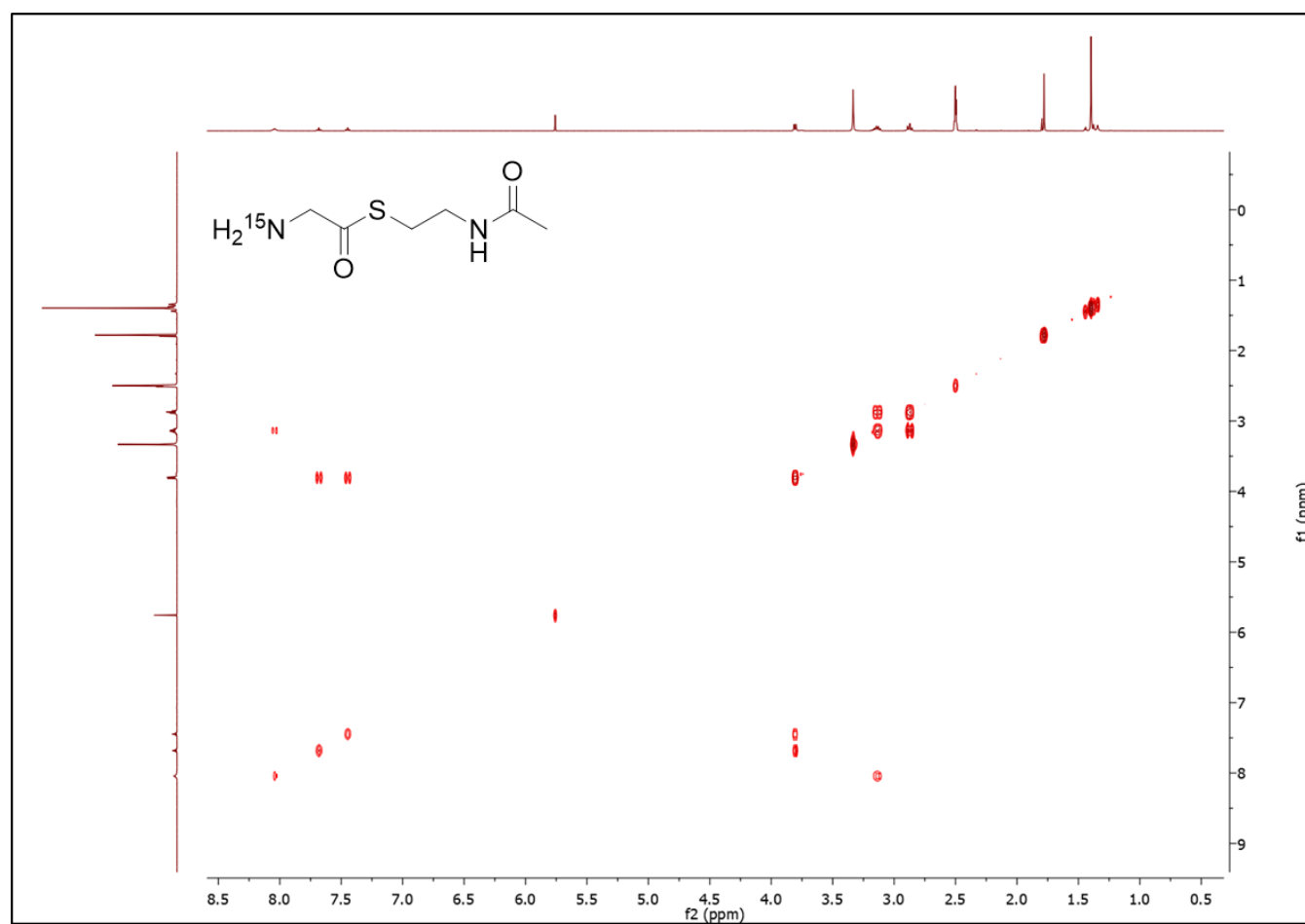


Figure S64.  $^1\text{H}$  NMR spectrum of  $^{15}\text{N}$ -Gly-SNAC (**S1**) in  $\text{DMSO-}d_6$  at 400 MHz.

## SUPPORTING INFORMATION



## SUPPORTING INFORMATION

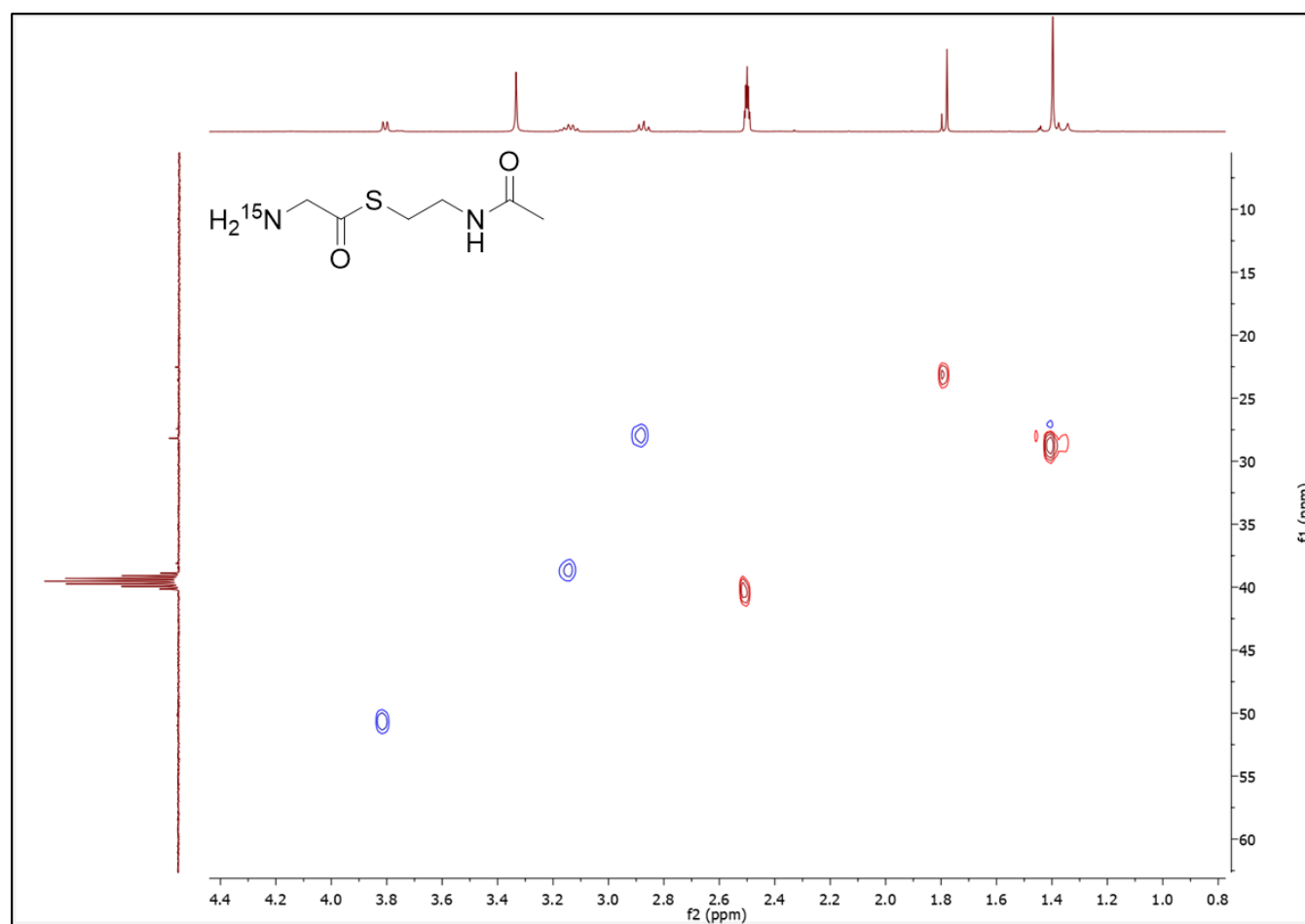


Figure S66. HSQC NMR spectrum of  $^{15}\text{N}$ -Gly-SNAC (**S1**) in  $\text{DMSO}-d_6$  at 400 MHz.

## SUPPORTING INFORMATION

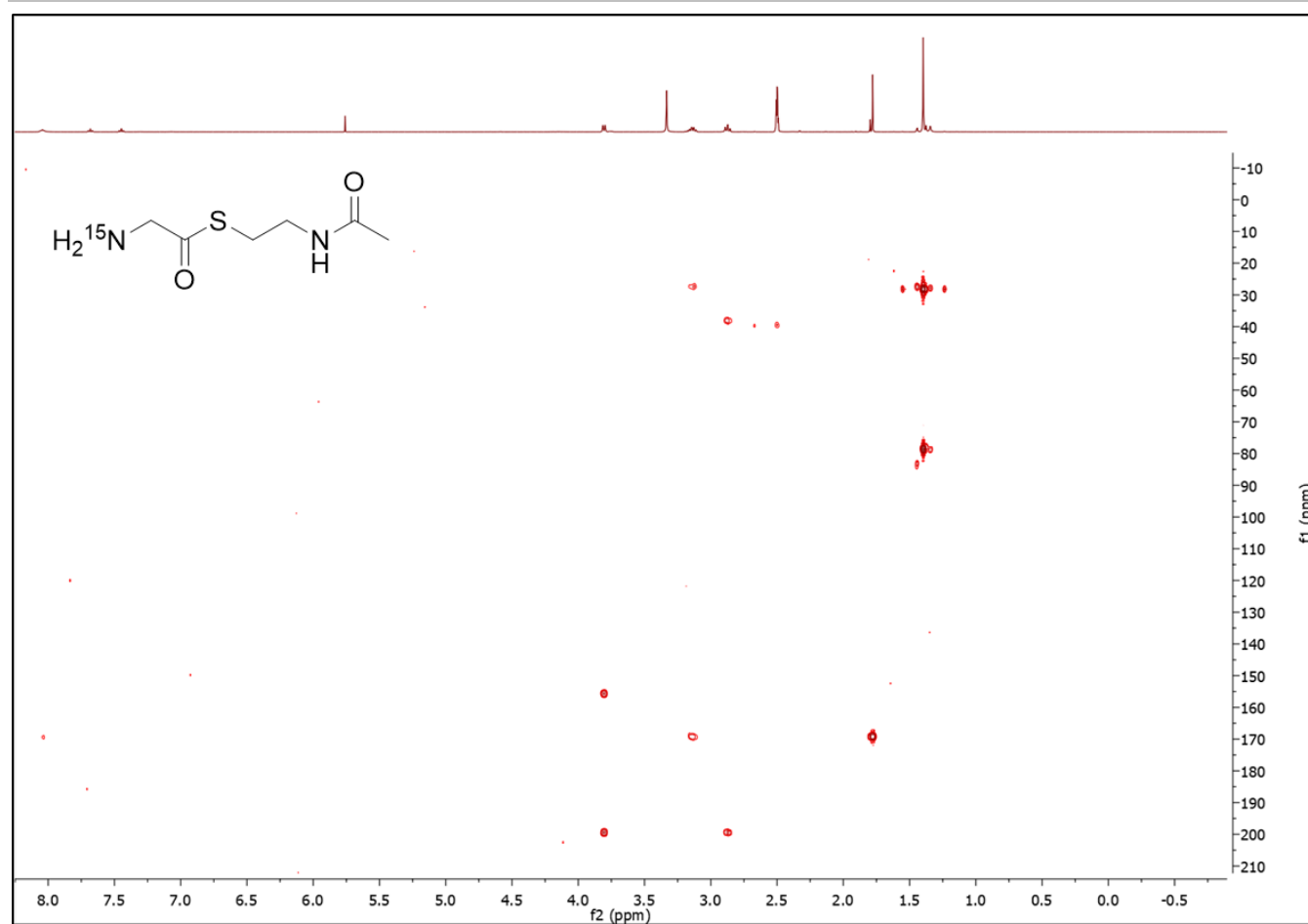


Figure S67. HMBC NMR spectrum of  $^{15}\text{N}$ -Gly-SNAC (**S1**) in  $\text{DMSO}-d_6$  at 400 MHz.



## SUPPORTING INFORMATION

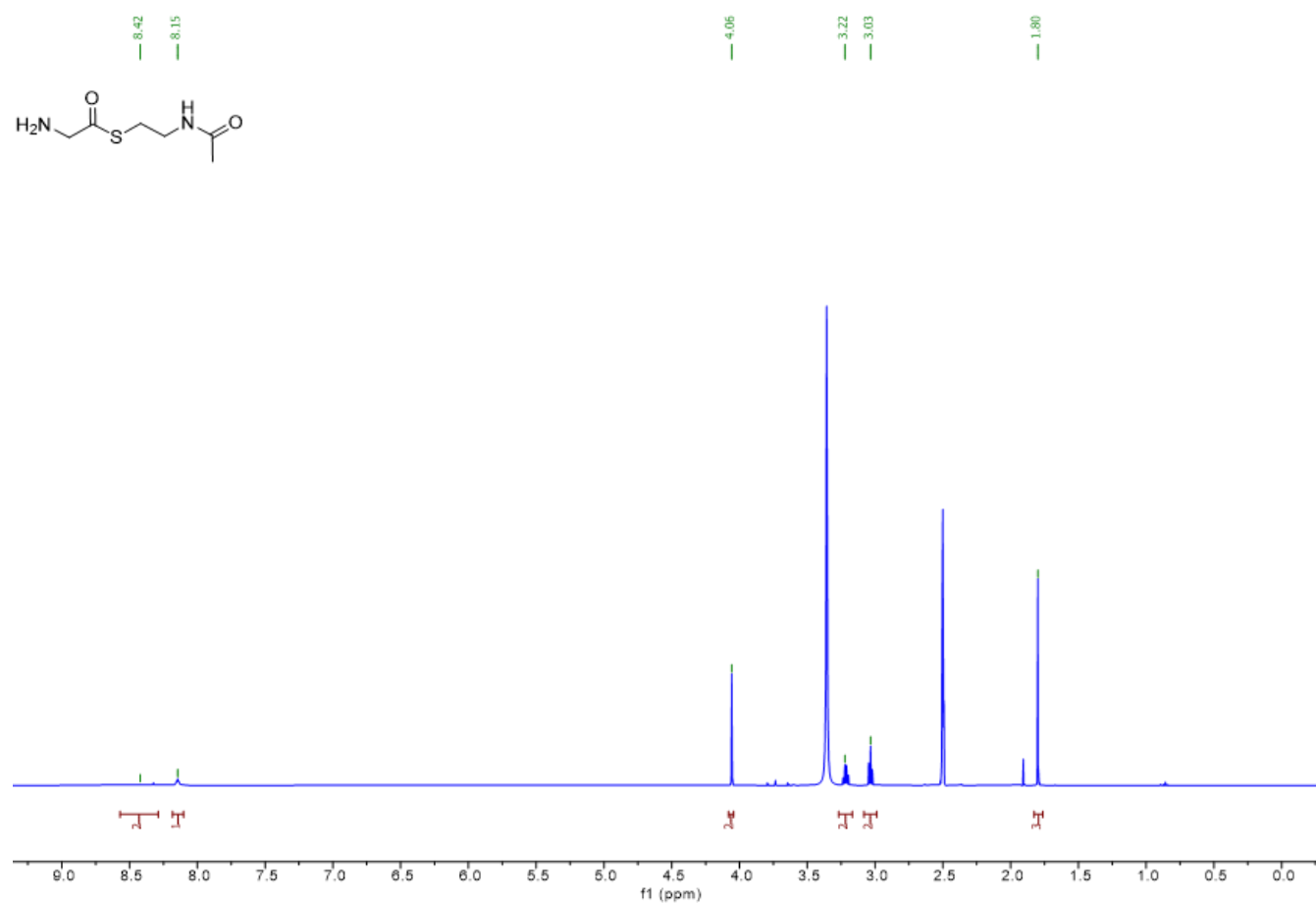
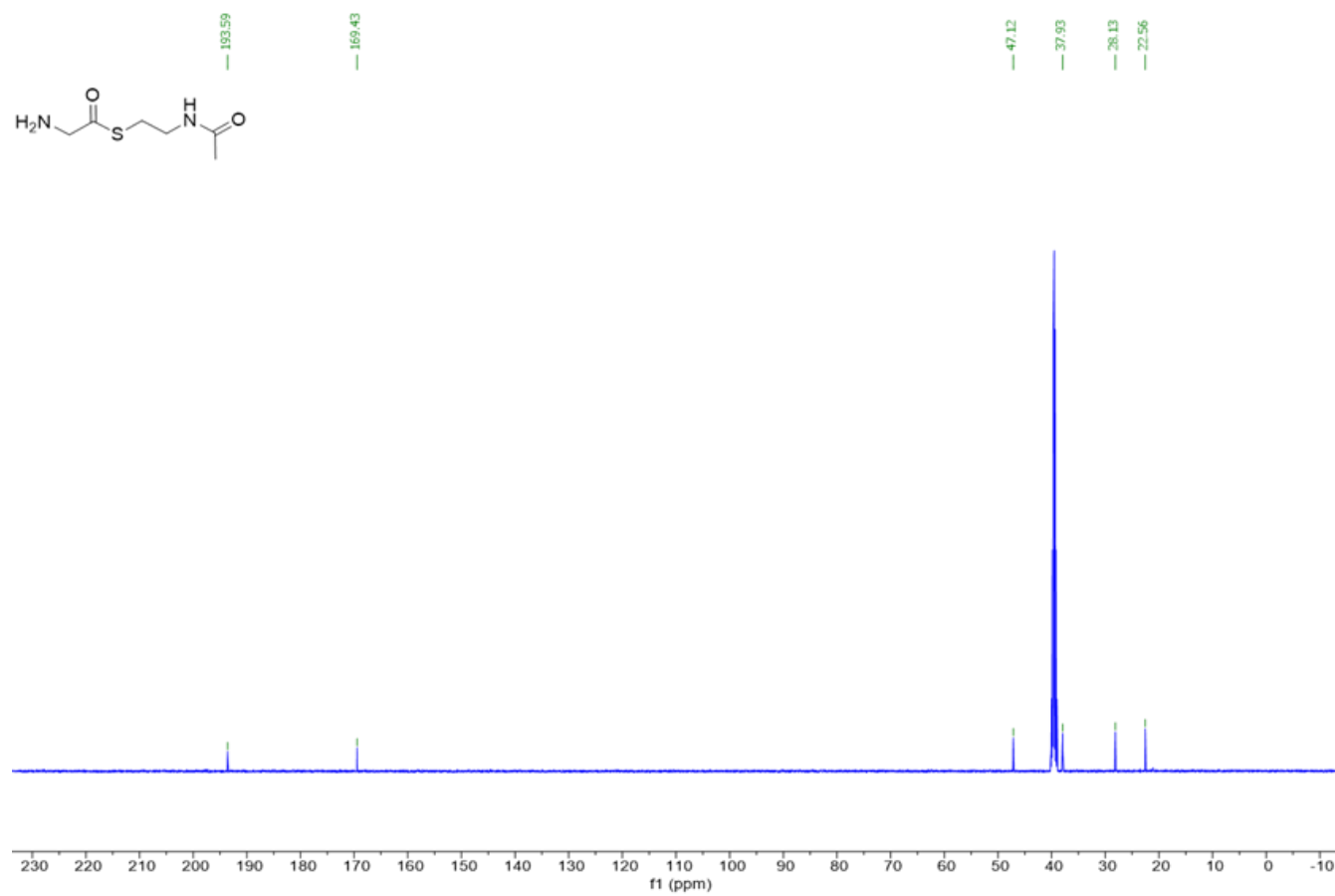


Figure S68. <sup>1</sup>H NMR spectrum of Gly-SNAC (5) in DMSO-d<sub>6</sub> at 500 MHz<sup>[9]</sup>.

## SUPPORTING INFORMATION



**Figure S69.** <sup>13</sup>C NMR spectrum of Gly-SNAC (5) in DMSO-*d*<sub>6</sub> at 500 MHz.<sup>[8]</sup>

## SUPPORTING INFORMATION

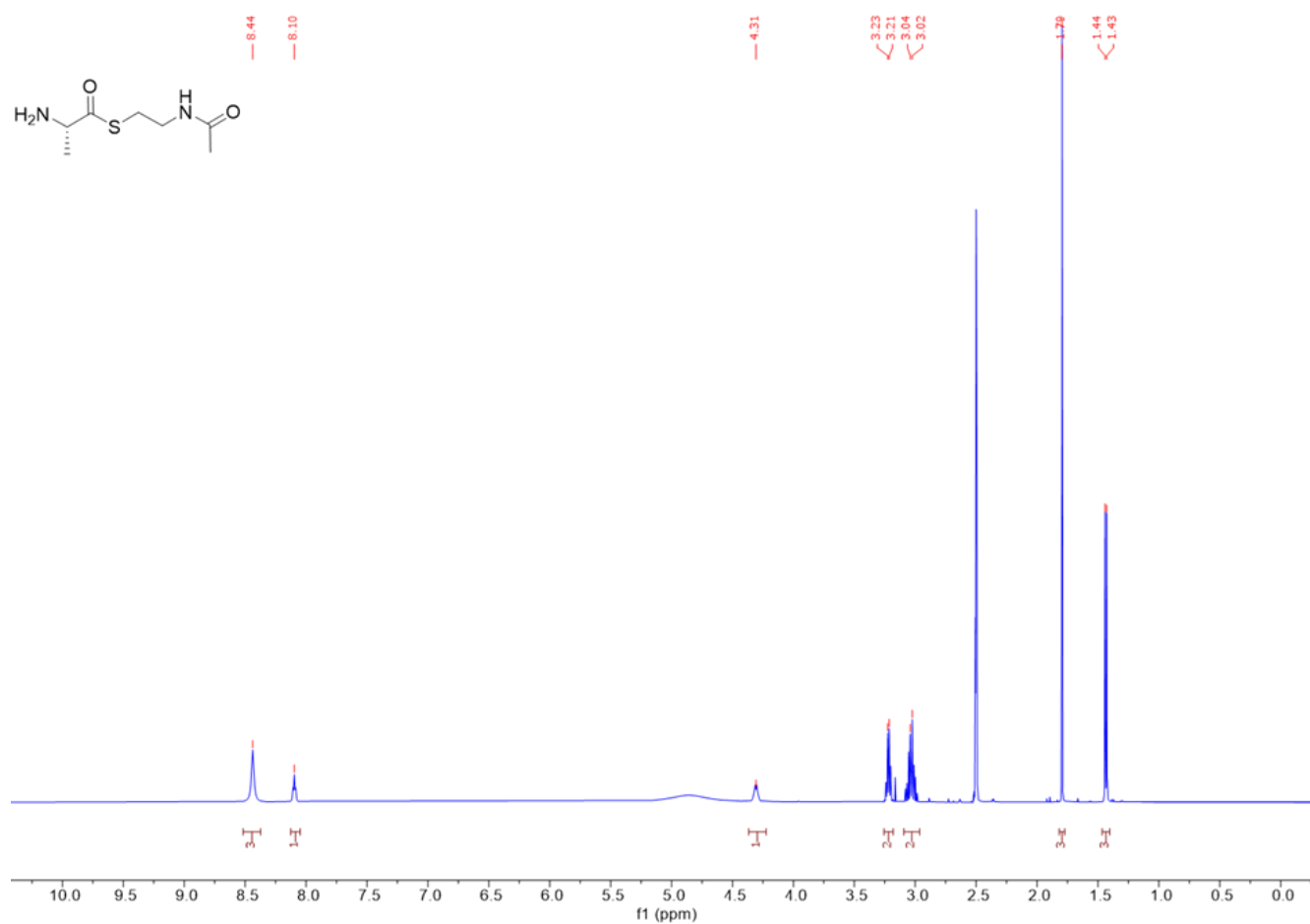
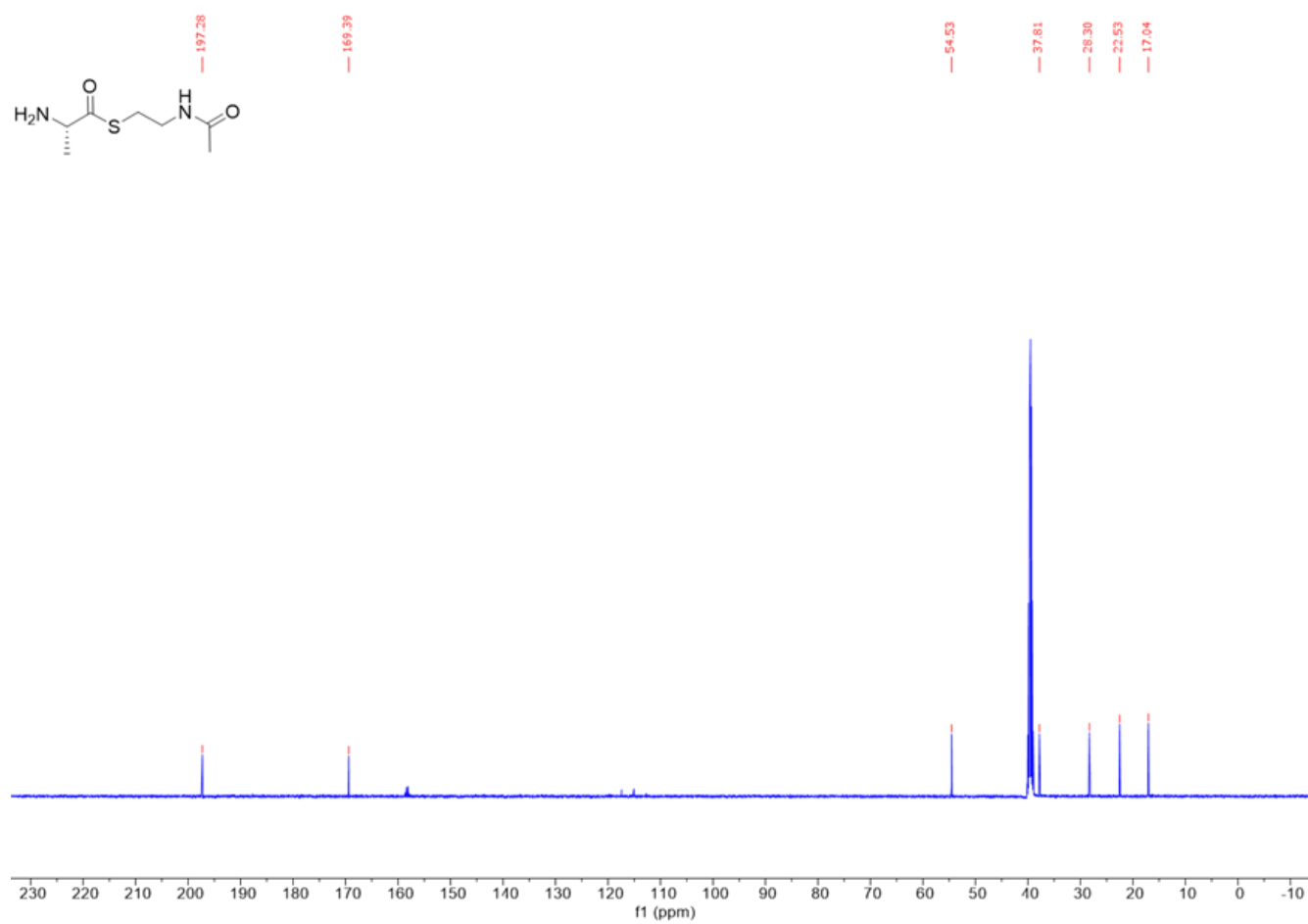


Figure S70. <sup>1</sup>H NMR spectrum of Ala-SNAC (7) in DMSO-*d*<sub>6</sub> at 500 MHz.<sup>[8]</sup>

## SUPPORTING INFORMATION



**Figure S71.** <sup>13</sup>C NMR spectrum of Ala-SNAC (7) in DMSO-*d*<sub>6</sub> at 500 MHz.<sup>[8]</sup>

## SUPPORTING INFORMATION

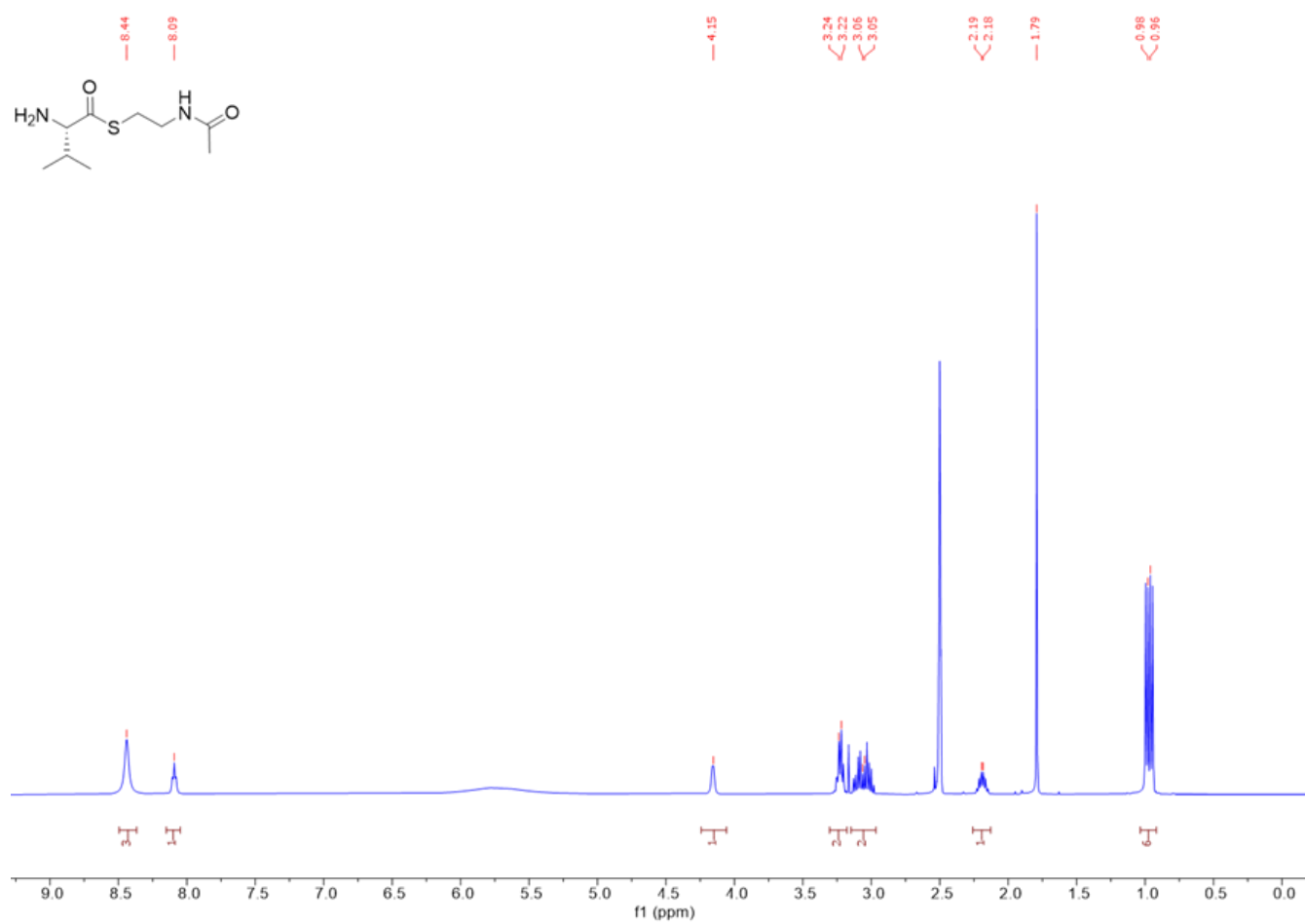


Figure S72.  $^1\text{H}$  NMR spectrum of Val-SNAC (9) in  $\text{DMSO-}d_6$  at 500 MHz.<sup>[8]</sup>

## SUPPORTING INFORMATION

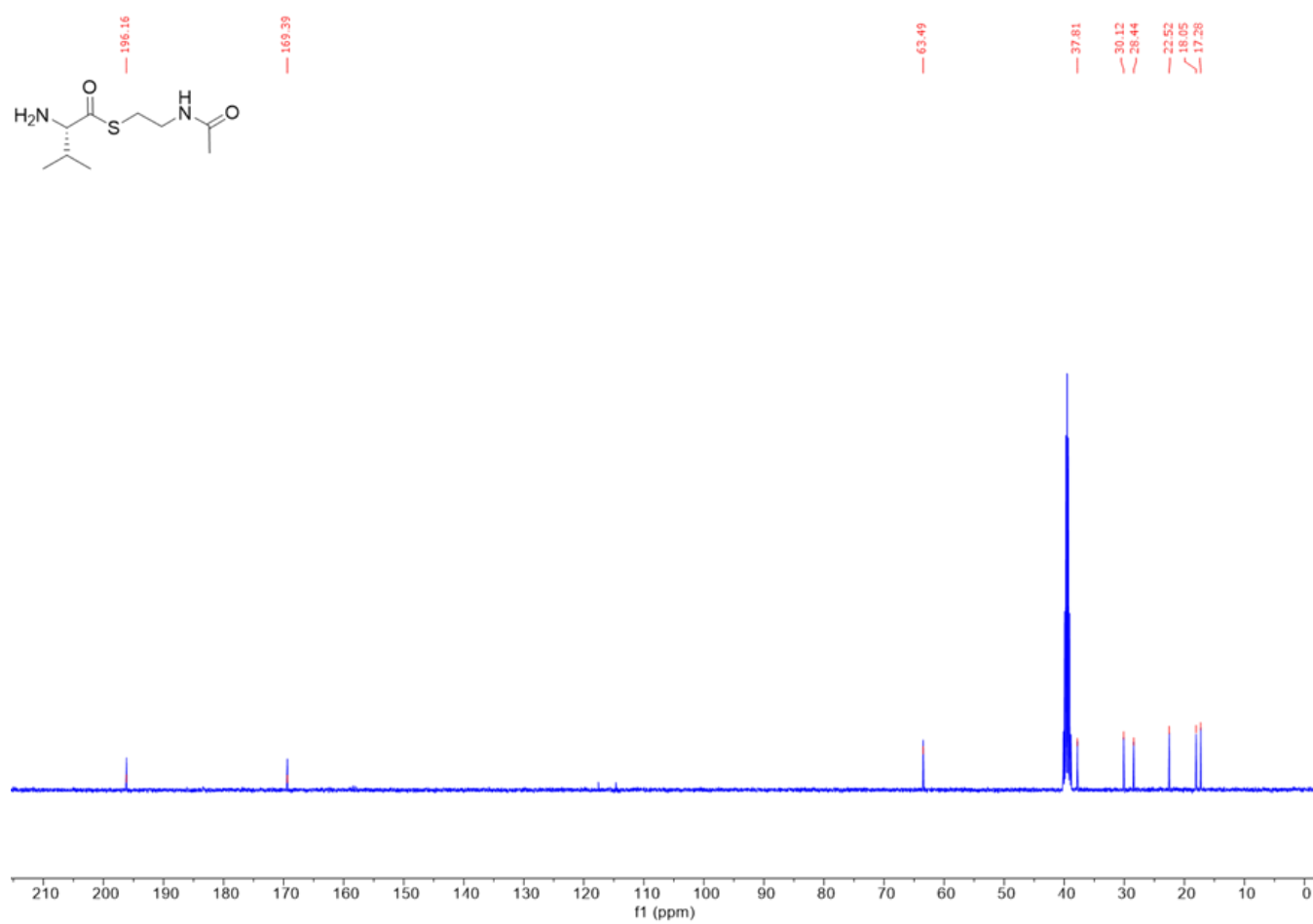


Figure S73. <sup>13</sup>C NMR spectrum of Val-SNAC (9) in DMSO-*d*<sub>6</sub> at 500 MHz.<sup>[6]</sup>

## SUPPORTING INFORMATION

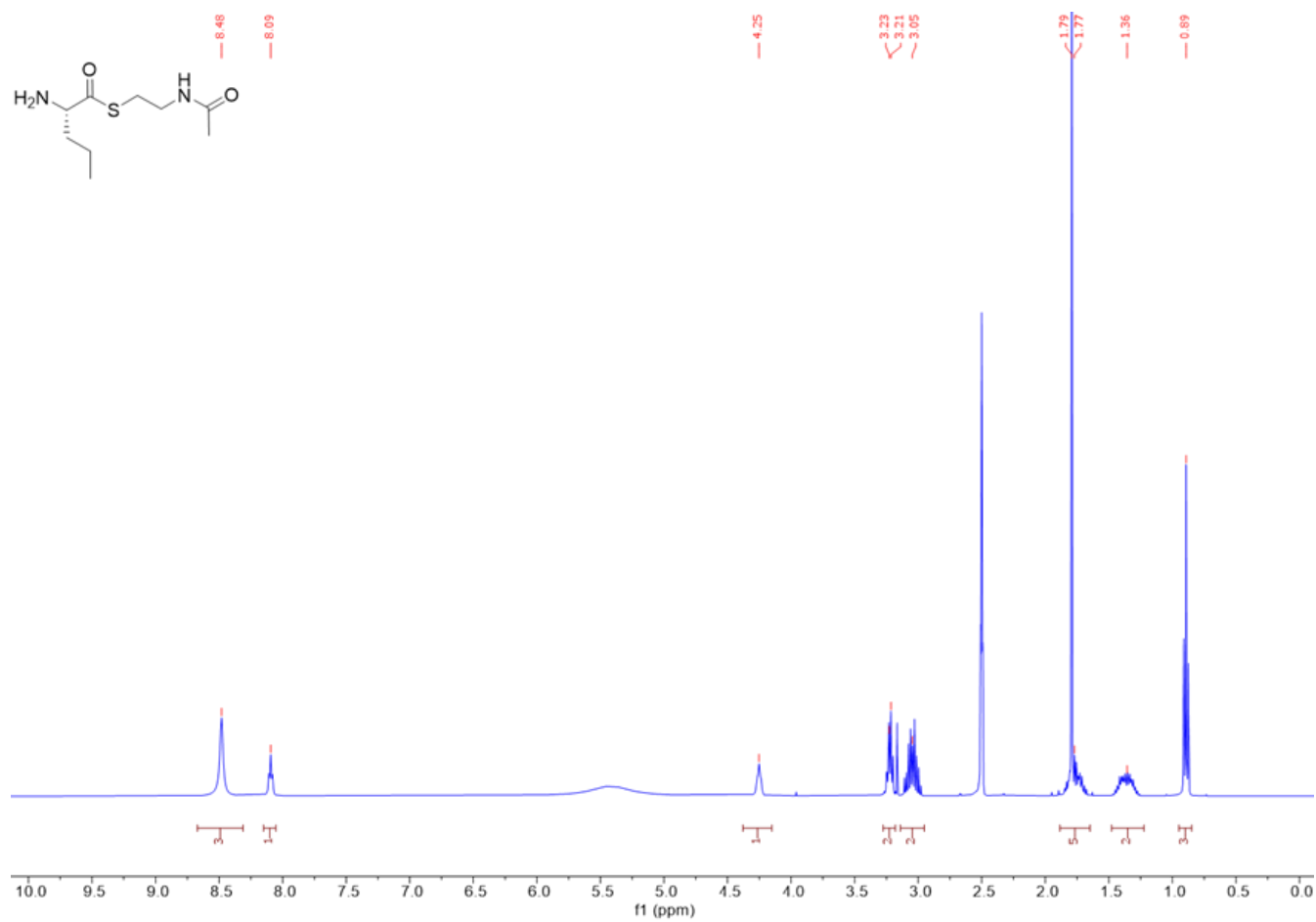


Figure S74. <sup>1</sup>H NMR spectrum of Nva-SNAC (10) in DMSO-*d*<sub>6</sub> at 400 MHz.<sup>[8]</sup>

## SUPPORTING INFORMATION

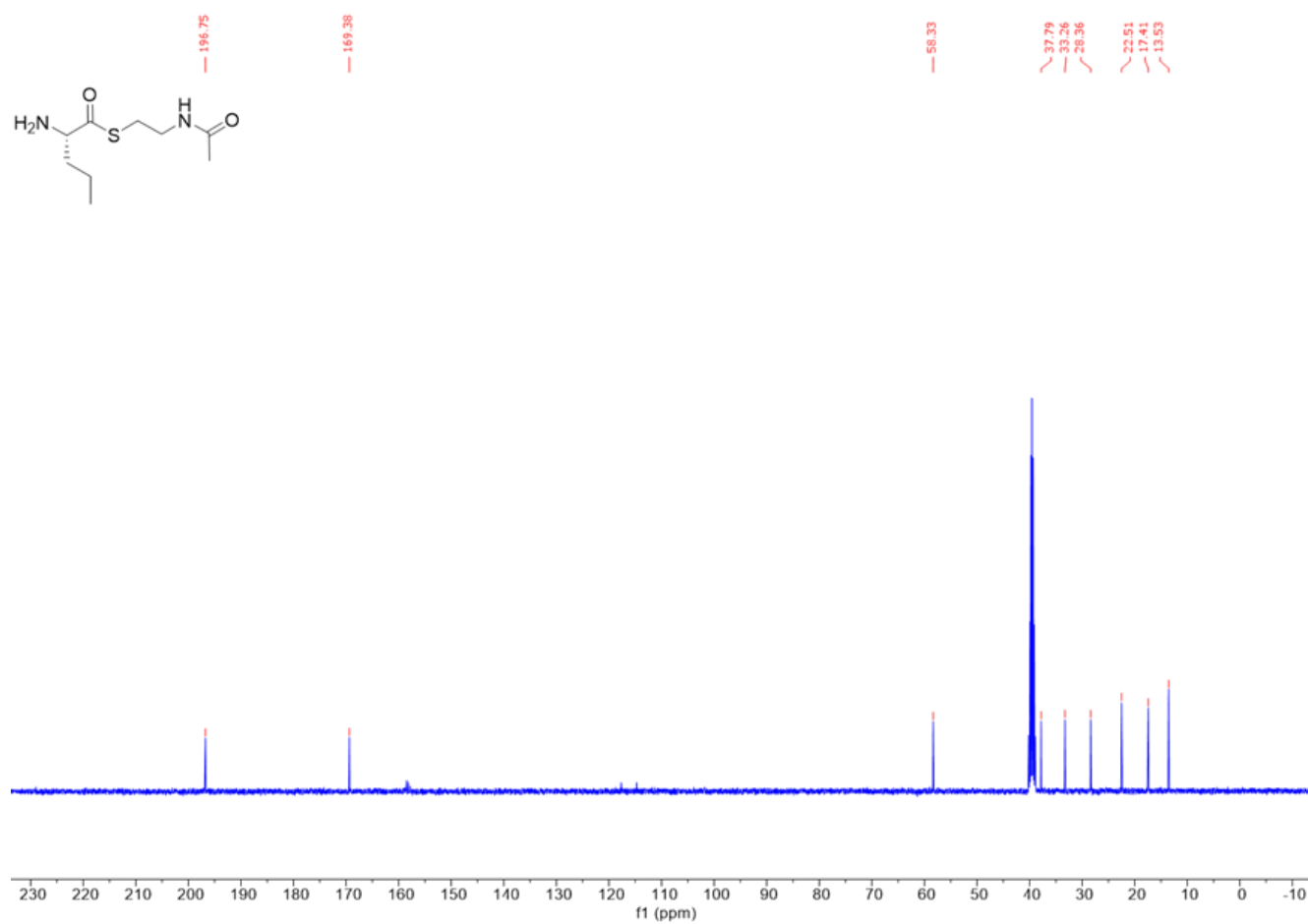


Figure S75.  $^{13}\text{C}$  NMR spectrum of Nva-SNAC (10) in DMSO- $d_6$  at 400 MHz.<sup>[8]</sup>



## SUPPORTING INFORMATION

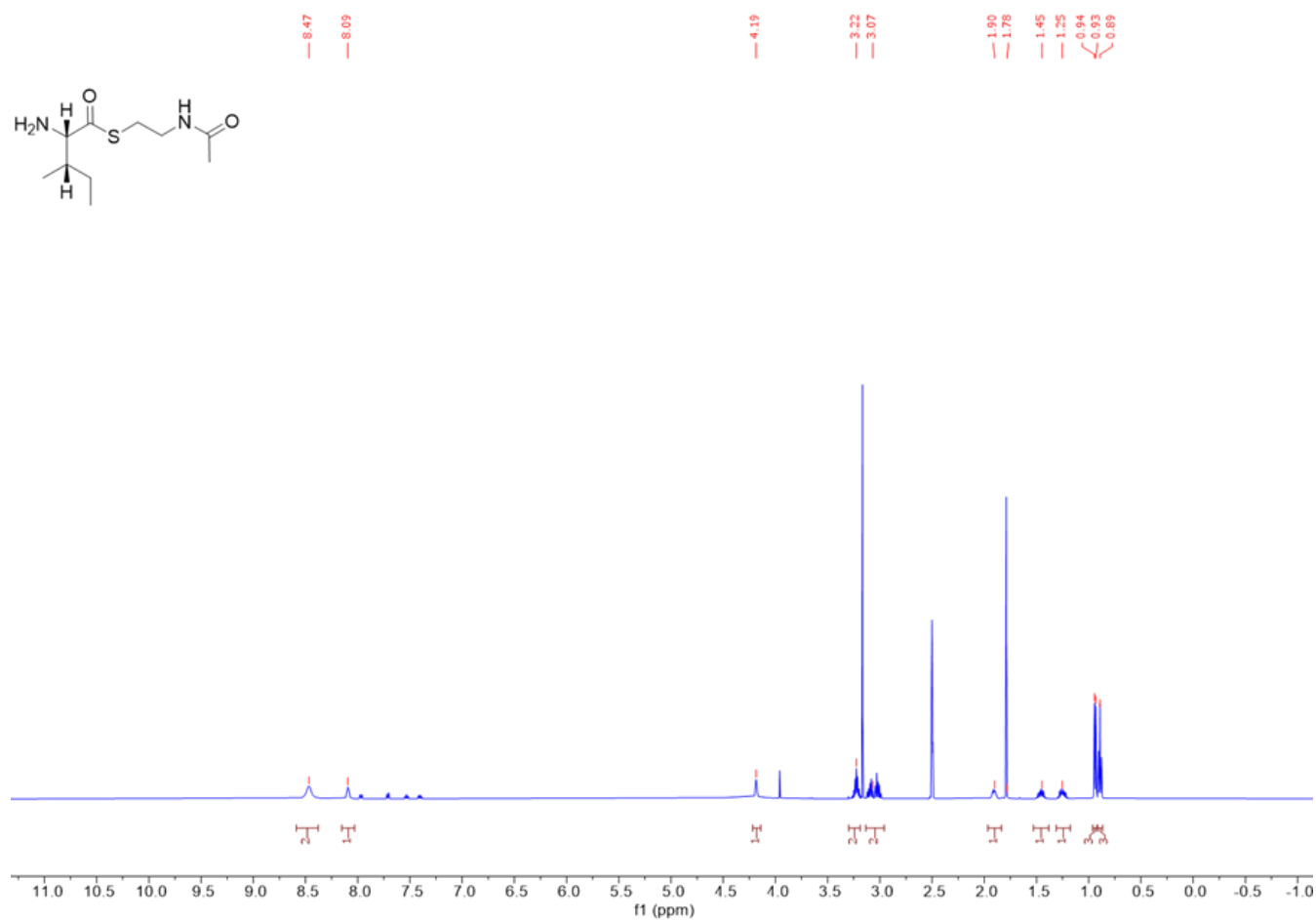
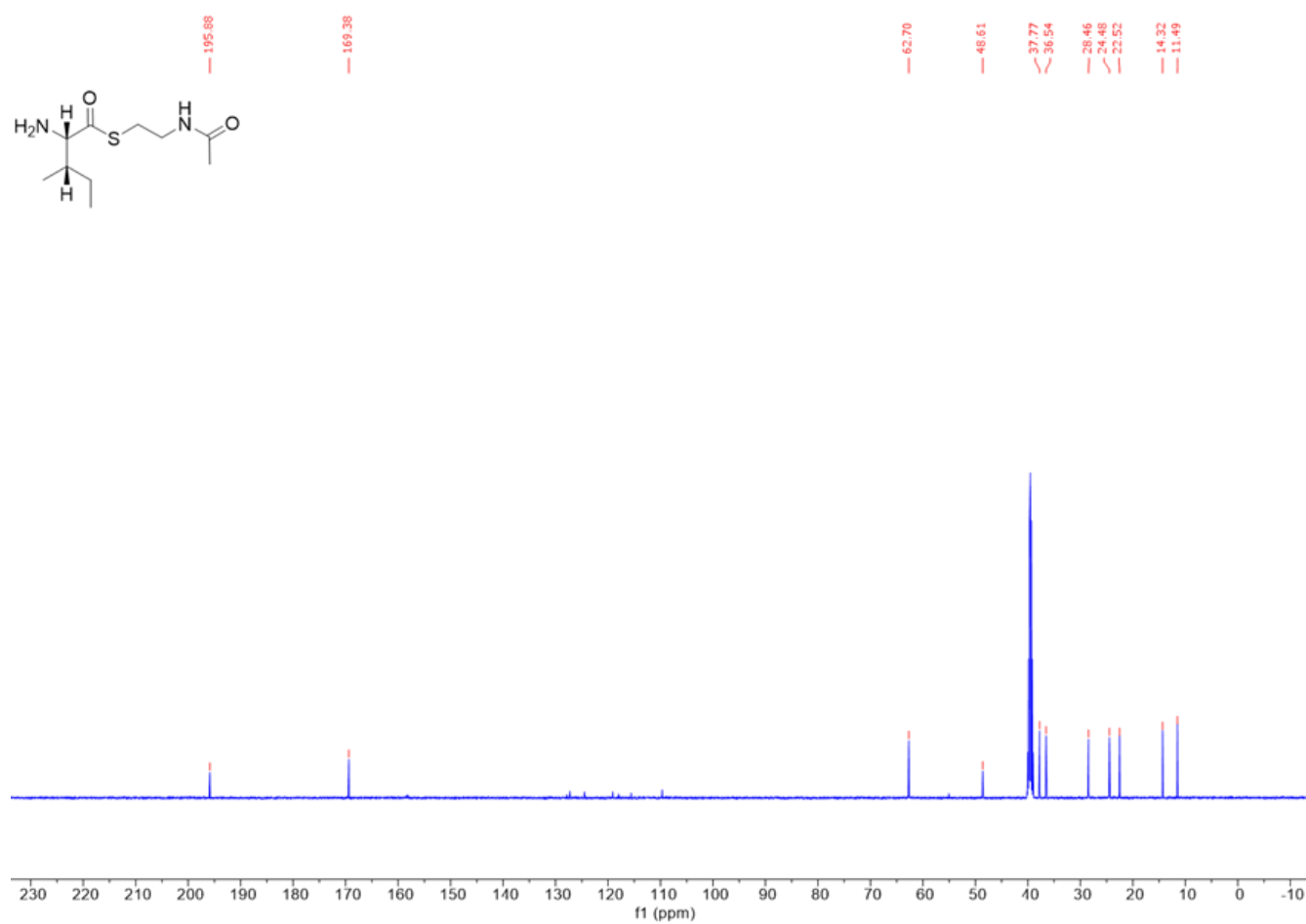


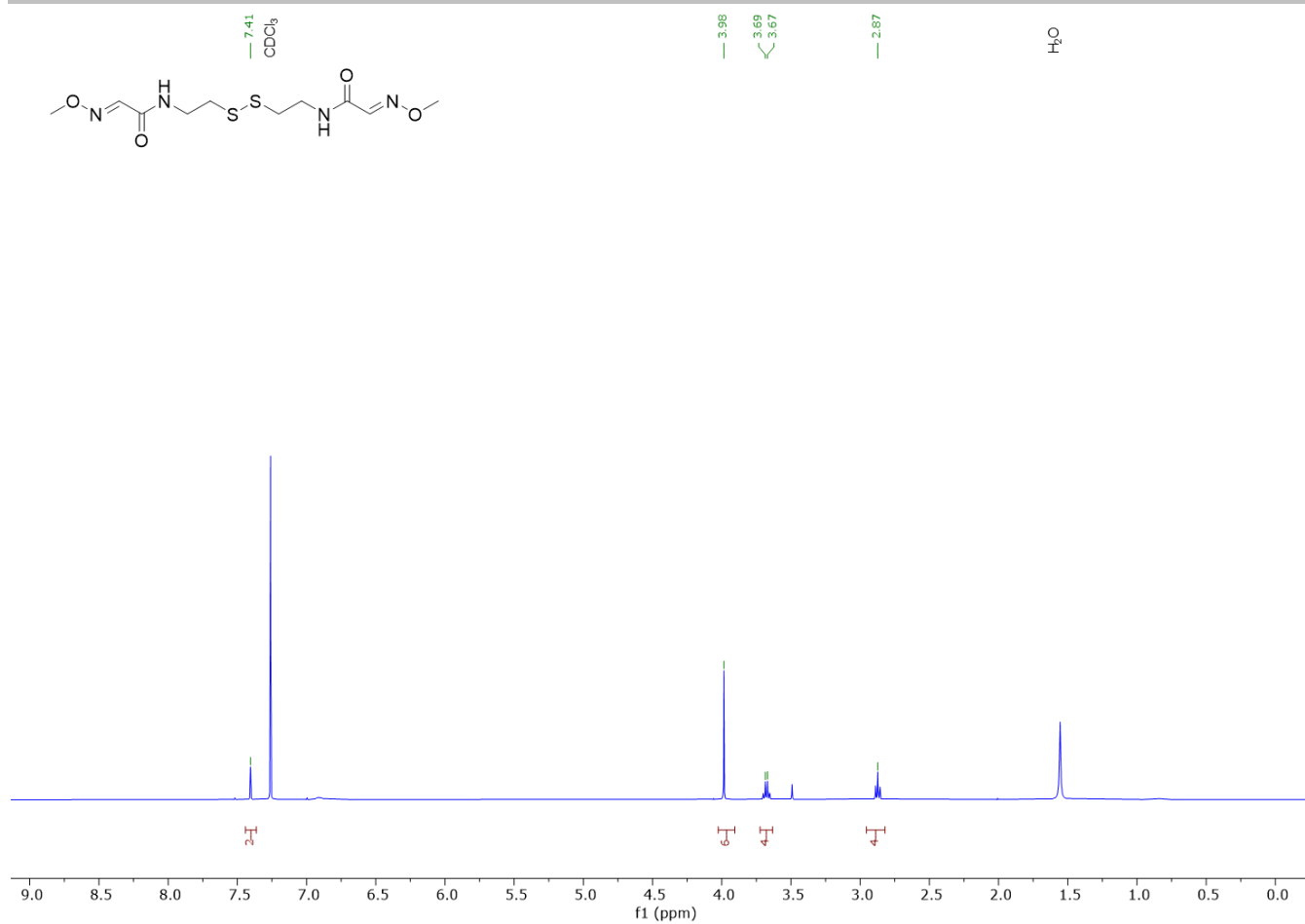
Figure S76. <sup>1</sup>H NMR spectrum of Ile-SNAC (11) in DMSO-d<sub>6</sub> at 400 MHz.<sup>[8]</sup>

## SUPPORTING INFORMATION



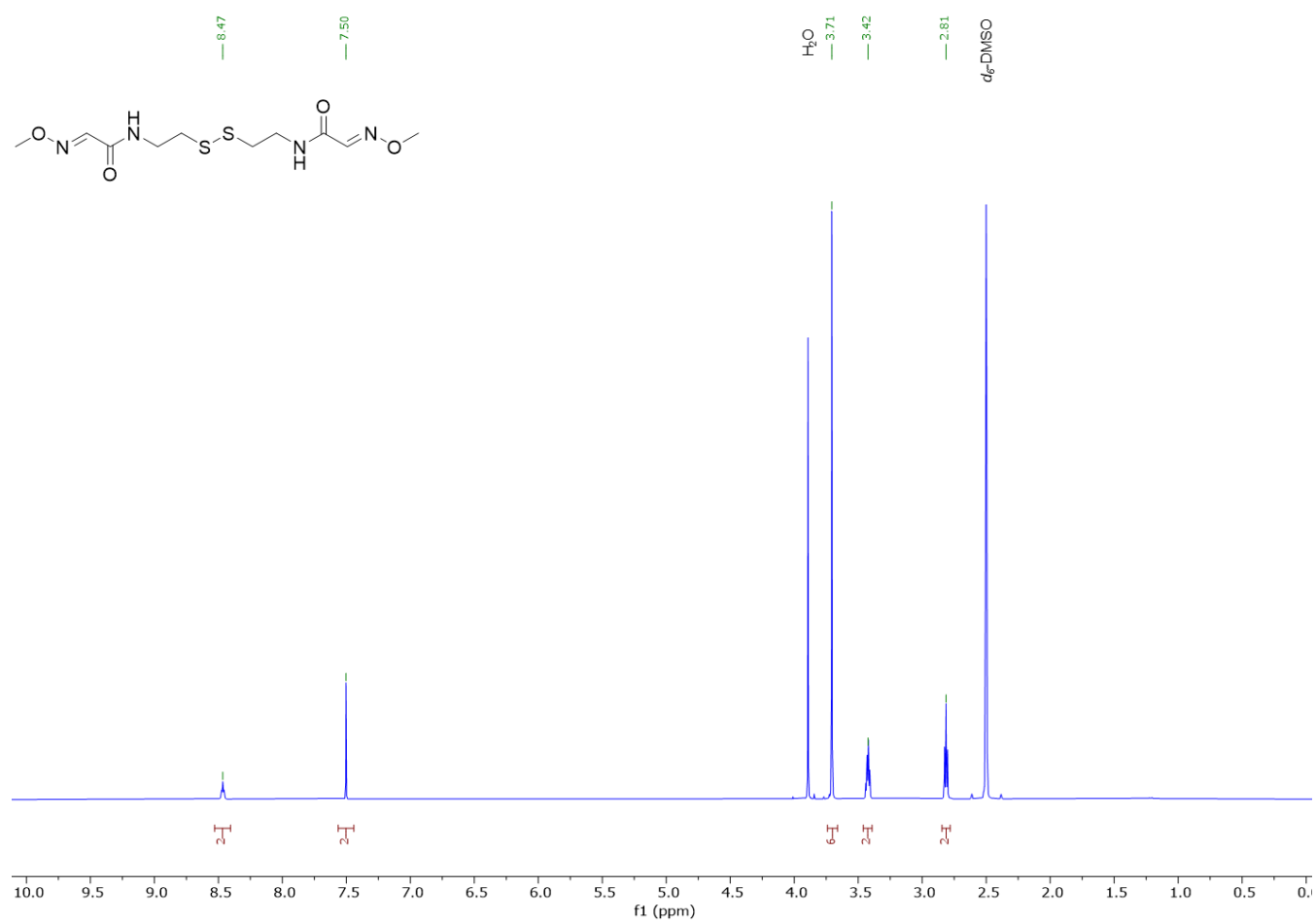
**Figure S77.** <sup>13</sup>C NMR spectrum of Ile-SNAC (11) in DMSO-*d*<sub>6</sub> at 400 MHz.<sup>[8]</sup>

## SUPPORTING INFORMATION



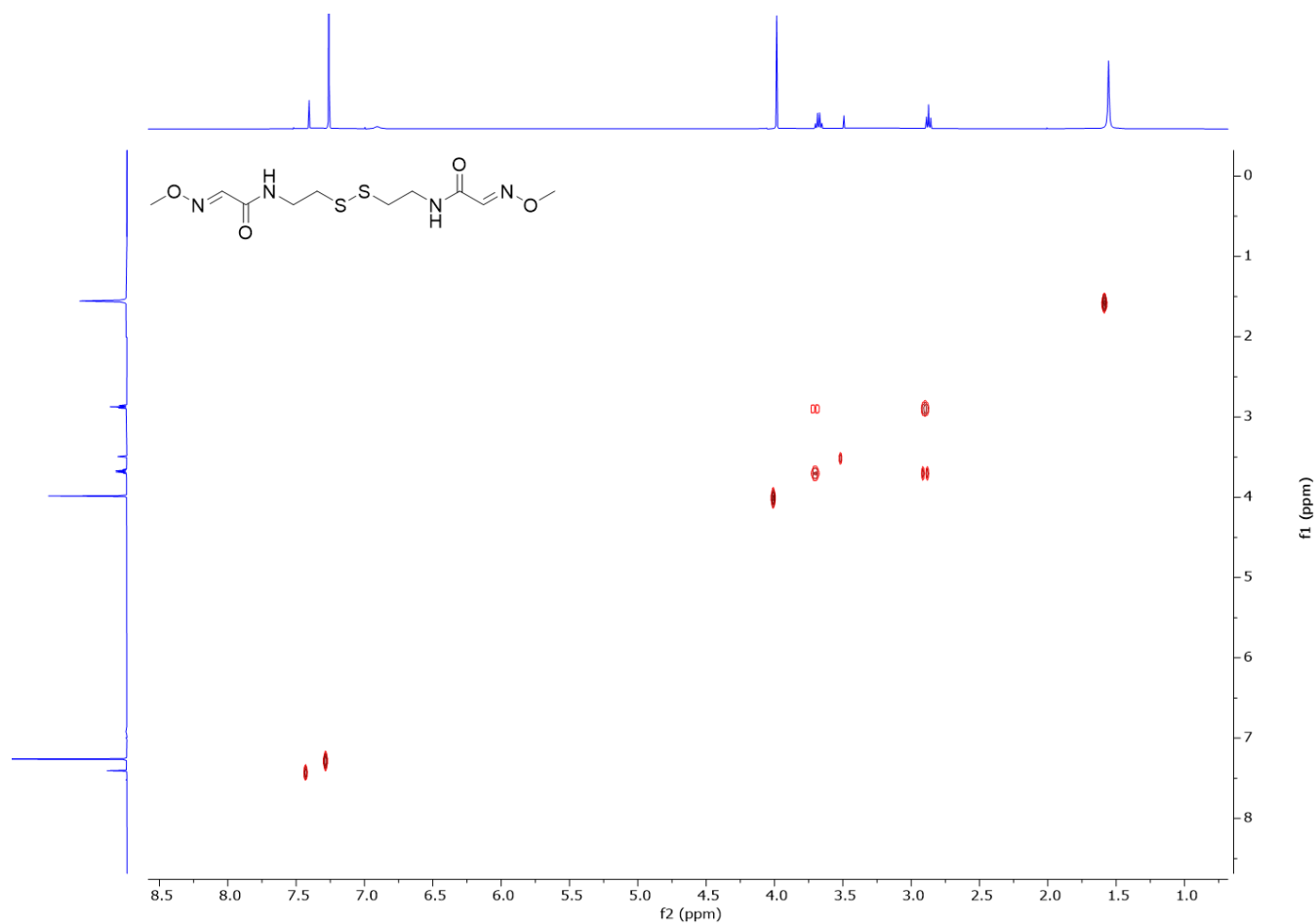
**Figure S78.** <sup>1</sup>H NMR spectrum of (2E,2'E)-N,N-(disulfaneyldiylbis(ethane-2,1-diyl))bis(2-(methoxyimino)acetamide) (**16**) in CDCl<sub>3</sub> at 400 MHz.

## SUPPORTING INFORMATION



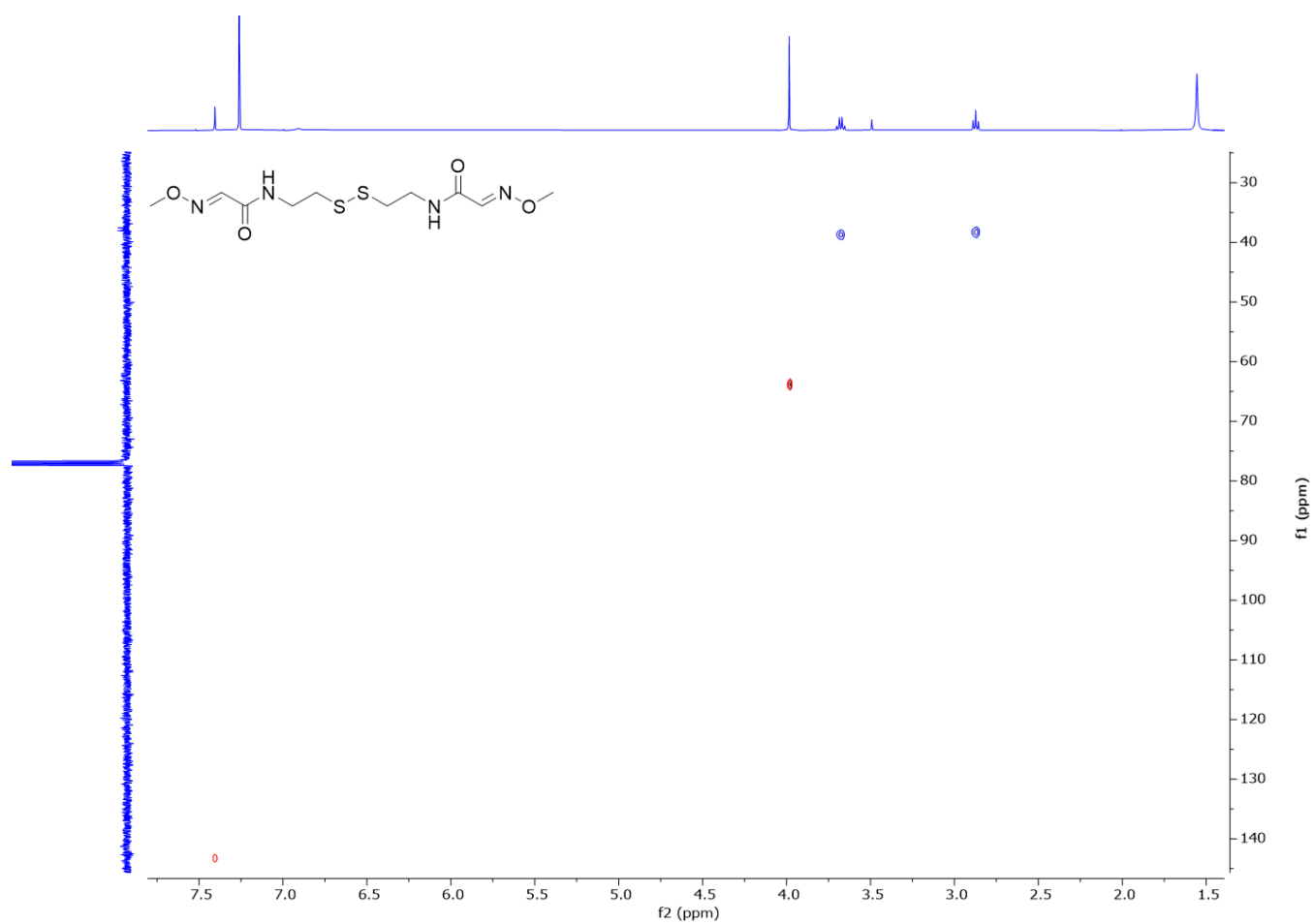
**Figure S79.** <sup>1</sup>H NMR spectrum of (2E,2'E)-N,N-(disulfanediy)bis(ethane-2,1-diy))bis(2-(methoxyimino)acetamide) (**16**) in 9:1 DMSO-*d*<sub>6</sub>:D<sub>2</sub>O at 400 MHz.

## SUPPORTING INFORMATION



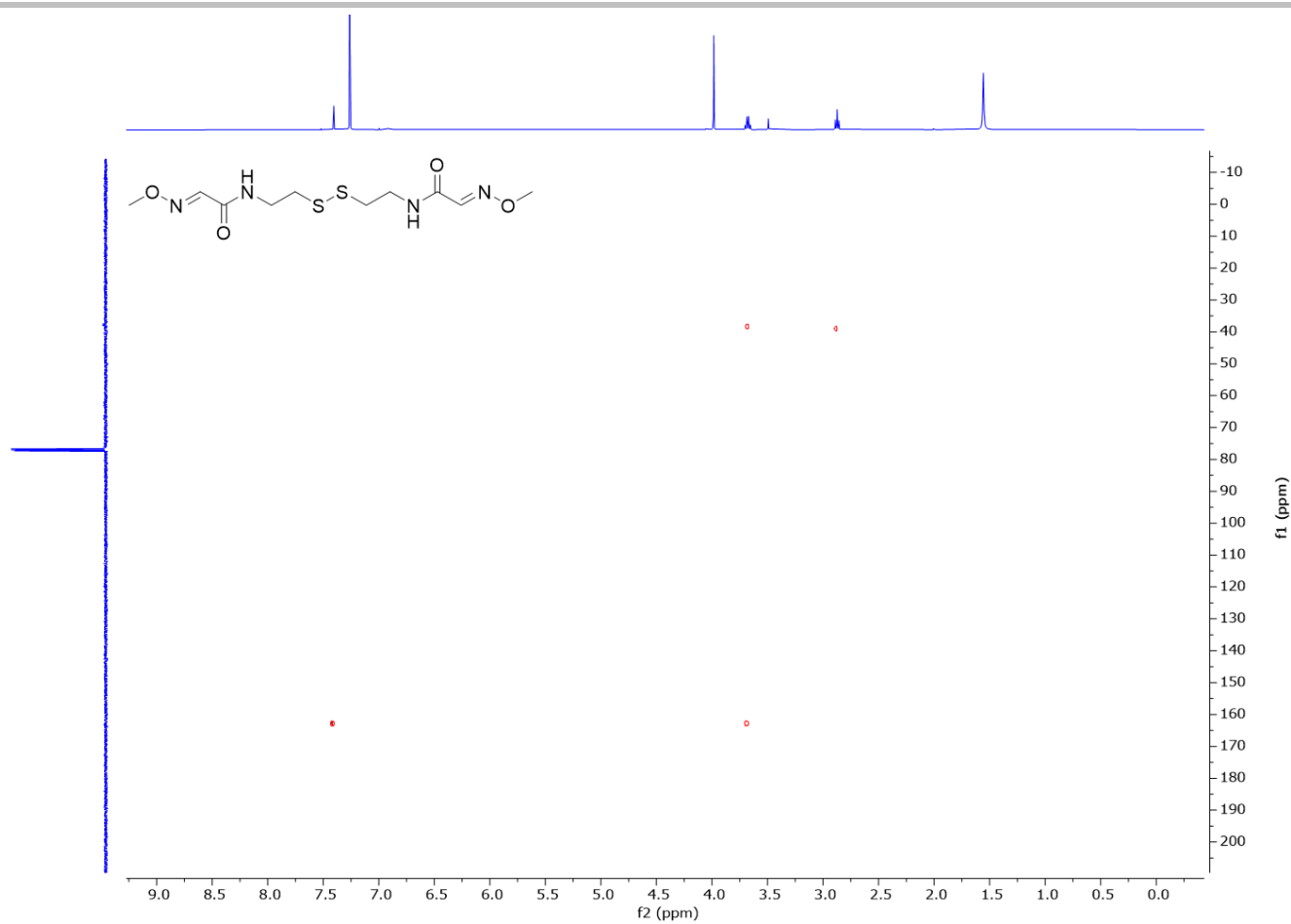
**Figure S80.** COSY spectrum of (2E,2'E)-N,N-(disulfanediybis(ethane-2,1-diy))bis(2-(methoxyimino)acetamide) (**16**) in CDCl<sub>3</sub> at 400 MHz.

## SUPPORTING INFORMATION



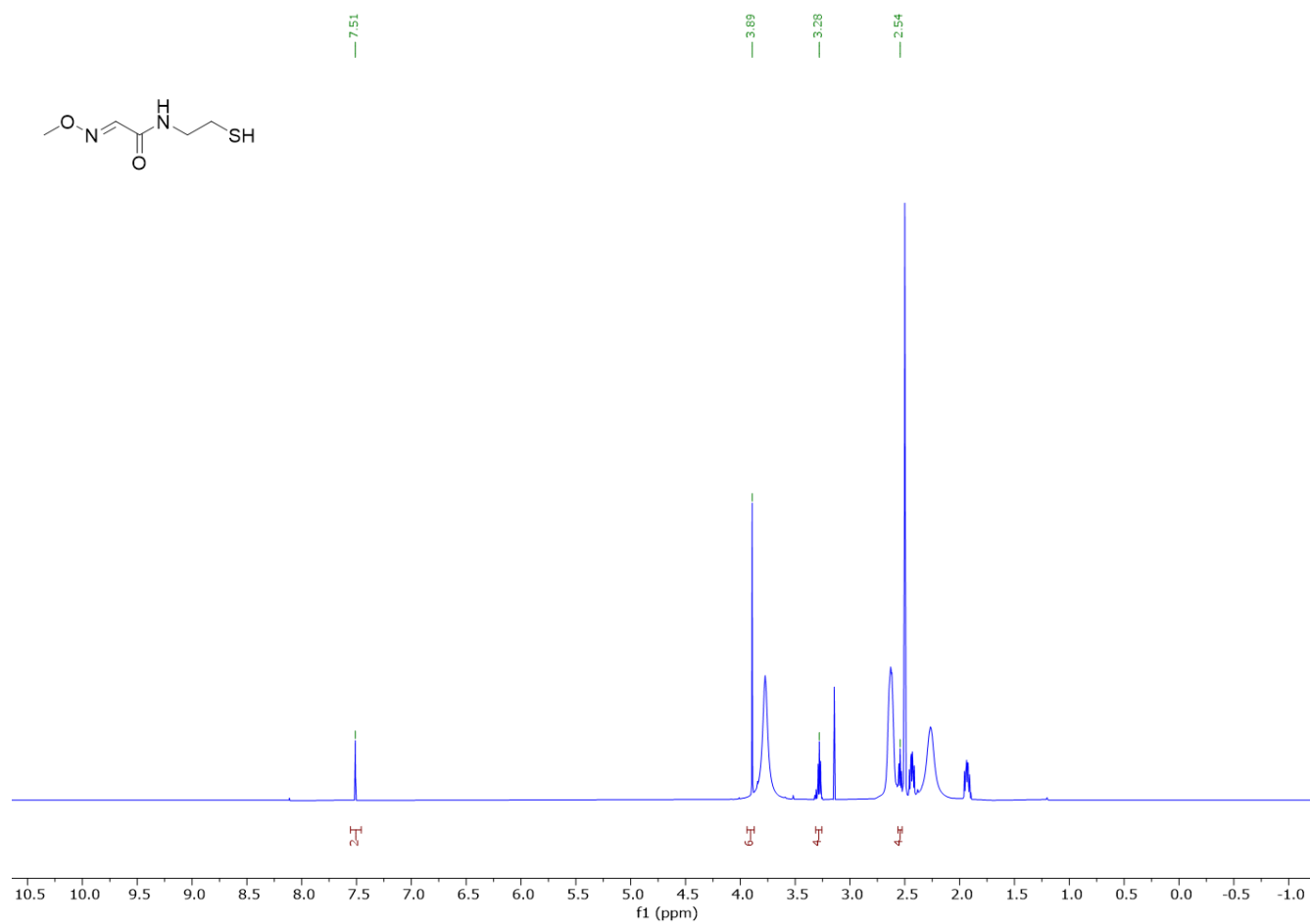
**Figure S81.** HSQC spectrum of *(2E,2'E)-N,N*-(disulfanediy)bis(ethane-2,1-diy)bis(2-(methoxyimino)acetamide) (**16**) in CDCl<sub>3</sub> at 400 MHz.

## SUPPORTING INFORMATION



**Figure S82.** HMBC spectrum of (2E,2'E)-N,N'-(disulfanediy)bis(ethane-2,1-diyl)bis(2-(methoxyimino)acetamide) (**16**) in CDCl<sub>3</sub> at 400 MHz.

## SUPPORTING INFORMATION



**Figure S83.** <sup>1</sup>H NMR spectrum (*E*)-*N*-(2-mercaptoethyl)-2-(methoxyimino)acetamide (**15**) in 9:1 DMSO-*d*<sub>6</sub>:D<sub>2</sub>O at 400 MHz.



## SUPPORTING INFORMATION

## References

- [1] G. R. Fulmer, A. J. M. Miller, N. H. Sherden, H. E. Gottlieb, A. Nudelman, B. M. Stoltz, J. E. Bercaw, K. I. Goldberg, *Organometallics* **2010**, *29*, 2176–2179.
- [2] R. C. Edgar, *Nucleic Acids Res.* **2004**, *32*, 1792–1797.
- [3] M. N. Price, P. S. Dehal, A. P. Arkin, *Plos One* **2010**, *5*, e9490.
- [4] V. I. S. Morariu, B. V.; Raykar, V. C.; Duraiswami, R.; Davis, L. S., *The Conference and Workshop on Neural Information Processing Systems*, **2008**.
- [5] D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, *Nat. Methods* **2009**, *6*, 343–341.
- [6] B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, C. Khosla, *Science* **2001**, *291*, 1790–1792.
- [7] G. W. Heberlig, J. T. C. Brown, R. D. Simard, M. Wirz, W. Zhang, M. Wang, L. I. Susser, M. E. Horsman, C. N. Boddy, *Org. Biomol. Chem.* **2018**, *16*, 5771–5779.
- [8] C. L. Dieterich, S. I. Probst, R. Ueoka, I. Sandu, D. Schäfle, M. Dal Molin, H. A. Minas, R. Costa, A. Oxenius, P. Sander, J. Piel, *Angew. Chem. Int. Ed.* **2022**, *61*, e202115802.
- [9] S. W. Haynes, B. D. Ames, X. Gao, Y. Tang, C. T. Walsh, *Biochemistry-U.S.* **2011**, *50*, 5668–5679.
- [10] H. Xu, X. M. Zou, Y. Q. Zhu, B. Liu, H. L. Tao, X. H. Hu, H. B. Song, F. Z. Hu, Y. Wang, H. Z. Yang, *Pest Manag. Sci.* **2006**, *62*, 522–530.
- [11] M. G. J. Baud, T. Leiser, P. Haus, S. Samlal, A. C. Wong, R. J. Wood, V. Petrucci, M. Gunaratnam, S. M. Hughes, L. Buluwela, F. Turlais, S. Neidle, F. J. Meyer-Almes, A. J. P. White, M. J. Fuchter, *J. Med. Chem.* **2012**, *55*, 1731–1750.
- [12] V. Sklenar, M. Piotto, R. Leppik, V. Saudek, *J. Magn. Res. Ser. A* **1993**, *102*, 241–245.
- [13] M. Piotto, V. Saudek, V. Sklenar, *J. Biomol. NMR* **1992**, *2*, 661–665.
- [14] M. Witanowski, *Pure Appl. Chem.* **1974**, *37*, 225–233.
- [15] a) P. Evans, *Acta Crystallogr. D Biol. Crystallogr.* **2006**, *62*, 72–82; b) P. R. Evans, *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 282–292; c) S. Bailey, *Acta Crystallogr. D* **1994**, *50*, 760–763.
- [16] A. Vagin, A. Teplyakov, *Acta Crystallogr. D* **2010**, *66*, 22–25.
- [17] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589.
- [18] K. Cowtan, *Acta Crystallogr. D* **2006**, *62*, 1002–1011.
- [19] P. Emsley, K. Cowtan, *Acta Crystallogr. D* **2004**, *60*, 2126–2132.
- [20] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 213–221.
- [21] Y. Djoumbou-Feunang, A. Pon, N. Karu, J. M. Zheng, C. Li, D. Arndt, M. Gautam, F. Allen, D. S. Wishart, *Metabolites* **2019**, *9*, 72–95.

## Author Contributions

HAM, RMMF, FH, RAM, AEF, AG, KJW, and JP designed the research, interpreted the data and wrote the manuscript; HAM, RAM and CLD carried out the syntheses; HAM, RMMF, FH, AEF, and SC performed the molecular biology, expressed and purified recombinant proteins and performed the enzyme work; HAM computed the phylogenetic trees; SHR carried out kinetic and <sup>15</sup>N NMR; RMMF and AG performed the X-ray crystallography and the structural biology work; HAM, RMMF, FH, AEF, CLD, SHR, AG, and JP wrote the supporting information. KJW and JP funded the overall project.



---

# *Bibliographie*

---



1. Fraenkel, G. S. The Raison d'Être of Secondary Plant Substances: These odd chemicals arose as a means of protecting plants from insects and now guide insects to food. *Science* **129**, 1466–1470 (1959).
2. Martin, C. & Gerats, T. Control of Pigment Biosynthesis Genes during Petal Development. *Plant Cell* **5**, 1253–1264 (1993).
3. Koes, R. E., Quattrocchio, F. & Mol, J. N. M. The flavonoid biosynthetic pathway in plants: Function and evolution. *BioEssays* **16**, 123–132 (1994).
4. Pichersky, E. & Raguso, R. A. Why do plants produce so many terpenoid compounds? *New Phytol.* **220**, 692–702 (2018).
5. Davies, J. Specialized microbial metabolites: functions and origins. *J. Antibiot. (Tokyo)* **66**, 361–364 (2013).
6. Erb, M. & Kliebenstein, D. J. Plant Secondary Metabolites as Defenses, Regulators, and Primary Metabolites: The Blurred Functional Trichotomy. *Plant Physiol.* **184**, 39–52 (2020).
7. Rosenthal, G. A. The Biological Effects and Mode of Action of L-Canavanine, a Structural Analogue of L-Arginine. *Q. Rev. Biol.* **52**, 155–178 (1977).
8. Chen, H. & Du, L. Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Appl. Microbiol. Biotechnol.* **100**, 541–557 (2016).
9. Cox, R. J. & Simpson, T. J. Fungal Type I Polyketide Synthases. in *Methods in Enzymology* vol. 459 49–78 (Academic Press, 2009).
10. Flores-Sanchez, I. J. & Verpoorte, R. Plant polyketide synthases: a fascinating group of enzymes. *Plant Physiol. Biochem. PPB* **47**, 167–174 (2009).
11. Torres, J. P. & Schmidt, E. W. The biosynthetic diversity of the animal world. *J. Biol. Chem.* **294**, 17684–17692 (2019).
12. Lopanik, N. B., Targett, N. M. & Lindquist, N. Ontogeny of a symbiont-produced chemical defense in *Bugula neritina* (Bryozoa). *Mar. Ecol. Prog. Ser.* **327**, 183–191 (2006).
13. Murata, M. & Yasumoto, T. The structure elucidation and biological activities of high

## Bibliographie

- molecular weight algal toxins: maitotoxin, prymnesins and zooxanthellatoxins (1993 to 1999). *Nat. Prod. Rep.* **17**, 293–314 (2000).
14. Wagstaff, B. A. *et al.* Insights into toxic *Prymnesium parvum* blooms: the role of sugars and algal viruses. *Biochem. Soc. Trans.* **46**, 413–421 (2018).
  15. Lenz, K. D., Klosterman, K. E., Mukundan, H. & Kubicek-Sutherland, J. Z. Macrolides: From Toxins to Therapeutics. *Toxins* **13**, 347 (2021).
  16. Barka, E. A. *et al.* Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol. Mol. Biol. Rev. MMBR* **80**, 1–43 (2016).
  17. Davies, J., Spiegelman, G. B. & Yim, G. The world of subinhibitory antibiotic concentrations. *Curr. Opin. Microbiol.* **9**, 445–453 (2006).
  18. Romero, D., Traxler, M. F., López, D. & Kolter, R. Antibiotics as Signal Molecules. *Chem. Rev.* **111**, 5492–5505 (2011).
  19. Krespach, M. K. C. *et al.* Streptomyces polyketides mediate bacteria–fungi interactions across soil environments. *Nat. Microbiol.* 1–14 (2023) doi:10.1038/s41564-023-01382-2.
  20. Schroeckh, V. *et al.* Intimate bacterial-fungal interaction triggers biosynthesis of archetypal polyketides in *Aspergillus nidulans*. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 14558–14563 (2009).
  21. Cooke, T. F. *et al.* Genetic Mapping and Biochemical Basis of Yellow Feather Pigmentation in Budgerigars. *Cell* **171**, 427–439.e21 (2017).
  22. Calestani, C., Rast, J. P. & Davidson, E. H. Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening. *Development* **130**, 4587–4596 (2003).
  23. Liu, D., Awazu, A., Sakuma, T., Yamamoto, T. & Sakamoto, N. Establishment of knockout adult sea urchins by using a CRISPR-Cas9 system. *Dev. Growth Differ.* **61**, 378–388 (2019).
  24. Krespach, M. K. C. *et al.* Bacterial marginolactones trigger formation of algal gloeocapsoids, protective aggregates on the verge of multicellularity. *Proc. Natl. Acad. Sci.* **118**, e2100892118 (2021).

25. Stubbendieck, R. M. & Straight, P. D. Linear mycins Activate a Two-Component Signaling System Involved in Bacterial Competition and Biofilm Morphology. *J. Bacteriol.* **199**, 10.1128/jb.00186-17 (2017).
26. Krespach, M. K. C. *et al.* Lichen-like association of *Chlamydomonas reinhardtii* and *Aspergillus nidulans* protects algal cells from bacteria. *ISME J.* **14**, 2794–2805 (2020).
27. Günther, M. *et al.* Yellow polyketide pigment suppresses premature hatching in social amoeba. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2116122119 (2022).
28. Yamasaki, D. T., Araki, T. & Narita, T. B. The polyketide synthase StIA is involved in inducing aggregation in *Polysphondylium violaceum*. *Biosci. Biotechnol. Biochem.* **86**, 1590–1598 (2022).
29. Gemperlein, K., Rachid, S., Garcia, R. O., Wenzel, S. C. & Müller, R. Polyunsaturated fatty acid biosynthesis in myxobacteria: different PUFA synthases and their product diversity. *Chem. Sci.* **5**, 1733–1741 (2014).
30. Russell, N. J. & Nichols, D. S. Polyunsaturated fatty acids in marine bacteria--a dogma rewritten. *Microbiol. Read. Engl.* **145 ( Pt 4)**, 767–779 (1999).
31. Hauvermale, A. *et al.* Fatty acid production in *Schizochytrium* sp.: Involvement of a polyunsaturated fatty acid synthase and a type I fatty acid synthase. *Lipids* **41**, 739–747 (2006).
32. Kaulmann, U. & Hertweck, C. Biosynthesis of Polyunsaturated Fatty Acids by Polyketide Synthases. *Angew. Chem. Int. Ed.* **41**, 1866–1869 (2002).
33. Hill, P., Heberlig, G. W. & Boddy, C. N. Sampling Terrestrial Environments for Bacterial Polyketides. *Mol. Basel Switz.* **22**, 707 (2017).
34. Lopera, J., Miller, I. J., McPhail, K. L. & Kwan, J. C. Increased Biosynthetic Gene Dosage in a Genome-Reduced Defensive Bacterial Symbiont. *mSystems* **2**, 10.1128/msystems.00096-17 (2017).
35. Weissman, K. J. Introduction to polyketide biosynthesis. *Methods Enzymol.* **459**, 3–16 (2009).
36. Birch, A. J., Massy-Westropp, R. A. & Moye, C. J. Studies in relation to biosynthesis. VII.

## Bibliographie

- 2-Hydroxy-6-methylbenzoic acid in *Penicillium griseofulvum* Dierckx. *Aust. J. Chem.* **8**, 539–544 (1955).
37. Eckermann, S. *et al.* New pathway to polyketides in plants. *Nature* **396**, 387–390 (1998).
38. Murata, M., Naoki, H., Matsunaga, S., Satake, M. & Yasumoto, T. Structure and Partial Stereochemical Assignments for Maitotoxin, the Most Toxic and Largest Natural Non-Biopolymer. *J. Am. Chem. Soc.* **116**, 7098–7107 (1994).
39. Kuyama, S. & Tamura, T. Cercosporin. A Pigment of *Cercosporina Kikuchii* Matsumoto *et* Tomoyasu. I. Cultivation of Fungus, Isolation and Purification of Pigment. *J. Am. Chem. Soc.* **79**, 5725–5726 (1957).
40. Helaly, S. E. *et al.* Langkolide, a 32-Membered Macrolactone Antibiotic Produced by *Streptomyces* sp. Acta 3062. *J. Nat. Prod.* **75**, 1018–1024 (2012).
41. Jambor, W. P., Steinberg, B. A. & Suydam, L. O. Amphotericins A and B: two new antifungal antibiotics possessing high activity against deep-seated and superficial mycoses. *Antibiot. Annu.* **3**, 574–578 (1955).
42. McGuire, J. M. *et al.* Ilotycin, a new antibiotic. *Antibiot. Chemother. Northfield III* **2**, 281–283 (1952).
43. Newman, D. J., Cragg, G. M. & Snader, K. M. The influence of natural products upon drug discovery. *Nat. Prod. Rep.* **17**, 215–234 (2000).
44. Zhang, L. *et al.* The strategies and techniques of drug discovery from natural products. *Pharmacol. Ther.* **216**, 107686 (2020).
45. Cragg, G. M. & Newman, D. J. Biodiversity: A continuing source of novel drug leads. *Pure Appl. Chem.* **77**, 7–24 (2005).
46. Unschuld, P. U. *Medecine in China: A History of Pharmaceutics*. (University of California Press, 1986).
47. Kapoor, L. D. *CRC Handbook of Ayurvedic Medicinal Plants*. (CRC Press, 2017). doi:10.1201/9781351070997.
48. Schultes, R. E. & Raffauf, R. F. *The Healing Forest: Medicinal and Toxic Plants of the Northwest Amazonia*. (Dioscorides Press, 1990).



49. Iwu, M. M. *Handbook of African Medicinal Plants*. (CRC Press, 2013). doi:10.1201/b16292.
50. Ayensu, E. S. *Medicinal plants of the West Indies*. (Reference Publications, 1981).
51. Arvigo, R. & Balick, M. *Rainforest Remedies: One Hundred Healing Herbs of Belize*. (Lotus Press, 1993).
52. Gupta, M. P. *270 plantas medicinales iberoamericanas*. (CYTED, 1995).
53. Dias, D. A., Urban, S. & Roessner, U. A Historical Overview of Natural Products in Drug Discovery. *Metabolites* **2**, 303–336 (2012).
54. Sertürner, F. W. Ueber das Morphinum, eine neue salzfähige Grundlage, und die Mekonsäure, als Hauptbestandtheile des Opiums. *Ann. Phys.* **55**, 56–89 (1817).
55. Lockermann, G. Friedrich Wilhelm Serturmer, the discoverer of morphine. *J. Chem. Educ.* **28**, 277 (1951).
56. Sneader, W. The discovery of aspirin: a reappraisal. *BMJ* **321**, 1591–1594 (2000).
57. Fleming, A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzae. *Br. J. Exp. Pathol.* **10**, 226–236 (1929).
58. Antibiotics: past, present and future. *Curr. Opin. Microbiol.* **51**, 72–80 (2019).
59. Woodruff, H. B. Selman A. Waksman, winner of the 1952 Nobel Prize for physiology or medicine. *Appl. Environ. Microbiol.* **80**, 2–8 (2014).
60. Waksman, S. A. & Henrici, A. T. The Nomenclature and Classification of the Actinomycetes. *J. Bacteriol.* **46**, 337–341 (1943).
61. Dubos, R. J. Bactericidal Effect of an Extract of a Soil Bacillus on Gram Positive Cocci. *Proc. Soc. Exp. Biol. Med.* **40**, 311–312 (1939).
62. Waksman, S. A. & Woodruff, H. B. The Soil as a Source of Microorganisms Antagonistic to Disease-Producing Bacteria. *J. Bacteriol.* **40**, 581–600 (1940).
63. Waksman, S. A., Horning, E. S. & Spencer, E. L. Two Antagonistic Fungi, *Aspergillus fumigatus* and *Aspergillus clavatus*, and Their Antibiotic Substances. *J. Bacteriol.* **45**, 233–248 (1943).

## Bibliographie

64. Waksman, S. A. & Woodruff, H. B. Streptothricin, a New Selective Bacteriostatic and Bactericidal Agent, Particularly Active Against Gram-Negative Bacteria. *Proc. Soc. Exp. Biol. Med.* **49**, 207–210 (1942).
65. Schatz, A., Bugle, E. & Waksman, S. A. Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.\*†. *Proc. Soc. Exp. Biol. Med.* **55**, 66–69 (1944).
66. Schatz, A. & Waksman, S. A. Effect of Streptomycin and Other Antibiotic Substances upon *Mycobacterium tuberculosis* and Related Organisms.,. *Proc. Soc. Exp. Biol. Med.* **57**, 244–248 (1944).
67. Ekinci, D., Şentürk, M. & Küfrevioğlu, Ö. İ. Salicylic acid derivatives: synthesis, features and usage as therapeutic tools. *Expert Opin. Ther. Pat.* **21**, 1831–1841 (2011).
68. Duggar, B. M. Aureomycin: A Product of the Continuing Search for New Antibiotics. *Ann. N. Y. Acad. Sci.* **51**, 177–181 (1948).
69. Nelson, M. L. & Levy, S. B. The history of the tetracyclines. *Ann. N. Y. Acad. Sci.* **1241**, 17–32 (2011).
70. Finlay, A. C. *et al.* Terramycin, a New Antibiotic. *Science* **111**, 85–85 (1950).
71. Stephens, C. R. *et al.* TERRAMYCIN. VIII. STRUCTURE OF AUREOMYCIN AND TERRAMYCIN. *J. Am. Chem. Soc.* **74**, 4976–4977 (1952).
72. Brockmann, H. & Henkel, W. Pikromycin, ein bitter schmeckendes Antibioticum aus Actinomyceten (Antibiotica aus Actinomyceten, VI. Mitteil. *Chem. Ber.* **84**, 284–288 (1951).
73. Kudoh, S., Azuma, A., Yamamoto, M., Izumi, T. & Ando, M. Improvement of survival in patients with diffuse panbronchiolitis treated with low-dose erythromycin. *Am. J. Respir. Crit. Care Med.* **157**, 1829–1832 (1998).
74. Emmet O'Brien, M., Restrepo, M. I. & Martin-Loeches, I. Update on the combination effect of macrolide antibiotics in community-acquired pneumonia. *Respir. Investig.* **53**, 201–209 (2015).
75. Hazen, E. L. & Brown, R. Fungicidin, an antibiotic produced by a soil actinomycete. *Proc.*

- Soc. Exp. Biol. Med. Soc. Exp. Biol. Med. N. Y. N* **76**, 93–97 (1951).
76. Carolus, H., Pierson, S., Lagrou, K. & Van Dijck, P. Amphotericin B and Other Polyenes—Discovery, Clinical Use, Mode of Action and Drug Resistance. *J. Fungi* **6**, 321 (2020).
77. Burg, R. W. *et al.* Avermectins, New Family of Potent Anthelmintic Agents: Producing Organism and Fermentation. *Antimicrob. Agents Chemother.* **15**, 361–367 (1979).
78. Crump, A. & Ōmura, S. Ivermectin, ‘Wonder drug’ from Japan: the human use perspective. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **87**, 13–28 (2011).
79. Vézina, C., Kudelski, A. & Sehgal, S. N. Rapamycin (AY-22,989), a new antifungal antibiotic. I. Taxonomy of the producing streptomycete and isolation of the active principle. *J. Antibiot. (Tokyo)* **28**, 721–726 (1975).
80. Mukherjee, S. & Mukherjee, U. A Comprehensive Review of Immunosuppression Used for Liver Transplantation. *J. Transplant.* **2009**, e701464 (2009).
81. Gerth, K., Bedorf, N., Höfle, G., Irschik, H. & Reichenbach, H. Epothilons A and B: Antifungal and Cytotoxic Compounds from *Sorangium cellulosum* (Myxobacteria) Production, Physico-chemical and Biological Properties. *J. Antibiot. (Tokyo)* **49**, 560–563 (1996).
82. Bollag, D. M. *et al.* Epothilones, a new class of microtubule-stabilizing agents with a taxol-like mechanism of action. *Cancer Res.* **55**, 2325–2333 (1995).
83. Kowalski, R. J., Giannakakou, P. & Hamel, E. Activities of the Microtubule-stabilizing Agents Epothilones A and B with Purified Tubulin and in Cells Resistant to Paclitaxel (Taxol®) \*. *J. Biol. Chem.* **272**, 2534–2541 (1997).
84. Egerton, N. Ixabepilone (Ixempra), a Therapeutic Option for Locally Advanced or Metastatic Breast Cancer. *Pharm. Ther.* **33**, 523–531 (2008).
85. Shushni, M. A. M., Singh, R., Mentel, R. & Lindequist, U. Balticolid: a new 12-membered macrolide with antiviral activity from an ascomycetous fungus of marine origin. *Mar. Drugs* **9**, 844–851 (2011).
86. Giamarellos-Bourboulis, E. J. Macrolides beyond the conventional antimicrobials: a class of potent immunomodulators. *Int. J. Antimicrob. Agents* **31**, 12–20 (2008).

## Bibliographie

87. Pasquale, T. R. & Tan, J. S. Nonantimicrobial Effects of Antibacterial Agents. *Clin. Infect. Dis.* **40**, 127–135 (2005).
88. Ferrara, G. *et al.* Macrolides in the treatment of asthma and cystic fibrosis. *Respir. Med.* **99**, 1–10 (2005).
89. Endo, A., Kuroda, M. & Tsujita, Y. ML-236A, ML-236B, AND ML-236C, NEW INHIBITORS OF CHOLESTEROGENESIS PRODUCED BY PENICILLIUM CITRINUM. *J. Antibiot. (Tokyo)* **29**, 1346–1348 (1976).
90. Endo, A., Kuroda, M. & Tanzawa, K. Competitive inhibition of 3-hydroxy-3-methylglutaryl coenzyme a reductase by ML-236A and ML-236B fungal metabolites, having hypocholesterolemic activity. *FEBS Lett.* **72**, 323–326 (1976).
91. Alberts, A. W. Discovery, biochemistry and biology of lovastatin. *Am. J. Cardiol.* **62**, J10–J15 (1988).
92. Alberts, A. W. *et al.* Mevinolin: a highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme A reductase and a cholesterol-lowering agent. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 3957–3961 (1980).
93. Endo, A. MONACOLIN K, A NEW HYPOCHOLESTEROLEMIC AGENT PRODUCED BY A MONASCUS SPECIES. *J. Antibiot. (Tokyo)* **32**, 852–854 (1979).
94. Nathan, C. & Cars, O. Antibiotic Resistance — Problems, Progress, and Prospects. *N. Engl. J. Med.* **371**, 1761–1763 (2014).
95. Atanasov, A. G., Zotchev, S. B., Dirsch, V. M. & Supuran, C. T. Natural products in drug discovery: advances and opportunities. *Nat. Rev. Drug Discov.* **20**, 200–216 (2021).
96. Hemmerling, F. & Piel, J. Strategies to access biosynthetic novelty in bacterial genomes for drug discovery. *Nat. Rev. Drug Discov.* **21**, 359–378 (2022).
97. Zarins-Tutt, J. S. *et al.* Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. *Nat. Prod. Rep.* **33**, 54–72 (2015).
98. Bode, H. B., Bethe, B., Höfs, R. & Zeeck, A. Big effects from small changes: possible ways to explore nature's chemical diversity. *Chembiochem Eur. J. Chem. Biol.* **3**, 619–

- 627 (2002).
99. Fuchser, J. & Zeeck, A. Secondary Metabolites by Chemical Screening, 34. – Aspinolides and Aspinonene/Aspyrone Co-Metabolites, New Pentaketides Produced by *Aspergillus ochraceus*. *Liebigs Ann.* **1997**, 87–95 (1997).
  100. Schiewe, H.-J. & Zeeck, A. Cineromycins,  $\gamma$ -Butyrolactones and Ansamycins by Analysis of the Secondary Metabolite Pattern Created by a Single Strain of *Streptomyces*. *J. Antibiot. (Tokyo)* **52**, 635–642 (1999).
  101. Höfs, R., Walker, M. & Zeeck, A. Hexacyclinic acid, a Polyketide from *Streptomyces* with a Novel Carbon Skeleton. *Angew. Chem. Int. Ed.* **39**, 3258–3261 (2000).
  102. Bode, H. B., Walker, M. & Zeeck, A. Cladospirones B to I from *Sphaeropsidales* sp. F-24'707 by Variation of Culture Conditions. *Eur. J. Org. Chem.* **2000**, 3185–3193 (2000).
  103. Hemphill, C. F. P. *et al.* OSMAC approach leads to new fusarielin metabolites from *Fusarium tricinctum*. *J. Antibiot. (Tokyo)* **70**, 726–732 (2017).
  104. Scherlach, K. & Hertweck, C. Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **7**, 1753–1760 (2009).
  105. Derewacz, D. K., Covington, B. C., McLean, J. A. & Bachmann, B. O. Mapping Microbial Response Metabolomes for Induced Natural Product Discovery. *ACS Chem. Biol.* **10**, 1998–2006 (2015).
  106. Cueto, M. *et al.* Pestalone, a New Antibiotic Produced by a Marine Fungus in Response to Bacterial Challenge. *J. Nat. Prod.* **64**, 1444–1446 (2001).
  107. Hofer, U. The majority is uncultured. *Nat. Rev. Microbiol.* **16**, 716–717 (2018).
  108. Winterberg, H. Zur Methodik der Bakterienzählung. *Z. Für Hyg. Infect.* **29**, 75–93 (1898).
  109. Lewis, K., Epstein, S., D'Onofrio, A. & Ling, L. L. Uncultured microorganisms as a source of secondary metabolites. *J. Antibiot. (Tokyo)* **63**, 468–476 (2010).
  110. Kaeberlein, T., Lewis, K. & Epstein, S. S. Isolating 'uncultivable' microorganisms in pure culture in a simulated natural environment. *Science* **296**, 1127–1129 (2002).
  111. Nichols, D. *et al.* Use of Ichip for High-Throughput In Situ Cultivation of "Uncultivable" Microbial Species. *Appl. Environ. Microbiol.* **76**, 2445–2450 (2010).

## Bibliographie

112. Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455–459 (2015).
113. Masson, F. & Lemaitre, B. Growing Ungrowable Bacteria: Overview and Perspectives on Insect Symbiont Culturability. *Microbiol. Mol. Biol. Rev. MMBR* **84**, e00089-20 (2020).
114. Ziemert, N., Alanjary, M. & Weber, T. The evolution of genome mining in microbes – a review. *Nat. Prod. Rep.* **33**, 988–1005 (2016).
115. Zhang, H., Boghigian, B. A., Armando, J. & Pfeifer, B. A. Methods and options for the heterologous production of complex natural products. *Nat. Prod. Rep.* **28**, 125–151 (2010).
116. Rust, M. *et al.* A multiproducer microbiome generates chemical diversity in the marine sponge *Mycale hentscheli*. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 9508–9518 (2020).
117. Blin, K. *et al.* antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **51**, W46–W50 (2023).
118. Helfrich, E. J. N. *et al.* Automated structure prediction of trans-acyltransferase polyketide synthase products. *Nat. Chem. Biol.* **15**, 813–821 (2019).
119. Zerikly, M. & Challis, G. L. Strategies for the Discovery of New Natural Products by Genome Mining. *ChemBioChem* **10**, 625–633 (2009).
120. Laureti, L. *et al.* Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6258–6263 (2011).
121. Zhong, H. A. ADMET Properties: Overview and Current Topics. *Drug Des. Princ. Appl.* 113–133 (2017) doi:10.1007/978-981-10-5187-6\_8.
122. Omura, S., Morimoto, S., Nagate, T., Adachi, T. & Kohno, Y. [Research and development of clarithromycin]. *Yakugaku Zasshi* **112**, 593–614 (1992).
123. Girard, A. E. *et al.* Pharmacokinetic and in vivo studies with azithromycin (CP-62,993), a new macrolide with an extended half-life and excellent tissue distribution. *Antimicrob. Agents Chemother.* **31**, 1948–1954 (1987).
124. Katz, L. & Ashley, G. W. Translation and Protein Synthesis: Macrolides. *Chem. Rev.* **105**,

- 499–528 (2005).
125. Rubinstein, E. Comparative safety of the different macrolides. *Int. J. Antimicrob. Agents* **18**, 71–76 (2001).
126. Denis, A. *et al.* Synthesis and antibacterial activity of HMR 3647 a new ketolide highly potent against erythromycin-resistant and susceptible pathogens. *Bioorg. Med. Chem. Lett.* **9**, 3075–3080 (1999).
127. Dinos, G. P. The macrolide antibiotic renaissance. *Br. J. Pharmacol.* **174**, 2967–2983 (2017).
128. Zhong, P. *et al.* Induction of ribosome methylation in MLS-resistant *Streptococcus pneumoniae* by macrolides and ketolides. *Microb. Drug Resist. Larchmt. N* **5**, 183–188 (1999).
129. Paterson, I. & Lam, N. Y. S. Challenges and discoveries in the total synthesis of complex polyketide natural products. *J. Antibiot. (Tokyo)* **71**, 215–233 (2018).
130. Aicher, T. D. *et al.* Total synthesis of halichondrin B and norhalichondrin B. *J. Am. Chem. Soc.* **114**, 3162–3164 (1992).
131. Littlefield, B. A. *et al.* Macrocyclic analogs and methods of their use and preparation. (2001).
132. Towle, M. J. *et al.* In Vitro and In Vivo Anticancer Activities of Synthetic Macrocyclic Ketone Analogues of Halichondrin B1. *Cancer Res.* **61**, 1013–1021 (2001).
133. Bauer, A. Story of Eribulin Mesylate: Development of the Longest Drug Synthesis. in *Synthesis of Heterocycles in Contemporary Medicinal Chemistry* (ed. Časar, Z.) 209–270 (Springer International Publishing, 2016). doi:10.1007/7081\_2016\_201.
134. Wang, Z. *et al.* A naturally inspired antibiotic to target multidrug-resistant pathogens. *Nature* **601**, 606–611 (2022).
135. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **348**, 176–178 (1990).
136. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. Modular

## Bibliographie

- organization of genes required for complex polyketide biosynthesis. *Science* **252**, 675–679 (1991).
137. Witkowski, A., Joshi, A. K. & Smith, S. Mechanism of the  $\beta$ -Ketoacyl Synthase Reaction Catalyzed by the Animal Fatty Acid Synthase. *Biochemistry* **41**, 10877–10887 (2002).
138. Weissman, K. J. Polyketide stereocontrol: a study in chemical biology. *Beilstein J. Org. Chem.* **13**, 348–371 (2017).
139. Abe, I. & Morita, H. Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases. *Nat. Prod. Rep.* **27**, 809–838 (2010).
140. Gross, F. *et al.* Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. *Arch. Microbiol.* **185**, 28–38 (2006).
141. Hashimoto, M., Nonaka, T. & Fujii, I. Fungal type III polyketide synthases. *Nat. Prod. Rep.* **31**, 1306–1317 (2014).
142. Austin, M. B. *et al.* Biosynthesis of *Dictyostelium discoideum* differentiation-inducing factor by a hybrid type I fatty acid-type III polyketide synthase. *Nat. Chem. Biol.* **2**, 494–502 (2006).
143. Austin, M. & Noel, J. The chalcone synthase superfamily of type III polyketide synthases. *Nat. Prod. Rep.* **20**, 79–110 (2003).
144. Shimizu, Y., Ogata, H. & Goto, S. Type III Polyketide Synthases: Functional Classification and Phylogenomics. *Chembiochem Eur. J. Chem. Biol.* **18**, 50–65 (2017).
145. Kreuzaler, F. & Hahlbrock, K. Enzymatic synthesis of aromatic compounds in higher plants: formation of naringenin (5,7,4'-trihydroxyflavanone) from p-coumaroyl coenzyme A and malonyl coenzyme A. *FEBS Lett.* **28**, 69–72 (1972).
146. Helariutta, Y. *et al.* Chalcone synthase-like genes active during corolla development are differentially expressed and encode enzymes with different catalytic properties in *Gerbera hybrida* (Asteraceae). *Plant Mol. Biol.* **28**, 47–60 (1995).
147. Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **24**, 162–190 (2007).



148. Burson, K. K. & Khosla, C. Dissecting the Chain Length Specificity in Bacterial Aromatic Polyketide Synthases using Chimeric Genes. *Tetrahedron* **56**, 9401–9408 (2000).
149. Chen, A., Re, R. N. & Burkart, M. D. Type II fatty acid and polyketide synthases: deciphering protein–protein and protein–substrate interactions. *Nat. Prod. Rep.* **35**, 1029–1045 (2018).
150. Lambalot, R. H. *et al.* A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923–936 (1996).
151. Rix, U., Fischer, C., Remsing, L. L. & Rohr, J. Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Nat. Prod. Rep.* **19**, 542–580 (2002).
152. Taguchi, T. *et al.* Bifunctionality of ActIV as a Cyclase-Thioesterase Revealed by in Vitro Reconstitution of Actinorhodin Biosynthesis in *Streptomyces coelicolor* A3(2). *Chembiochem Eur. J. Chem. Biol.* **18**, 316–323 (2017).
153. Hashimoto, M. *et al.* Actinorhodin Biosynthesis Terminates with an Unprecedented Biaryl Coupling Reaction. *Angew. Chem. Int. Ed.* **62**, e202214400 (2023).
154. Du, D., Katsuyama, Y., Shin-ya, K. & Ohnishi, Y. Reconstitution of a Type II Polyketide Synthase that Catalyzes Polyene Formation. *Angew. Chem. Int. Ed.* **57**, 1954–1957 (2018).
155. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416 (2001).
156. Cox, R. J. Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Org. Biomol. Chem.* **5**, 2010–2026 (2007).
157. Castoe, T. A., Stephens, T., Noonan, B. P. & Calestani, C. A novel group of type I polyketide synthases (PKS) in animals and the complex phylogenomics of PKSs. *Gene* **392**, 47–58 (2007).
158. Herbst, D. A., Townsend, C. A. & Maier, T. The architectures of iterative type I PKS and FAS. *Nat. Prod. Rep.* **35**, 1046–1069 (2018).
159. Metz, J. G. *et al.* Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. *Science* **293**, 290–293 (2001).

## Bibliographie

160. Weissman, K. J. Peering into the Black Box of Fungal Polyketide Biosynthesis. *ChemBioChem* **11**, 485–488 (2010).
161. Parascandolo, J. S. *et al.* Insights into 6-Methylsalicylic Acid Bio-assembly by Using Chemical Probes. *Angew. Chem. Int. Ed.* **55**, 3463–3467 (2016).
162. Wang, J. *et al.* Structural basis for the biosynthesis of lovastatin. *Nat. Commun.* **12**, 867 (2021).
163. Zhang, Q., Pang, B., Ding, W. & Liu, W. Aromatic Polyketides Produced by Bacterial Iterative Type I Polyketide Synthases. *ACS Catal.* **3**, 1439–1447 (2013).
164. Liu, W., Christenson, S. D., Standage, S. & Shen, B. Biosynthesis of the Eneidyne Antitumor Antibiotic C-1027. *Science* **297**, 1170–1173 (2002).
165. Blodgett, J. A. V. *et al.* Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. *Proc. Natl. Acad. Sci.* **107**, 11692–11697 (2010).
166. Weissman, K. J. & Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol.* **3**, 925–936 (2005).
167. Weissman, K. J. Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Nat. Prod. Rep.* **33**, 203–230 (2016).
168. Scotti, C. *et al.* A *Bacillus subtilis* large ORF coding for a polypeptide highly similar to polyketide synthases. *Gene* **130**, 65–71 (1993).
169. Paitan, Y., Alon, G., Orr, E., Ron, E. Z. & Rosenberg, E. The first gene in the biosynthesis of the polyketide antibiotic TA Of *Myxococcus xanthus* codes for a unique PKS module coupled to a peptide synthetase<sup>11</sup>Edited by M. Yaniv. *J. Mol. Biol.* **286**, 465–474 (1999).
170. Huang, G., Zhang, L. & Birch, R. G. A multifunctional polyketide-peptide synthetase essential for albicidin biosynthesis in *Xanthomonas albilineans*. *Microbiol. Read. Engl.* **147**, 631–642 (2001).
171. Piel, J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14002–14007 (2002).

172. Cheng, Y.-Q., Tang, G.-L. & Shen, B. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3149–3154 (2003).
173. Musiol, E. M. *et al.* Supramolecular Templating in Kirromycin Biosynthesis: The Acyltransferase KirCII Loads Ethylmalonyl-CoA Extender onto a Specific ACP of the trans-AT PKS. *Chem. Biol.* **18**, 438–444 (2011).
174. Zhao, C. *et al.* Oxazolomycin biosynthesis in *Streptomyces albus* JA3453 featuring an ‘acyltransferase-less’ type I polyketide synthase that incorporates two distinct extender units. *J. Biol. Chem.* **285**, 20097–20108 (2010).
175. Menche, D. *et al.* Stereochemical Determination and Complex Biosynthetic Assembly of Etnangien, a Highly Potent RNA Polymerase Inhibitor from the Myxobacterium *Sorangium cellulosum*. *J. Am. Chem. Soc.* **130**, 14234–14243 (2008).
176. Helfrich, E. J. N. & Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **33**, 231–316 (2016).
177. Nguyen, T. *et al.* Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* **26**, 225–233 (2008).
178. Helfrich, E. J. N. *et al.* Evolution of combinatorial diversity in trans-acyltransferase polyketide synthase assembly lines across bacteria. *Nat. Commun.* **12**, 1422 (2021).
179. Mattheus, W. *et al.* Isolation and Purification of a New Kalimantacin/Batumin-Related Polyketide Antibiotic and Elucidation of Its Biosynthesis Gene Cluster. *Chem. Biol.* **17**, 149–159 (2010).
180. Gaisser, S. *et al.* Direct production of ivermectin-like drugs after domain exchange in the avermectin polyketide synthase of *Streptomyces avermitilis* ATCC31272. *Org. Biomol. Chem.* **1**, 2840–2847 (2003).
181. Tang, L., Fu, H. & McDaniel, R. Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases. *Chem. Biol.* **7**, 77–84 (2000).

## Bibliographie

182. Till, M. & Race, P. R. Progress challenges and opportunities for the re-engineering of trans-AT polyketide synthases. *Biotechnol. Lett.* **36**, 877–888 (2014).
183. Traitcheva, N., Jenke-Kodama, H., He, J., Dittmann, E. & Hertweck, C. Non-Colinear Polyketide Biosynthesis in the Aureothin and Neoaureothin Pathways: An Evolutionary Perspective. *ChemBioChem* **8**, 1841–1849 (2007).
184. Nivina, A., Yuet, K. P., Hsu, J. & Khosla, C. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem. Rev.* **119**, 12524–12547 (2019).
185. Tang, Y., Chen, A. Y., Kim, C.-Y., Cane, D. E. & Khosla, C. Structural and Mechanistic Analysis of Protein Interactions in Module 3 of the 6-Deoxyerythronolide B Synthase. *Chem. Biol.* **14**, 931–943 (2007).
186. Tang, Y., Kim, C.-Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11124–11129 (2006).
187. Miyanaga, A., Iwasawa, S., Shinohara, Y., Kudo, F. & Eguchi, T. Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1802–1807 (2016).
188. Miyanaga, A., Kawada, K., Chisuga, T., Kudo, F. & Eguchi, T. Structural Basis of Transient Interactions of Acyltransferase Vink with the Loading Acyl Carrier Protein of the Vicenistatin Modular Polyketide Synthase. *Biochemistry* **62**, 17–21 (2023).
189. Wong, F. T., Jin, X., Mathews, I. I., Cane, D. E. & Khosla, C. Structure and mechanism of the trans-acting acyltransferase from the disorazole synthase. *Biochemistry* **50**, 6539–6548 (2011).
190. Hedstrom, L. Serine Protease Mechanism and Specificity. *Chem. Rev.* **102**, 4501–4524 (2002).
191. Li, Y. *et al.* Structural Basis of a Broadly Selective Acyltransferase from the Polyketide Synthase of Splenocin. *Angew. Chem. Int. Ed.* **57**, 5823–5827 (2018).
192. Tsai, S.-C. (Sheryl) & Ames, B. D. Structural enzymology of polyketide synthases. *Methods Enzymol.* **459**, 17–47 (2009).

193. Lau, J., Cane, D. E. & Khosla, C. Substrate Specificity of the Loading Didomain of the Erythromycin Polyketide Synthase. *Biochemistry* **39**, 10514–10520 (2000).
194. Smith, S. & Tsai, S.-C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* **24**, 1041–1072 (2007).
195. Haydock, S. F. *et al.* Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* **374**, 246–248 (1995).
196. Chang, C. *et al.* Uncovering the Formation and Selection of Benzylmalonyl-CoA from the Biosynthesis of Splenocin and Enterocin Reveals a Versatile Way to Introduce Amino Acids into Polyketide Carbon Scaffolds. *J. Am. Chem. Soc.* **137**, 4183–4190 (2015).
197. Bravo-Rodriguez, K. *et al.* Substrate Flexibility of a Mutated Acyltransferase Domain and Implications for Polyketide Biosynthesis. *Chem. Biol.* **22**, 1425–1430 (2015).
198. Reeves, C. D. *et al.* Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations. *Biochemistry* **40**, 15464–15470 (2001).
199. Englund, E. *et al.* Expanding Extender Substrate Selection for Unnatural Polyketide Biosynthesis by Acyltransferase Domain Exchange within a Modular Polyketide Synthase. *J. Am. Chem. Soc.* **145**, 8822–8832 (2023).
200. Moretto, L., Heylen, R., Holroyd, N., Vance, S. & Broadhurst, R. W. Modular type I polyketide synthase acyl carrier protein domains share a common N-terminally extended fold. *Sci. Rep.* **9**, 2325 (2019).
201. Collin, S. *et al.* Decrypting the programming of  $\beta$ -methylation in virginiamycin M biosynthesis. *Nat. Commun.* **14**, 1327 (2023).
202. Gay, D. C. *et al.* A Close Look at a Ketosynthase from a Trans-Acyltransferase Modular Polyketide Synthase. *Structure* **22**, 444–451 (2014).
203. Tang, Y., Chen, A. Y., Kim, C.-Y., Cane, D. E. & Khosla, C. Structural and Mechanistic Analysis of Protein Interactions in Module 3 of the 6-Deoxyerythronolide B Synthase. *Chem. Biol.* **14**, 931–943 (2007).

## Bibliographie

204. Bisang, C. *et al.* A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature* **401**, 502–505 (1999).
205. Al-Dhelaan, R. *et al.* Condensation-Incompetent Ketosynthase Inhibits trans-Acyltransferase Activity. *ACS Chem. Biol.* **14**, 304–312 (2019).
206. Keatinge-Clay, A. T. & Stroud, R. M. The Structure of a Ketoreductase Determines the Organization of the  $\beta$ -Carbon Processing Enzymes of Modular Polyketide Synthases. *Structure* **14**, 737–748 (2006).
207. Piasecki, S. K., Zheng, J., Axelrod, A. J., Detelich, M. E. & Keatinge-Clay, A. T. Structural and functional studies of a trans-acyltransferase polyketide assembly line enzyme that catalyzes stereoselective  $\alpha$ - and  $\beta$ -ketoreduction. *Proteins* **82**, 2067–2077 (2014).
208. Weissman, K. J. 1.02 - Bacterial Type I Polyketide Synthases. in *Comprehensive Natural Products III* (eds. Liu, H.-W. (Ben) & Begley, T. P.) 4–46 (Elsevier, 2020). doi:10.1016/B978-0-12-409547-2.14644-X.
209. Kavanagh, K. L., Jörnvall, H., Persson, B. & Oppermann, U. Medium- and short-chain dehydrogenase/reductase gene and protein families. *Cell. Mol. Life Sci.* **65**, 3895 (2008).
210. Zheng, J. & Keatinge-Clay, A. T. The status of type I polyketide synthase ketoreductases. *MedChemComm* **4**, 34–40 (2012).
211. Schäfer, M., Stevenson, C. E. M., Wilkinson, B., Lawson, D. M. & Buttner, M. J. Substrate-Assisted Catalysis in Polyketide Reduction Proceeds via a Phenolate Intermediate. *Cell Chem. Biol.* **23**, 1091–1097 (2016).
212. McPherson, M., Khosla, C. & Cane, D. E. Erythromycin Biosynthesis: The  $\beta$ -Ketoreductase Domains Catalyze the Stereospecific Transfer of the 4-pro-S Hydride of NADPH. *J. Am. Chem. Soc.* **120**, 3267–3268 (1998).
213. Yin, Y., Gokhale, R., Khosla, C. & Cane, D. E. Erythromycin biosynthesis. The 4-pro-S hydride of NADPH is utilized for ketoreduction by both module 5 and module 6 of the 6-deoxyerythronolide B synthase. *Bioorg. Med. Chem. Lett.* **11**, 1477–1479 (2001).
214. Liu, C. *et al.* Substrate-bound structures of a ketoreductase from amphotericin modular polyketide synthase. *J. Struct. Biol.* **203**, 135–141 (2018).

215. Caffrey, P. Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *Chembiochem Eur. J. Chem. Biol.* **4**, 654–657 (2003).
216. Keatinge-Clay, A. T. A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. *Chem. Biol.* **14**, 898–908 (2007).
217. Bonnett, S. A. *et al.* Structural and Stereochemical Analysis of a Modular Polyketide Synthase Ketoreductase Domain Required for the Generation of a cis-Alkene. *Chem. Biol.* **20**, 10.1016/j.chembiol.2013.04.014 (2013).
218. Xie, X., Garg, A., Keatinge-Clay, A. T., Khosla, C. & Cane, D. E. Epimerase and Reductase Activities of Polyketide Synthase Ketoreductase Domains Utilize the Same Conserved Tyrosine and Serine Residues. *Biochemistry* **55**, 1179–1186 (2016).
219. Drufva, E. E., Spengler, N. R., Hix, E. G. & Bailey, C. B. Site-Directed Mutagenesis of Modular Polyketide Synthase Ketoreductase Domains for Altered Stereochemical Control. *Chembiochem Eur. J. Chem. Biol.* **22**, 1122–1150 (2021).
220. Valenzano, C. R. *et al.* Stereospecificity of the Dehydratase Domain of the Erythromycin Polyketide Synthase. *J. Am. Chem. Soc.* **132**, 14697–14699 (2010).
221. Keatinge-Clay, A. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *J. Mol. Biol.* **384**, 941–953 (2008).
222. Akey, D. L. *et al.* Crystal Structures of Dehydratase Domains from the Curacin Polyketide Biosynthetic Pathway. *Structure* **18**, 94–105 (2010).
223. Barajas, J. F. *et al.* Structural insights into dehydratase substrate selection for the borrelidin and fluvirucin polyketide synthases. *J. Ind. Microbiol. Biotechnol.* **46**, 1225–1235 (2019).
224. Labonte, J. W. & Townsend, C. A. Active Site Comparisons and Catalytic Mechanisms of the Hot Dog Superfamily. *Chem. Rev.* **113**, 2182–2204 (2013).
225. Palaniappan, N., Alhamadsheh, M. M. & Reynolds, K. A. cis- $\Delta^{2,3}$ -Double Bond of Phoslactomycins Is Generated by a Post-PKS Tailoring Enzyme. *J. Am. Chem. Soc.* **130**, 12236–12237 (2008).

## Bibliographie

226. Kandziora, N. *et al.* Uncovering the origin of Z-configured double bonds in polyketides: intermediate E-double bond formation during borrelidin biosynthesis. *Chem. Sci.* **5**, 3563–3567 (2014).
227. Xie, X. & Cane, D. E. pH-Rate profiles establish that polyketide synthase dehydratase domains utilize a single-base mechanism. *Org. Biomol. Chem.* **16**, 9165–9170 (2018).
228. Wagner, D. T. *et al.* Structural and Functional Trends in Dehydrating Bimodules from trans-Acyltransferase Polyketide Synthases. *Struct. Lond. Engl.* **1993** **25**, 1045-1055.e2 (2017).
229. Hobson, C. *et al.* Diene incorporation by a dehydratase domain variant in modular polyketide synthases. *Nat. Chem. Biol.* **18**, 1410–1416 (2022).
230. He, H.-Y., Tang, M.-C., Zhang, F. & Tang, G.-L. Cis-Double Bond Formation by Thioesterase and Transfer by Ketosynthase in FR901464 Biosynthesis. *J. Am. Chem. Soc.* **136**, 4488–4491 (2014).
231. Zheng, J., Gay, D. C., Demeler, B., White, M. A. & Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol.* **8**, 615–621 (2012).
232. McCullough, T. M. *et al.* Structure of a modular polyketide synthase reducing region. *Struct. Lond. Engl.* **1993** S0969-2126(23)00195–8 (2023) doi:10.1016/j.str.2023.05.019.
233. Kwan, D. H. *et al.* Prediction and Manipulation of the Stereochemistry of Enoylreduction in Modular Polyketide Synthases. *Chem. Biol.* **15**, 1231–1240 (2008).
234. Kwan, D. H. & Leadlay, P. F. Mutagenesis of a Modular Polyketide Synthase Enoylreductase Domain Reveals Insights into Catalysis and Stereospecificity. *ACS Chem. Biol.* **5**, 829–838 (2010).
235. Rosenthal, R. G. *et al.* The use of ene adducts to study and engineer enoyl-thioester reductases. *Nat. Chem. Biol.* **11**, 398–400 (2015).
236. Bukhari, H. S. T., Jakob, R. P. & Maier, T. Evolutionary origins of the multienzyme architecture of giant fungal fatty acid synthase. *Struct. Lond. Engl.* **1993** **22**, 1775–1785 (2014).



237. Ames, B. D. *et al.* Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis. *Proc. Natl. Acad. Sci.* **109**, 11144–11149 (2012).
238. Zhou, Y., Prediger, P., Dias, L. C., Murphy, A. C. & Leadlay, P. F. Macrodilide Formation by the Thioesterase of a Modular Polyketide Synthase. *Angew. Chem. Int. Ed.* **54**, 5232–5235 (2015).
239. Heathcote, M. L., Staunton, J. & Leadlay, P. F. Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem. Biol.* **8**, 207–220 (2001).
240. Kotowska, M. & Pawlik, K. Roles of type II thioesterases and their application for secondary metabolite yield improvement. *Appl. Microbiol. Biotechnol.* **98**, 7735–7746 (2014).
241. Du, L. & Lou, L. PKS and NRPS release mechanisms. *Nat. Prod. Rep.* **27**, 255–278 (2010).
242. Korman, T. P. *et al.* Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 6246–6251 (2010).
243. Gaudelli, N. M. & Townsend, C. A. Epimerization and substrate gating by a TE domain in  $\beta$ -lactam antibiotic biosynthesis. *Nat. Chem. Biol.* **10**, 251–258 (2014).
244. Scaglione, J. B. *et al.* Biochemical and Structural Characterization of the Tautomycetin Thioesterase: Analysis of a Stereoselective Polyketide Hydrolase. *Angew. Chem. Int. Ed.* **49**, 5726–5730 (2010).
245. Pinto, A., Wang, M., Horsman, M. & Boddy, C. N. 6-Deoxyerythronolide B Synthase Thioesterase-Catalyzed Macrocyclization Is Highly Stereoselective. *Org. Lett.* **14**, 2278–2281 (2012).
246. Horsman, M. E., Hari, T. P. A. & Boddy, C. N. Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: logic gate or a victim of fate? *Nat. Prod. Rep.* **33**, 183–202 (2016).

## Bibliographie

247. Tsai, S.-C. *et al.* Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: Versatility from a unique substrate channel. *Proc. Natl. Acad. Sci.* **98**, 14808–14813 (2001).
248. Giraldes, J. W. *et al.* Structural and mechanistic insights into polyketide macrolactonization from polyketide-based affinity labels. *Nat. Chem. Biol.* **2**, 531–536 (2006).
249. Zhou, Y. *et al.* Structural and Mechanistic Insights into Chain Release of the Polyene PKS Thioesterase Domain. *ACS Catal.* **12**, 762–776 (2022).
250. Claxton, H. B., Akey, D. L., Silver, M. K., Admiraal, S. J. & Smith, J. L. Structure and Functional Analysis of RifR, the Type II Thioesterase from the Rifamycin Biosynthetic Pathway. *J. Biol. Chem.* **284**, 5021–5029 (2009).
251. Curran, S. C. *et al.* Structure and Function of BorB, the Type II Thioesterase from the Borrelidin Biosynthetic Gene Cluster. *Biochemistry* **59**, 1630–1639 (2020).
252. Whicher, J. R. *et al.* Structure and Function of the RedJ Protein, a Thioesterase from the Prodiginine Biosynthetic Pathway in *Streptomyces coelicolor*. *J. Biol. Chem.* **286**, 22558–22569 (2011).
253. Weissman, K. J. & Müller, R. Protein–Protein Interactions in Multienzyme Megasynthetases. *ChemBioChem* **9**, 826–848 (2008).
254. Miyanaga, A. *et al.* Structural Basis of Protein–Protein Interactions between a trans-Acting Acyltransferase and Acyl Carrier Protein in Polyketide Disorazole Biosynthesis. *J. Am. Chem. Soc.* **140**, 7970–7978 (2018).
255. Feng, Y. *et al.* Structural visualization of transient interactions between the cis-acting acyltransferase and acyl carrier protein of the salinomycin modular polyketide synthase. *Acta Crystallogr. Sect. Struct. Biol.* **78**, 779–791 (2022).
256. Kapur, S., Chen, A. Y., Cane, D. E. & Khosla, C. Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci.* **107**, 22066–22071 (2010).
257. Miyanaga, A., Ouchi, R., Kudo, F. & Eguchi, T. Complex structure of the acyltransferase

- VinK and the carrier protein VinL with a pantetheine cross-linking probe. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **77**, 294–302 (2021).
258. Moretto, L., Vance, S., Heames, B. & Broadhurst, R. W. Dissecting how modular polyketide synthase ketoreductases interact with acyl carrier protein-attached substrates †Electronic supplementary information (ESI) available: Complete experimental details and additional figures. See DOI: 10.1039/c7cc04625a. *Chem. Commun. Camb. Engl.* **53**, 11457–11460 (2017).
259. Nguyen, C. *et al.* Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **505**, 427–431 (2014).
260. Dodge, G. J., Maloney, F. P. & Smith, J. L. Protein-protein interactions in 'cis-AT' polyketide synthases. *Nat. Prod. Rep.* **35**, 1082–1096 (2018).
261. Tran, L., Broadhurst, R. W., Tosin, M., Cavalli, A. & Weissman, K. J. Insights into Protein-Protein and Enzyme-Substrate Interactions in Modular Polyketide Synthases. *Chem. Biol.* **17**, 705–716 (2010).
262. Tran, L., Tosin, M., Spencer, J. B., Leadlay, P. F. & Weissman, K. J. Covalent Linkage Mediates Communication between ACP and TE Domains in Modular Polyketide Synthases. *ChemBioChem* **9**, 905–915 (2008).
263. Maloney, F. P., Gerwick, L., Gerwick, W. H., Sherman, D. H. & Smith, J. L. Anatomy of the  $\beta$ -branching enzyme of polyketide biosynthesis and its interaction with an acyl-ACP substrate. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10316–10321 (2016).
264. Maier, T., Jenni, S. & Ban, N. Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* **311**, 1258–1262 (2006).
265. Maier, T., Leibundgut, M. & Ban, N. The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science* **321**, 1315–1322 (2008).
266. Weissman, K. J. Uncovering the structures of modular polyketide synthases. *Nat. Prod. Rep.* **32**, 436–453 (2015).
267. Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512–517 (2014).
268. Edwards, A. L., Matsui, T., Weiss, T. M. & Khosla, C. Architectures of Whole-Module and

## Bibliographie

- Bimodular Proteins from the 6-Deoxyerythronolide B Synthase. *J. Mol. Biol.* **426**, 2229–2245 (2014).
269. Whicher, J. R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560–564 (2014).
270. Bagde, S. R., Mathews, I. I., Fromme, J. C. & Kim, C.-Y. Modular polyketide synthase contains two reaction chambers that operate asynchronously. *Science* **374**, 723–729 (2021).
271. Cogan, D. P. *et al.* Mapping the catalytic conformations of an assembly-line polyketide synthase module. *Science* **374**, 729–734 (2021).
272. Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **11**, 660–670 (2015).
273. Davison, J. *et al.* Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module. *Chem. Sci.* **5**, 3081–3095 (2014).
274. Dorival, J. *et al.* Characterization of Intersubunit Communication in the Virginiamycin trans-Acyl Transferase Polyketide Synthase. *J. Am. Chem. Soc.* **138**, 4155–4167 (2016).
275. Skiba, M. A. *et al.* A Mononuclear Iron-Dependent Methyltransferase Catalyzes Initial Steps in Assembly of the Apratoxin A Polyketide Starter Unit. *ACS Chem. Biol.* **12**, 3039–3048 (2017).
276. Skiba, M. A. *et al.* Biosynthesis of t-Butyl in Apratoxin A: Functional Analysis and Architecture of a PKS Loading Module. *ACS Chem. Biol.* **13**, 1640–1650 (2018).
277. Wang, J. *et al.* C–N bond formation by a polyketide synthase. *Nat. Commun.* **14**, 1319 (2023).
278. Smith, H. G., Beech, M. J., Lewandowski, J. R., Challis, G. L. & Jenner, M. Docking domain-mediated subunit interactions in natural product megasynth(et)ases. *J. Ind. Microbiol. Biotechnol.* **48**, kuab018 (2021).
279. Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F. & Weissman, K. J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723–731 (2003).

280. Buchholz, T. J. *et al.* Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.* **4**, 41–52 (2009).
281. Risser, F. *et al.* Towards improved understanding of intersubunit interactions in modular polyketide biosynthesis: Docking in the enacyloxin IIa polyketide synthase. *J. Struct. Biol.* **212**, 107581 (2020).
282. Whicher, J. R. *et al.* Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem. Biol.* **20**, 1340–1351 (2013).
283. Zeng, J. *et al.* Portability and Structure of the Four-Helix Bundle Docking Domains of trans-Acyltransferase Modular Polyketide Synthases. *ACS Chem. Biol.* **11**, 2466–2474 (2016).
284. Jenner, M. *et al.* Mechanism of intersubunit ketosynthase-dehydratase interaction in polyketide synthases. *Nat. Chem. Biol.* **14**, 270–275 (2018).
285. Richter, C. D., Nietlispach, D., Broadhurst, R. W. & Weissman, K. J. Multienzyme docking in hybrid megasynthetases. *Nat. Chem. Biol.* **4**, 75–81 (2008).
286. Dowling, D. P. *et al.* Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl. Acad. Sci.* **113**, 12432–12437 (2016).
287. Kosol, S. *et al.* Structural basis for chain release from the enacyloxin polyketide synthase. *Nat. Chem.* **11**, 913–923 (2019).
288. Hacker, C. *et al.* Structure-based redesign of docking domain interactions modulates the product spectrum of a rhabdopeptide-synthesizing NRPS. *Nat. Commun.* **9**, 4366 (2018).
289. Keatinge-Clay, A. T. The Uncommon Enzymology of Cis-Acyltransferase Assembly Lines. *Chem. Rev.* **117**, 5334–5366 (2017).
290. Skiba, M. A. *et al.* Domain Organization and Active Site Architecture of a Polyketide Synthase C-methyltransferase. *ACS Chem. Biol.* **11**, 3319–3327 (2016).
291. Skiba, M. A. *et al.* Structural Basis of Polyketide Synthase O-Methylation. *ACS Chem. Biol.* **13**, 3221–3228 (2018).
292. Gay, D. C., Spear, P. J. & Keatinge-Clay, A. T. A double-hotdog with a new trick: structure and mechanism of the trans-acyltransferase polyketide synthase enoyl-isomerase. *ACS*

## Bibliographie

- Chem. Biol.* **9**, 2374–2381 (2014).
293. Wagner, D. T. *et al.* Structural and Functional Studies of a Pyran Synthase Domain from a trans-Acyltransferase Assembly Line. *ACS Chem. Biol.* **13**, 975–983 (2018).
294. Fraley, A. E. *et al.* Heterocomplex structure of a polyketide synthase component involved in modular backbone halogenation. *Struct. Lond. Engl.* **1993** **31**, 565-572.e4 (2023).
295. Bretschneider, T. *et al.* Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* **502**, 124–128 (2013).
296. Rudolf, J. D., Chang, C.-Y., Ma, M. & Shen, B. Cytochromes P450 for natural product biosynthesis in Streptomyces: sequence, structure, and function. *Nat. Prod. Rep.* **34**, 1141–1172 (2017).
297. Devine, A. J. *et al.* The Role of Cytochrome P450AbyV in the Final Stages of Abyssomicin C Biosynthesis. *Angew. Chem. Int. Ed Engl.* **62**, e202213053 (2023).
298. Richter, M. E. A., Traitcheva, N., Knüpfer, U. & Hertweck, C. Sequential asymmetric polyketide heterocyclization catalyzed by a single cytochrome P450 monooxygenase (AurH). *Angew. Chem. Int. Ed Engl.* **47**, 8872–8875 (2008).
299. Moebius, N. *et al.* Biosynthesis of the Respiratory Toxin Bongkrekic Acid in the Pathogenic Bacterium *Burkholderia gladioli*. *Chem. Biol.* **19**, 1164–1174 (2012).
300. Partida-Martinez, L. P. & Hertweck, C. A Gene Cluster Encoding Rhizoxin Biosynthesis in “*Burkholderia rhizoxina*”, the Bacterial Endosymbiont of the Fungus *Rhizopus microsporus*. *ChemBioChem* **8**, 41–45 (2007).
301. Seo, J.-W. *et al.* Comparative Characterization of the Lactimidomycin and iso-Migrastatin Biosynthetic Machineries Revealing Unusual Features for Acyltransferase-less Type I Polyketide Synthases and Providing an Opportunity To Engineer New Analogues. *Biochemistry* **53**, 7854–7865 (2014).
302. Moldenhauer, J., Chen, X.-H., Borriss, R. & Piel, J. Biosynthesis of the Antibiotic Bacillaene, the Product of a Giant Polyketide Synthase Complex of the trans-AT Family. *Angew. Chem. Int. Ed.* **46**, 8195–8197 (2007).
303. Huijbers, M. M. E., Montersino, S., Westphal, A. H., Tischler, D. & van Berkel, W. J. H.

- Flavin dependent monooxygenases. *Arch. Biochem. Biophys.* **544**, 2–17 (2014).
304. van Berkel, W. J. H., Kamerbeek, N. M. & Fraaije, M. W. Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J. Biotechnol.* **124**, 670–689 (2006).
305. Kubitza, C. *et al.* Crystal structure of pyrrolizidine alkaloid N-oxygenase from the grasshopper *Zonocerus variegatus*. *Acta Crystallogr. Sect. Struct. Biol.* **74**, 422–432 (2018).
306. Zhou, Q. *et al.* A Flavin-Dependent Monooxygenase Mediates Divergent Oxidation of Rifamycin. *Org. Lett.* **23**, 2342–2346 (2021).
307. Ueoka, R. *et al.* Genome Mining of Oxidation Modules in trans-Acyltransferase Polyketide Synthases Reveals a Culturable Source for Lobatamides. *Angew. Chem. Int. Ed Engl.* **59**, 7761–7765 (2020).
308. Meoded, R. A. *et al.* A Polyketide Synthase Component for Oxygen Insertion into Polyketide Backbones. *Angew. Chem. Int. Ed Engl.* **57**, 11644–11648 (2018).
309. Kačar, D. *et al.* Identification of trans-AT polyketide clusters in two marine bacteria reveals cryptic similarities between distinct symbiosis factors. *Environ. Microbiol.* **23**, 2509–2521 (2021).
310. Ueoka, R., Bortfeld-Miller, M., Morinaka, B. I., Vorholt, J. A. & Piel, J. Toblerols: Cyclopropanol-Containing Polyketide Modulators of Antibiosis in *Methylobacteria*. *Angew. Chem.* **130**, 989–993 (2018).
311. Hemmerling, F. *et al.* Modular Halogenation,  $\alpha$ -Hydroxylation, and Acylation by a Remarkably Versatile Polyketide Synthase. *Angew. Chem. Int. Ed Engl.* **61**, e202116614 (2022).
312. Zhu, Y. *et al.* Insights into Caerulomycin A Biosynthesis: A Two-Component Monooxygenase CrmH-Catalyzed Oxime Formation. *J. Am. Chem. Soc.* **135**, 18750–18753 (2013).
313. Skiba, M. A. *et al.* PKS–NRPS Enzymology and Structural Biology: Considerations in Protein Production. *Methods Enzymol.* **604**, 45–88 (2018).
314. Huang, Y. & Zhang, L. An In Vitro Single-Primer Site-Directed Mutagenesis Method for

## Bibliographie

- Use in Biotechnology. *Methods Mol. Biol.* **1498**, 375–383 (2017).
315. Kim, H.-A., Seo, J.-K., Kim, T. & Lee, B.-T. Nanometrology and its perspectives in environmental research. *Environ. Health Toxicol.* **29**, (2014).
316. Audic, S., Lopez, F., Claverie, J. M., Poirot, O. & Abergel, C. SAmBA: an interactive software for optimizing the design of biological macromolecules crystallization experiments. *Proteins* **29**, 252–257 (1997).
317. Armstrong, N., Lencastre, A. D. & Gouaux, E. A new protein folding screen: Application to the ligand binding domains of a glutamate and kainate receptor and to lysozyme and carbonic anhydrase. *Protein Sci.* **8**, 1475–1483 (1999).
318. Lindwall, G., Chau, M., Gardner, S. R. & Kohlstaedt, L. A. A sparse matrix approach to the solubilization of overexpressed proteins. *Protein Eng.* **13**, 67–71 (2000).
319. Jancarik, J., Pufan, R., Hong, C., Kim, S. H. & Kim, R. Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1670–1673 (2004).
320. Leibly, D. J. *et al.* Stabilizing Additives Added during Cell Lysis Aid in the Solubilization of Recombinant Proteins. *PLOS ONE* **7**, e52482 (2012).
321. Wang, Y. *et al.* A Systematic Protein Refolding Screen Method using the DGR Approach Reveals that Time and Secondary TSA are Essential Variables. *Sci. Rep.* **7**, 9355 (2017).
322. Wang, H. *et al.* CrysAlis: an integrated server for computational analysis and design of protein crystallization. *Sci. Rep.* **6**, 21383 (2016).
323. Elbasir, A. *et al.* DeepCrystal: a deep learning framework for sequence-based protein crystallization prediction. *Bioinformatics* **35**, 2216–2225 (2019).
324. Elbasir, A. *et al.* BCrystal: an interpretable sequence-based protein crystallization predictor. *Bioinforma. Oxf. Engl.* **36**, 1429–1438 (2020).
325. Wang, H. *et al.* Critical evaluation of bioinformatics tools for the prediction of protein crystallization propensity. *Brief. Bioinform.* **19**, 838–852 (2017).
326. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **70**, 2–20 (2013).



327. Chayen, N. E. Turning protein crystallisation from an art into a science. *Curr. Opin. Struct. Biol.* **14**, 577–583 (2004).
328. Chayen, N. E. & Saridakis, E. Protein crystallization: from purified protein to diffraction-quality crystal. *Nat. Methods* **5**, 147–153 (2008).
329. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 (1968).
330. Garman, E. F. & Schneider, T. R. Macromolecular Cryocrystallography. *J. Appl. Crystallogr.* **30**, 211–237 (1997).
331. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
332. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133–144 (2010).
333. Rossmann, M. G. & Blow, D. M. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.* **15**, 24–31 (1962).
334. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25 (2010).
335. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
336. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
337. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
338. Casañal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci. Publ. Protein Soc.* **29**, 1069–1078 (2020).
339. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
340. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
341. Röttig, M. *et al.* NRPSpredictor2—a web server for predicting NRPS adenylation domain

## Bibliographie

- specificity. *Nucleic Acids Res.* **39**, W362–W367 (2011).
342. Holm, L. Using Dali for Protein Structure Comparison. in *Structural Bioinformatics: Methods and Protocols* (ed. Gáspári, Z.) 29–42 (Springer US, 2020). doi:10.1007/978-1-0716-0270-6\_3.
343. Yan, Y., Zhang, D., Zhou, P., Li, B. & Huang, S.-Y. HDock: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* **45**, W365–W373 (2017).
344. Minas, H. A. *et al.* Modular Oxime Formation by a trans-AT Polyketide Synthase. *Angew. Chem. Int. Ed.* **62**, e202304481 (2023).
345. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
346. Combet, C., Blanchet, C., Geourjon, C. & Deléage, G. NPS@: network protein sequence analysis. *Trends Biochem. Sci.* **25**, 147–150 (2000).
347. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
348. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinforma. Oxf. Engl.* **16**, 404–405 (2000).
349. Romero, P., Obradovic, Z. & Dunker, A. K. Natively Disordered Proteins. *Appl. Bioinformatics* **3**, 105–113 (2004).
350. Mareš, J., Hájek, J., Urajová, P., Kopecký, J. & Hrouzek, P. A Hybrid Non-Ribosomal Peptide/Polyketide Synthetase Containing Fatty-Acyl Ligase (FAAL) Synthesizes the  $\beta$ -Amino Fatty Acid Lipopeptides Puwainaphycins in the Cyanobacterium *Cylindrospermum alatosporum*. *PLoS ONE* **9**, e111904 (2014).
351. Fisher, A. J., Thompson, T. B., Thoden, J. B., Baldwin, T. O. & Rayment, I. The 1.5-Å Resolution Crystal Structure of Bacterial Luciferase in Low Salt Conditions \*. *J. Biol. Chem.* **271**, 21956–21968 (1996).
352. Alali, A. *et al.* Biosynthesis of the Tricyclic Aromatic Type II Polyketide Rishirilide: New Potential Third Ring Oxygenation after Three Cyclization Steps. *Mol. Biotechnol.* **63**, 502–

- 514 (2021).
353. Sibbesen, O., Koch, B., Halkier, B. A. & Møller, B. L. Cytochrome P-450TYR Is a Multifunctional Heme-Thiolate Enzyme Catalyzing the Conversion of L-Tyrosine to p-Hydroxyphenylacetaldehyde Oxime in the Biosynthesis of the Cyanogenic Glucoside Dhurrin in *Sorghum bicolor* (L.) Moench (\*). *J. Biol. Chem.* **270**, 3506–3511 (1995).
354. Kelly, W. L. & Townsend, C. A. Role of the Cytochrome P450 NocL in Nocardicin A Biosynthesis. *J. Am. Chem. Soc.* **124**, 8186–8187 (2002).
355. Garcia, I. *et al.* Engineering the Biosynthesis of the Polyketide-Nonribosomal Peptide Collismycin A for Generation of Analogs with Neuroprotective Activity. *Chem. Biol.* **20**, 1022–1032 (2013).
356. Cortina, N. S., Revermann, O., Krug, D. & Müller, R. Identification and Characterization of the Althiomycin Biosynthetic Gene Cluster in *Myxococcus xanthus* DK897. *ChemBioChem* **12**, 1411–1416 (2011).
357. Hsieh, Y.-C., Chia, T. S., Fun, H.-K. & Chen, C.-J. Crystal Structure of Dimeric Flavodoxin from *Desulfovibrio gigas* Suggests a Potential Binding Region for the Electron-Transferring Partner. *Int. J. Mol. Sci.* **14**, 1667–1683 (2013).
358. Inami, K. & Shiba, T. Syntheses of Althiomycin Analogs in Relation to Antibacterial Activities. *Bull. Chem. Soc. Jpn.* **59**, 2185–2189 (1986).
359. Cornish, V. W., Hahn, K. M. & Schultz, P. G. Site-Specific Protein Modification Using a Ketone Handle. *J. Am. Chem. Soc.* **118**, 8150–8151 (1996).
360. Rabuka, D., Rush, J. S., deHart, G. W., Wu, P. & Bertozzi, C. R. Site-specific chemical protein conjugation using genetically encoded aldehyde tags. *Nat. Protoc.* **7**, 1052–1067 (2012).
361. Vriens, E. *et al.* Polyketide Synthase-Mediated O-Methyloxime Formation in the Biosynthesis of the Oximidine Anticancer Agents. *Angew. Chem. Int. Ed.* **62**, e202304476 (2023).
362. Kao, C. M., Pieper, R., Cane, D. E. & Khosla, C. Evidence for Two Catalytically Independent Clusters of Active Sites in a Functional Modular Polyketide Synthase.

## Bibliographie

*Biochemistry* **35**, 12363–12368 (1996).

363. Tosin, M., Spiteller, D. & Spencer, J. B. Malonyl carba(dethia)- and Malonyl oxa(dethia)-coenzyme A as Tools for Trapping Polyketide Intermediates. *ChemBioChem* **10**, 1714–1723 (2009).



Les polycétides sont des métabolites spécialisés synthétisés par une grande variété d'organismes incluant les bactéries, les champignons, les végétaux ou encore certains animaux. Bien que leurs fonctions écologiques demeurent souvent mystérieuses, leurs activités biologiques permettent de les exploiter comme médicaments chez l'homme et l'animal (antibiotiques, anthelminthiques, anticancéreux, etc.). Si, dans certains cas, les molécules naturelles peuvent être utilisées, dans d'autres cas, il est nécessaire d'adapter les structures des métabolites pour optimiser leurs propriétés médicinales. Une stratégie prometteuse pour obtenir de telles molécules consiste à modifier les enzymes polycétide synthases (PKS) par génie génétique. Cependant, cette approche nécessite une connaissance détaillée du fonctionnement de ces enzymes au niveau moléculaire. Les PKS sont des systèmes complexes de haut poids moléculaire divisés en plusieurs polypeptides appelés sous-unités, eux-mêmes composés de modules, chacun catalysant un cycle d'extension et de modification des polycétides. Les activités de ces modules sont divisées en domaines, chacun ayant un rôle biosynthétique défini.

La découverte de nouvelles fonctions enzymatiques portées par ces domaines élargit la gamme connue des réactions catalysées par les PKS et, par conséquent, la diversité chimique des polycétides synthétisés. En principe, ces activités peuvent ensuite être incorporées dans d'autres PKS, conduisant à la synthèse de dérivés de polycétides aux propriétés potentiellement améliorées. Plusieurs réactions non canoniques intéressantes ont été récemment identifiées dans les PKS synthétisant l'oocydine et la lobatamide A, catalysées par diverses activités monooxygénases. Plus précisément, trois types de monooxygénases ont été identifiés : les monooxygénases de type Baeyer-Villiger (BVMO), les  $\alpha$ -hydroxylases et les monooxygénases génératrices d'oxime (Ox).

Afin d'étayer les efforts visant à exploiter cette chimie dans l'ingénierie, le travail décrit dans cette thèse visait à établir des relations structure-fonction détaillées pour ces enzymes. Le principal accomplissement de ce doctorat a été de caractériser en détail le domaine Ox de la PKS de la lobatamide en utilisant une approche combinant la cristallographie aux rayons X et la mutagenèse dirigée, couplée à la spectrométrie de masse. Sa structure tridimensionnelle a été élucidée à haute résolution et les acides aminés impliqués dans la catalyse enzymatique ont été identifiés, ce qui nous a permis de proposer un mécanisme catalytique. De plus, des expériences de *docking in silico* et de mutagenèse ont fourni des informations supplémentaires sur l'ordre des événements biosynthétiques au sein du module et sur son agencement spatial au cours de la catalyse. Les résultats de cette partie du travail ont été publiés dans la revue à fort impact *Angew. Chemie* en tant qu'article complet.

Des études bioinformatiques préliminaires, suivies de la production et de la purification des BVMO et de l' $\alpha$ -hydroxylase, nous ont permis d'initier les premiers essais de cristallisation de ces enzymes, dans le but de résoudre leurs structures et d'établir les bases structurales de ces activités monooxygénases. Ces tentatives n'ont pas été couronnées de succès, mais des expériences supplémentaires sont proposées.

Mots-clefs : polycétide, PKS, monooxygénase à flavine, cristallographie des protéines, mutagenèse dirigée.

Polyketides are specialized metabolites synthesized by a wide range of organisms, including bacteria, fungi, plants, and some animals. Although their ecological functions remain poorly understood, their biological activities allow them to be exploited as drugs in humans and animals (as e.g. antibiotics, anthelmintics, anticancer agents, etc.). While in some cases the natural molecules can be utilized, in others it is necessary to adapt the structures of the metabolites to optimize their medicinal properties. A promising strategy for obtaining such molecules is to modify the polyketide synthase (PKS) enzymes by genetic engineering. However, this approach requires detailed knowledge of how these enzymes function on the molecular level. PKSs are complex high-molecular-weight systems divided into several polypeptides called subunits, which are themselves composed of modules, each catalyzing a cycle of polyketide extension and modification. The activities of these modules are divided into domains, each with a defined biosynthetic role.

The discovery of new enzymatic functions carried by these domains extends the known range of reactions catalyzed by PKSs and, consequently, the chemical diversity of the polyketides synthesized. In principle, these activities can then be incorporated into other PKSs, leading to the synthesis of polyketide derivatives with potentially improved properties. Several non-canonical reactions of interest have recently been identified in the PKSs synthesizing oocydine and lobatamide A catalyzed by various monooxygenase activities. Specifically, three types of monooxygenases were identified: Baeyer-Villiger type monooxygenases (BVMOs),  $\alpha$ -hydroxylases and oxime-generating monooxygenases (Ox).

In order to underpin efforts to exploit this chemistry in engineering, the work describes in this thesis aimed to establish detailed structure-function relationships for these enzymes. The principal achievement of this PhD was to characterize in detail the Ox domain of the lobatamide PKS using an approach combining X-ray crystallography and site-directed mutagenesis coupled with mass spectrometry. Its three-dimensional structure was elucidated at high resolution and the amino acids involved in enzymatic catalysis were identified, allowing us to propose a catalytic mechanism. In addition, *in silico docking* and mutagenesis experiments provided further information on the order of biosynthetic events within the module and its spatial arrangement during catalysis. The results of this portion of the work were published in the high-impact journal *Angew. Chemie* as a full article.

Preliminary bioinformatics studies, followed by the production and purification of the BVMOs and  $\alpha$ -hydroxylase, enabled us to initiate the first crystallisation trials of these enzymes, with the aim of solving their structures and establishing the structural basis of these monooxygenase activities. These attempts were not successful, but follow-up experiments are proposed.

Keywords: polyketide, PKS, flavin-dependent monooxygenase, protein crystallography, site-directed mutagenesis.