UNIVERSITÉ DE LORRAINE | BIBLIOTHÈQUES UNIVERSITAIRES

AVERTISSEMENT
DROIT D'AUTEUR

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. Toute contrefaçon, plagiat ou reproduction illicite encourt une poursuite pénale.

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

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

UNIVERSITÉ DE LORRAINE
34 COURS LÉOPOLD — CS 25233
54052 NANCY CEDEX
TÉL : 03 54 50 54 00
CONTACT@UNIV-LORRAINE.FR
WWW.UNIV-LORRAINE.FR

Université de Lorraine (UL)
École doctorale Sciences et Ingénierie des Molécules, des Produits, des Procédés et de l'Énergie (SIMPPE)
Laboratoire Réactions et Génie des Procédés (LRGP)

THESE

En vue de l'obtention du titre de
DOCTEUR DE L'UNIVERSITÉ DE LORRAINE
Mention : Procédés biotechnologiques

Présentée par **Laureline GENNESSEAUX**
Sous la direction de **Catherine VIROT HUMEAU** et **Yann GUIAVARC'H**

Production et caractérisation de nouvelles amino-acylases par une approche combinée expérimentale et numérique – Mise en œuvre dans des procédés verts de N-acylation

Présentée et soutenue publiquement le **lundi 1er décembre 2025**

Membres du jury :

Directeurs de thèse	Mme Catherine VIROT HUMEAU	Professeur, Université de Lorraine – LRGP, Vandœuvre-lès-Nancy
	M. Yann GUIAVARC'H	Maître de conférences, HDR, Université de Lorraine – LRGP, Vandœuvre-lès-Nancy
Président du jury	M. Bertrand AIGLE	Professeur, Université de Lorraine – DynAMic, Vandœuvre-lès-Nancy
Rapporteurs	M. Eric DUBREUCQ	Professeur, Université de Montpellier – Institut Agro, Montpellier
	Mme Anne LOMASCOLO	Maître de conférences, HDR, Université de Aix Marseille – BBF, Marseille
Examinateur	M. Mohamed Chafik BOURKAIB	Docteur – Laboratoire LCQ Bourkaib
Membres invités	Mme Isabelle CHEVALOT	Professeur, Université de Lorraine – LRGP, Vandœuvre-lès-Nancy
	M. Stéphane DELAUNAY	Professeur, Université de Lorraine – LRGP, Vandœuvre-lès-Nancy

« *La curiosité mène à tout :*
parfois à écouter aux portes,
parfois à découvrir l'Amérique. »

José Maria Eça de Queiros (1845-1900)

« *Do what you can,*
with what you've got,
where you are. »

Theodore Roosevelt (1898-1960)

Le Covid long provoque des problèmes de santé multiples affectant les capacités cognitives comme physiques : problèmes neurologiques, cardiaques, pulmonaires, digestifs, gynécologiques, articulaires, musculaires, immunologiques, psychologiques, etc.
Prenez soin de vous et de vos proches.

« *Si vous êtes neutre dans les situations d'injustice,*
vous avez choisi le camp de l'agresseur. »

Desmond Tutu (1931-2021)

Remerciements

Au terme de ce travail de thèse, je tiens à exprimer ma profonde gratitude à toutes les personnes qui, de près ou de loin, ont contribué à son aboutissement.

Je tiens tout d'abord à remercier chaleureusement les membres du jury pour avoir accepté d'examiner ce travail. Je remercie M. Eric Dubreucq et Mme Anne Lomascolo pour leur rôle de rapporteurs, ainsi que M. Bertrand Aigle et M. Chafik Bourkaib pour leur présence en tant qu'examinateurs. Je leur suis très reconnaissante d'avoir consacré de leur temps à la lecture de cette thèse et d'avoir accepté de prendre part à son évaluation.

Je souhaite adresser mes remerciements les plus sincères à mes directeurs de thèse, Mme Catherine Humeau et M. Yann Guiavarc'h, pour leur accompagnement tout au long de ces années. Leur confiance, leur rigueur scientifique, leur disponibilité et leurs encouragements ont été essentiels dans la réalisation de ce travail.

Je tiens également à remercier l'ensemble de l'équipe encadrante de la thèse, notamment Mme Isabelle Chevalot, M. Stéphane Delaunay et M. Xavier Framboisier, pour leur soutien technique, leurs conseils avisés et les nombreuses discussions enrichissantes qui ont jalonné mon parcours.

Je pense aussi à mes collègues, proches et moins proches, avec qui j'ai eu le plaisir de partager ce chemin. Merci à Delphine, Eusèbe, Xavier et Pilar pour leur présence quotidienne, leur bonne humeur, les échanges scientifiques (et non scientifiques !) qui ont rythmé nos journées. Merci également à tous les autres membres du laboratoire, passés ou présents, sans citer de nom afin de n'oublier personne, pour leur accueil et l'ambiance conviviale.

Je souhaite également remercier l'ensemble de mes collègues enseignants, avec qui j'ai eu l'honneur de collaborer dans le cadre de mes activités d'enseignement. Leur engagement, leur professionnalisme et la qualité des échanges que nous avons pu entretenir ont grandement contribué à enrichir ma réflexion pédagogique et à faire de cette expérience une composante précieuse de mon parcours doctoral.

Un grand merci à mes deux stagiaires, Loïc et Enzo, pour leur implication, leur motivation et leur contribution au projet. Travailler avec vous a été une expérience très enrichissante, tant sur le plan humain que professionnel.

Je souhaite adresser des remerciements tout particuliers à Boris, pour ton amour, ta patience et ton soutien sans faille. Tu as su m'encourager dans les moments de doute, m'écouter dans les périodes de stress, et célébrer chaque petite victoire à mes côtés. Merci d'avoir cru en moi, parfois plus que moi-même, et d'avoir su me rappeler l'essentiel quand je m'y perdais.

Enfin, je souhaite remercier ma famille et mes amis, anciens et nouveaux, pour leur soutien inconditionnel, leur patience, leurs encouragements dans les moments de doute et leur enthousiasme dans les moments de joie. Merci pour votre présence constante, malgré la distance ou les emplois du temps chargés. Ce travail est aussi le fruit de votre soutien.

À toutes et à tous, merci du fond du cœur.

Table des matières

Index des figures 12

Index des tableaux 18

Liste des abréviations 20

Introduction générale 23

Chapitre I : Etude bibliographique 26

Dérivés acylés d'acides aminés et de peptides 26
Définition et classification 26
Propriétés et activités 28
Propriétés de surface 28
Propriétés d'agrégation 29
Propriétés moussantes et gélifiantes 29
Activités hémolytiques 30
Propriétés antimicrobiennes 30
Biodégradabilité 31
Toxicité 31
Applications de dérivés acylés d'acides aminés 31
Dans les produits d'hygiène et de soin 32
En agriculture 32
En biologie et en médecine 33
En agroalimentaire 34
Dans les autres industries 34
Synthèse chimique et enzymatique en milieux solvant des dérivés acylés d'acides aminés 36
Synthèse chimique 36
Synthèse chimio-enzymatique 38
Synthèse enzymatique en milieu solvant 40
Conclusions 44
Enzymes catalysant l'hydrolyse ou la synthèse de liaisons amides en milieu aqueux 44
Enzymes catalysant l'hydrolyse ou la synthèse de liaisons peptidiques en milieu aqueux 45
Enzymes catalysant la N-acylation/déacylation en milieu aqueux 46
Les aminoacyltransférases 46
Les amidohydrolases 46
Les N-succinyl-L,L-diaminopimelate désuccinylases 47
Les pénicillines acylases (ou pénicillines amidases) 48
Les ε-lysine acylases 49
Les aminoacylases I 50
Les aminoacylases de *Streptomyces ambofaciens* 53
Structure des aminoacylases 59
Mécanismes catalytiques 62

Mécanisme catalytique sans ion métallique 62
Mécanismes catalytiques avec ion(s) métallique(s) 63
Conclusions 70
Production hétérologue des aminoacylases 71
Escherichia coli 72
E. coli K12 75
E. coli BL21 (DE3) 76
E. coli Origami B (DE3) 78
E. coli Rosetta (DE3) et Rosetta 2 (DE3) 82
E. coli BL21 Gold (DE3) 86
E. coli Arctic Express (DE3) 87
Refolding 90
Streptomyces lividans 91
Vibrio natriegens 94
Komagataella phaffi (syn. *Pichia Pastoris*) 95
Cellules d'insectes 95
Conclusions 96
Modélisation moléculaire appliquée à la biocatalyse 97
Modélisation par homologie 98
Identification des templates 99
Alignements de séquences et corrections d'alignement 100
Modélisation de la structure tridimensionnelle des protéines 101
Validation de la structure tridimensionnelle modélisée 104
Optimisation du modèle 106
Simulations de docking moléculaire 107
Conclusions 109

Chapitre 2 : Matériels et méthodes 111

Etude numérique : Modélisation moléculaire 111
Identification des templates 111
Construction des modèles par modélisation par homologie 112
Modeller (Discovery Studio®) 112
Swiss Model 113
ColabFold (AlphaFold2) 113
AlphaFold3 114
Validation des modèles d'homologie 114
Scoring 114
Etude de la conservation des éléments de spécificité 115
Optimisation des modèles d'homologie 115
Ajout des ions métalliques et de la molécule d'eau catalytique 116
Solvatation et minimisation d'énergie progressive 116
Simulations de docking et scoring 117
Structure des ligands 117
Définition du site actif 117
Simulations de docking 117
Analyses des poses de docking 118
Etude expérimentale : Production par expression hétérologue 119

Systèmes d'expression choisis et milieux de culture 119
Transformations bactériennes d'*E. coli* 124
Cultures bactériennes 125
Escherichia coli 126
Streptomyces lividans 128
Obtention des spores 128
En erlenmeyers 128
En bioréacteur 128
Extraction et lyse cellulaire 129
Protocole de refolding 130
Dosages protéiques 130
Dosage de Bradford 130
Dosage BCA 131
Concentration par ultrafiltration 131
Electrophorèse SDS-PAGE 131
Synthèse enzymatique 132
Hydrolyse enzymatique 133
Méthodes analytiques 133
Chromatographie sur couche mince (CCM) 133
Analyse HPLC 133
Purification de la protéine recombinante SamAA 134
Système seringue 134
ÄKTA Pure 135
Dialyse – Concentration protéique 135
Dot Blot 135

Chapitre 3 : Modélisation de la synthèse et de l'hydrolyse de dérivés N-acylés de la L-lysine catalysées par SamAA 137

Introduction 137
Etude par modélisation moléculaire de la structure de l'aminoacylase SamAA et sa sélectivité envers la N-acylation de la L-lysine 138
Introduction 140
Materiels and methods 143
Search for templates 143
Homology modelling 143
Evaluation of homology models 144
Optimisation of the model 144
Docking simulations and scoring 145
Results and discussion 145
Identification of the templates and main structural features 146
Protein sequence analysis 147
Construction of homology models 150
Structural evaluation of the model 152
Description of SamAA predicted 3D structure 154
Docking results 159
Hydrolysis reaction 160
Docking of N-α-acetyl-L-lysine 160

Docking of N-ε-acetyl-L-lysine 161
Docking of N-α-caproyl-L-lysine 162
Docking of N-ε-caproyl-L-lysine 163
Synthesis reaction 164
Docking of acyl donor first 164
Docking of L-lysine within SamAA in complex with acyl donor substrate 165
Docking of L-lysine in first 166
Docking of acyl donor substrate within SamAA catalytic pocket in complex with L-lysine 167
Proposed catalytic mechanism of N-α-caproyl-L-lysine hydrolysis or synthesis 169
Conclusions 171
Contribution de l'article 172
Résultats complémentaires 174
Construction du modèle de la structure de SamAA avec Modeller 174
Optimisation de l'alignement de séquences 175
Optimisation de la structure du modèle de SamAA 176
Conclusions 177
Modélisation de la synthèse/hydrolyse de la N-lauroyl-L-lysine 178
Etude de l'hydrolyse de la N-lauroyl-L-lysine 179
Etude de la synthèse de la N-lauroyl-L-lysine 182
Docking de l'acide laurique puis de la L-lysine 182
Docking de la L-lysine puis de l'acide laurique 183
Conclusions 185
Prédiction de la structure de SamAA en conformation ouverte 188
Description de la structure tridimensionnelle de SamAA modélisée sur la base du template 3IC1 188
Description de la structure de SamAA modélisée par AlphaFold3 190
Résultats de docking avec le modèle de SamAA obtenu avec AlphaFold3 191
Etude de l'hydrolyse de la N-lauroyl-L-lysine 191
Etude de la synthèse de la N-lauroyl-L-lysine 194
Docking de l'acide laurique puis de la L-lysine 194
Docking de la L-lysine puis de l'acide laurique 195
Conclusions 198

Chapitre IV : Modélisation de la N-acylation de la L-lysine catalysée par SamELA 200

Introduction 200
Identification des templates de SamELA et étude de leurs principales caractéristiques structurelles 201
Analyse de la séquence protéique de SamELA 202
Modélisation de la structure tridimensionnelle de SamELA 207
Description de la structure de SamELA prédite par ColabFold 207
Description de la structure de SamELA prédite par AlphaFold3 209
Modélisation de la synthèse de la N-capryl- et de la N-lauroyl-L-lysine 210
Conclusions 214

Chapitre V : Production hétérologue des deux aminoacylases de *Streptomyces ambofaciens*, SamAA et SamELA 216

Introduction 216
Production hétérologue de SamAA 218
Production chez *Escherichia coli* 218
Production avec *E. coli* Origami B (DE3) 220
Production avec *E. coli* Rosetta 2 pLysS (DE3) 223
Production avec *E. coli* Arctic Express (DE3) 225
Production avec *E. coli* Rosetta 2 (DE3) 229
Production avec *E. coli* BL21 Gold pLysS (DE3) 234
Production avec *E. coli* BL21 Gold (DE3) 237
Conclusions de la production de SamAA chez *E. coli* 242
Production chez *Streptomyces lividans* 242
Valorisation des corps d'inclusion de SamAA 245
Purification de SamAA recombinante 247
Production hétérologue de SamELA chez *Escherichia coli* 250
Conclusions sur la production de SamAA et de SamELA 254

Conclusion générale et perspectives 256

Références bibliographiques 263

Annexes 298

Résumé de la thèse 327

Thesis abstract 329

Index des figures

Chapitre I

Figure I.1 : Classification des dérivés acylés à base d'acides aminés (Bourkaib, 2020).
Figure I.2 : Utilisation d'un géminoïde pour le transfert d'un fragment nucléotidique au sein d'une cellule (Damen et al., 2010).
Figure I.3 : Voies principales de synthèse chimique des dérivés d'acides aminés.
Figure I.4 : A : Réaction chimique de Schotten-Baumann ; B : Production des dérivés N-acylés selon la réaction chimique de Schotten-Baumann à l'échelle industrielle.
Figure I.5 : Exemple de structure des tensio-actifs cationiques gémini à base d'arginine (Piera et al., 2000).
Figure I.6 : Réaction catalysée par la désuccinylase de l'acide N-succinyl-L,L-diaminopimélique (Bienvenue et al., 2003).
Figure I.7 : Taux de conversion des acides aminés lors de la N-acylation par l'acide laurique (Koreishi et al., 2006).
Figure I.8 : Rendements de synthèse de dérivés N-acylés d'acides aminés catalysée par l'aminoacylase de *P. monticola* (Haeger et al., 2024a).
Figure I.9 : Activité spécifique du surnageant de culture de *S. ambofaciens* pour l'hydrolyse de la N-α- ou N-ε-acétyl-lysine, au cours du temps (Ferrari, 2014).
Figure I.10 : Rendements de conversion pour l'acylation de différents acides aminés par l'acide undécénoïque (Bourkaib et al., 2020a).
Figure I.11 : Pourcentage de conversion des acides gras lors de la N-acylation de la lysine. (Bourkaib, 2020).
Figure I.12 : Exemples de stratégies d'immobilisation d'enzymes (inspirée de Sheldon et al., 2021).
Figure I.13 : Structures tridimensionnelles des monomères constitutifs des aminoacylases (Hsieh et al., 2010).
Figure I.14 : Proposition de mécanisme de l'hydrolyse de la pénicilline G en acide phénylacétique et en acide 6-aminopenicillianique, catalysée par la pénicilline G acylase (Srirangan et al., 2013).
Figure I.15 : Changement conformationnel observé chez l'enzyme HiDapE (Nocek et al., 2018).
Figure I.16 : Proposition de mécanisme catalytique pour l'hydrolyse du L,L-SDAP par la succinyl diaminopimelate desuccinylase d'*H. influenzae* (Nocek et al., 2018).
Figure I.17 : Proposition du mécanisme catalytique des peptidases M38 (Cummings et al., 2010).
Figure I.18 : Proposition du mécanisme catalytique d'hydrolyse de l'aspartoacylase humaine appartenant à la famille des peptidases M14 (Le Coq et al., 2008).
Figure I.19 : Système de production de protéine hétérologue en utilisant *E. coli* et un plasmide pET (Bourgeois, 2011).
Figure I.20 : Analyse par SDS-PAGE de l'aminoacylase recombinante de *S. maltophilia*, après purification (Cao et al., 2020).
Figure I.21 : Analyse par SDS-PAGE des protéines issues des souches *E. coli* Origami B (DE3) (A) et *E. coli* BL21 (DE3) (B) (George, 2017).

Figure I.22 : Analyse par SDS-PAGE des protéines issues des cultures d'*E.coli* Origami B (DE3) non transformée (A), transformée avec le plasmide pET-15b/SamELA (B), transformée avec le plasmide pET-15b/SamPVA (C), et transformée avec le plasmide pET-15b/SamAA (D) (Nili, 2018).
Figure I.23 : Représentation graphique de « l'espace d'utilisation des codons » (Gustafsson et al., 2004).
Figure I.24 : Schémas des plasmides pRARE et pRARE 2 permettant la surexpression des ARNt pour les codons AUA (*IleX tRNA*), AGG (*ArgW tRNA*), AGA (*ArgU tRNA*), CUA (*LeuW tRNA*), CCC (*ProL tRNA*), GGA (*GlyT RNAt*) et CGG (*ArgX tRNA*) (Novagen).
Figure I.25 : Analyse SDS-PAGE des différentes fractions obtenues lors de la purifiation de PmAcy (Haeger et al, 2024a).
Figure I.26 : Analyse SDS/PAGE des fractions obtenues lors de la purification des aminoacylases recombinantes de *S. griseus* (Haeger et al., 2023a).
Figure I.27 : Les sept étapes de la modélisation par homologie (Murmu et al., 2022).

Chapitre II

Figure II.1 : Représentation schématique du site actif de la succinyl-diaminopimelate desuccinylase de *H. influenzae* mentionnant les distances entre les différents éléments (Nocek et al., 2010 ; Nocek et al., 2018).
Figure II.2 : Schéma du plasmide pET-15b (Novagen).
Figure II.3 : Schéma du plasmide pET-28a (Novagen).
Figure II.4 : Récapitulatif des paramètres et des conditions étudiés pour la sur-expression de SamAA sous une forme soluble et active chez *E. coli*.

Chapitre III

Figure III.1: N-acylation of L-lysine; A: N-α-acylation and B: N-ε-acylation.
Figure III.2: Multiple sequence alignment of SamAA and templates.
Figure III.3: A and B: molecular overlay of the best model obtained with Modeller and HiDapE structure, B: focus on the metal ion binding sites, C and D: molecular overlay of the best model obtained with Swiss Model from hCN2 as a template and hCN2 structure, D: focus on the metal ion binding sites, E and F: molecular overlay of the best model obtained with Swiss Model from hiDapE as a template and HiDapE structure, F: focus on the metal ion binding sites, G and H: molecular overlay of the best model obtained with ColabFold and HiDapE structure, H: focus on the metal ion binding sites.
Figure III.4: A: Predicted 3D structure of SamAA in its dimeric form. B: Predicted architecture of SamAA active site with the catalytic water molecule, the catalytic residue, the metal ion-binding residues and the two Zn^{2+} ions ; C: Van der Walls-type surface of all on atoms present in a 10 Å sphere around the catalytic water molecule showing the catalytic cavity of the monomer.
Figure III.5: Graphical representation of the secondary structure of SamAA.
Figure III.6: Preferential orientations of N-α-acetyl-L-lysine (A) and N-ε-acetyl-L-lysine (B) within SamAA catalytic pocket.
Figure III.7: Some poses from docking of N-α-caproyl-L-lysine (A) and N-ε-caproyl-L-lysine (B) where the amide bond to be hydrolyzed comes near SamAA catalytic machinery.

Figure III.8: Main orientations and interactions of acyl donors acetate (A) and caproate (B) within SamAA active site.
Figure III.9: Main orientations and interactions of L-lysine within SamAA active site in complex with acyl donor substrates acetate (A) or caproate (B).
Figure III.10: L-lysine in interaction with SamAA active site.
Figure III.11: Main orientations and interactions of the acyl donors acetate (A) and caproate (B) within SamAA active site in complex with L-lysine.
Figure III.12: Proposal for SamAA catalytic mechanism during N-α-caproyl-L-lysine hydrolysis.
Figure III.13: Proposal for SamAA catalytic mechanism leading to the synthesis of N-α-caproyl-L-lysine.
Figure III.14 : Architecture du site actif prédit de SamAA avec la molécule d'eau catalytique, les deux ions métalliques, le résidu catalytique acide-base et les résidus fixateurs des ions métalliques.
Figure III.15 : Orientations préférentielles de la N-α-lauroyl-L-lysine dans le site actif de SamAA.
Figure III.16 : Orientations préférentielles de la N-ε-lauroyl-L-lysine dans le site actif de SamAA.
Figure III.17 : A : Orientation préférentielle de l'acide laurique dans le site actif de SamAA. B et C : Exemples de deux orientations préférentielles de la L-lysine dans le site actif de SamAA complexé avec le donneur d'acyle lauroate.
Figure III.18 : Orientation et interactions de la L-lysine dans le site actif de SamAA.
Figure III.19 : Orientations préférentielles (A et B) et interactions (C) de l'acide laurique dans le site actif de SamAA.
Figure III.20 : Comparaison des structures de la succinyl diaminopimelate de *H. influenzae* dans sa conformation « ouverte » ou « fermée ». (A) et (B) : Diagramme en ruban de la structure de l'enzyme dans sa conformation « ouverte » (entrée PDB 3IC1) et « fermée » en interaction avec les produits acide succinique et acide L-diaminopimélique (entrée PDB 5VO3). (C) et (D) Mise en évidence des changements conformationnels par superposition des domaines catalytiques de dimérisation des monomères dans la structure ouverte ou fermée. (E) : Zoom sur le passage d'une position distale à une position proximale du résidu His194 après l'arrivée du substrat dans le site actif (Nocek et al., 2018).
Figure III.21 : A : Superposition des structures 3D de SamAA prédite par ColabFold et Modeller en utilisant 3IC1 comme template. B : Site actif avec la molécule d'eau catalytique, le résidu catalytique et les résidus fixateurs des ions métalliques dans le modèle de SamAA prédit par Modeller sur la base de 3IC1 comme template.
Figure III.22 : A : Superposition des structures 3D des modèles de SamAA issus de ColabFold et AlphaFold3. B : Site actif avec la molécule d'eau catalytique, le résidu catalytique et les résidus fixateurs des ions métalliques du modèle de SamAA généré par AlphaFold3.
Figure III.23 (A et B) : Orientations préférentielles de la N-α-lauroyl-L-lysine dans le site actif de SamAA.
Figure III.24 (A, B et C) : Orientations préférentielles de la N-ε-lauroyl-L-lysine dans le site actif de SamAA.
Figure III.25 : Orientation préférentielle de l'acide laurique dans le site actif de SamAA, choisie comme cible de docking de la L-lysine.
Figure III.26 (A et B) : Orientations préférentielles de la L-lysine dans le complexe SamAA-acide laurique.

Figure III.27 : A : Orientation préférentielle de la L-lysine dans le site actif de SamAA. B : Exemple d'une orientation de l'acide laurique au sein du complexe SamAA-L-lysine.

Chapitre IV

Figure IV.1 : Alignement des séquences de SamELA et du template PfH (Robert et al., 2014).
Figure IV.2 : Alignement des séquences de SamELA et de EcIAD (Robert et al., 2014).
Figure IV.3 : A : Structure tridimensionnelle monomérique de SamELA prédite par ColabFold. B : Site actif de SamELA prédit par ColabFold, avec la molécule d'eau catalytique, le résidu catalytique, les résidus fixateurs des ions métalliques et les deux ions métalliques.
Figure IV.4 : A : Superposition des structures tridimensionnelles de SamELA prédites par ColabFold et par AlphaFold3. B : Site actif de SamELA prédit par AlphaFold3, avec la molécule d'eau catalytique, le résidu catalytique, les résidus fixateurs des ions métalliques et les deux ions métalliques.
Figure IV.5 : Orientations préférentielles de la L-lysine dans la cavité catalytique de SamELA. A : orientation ε dans laquelle la fonction amine ε de la chaine latérale pointe vers le site actif. B : orientation α où la fonction amine α de la L-lysine est dirigée vers le site actif.
Figure IV.6 : Orientations préférentielles de l'acide caprylique dans la cavité catalytique de SamELA en présence de la L-lysine en orientation ε (A) ou α (B).
Figure IV.7 : Orientation préférentielle de l'acide caprylique seul (A) et en présence de la L-lysine (B) dans le site actif de SamELA.
Figure IV.8 : Orientations préférentielles des substrats de la réaction de N-acylation de la L-lysine par l'acide laurique, catalysée par SamELA. A et B : Modes d'interaction de l'acide laurique en présence de la L-lysine en orientation ε (A) ou en orientation α (B). C : Orientation de l'acide laurique seul. D : Mode d'interaction majoritaire de la L-lysine en présence de l'acide laurique.

Chapitre V

Figure V.1 : Récapitulatif des conditions étudiées pour la surexpression de SamAA chez *E. coli*.
Figure V.2 : Récapitulatif des conditions étudiées pour la surexpression de SamAA chez *E. coli* Origami B (DE3).
Figure V.3 : Suivi de la densité optique du milieu de culture LB, à 600 nm, lors de l'expression protéique chez les souches *E. coli* Origami B (DE3) transformées par les plasmides pET-15b vide ou pET-15b-SamAA.
Figure V.4 : Analyse par SDS-PAGE des protéines issues des cultures de 24 h d'*E. coli* Origami B (DE3) avec le plasmide pET-15b vide et pET-15b-SamAA en milieu LB.
Figure V.5 : Récapitulatif des conditions étudiées pour la surexpression de SamAA chez *E. coli* Rosetta 2 pLysS (DE3).
Figure V.6 : Suivi de la densité optique du milieu de culture LB, à 600 nm, lors de l'expression protéique chez les souches *E. coli* Rosetta 2 pLysS (DE3) transformées par les plasmides pET-15b vide ou pET-15b-SamAA.
Figure V.7 : Analyse par SDS-PAGE des protéines issues des cultures de 24 h d'*E. coli* Rosetta 2 pLysS (DE3) avec le plasmide pET-15b vide et pET-15b-SamAA, en milieu LB.
Figure V.8 : Récapitulatif des conditions étudiées pour la surexpression de SamAA chez *E. coli* Arctic Express (DE3).

Figure V.9 : Suivi de la densité optique du milieu de culture LB-Co, à 600 nm, lors de l'expression protéique chez les souches *E. coli* Arctic Express (DE3) transformées par les plasmides pET-28a-SamAA 6-His N-ter ou Strep-Tag N-ter, à 11 °C après l'induction.
Figure V.10 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Arctic Express (DE3) avec le plasmide pET-28a-SamAA 6-His ou Strep-Tag N-ter en milieu LB supplémenté en cobalt (0,25 mM).
Figure V.11 : Récapitulatif des conditions étudiées pour la surexpression de SamAA chez *E. coli* Rosetta 2 (DE3).
Figure V.12 : Suivi de la densité optique du milieu de culture LB, avec et sans cobalt, à 600 nm, lors de l'expression protéique chez les souches *E. coli* Rosetta 2 (DE3) transformées par les plasmides pET-15b-SamAA, pET-28a-SamAA 6-His N-ter ou pET-28A-SamAA 6-His C-ter.
Figure V.13 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu LB supplémenté ou non en cobalt (0,25 mM).
Figure V.14 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-15b-SamAA 6-His N-ter en milieu LB supplémenté ou non en cobalt (0,25 mM).
Figure V.15 : Récapitulatif des conditions étudiées la surexpression de SamAA chez *E. coli* BL21 Gold pLysS (DE3).
Figure V.16 : Suivi de la densité optique à 600 nm du milieu de culture LB, supplémenté ou non en cobalt, lors de l'expression protéique chez les souches *E. coli* BL21 Gold pLysS (DE3) transformées par les plasmides pET-28a-SamAA 6-His N-ter ou pET-28A-SamAA 6-His C-ter.
Figure V.17 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold pLysS (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu LB supplémenté ou non en cobalt (0,25 mM).
Figure V.18 : Récapitulatif des conditions étudiées pour la surexpression de SamAA chez *E. coli* BL21 Gold (DE3).
Figure V.19 : Suivi de la densité optique du milieu de culture LB, avec et sans cobalt, à 600 nm, lors de l'expression protéique chez les souches *E. coli* BL21 Gold (DE3) transformées par les plasmides pET-28a-SamAA 6-His N-ter ou C-ter.
Figure V.20 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu LB supplémenté ou non en cobalt (0,25 mM).
Figure V.21 : Analyse par chromatographie du milieu de synthèse de la N-α-lauroyl-L-lysine en présence de la fraction soluble concentrée issue de la culture d'E. coli BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter, en milieu LB supplémenté en cobalt dès le début de la culture.
Figure V.22 : Analyse par SDS-PAGE des protéines issues des cultures de *S. lividans* avec le plasmide pIB139-SamAA StrepTag N-ter.
Figure V.23 : Analyse par SDS-PAGE des protéines issues du refolding des fractions insolubles issues de la culture avec *E. coli* Rosetta 2 (DE3), le plasmide pET-28a-SamAA 6-His C-ter, et le milieu LB ou OV sans cobalt.
Figure V.24 : Analyse par électrophorèse SDS-PAGE des protéines issues des élutions ultrafiltrées sur colonne IMAC (à la seringue) des fractions solubles issues des cultures avec *E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu LB avec cobalt.

Figure V.25 : Chromatogramme obtenu lors de la purification de SamAA sur ÄKTA Pure.
Figure V.26 : Récapitulatif des conditions étudiées pour la surexpression de SamELA chez *E. coli*.
Figure V.27 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter, en milieu LB ou OV sans cobalt.
Figure V.28 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter, en milieu LB ou OV supplémenté en cobalt.
Figure V.29 : Analyse par CCM du milieu de synthèse de la N-ε-lauroyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter, en milieu LB ou Overnight, avec ou sans cobalt.
Figure V.30 : Analyse par CCM du milieu d'hydrolyse de la N-ε-acétyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter, en milieu LB ou OV, avec ou sans cobalt.

Index des tableaux

Chapitre I

Tableau I.1 : Activité hydrolytique de l'aminoacylase au nom inconnu de *S. mobaraensis*. (Koreishi et al., 2005b).
Tableau I.2 : A : Spécificité de substrat de Sm-AA recombinante envers différents acides aminés N-acétylés et N-lauroylés ; B : Spécificité de substrats de Sm-AA recombinante envers différents dérivés de type N-acyl-L-méthionine (Koreishi et al., 2009b).
Tableau I.3 : Activités spécifiques d'hydrolyse de la N-α-acétyl-L-lysine et de la N-ε-acétyl-L-lysine relevées dans des fractions issues de la séparation des protéines des extraits bruts des souches sauvage ou mutées de *S. ambofaciens* (Dettori et al., 2018a).
Tableau I.4 : Principales caractéristiques des souches commerciales *E. coli*.
Tableau I.5 : Résultats relatifs à la purification de l'aminoacylase recombinante d'*A. xylosoxidans* (Wang et al., 2013).
Tableau I.6 : Activités spécifiques apparentes des vitesses de réactions hydrolytiques en présence des fractions solubles des cultures d'*E. coli* Origami B (DE3) transformées (Nili, 2018).
Tableau I.7 : Expression et activité spécifique de l'aminoacylase d'*A. xylosoxydans* selon le système d'expression (Hou et al., 2016).
Tableau I.8 : Résumé de la purification de l'aminoacylase SmAA recombinante issue de *S. lividans* TK24 (Koreishi et al., 2009b).

Chapitre II

Tableau II.1 : Récapitulatif des caractéristiques des systèmes *E. coli* étudiés pour l'expression de SamAA.
Tableau II.2 : Récapitulatif des caractéristiques des systèmes *S. lividans* étudiés pour l'expression de SamAA.
Tableau II.3 : Récapitulatif des caractéristiques des systèmes *E. coli* étudiés pour l'expression de SamELA.
Tableau II.4 : Antibiotiques et leurs concentrations utilisées selon les souches bactériennes et les plasmides étudiés.
Tableau II.5 : Composition et gradient des solvants d'élution utilisés lors des analyses HPLC.

Chapitre III

Table III.1: Percentages of sequence identity and similarity between SamAA and its templates.
Table III.2: List and role of conserved residues.
Table III.3: Angle measurements around each metal ion after energy minimization.
Tableau III.4 : Caractéristiques et pourcentages de similarité et d'identité des alignements de séquences réalisés et valeurs du DOPE Score normalisé des modèles obtenus.
Tableau III.5 : Caractéristiques des alignements de séquences choisis et valeurs du DOPE Score normalisé des modèles obtenus.

Chapitre IV

Tableau IV.1 : Pourcentages d'identité et de similarité de séquences entre les templates identifiés et SamELA.

Chapitre V

Tableau V.1 : Activités spécifiques apparentes d'hydrolyse de la N-α-acétyl-L-lysine et de la N-ε-acétyl-L-lysine en présence des fractions solubles des cultures d'*E. coli* Origami B (DE3) transformées (Nili, 2018).
Tableau V.2 : Activités spécifiques apparentes pour la synthèse de la N-α-lauroyl-L-lysine en présence des fractions solubles concentrées des cultures d'*E. coli* Arctic Express (DE3) transformées par le plasmide pET-28a-SamAA 6-His ou StrepTag N-ter en milieu LB, OV ou TB-AIM, supplémenté ou non en cobalt.
Tableau V.3 : Mesure de la densité optique du milieu de culture Overnight (OV) avec et sans cobalt, en fin de culture (16 h), à 600 nm, lors de l'expression protéique chez les souches *E. coli* Rosetta 2 (DE3) transformées par les plasmides pET-15b-SamAA, pET-28a-SamAA 6-His N-ter ou pET-28a-SamAA 6-His C-ter.
Tableau V.4 : Activités spécifiques apparentes pour la synthèse de la N-α-lauroyl-L-lysine en présence des fractions solubles concentrées des cultures d'*E. coli* Rosetta 2 (DE3) transformées par le plasmide pET-15b-SamAA 6-His ou pET-28a-SamAA 6-His N-ter ou C-ter en milieu LB ou OV, supplémenté ou non en cobalt.
Tableau V.5 : Activités spécifiques apparentes pour la synthèse de la N-α-lauroyl-L-lysine en présence des fractions solubles concentrées des cultures d'*E. coli* BL21 Gold pLysS (DE3) transformées par le plasmide pET-28a-SamAA 6-His N-ter ou C-ter avec le milieu LB supplémenté ou non en cobalt.
Tableau V.6 : Mesure de la densité optique du milieu de culture Overnight (OV) en fin de culture (16 h) avec et sans cobalt, à 600 nm, lors de l'expression protéique chez les souches *E. coli* BL21 Gold (DE3) transformées par les plasmides pET-28a-SamAA 6-His N-ter ou pET-28a-SamAA 6-His C-ter.
Tableau V.7 : Activités spécifiques apparentes pour la synthèse de la N-α-lauroyl-L-lysine en présence des fractions solubles concentrées des cultures d'*E. coli* BL21 Gold (DE3) transformées par le plasmide pET-28a-SamAA 6-His N-ter ou C-ter, en milieu LB ou OV supplémenté ou non en cobalt.
Tableau V.8 : Principales caractéristiques des quatre plasmides étudiés pour l'expression de SamAA chez *S. lividans*.
Tableau V.9 : Activités spécifiques apparentes pour la synthèse de la N-α-lauroyl-L-lysine en présence des fractions solubles issues des cultures de *S. lividans* transformées par les plasmides pIB139, pIJ8600, pIJ86, pIJ8660 vides ou avec SamAA.

Conclusion gérérale

Tableau CCL.1 : Récapitulatif des templates identifiés et des résidus conservés chez SamAA et SamELA.

Liste des abréviations

2M2B : 2-méthyl-2-butanol

3D : tridimensionnelle

Å : Angström(s)

ADN : Acide désoxyribonucléique

AMBER : Assisted model building and energy refinement

APTES : 3-aminopropyltriethoxysilane

ARN(t) : Acide ribonucléique (de transfert)

BCA : Bicinchoninic acid assay

BLAST : Basic local alignment search tool

CAMEO : Continuous automated model evaluation

CASP : Critical assessment of protein structure prediction

Cbz : Carbobenzoxy

CCF : Consistent force field

CCM : Chromatographie sur couche mince

CHARMm : Chemistry at Harvard molecular mechanics

CLEAs : Agrégat d'enzymes réticulées (cross-linked enzyme aggregates)

CO : Cobalt

CO_2 : Dioxyde de carbone

DAP : L,L-2,6-diaminoheptanedioate

DEDL : Détecteur évaporatif à diffusion de lumière à faible température

DO : Densité optique

DOPE : Discrete optimized protein energy

DTT : Dithioréitol

EBNL : Extrait brut non-lysé

EDTA : Ethylène diamine tétra-acétique

EPR : Electron paramagnetic resonance

FI(R) : Fraction insoluble (refoldée)

FS(R)(c) : Fraction soluble (refoldée) (concentrée)

GMQE : Global model quality estimate

GST : Glutathion S transférase

HMM : Modèles de Markov cachés (hidden Markov models)

HPLC : High performance liquid chromatography

IMAC : Chromatographie d'affinité au nickel (immobilized metal affinity chromatography)

IMOPA : Ingénierie moléculaire et physiopathologie articulaire

IPTG : Isopropyl β-D-1-thiogalactopyranoside

(k)Da : (kilo)Dalton

LB : Lysogeny Broth ou Luria-Bertani

LC-MS/MS : Chromatographie en phase liquide couplée à de la spectrométrie de masse en tandem

lDDT : test de différence de distance local (local distance difference test)

LRGP : Laboratoire Réaction et Génie des Procédés

MBP : Protéine de liaison au maltose (maltose binding protein)

MM (ou MW) : marqueur de poids moléculaire

MSF : Mannitol soya flour

NCBI : National center for biotechnology information

Orn : Ornithine

OV (ou Overnight) : Milieu Overnight

PAE : Predicted align error

PBTSM : modélisation de l'état de transition basé sur le produit

PDB : Protein Data Bank

pH : Potentiel hydrogène

PLP : Piecewise linear potential

PVDF : Polyfluorure de vinylidène

QM/MM : quantum mechanics/molecular mechanics

QMEAN : Qualitative model energy analysis

QSAR : Relation quantitative structure à activité (quantitative structure-activity relationship)

RMN : Résonance magnétique nucléaire

RMSD : Déviation quadratique moyenne

SBA-15 : Santa Barbara amorphous 15

SDAP : Acide N-succinyl-L,L-diaminopimelique

SDS-PAGE : Electrophorèse en gel de polyacrylamide contenant du laurylsulfate de sodium (sodium dodecyl sulfate polyacrylamide gel electrophoresis)

SiHIPE : High internal phase emulsion

TB : Terrific Broth

TBS : Tris buffered saline

TFA : Acide trifluoroacétique

USD : Dollar américain

Introduction générale

Les dérivés acylés d'acides aminés ou acyl-aminoacides constituent l'une des catégories les plus importantes de molécules tensioactives utilisées dans de nombreuses applications notamment dans les domaines de l'hygiène et de la cosmétique. Ils sont obtenus par une réaction de N-acylation, qui consiste à greffer une entité acyle appelée donneur d'acyle (la plupart du temps un acide gras) sur une fonction amine d'un acide aminé, considéré comme un accepteur d'acyle. Ce sont des molécules à un fort potentiel applicatif en raison de leurs nombreuses propriétés technofonctionnelles. Elles sont amphiphiles présentant une activité de surface élevée et parfois une activité biologique spécifique. Constituées de briques d'origine naturelle, elles sont considérées comme biosourcées. Enfin, elles sont biodégradables et biocompatibles.
Ces dérivés sont produits à l'échelle industrielle par la réaction chimique de Schotten-Baumann qui est aujourd'hui maitrisée. Les chlorures d'acyle (sorte d'acides gras activé) et les acides aminés sont mis en solution dans un milieu réactionnel maintenu à un pH très basique grâce à l'ajout de soude. Des sels sont libérés lors cette réaction. Elle est arrêtée par l'ajout d'acide fort, entrainant la précipitation des acyl-aminoacides, des sels et des substrats en excès. L'abaissement du pH peut provoquer la réaction inverse (réaction d'hydrolyse), libérant des acides gras. Cette réaction s'accompagne de rendements compris entre 70 et 90 %. Les produits sont ensuite isolés par filtration et lavages, ce qui génère des volumes importants d'effluents organiques salins à traiter.
Les dérivés acylés d'acides aminés peuvent également être produits par la voie enzymatique, en utilisant des lipases, des acyltransférases, des carboxylestérases ou encore des amidohydrolases/aminoacylases. Cette catalyse est chimio-, régio- et énantio-sélective et se déroule dans des conditions réactionnelles douces, respectant ainsi les principes de la chimie verte (Sheldon et al., 2018a).
Entre 2005 et 2009, l'équipe de Koreishi a découvert et caractérisé quatre animoacylases chez *Streptomyces mobaraensis* :

- Sm-ELA, catalysant l'hydrolyse ou la synthèse de dérivés N-ε-acyl-L-lysine (Koreishi et al., 2005a ; Koreishi et al., 2009a).
- Sm-AA, catalysant l'hydrolyse ou la synthèse de dérivés N-α-acyl-L-lysine (Koreishi et al., 2009b).
- Sm-PVA, catalysant l'hydrolyse ou la synthèse de dérivés d'acides aminés ou de peptides méthylés (Koreishi et al., 2006 ; Koreishi et al., 2007).

- Une quatrième enzyme catalysant l'hydrolyse des dérivés acylés d'acides aminés (Koreishi et al. 2005b).

Le LRGP travaille depuis de nombreuses années sur la problématique de l'acylation des acides aminés et des peptides en utilisant des lipases commerciales, dans le but d'étudier la réaction de N-acylation. En 2014, dans le cadre de ces travaux, le LRGP a découvert une activité aminoacylase dans le surnageant de culture de *Streptomyces ambofaciens*. Une étude bioinformatique a permis alors d'identifier quatre gènes de cette bactérie, présentant une homologie avec les séquences des aminoacylases identifiées chez *S. mobaraensis* (Ferrari, 2014). La construction de mutants de *S. ambofaciens* ayant un ou plusieurs de ces gènes délétés a permis de confirmer l'activité des enzymes appelées SamAA et SamELA. SamAA catalyse spécifiquement l'acylation ou la déacylation de la fonction amine α de la L-lysine, ouvrant la voie à l'acylation de n'importe quel acide aminé. SamELA quant à elle, catalyse spécifiquement l'acylation ou la déacylation de la fonction amine ε de la L-lysine, ce qui est similaire à l'activité observée avec les lipases (Dettori et al., 2018a). Plusieurs études ont été menées pour confirmer ces résultats (Bourkaib et al., 2020a ; Bourkaib et al., 2021). Convaincu de l'intérêt de ces enzymes, le laboratoire a mené plusieurs tentatives de production hétérologue de ces quatre enzymes chez plusieurs souches d'*Escherichia coli*. Malheureusement, la production des enzymes sous forme soluble est faible, car celles-ci précipitent sous la forme de corps d'inclusion.

Les travaux de cette thèse se sont focalisés sur deux des quatre aminoacylases de *S. ambofaciens*, SamAA et SamELA, dont l'activité enzymatique a été caractérisée expérimentalement. Les principaux objectifs étaient la production, la purification et la caractérisation de ces deux enzymes. Pour les atteindre, deux axes d'investigation ont été suivis. Le premier axe concernait l'identification des conditions optimales de production des deux aminoacylases recombinantes de *S. ambofaciens* et leur purification. Il s'agissait notamment d'identifier un système d'expression approprié (souche et conditions de cultures) pour chacune des deux enzymes, limitant la formation des corps d'inclusion et de déterminer les conditions optimales permettant l'obtention de quantités suffisantes d'enzymes pures pour pouvoir mener des études cristallographiques et de caractérisation structurale.

Le deuxième axe de travail visait à acquérir davantage de connaissances sur les deux aminoacylases SamAA et SamELA. Il s'agissait notamment de caractériser l'activité catalytique de chacune de ces enzymes en termes de sélectivité et spécificité de substrat à l'échelle moléculaire, en vue d'établir des relations structure-fonction. Dans cette approche, la connaissance de la structure tridimensionnelle des enzymes est primordiale. Pour les enzymes

de structure non élucidée, comme c'est le cas pour SamAA et SamELA, des outils de prédiction de structure peuvent être mis en œuvre. Associant intelligence artificielle et consultation des bases de données protéiques, ces outils permettent de prédire la structure des protéines à partir de leur séquence, à des niveaux de précision de plus en plus élevés. Cette structure prédite peut alors être étudiée et comparée aux structures élucidées d'enzymes homologues. Des résidus et des domaines conservés peuvent être identifiés, et des hypothèses concernant leur rôle dans la catalyse et la reconnaissance des substrats peuvent être formulées. De par leur statut de modèles, les structures prédites demandent à être validées par des données expérimentales. Elles permettent néanmoins de continuer à avancer dans le processus de découverte de nouvelles enzymes, en attendant l'élucidation de leur structure.

Les travaux présentés ici s'articulent donc autour de deux volets, l'un expérimental, l'autre numérique. Après un rappel des propriétés et voies de synthèse des dérivés acylés d'acides aminés, les éléments bibliographiques relatifs aux enzymes catalysant les réactions de N-acylation et N-déacylation sont donnés, avec un focus particulier sur les aminoacylases. L'un des objectifs étant la production de deux enzymes appartenant à cette famille, une troisième partie reprend les principales données de la littérature concernant la production par voie hétérologue de ces enzymes. L'analyse de la bibliographie se termine par une présentation des éléments de la littérature ayant contribué à orienter la partie numérique du travail. Les résultats produits au cours de cette étude sont présentés sous la forme de trois chapitres. Le premier et le deuxième portent sur les travaux numériques relatifs à SamAA et SamELA, tandis que le troisième est focalisé sur la production hétérologue de ces deux enzymes. Le mémoire se termine par une synthèse des principales avancées obtenues à l'issue de ces travaux de recherche suivie des perspectives envisagées.

1. Chapitre I : Etude bibliographique

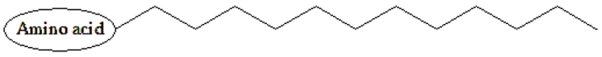
1.1. Dérivés acylés d'acides aminés et de peptides

1.1.1. Définition et classification

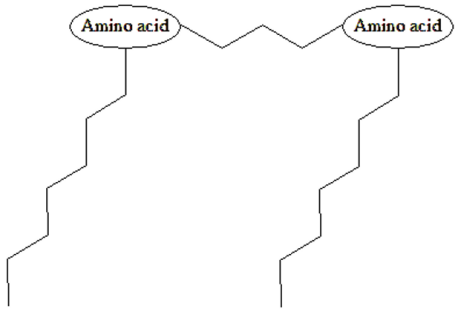
Les tensioactifs sont les produits chimiques les plus consommés quotidiennement à l'échelle mondiale. Ils ont un rôle important dans de nombreux domaines industriels, comme l'hygiène individuelle et collective et le nettoyage domestique et industriel. Le type d'applications dépend principalement de leur structure responsable, et par suite de leurs propriétés comme leur capacité à s'organiser en architectures supramoléculaires par exemple. La demande en tensioactifs à l'échelle mondiale est de plus en plus importante, ce qui provoque un intérêt croissant pour les tensioactifs dits renouvelables, car ils ont un faible impact sur l'environnement et n'ont pas une origine fossile. Dans cette catégorie figurent les biotensioactifs issus de processus fermentaires impliquant des flux métaboliques de micro-organismes et les tensioactifs biosourcés issus de procédés enzymatiques. Les tensioactifs à base d'acides aminés comme les dérivés acylés d'acides aminés font partie de cette dernière catégorie. En 2021, le marché des tensioactifs à base d'acide aminés était de 528,58 millions USD (dollar américain) ; il devrait plus que doubler d'ici 2028 pour atteindre 1163,05 millions USD, donnant un taux de croissance annuel de 12 % (Nagtode et al., 2023). Les dérivés acylés d'acides aminés sont constitués de deux entités moléculaires liées de façon covalente : un acide aminé et un acide gras. Cette structure fait d'eux des composés durables et respectueux de l'environnement, car ils sont peu toxiques et rapidement biodégradables. Leurs structures sont variées, ce qui confère de nombreuses propriétés et applications à cette famille de molécules (Le Guenic et al., 2018).

Selon leur origine et leur composition, il existe différentes classifications des dérivés acylés d'acides aminés. Il est ainsi possible de les regrouper selon le type de groupe fonctionnel libre en bout de chaine. Ils sont définis comme non-ioniques quand ils ne comptent pas de charge. A l'inverse, ils sont dits cationiques ou anioniques quand ils possèdent une charge positive ou négative libre sur la partie hydrophile de leur structure. Ils peuvent également être zwittérioniques ou amphotères quand ils contiennent à la fois des fonctions anioniques et cationiques (Guo et al., 2022). Par ailleurs, ils peuvent être classés selon le nombre de chaines hydrophobes greffées sur l'acide aminé. Ils sont appelés linéaires lorsqu'ils résultent de

l'acylation d'un acide aminé par un seul acide gras. Ce sont les dérivés acylés d'acides aminés les plus courants. Ils sont dits « dimériques » ou « gemini » quand ils résultent de l'association de deux chaines acylées linéaires séparées par une chaine aliphatique d'une longueur variable. La condensation de deux acides aminés à chaque extrémité d'une unique chaine hydrophobe conduit aux composés bolaamphiphiles. Enfin il existe des dérivés de type glycérolipides comme les mono- et di-acyl-glycérols à base d'acides aminés (Figure I.1) (Tripathy et al., 2018).



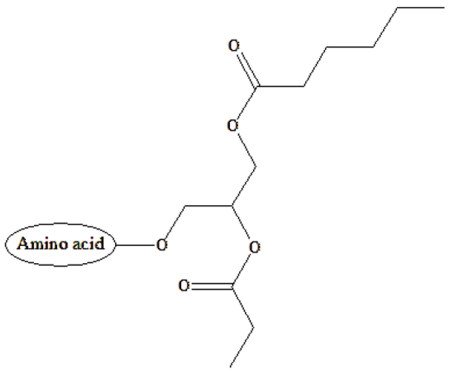
Dérivés acylés d'acides aminés linéaires



Dérivés acylés d'acides aminés dimériques (gemini)



Dérivés acylés d'acides aminés de type bolaamphiphile



Dérivés acylés d'acides aminés de type glycérolipide

Figure I.1 : Classification des dérivés acylés à base d'acides aminés (Bourkaib, 2020).

1.1.2. Propriétés et activités

La partie hydrophobe de ces composés est constituée la plupart du temps d'un acide gras tandis que l'acide aminé constitue la partie hydrophile. En fonction de la composition atomique de chacune de ces deux parties, les dérivés acylés d'acides aminés présentent des caractéristiques structurales leur conférant de diverses propriétés et un haut potentiel applicatif (Wang et al., 2021).

1.1.2.1. Propriétés de surface

La solubilité des tensioactifs augmente de façon brutale et discontinue à partir d'une certaine température appelée « température de Krafft ». La température de Krafft ou le point de Krafft des tensioactifs ioniques est la température minimum pour que les tensioactifs puissent former des micelles. La solubilité des tensioactifs ioniques non associés est limitée en dessous de la température de Krafft puis elle augmente progressivement jusqu'à atteindre le point de Krafft où les tensioactifs commencent à former des micelles (Malik et al., 2016).

Le changement de l'entité hydrophobe du tensioactif permet de modifier les propriétés de surface des émulsions qui les contiennent comme la concentration micellaire critique et la température de Krafft : la concentration micellaire critique est augmentée en présence d'insaturations sur la chaine carbonée du tensioactif tandis que la température de Krafft est diminuée dans ces conditions (Brito et al., 2011). La température de Krafft est également modifiée selon la nature de l'acide aminé présent au sein du tensioactif : la caractérisation de six dérivés N-hexadécanoyl de glycine, d'alanine, de valine, de leucine, d'isoleucine et de phénylalanine a montré que la température de Krafft augmentait avec la diminution de la taille de l'entité acide aminé, sauf dans le cas de la phénylalanine (Ohta et al., 2003).

L'équipe de Shi a étudié les propriétés de surface de la lysine acylée par des acides gras ayant des chaines de différentes longueurs. Il a été observé que l'utilisation de l'acide laurique (C12) à la place d'acides gras de plus courtes chaines (C8 et C10) permet d'augmenter l'efficacité des tensioactifs en diminuant la tension de surface, la concentration micellaire critique et la taille des gouttelettes au sein d'une émulsion huile dans eau (Shi et al., 2019). Cette tendance avait déjà été relevée dans une review en 2013 (Chandra et al., 2013) rapportant que la concentration micellaire critique et les propriétés tensioactives des dérivés acylés d'acides aminés étaient respectivement inférieures et supérieures à celles des tensioactifs conventionnels correspondants. Des auteurs s'accordent pour affirmer que les dérivés acylés d'acides aminés

ont une meilleure capacité émulsifiante et anticalcique et une meilleure stabilité à l'eau dure que les savons à base d'acides gras (Mao et al., 2015 ; Shi et al., 2019).

1.1.2.2. Propriétés d'agrégation

Les agrégats de micelles obtenus avec les dérivés acylés d'acides aminés ont un comportement d'assemblage et une morphologie dépendante de la structure hydrophobe et hydrophile de ces molécules. A titre d'exemple, des travaux sur les acides aminés acylés par des acides gras issus d'huiles végétales montrent que l'utilisation de l'huile de ricin et des acides aminés glycine ou alanine permet la formation d'agrégats tubulaires liés à la présence des groupes hydroxyles sur les chaines aliphatiques alors que l'utilisation de sérine provoque l'apparition d'agrégats sphériques liés à la présence d'une fonction hydroxyle supplémentaire (Liu et al., 2018). L'utilisation de l'huile de ricin à la place de l'huile de coton entraine une diminution de l'activité interfaciale et de l'adsorption des films formés car les longues chaines alcanes présentes dans l'huile de coton améliorent les interactions intermoléculaires de type Van der Waals alors que le groupe hydroxyle de l'acide ricinoléique limite les interactions entre les chaines hydrophobes (Chen et al., 2020).

1.1.2.3. Propriétés moussantes et gélifiantes

Les propriétés d'adsorption interfaciale, les caractéristiques de mousses et de gels contenant le lauroylglutamate de sodium ont été étudiées à différentes températures et à différents pH. Il a été montré que la formation et la stabilité des mousses et des gels dépendent de la température et du pH, à cause des différents états de protonation de la molécule tensioactive (Zhang et al., 2016). La moussabilité, la stabilité, la morphologie des mousses contenant du N-acylglycinate de sodium et du N-acylphénylalaninate de sodium ont été analysées par la méthode de balayage de la mousse pour étudier l'effet de la nature de la chaine hydrophobe du tensioactif et de la température sur leurs propriétés moussantes (Zhao et al., 2019). Dans le cadre de la recherche et du développement de gels d'huile comestible, les gels formés avec quatre huiles commerciales (huile d'olive, huile de sésame, huile de maïs et huile de soja) en présence de la N-lauroyl-L-alanine et de la N-lauroyl-L-alanine dodécylamide ont été caractérisés. L'étude a montré que les interactions de type liaison hydrogène entre les liaisons amides est la force motrice intermoléculaire de la formation du gel. Les deux molécules dérivées d'acides aminés

peuvent être utilisées comme gélifiant d'huile, la N-lauroyl-L-alanine dodécylamide étant la plus performante en raison des deux liaisons amides qu'elle comporte (Luo et al., 2020).

1.1.2.4. Activités hémolytiques

Certains dérivés acylés d'acides aminés sont de plus en plus utilisés comme excipients dans les produits pharmaceutiques sous la forme de vésicules cationiques à base de tensioactifs chargés positivement. Leur activité hémolytique est dépendante de la longueur de la chaine alkyle, de la densité de charge cationique, de la longueur de la chaine d'espacement entre les acides aminés et de la taille des agrégats en solution (Tavano et al., 2013 ; Pinazo et al., 2016). L'effet de la Nα,Nε-dioctanoyl lysine avec cinq contre ions différents sur des globules rouges murins a été observé au microscope électronique à balayage pour l'étude de la perturbation de la membrane plasmique et de l'activité hémolytique. Ces molécules sont capables de déformer les globules rouges et de détruire leur membrane cellulaire. Leur activité est concentration et pH dépendante (Nogueira et al., 2011). Plus récemment, l'activité d'un conjugué d'acide stéarique et de valine a été comparée à celle de tensioactifs commerciaux comme le Triton X-100 et le Tween® 80 sur les lignées cellulaires MCF-7 et HEK et sur des souris albinos suisses femelles. Les marqueurs sériques, l'histologie des organes et la morphologie des globules rouges ont été étudiés : le dérivé acylé de valine présente une biocompatibilité plus élevée *in vitro* et une biocompatibilité similaire *in vivo* aux tensioactifs commerciaux (Katiyar et al., 2019).

1.1.2.5. Propriétés antimicrobiennes

Certains dérivés acylés d'acides aminés présentent une activité antimicrobienne, qui peut être quantifiée par la concentration minimale inhibitrice (Infante et al., 1984). Cette activité dépend de la densité de la charge cationique, du ou des acides aminés formant l'extrémité polaire, de la longueur des chaines acyles et donc de l'équilibre lipophile/hydrophile de la molécule (HLB) (Pinazo et al., 2016). La structure du dérivé détermine la cible de l'activité antimicrobienne (Castillo et al., 2006). Ainsi, les acides aminés N-acylés montrent une meilleure activité contre *Staphylococcus aureus* MLS-16 et *Bacillus subtilis* que les sels de sodium des acides aminés N-acylés ou d'autres composés standards (Sreenu et al., 2015). Les dérivés acylés à base d'arginine sont parmi les molécules les plus étudiées pour leurs propriétés antimicrobiennes car elles sont efficaces tout en étant peu toxiques et hautement biodégradables (Clapés et al., 1998 ; Singare et al., 2012 ; Chandra et al., 2013).

1.1.2.6. Biodégradabilité

Plusieurs études ont montré le caractère biodégradable des dérivés acylés d'acides aminés. Déjà en 1975, l'équipe de Shida montrait la dégradation de la N-lauroyl-L-valine en acide laurique et en valine dans le sol (Shida et al., 1975). Des dérivés acylés à base d'aspartate ou de glutamate couplé au monomère ou au dimère de chitine sont biodégradés à hauteur de 27 à 43 % en 14 jours dans l'environnement (Abe et al., 2004). La biodégradation primaire de l'acyl glycine sodique ou potassique de noix de coco est de 99 % ; elle a été évaluée par un test de boues activées basé sur la méthode standard coréenne (KSM 2714) (Yea et al., 2018).

1.1.2.7. Toxicité

Les tensio-actifs conventionnels comme le dodécyl sulfate de sodium présentent une toxicité importante vis-à-vis de l'environnement. Les dérivés acylés d'acides aminés quant à eux sont présentés comme peu ou non toxiques pour l'homme et son environnement (Holmberg et al., 2025). Une étude sur la relation entre la structure, les propriétés physicochimiques et le profil toxicologique de tensioactifs à base d'acides aminés N-décanoylés de type leucine, méthionine, sérine et proline a été menée. La cytotoxicité de ces molécules sur différentes lignées cellulaires humaines (Caco-2, A549 et Calu-3) est inférieure à celle du dodécyl sulfate de sodium (Perinelli et al., 2016). Obata et al. ont étudié des assemblages cationiques à base de dérivés acylés de lysine, d'histidine et d'arginine sous forme de vésicules pour une utilisation dans des transferts de gènes. Ces vésicules montrent une efficacité supérieure d'expression génique et une cytotoxicité très faible par rapport à la lipofectamine2000, qui est le réactif transgénique conventionnel (Obata et al., 2008). Les systèmes liposomaux peuvent être utilisés pour l'administration de médicaments. Des travaux portant sur des vésicules constituées de dérivés acylés d'acides aminés (tyrosine, sérine, hydroproline, lysine) ont mis en évidence leur écotoxicité moindre sur *Daphnia magna* (petits crustacés planctoniques) ainsi qu'un caractère irritant oculaire égal ou inférieur comparativement à la molécule de référence (dodecyltrimethylammonium bromide - DTAB) (Brito et al., 2009).

1.1.3. Applications des dérivés acylés d'acides aminés

Comme mentionné précédemment, les dérivés acylés d'acides aminés possèdent des propriétés variées. L'intérêt porté à ces molécules ne cesse de croître, ce qui conduit à une amélioration

de leurs procédés de production et de leur potentiel applicatif dans différents secteurs comme l'agriculture, l'industrie alimentaire, la médecine ou encore la cosmétique.

1.1.3.1. Dans les produits d'hygiène et de soin

Les tensioactifs sont des molécules utilisées dans de nombreux produits pour leurs propriétés détergentes, biodégradables, biocompatibles et non-irritantes. Les dérivés acylés d'acides aminés sont utilisés dans les produits nettoyants de la peau, des cheveux et de la bouche et dans les produits cosmétiques. Les propriétés moussantes de dérivés N-acylglycinate de sodium et N-acylphénylalaninate de sodium comportant une chaine acyle d'une longueur comprise entre 10 et 18 atomes de carbone ont été étudiées. La morphologie et la stabilité de la mousse obtenue ont également été analysées. L'effet de la nature des groupes hydrophiles et hydrophobes, de la concentration en tensioactif et de la température a été mis en évidence par la méthode de balayage de mousse (Zhao et al., 2019). D'autres exemples d'applications peuvent être cités comme la N-dodecyl-sarcosine sodique ajoutée à la formulation des dentifrices en prévention de la formation des caries dentaires ou encore la N-10-undécénoyl-phénylalanine produite par des aminoacylases issues de cultures de *Streptomyces ambofaciens* qui permet d'obtenir une action éclaircissante de la peau (Dan et al., 2010 ; Bourkaib et al., 2020b).

1.1.3.2. En agriculture

En agriculture, les dérivés acylés d'acides aminés sont utilisés comme pesticides en remplacement des tensioactifs conventionnels. Mais ils peuvent aussi améliorer la qualité des sols agricoles en dégradant les polluants et favoriser la croissance des plantes en éliminant les bactéries pathogènes grâce à leurs propriétés antibactériennes ou en favorisant les interactions entre les plantes et les micro-organismes (Sachdev et al., 2013). En 2004, un lipopeptide a été découvert chez *Bacillus thuringiensis* CMB26 lors d'une étude contre l'agent responsable de l'anthracnose du poivre. Il est de structure cyclique et est composé des acides aminés glutamate, ornithine, tyrosine, thréonine, alanine, valine, proline et isoleucine. Cette molécule ressemble au fengycine et présente une activité fongicide, bactéricide et insecticide (Kim et al., 2004). En 2010, une seconde étude a permis d'identifier et de confirmer la présence de trois lipopeptides antifongiques chez *Bacillus subtilis* CMB32 : l'iturine est un lipopeptide cyclique à base d'asparagine, de glutamine, de proline, de sérine et d'un acide gras à 14 ou 15 atomes de carbone ; la surfactine est un lipopeptide cyclique incluant une leucine, un aspartate, une valine,

un glutamate et un acide gras β-hydroxylé de 13 à 15 atomes de carbone ; la fengycine (Kim et al., 2010). D'autres molécules semblables à la surfactine, l'iturine et la bacillomycine (lipopeptide cyclique constitué d'un aspartate, d'une tyrosine, d'une glutamine, d'une proline, d'un glutamate, d'une sérine, d'une thréonine et d'un acide gras) sont présentes chez *Bacillus spp.* et sont capables d'inhiber la croissance de champignons phytopathogènes, comme *Fusarium spp.*, *Aspergillus spp.* et *Bipolaris sorokiniana* (Velho et al., 2011). Enfin, des travaux ont montré que les dérivés acylés d'acides aminés et de peptides ont un large spectre d'action dans le cadre de la remédiation des sols (Mao et al., 2015).

1.1.3.3. En biologie et en médecine

Les dérivés acylés d'acides aminés sont beaucoup utilisés dans le domaine de la biologie et de la médecine. Déjà en 1994, une émulsion de vitamine E stabilisée par la N-lauroylsarcosinate avait été proposée pour améliorer la solubilisation de cette vitamine en vue de son admistration (Chiu et al., 1994). Plus récemment, des auto-assemblages à base de tensioactifs ont été utilisés sous la forme de micelles ou de vésicules pour l'administration de substances. Des auteurs ont montré que les charges portées par la partie acide aminé et dépendantes du pH ont un effet sur la formation ou la destruction de ces micelles et leur forme (Lv et al., 2016). Par exemple, un système tensioactif anionique biocompatible et biodégradable à base de N-tétradécanoyl phénylalanine de sodium est utilisé pour l'administration du médicament Ibuprofène. A un pH de 8,25, les micelles sphériques éclatent, permettant une libération rapide du médicament alors qu'à un pH de 8,65, un gel se forme car les micelles sont de type vers sans éclatement, entraînant une libération retardée et progressive de l'ibuprofène. Ce système permet une transition solution-gel réversible en fonction du pH, tout en étant simple, rapide et économique à mettre en place (Lone et al., 2017). En biologie, des assemblages cationiques à base de dérivés acylés d'acides aminés sont utilisés pour la délivrance d'ADN ou d'ARN en thérapie génique par exemple. Les molécules tensioactives se condensent autour du fragment nucléotidique en formant un lipoplexe lamellaire qui devient hexagonal et plus lysogène lors d'une baisse du pH (Figure I.2). L'équipe d'Obata a montré que le 1,5-ditetradécyl-N-lysyl-L-glutamate et le 1,5-ditetradécyl-N-arginyl-L-glutamate peuvent être utilisés comme des systèmes de distribution d'ADN plasmidique plus efficaces et moins cytotoxiques que la Lipofectamine2000 qui est la réactif transgénique conventionnel (Obata et al., 2008). Les travaux menés ont montré qu'au sein des peptides apparentés Ser-Pro-Lys-Arg couplés à deux queues oléoyl et/ou oléyl, la proline et l'un des résidus cationiques (lysine ou arginine) sont essentiels à l'activité biologique

et qu'il n'est pas nécessaire d'utiliser un lipide auxiliaire lysogène si les queues lipidiques présentent des insaturations (Damen et al., 2010). Un autre dérivé acylés de type gémini constitué du peptide Trp-Trp-Cys-Orn-Orn et d'un acide gras à 18 atomes de carbone forme un complexe avec le plasmide pLenti-CMV-GFP. Celui-ci apparaît comme un vecteur prometteur pour la délivrance de gènes au sein des cellules CHO-K1 (lignée cellulaire issue d'ovaires de hamster de Chine) car son activité est supérieure à celle des actifs commerciaux disponibles et l'absence de cytotoxicité a été démontrée (Pena et al., 2017).

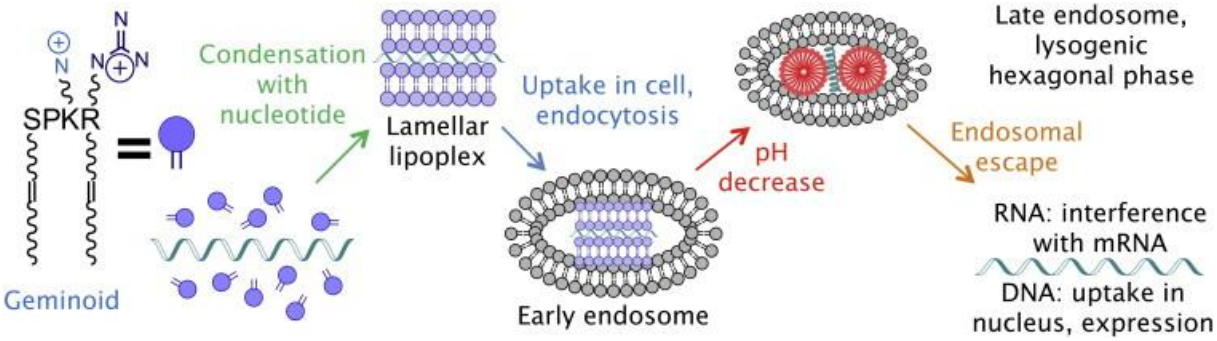


Figure I.2 : Utilisation d'un géminoïde pour le transfert d'un fragement nucléotidique au sein d'une cellule (Damen et al., 2010).

1.1.3.4. En agroalimentaire

Les dérivés acylés d'acides aminés sont utilisés comme émulsifiants et agents antimoussants dans le domaine alimentaire car ils sont biocompatibles et non-toxiques (Zhang et al., 2020). Dans cette étude, les lipases de *Candida cylindracea* et de *Rhizopus delemar* permettent de catalyser la O-acylation de la L-homosérine ou de la L-sérine avec des acides gras libres (acide myristique, acide palmitique, acide stéarique, acide palmitoléique, acide oléique ou acide linoléique) ou de l'huile végétale (huile de soja, huile de maïs ou huile d'olive) pour l'obtention de dérivés acylés d'acides aminés. Il a été montré que l'activité émulsifiante de l'O-oléoyl-L-homosérine est supérieure à celle des autres tensioactifs étudiés (Nagao et al., 1989).

1.1.3.5. Dans les autres industries

Les dérivés acylés d'acides aminés sont également employés dans d'autres domaines industriels tels que l'extraction minière et pétrolière, l'énergie, les nanomatériaux, etc. En effet, ils sont plus biodégradables et moins toxiques que les tensioactifs conventionnels. Des études ont par

ailleurs montré que ces composés sont plus performants que les molécules classiquement utilisées.

Dans l'extraction minière, les dérivés acylés à base d'acides aminés dicarboxyliques (Asp et Glu) sont des agents chélateurs efficaces et peuvent être utilisés comme collecteur pour la flottation minérale. La flottation minérale est une méthode de séparation et de concentration des minerais en modifiant leur surface pour les rendre hydrophobes ou hydrophiles, quand ces derniers sont agités dans l'eau (Bordes et al., 2015).

Dans l'extraction pétrolière, l'inondation par tensioactif est une méthode approuvée de récupération chimique du pétrole. Cette méthode repose sur la formation d'une microémulsion de pétrole en présence de tensioactifs. Deljooei et al. ont étudié l'efficacité d'un nouveau dérivé acylé obtenu à partir d'acide L-aspartique et de chlorure de lauryl en remplacement d'autres molécules (alcalis, polymères ou autres tensioactifs conventionnels) employées lors de la récupération chimique assistée de pétrole (Deljooei et al., 2021). L'utilisation couplée de polymères et de dérivés acylés de L-lysine et de chlorure de lauroyl pour la formation de bis-lauroyl lysine permet de récupérer 18 % de pétrole supplémentaire, permettant d'atteindre un rendement total de récupération du pétrole de 66 % (Ren et al., 2016). Une étude récente a été menée sur l'utilisation du cocoyl alaninate de sodium, de la carpyloyl glycine, de l'acide lauroyl glutamique et de la N-lauroyl sarcosine en remplacement du dodécyl sulfate de sodium pour des raisons environnementales. Le dérivé acylé à base d'alanine est une alternative intéressante aux trois autres dérivés acylés à base d'acides aminés et au dodécyl sulfate de sodium car il possède de meilleures propriétés pour la récupération assistée de pétrole telles qu'une réduction de la tension superficielle et interfaciale, une altération de la mouillabilité et une modification du comportement d'agrégation (Tackie-Otoo et al., 2020). La souche bactérienne *Brevibacterium aureum* MSA13 synthétise naturellement un dérivé composé d'acide octadécanoïque comme fraction hydrophobe et du tétrapeptide Pro-Leu-Gly-Gly comme fraction hydrophile. Ceci est un autre exemple de tensioactif pouvant être utilisé dans la récupération assistée du pétrole (Kiran et al., 2010).

Enfin, un dernier exemple d'application industrielle des dérivés acyles d'acides aminés est l'industrie textile dans laquelle ils sont utilisés en tant qu'émulsifiants, auxiliaires de teintures ou agents de graissage du cuir. Par exemple, la N-lauroyl-L-lysine en présence de CO_2 liquide est plus efficace que le perchloréthylène lors d'un procédé de nettoyage à sec à deux bains, le premier lavage à sec permettant l'élimination des salissures particulaires, le second lavage humide (où l'eau joue le rôle de co-solvant) conduisant à l'élimination des salissures non particulaires (van Roosmalen et al., 2004).

1.1.4. Synthèse chimique et enzymatique en milieux solvant des dérivés acylés d'acides aminés

1.1.4.1. Synthèse chimique

La synthèse chimique des dérivés d'acides aminés a commencé au XXème siècle. Selon la structure visée, quatre voies de synthèse peuvent être mise en œuvre (Figure I.3).

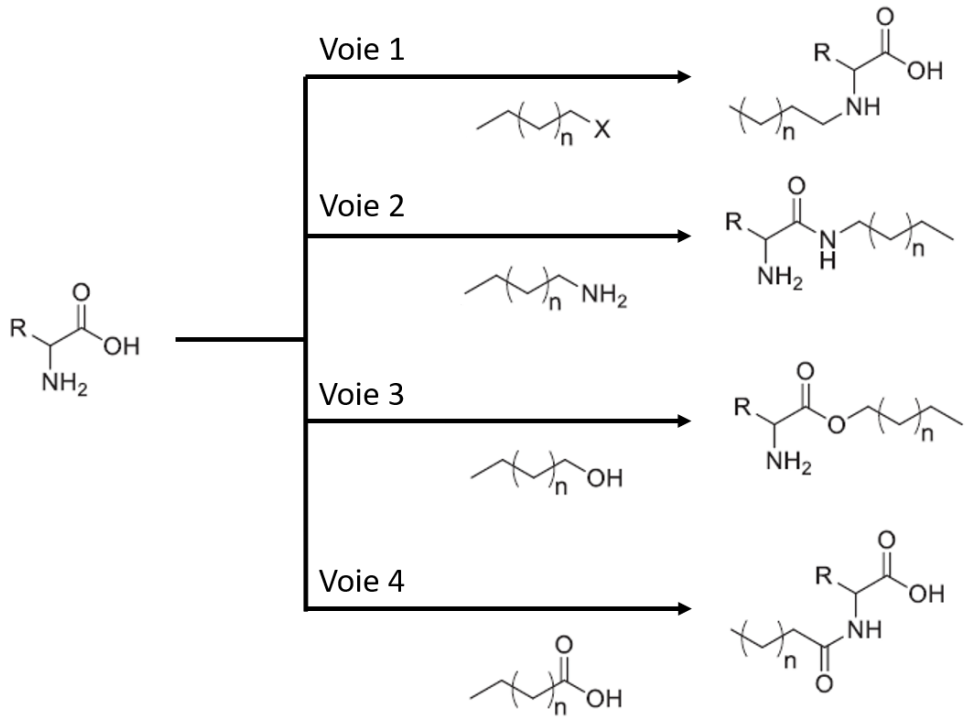


Figure I.3 : Voies principales de synthèse chimique des dérivés d'acides aminés.

En voie 1, l'amine secondaire de l'acide aminé peut être alkylée avec un halogénure d'alkyle. Les acides aminés, une fois activés, peuvent également réagir avec des amines alkylées pour la production d'amines aminées : c'est la voie 2. La voie 3 est la O-acylation où les acides aminés réagissent avec des alcools gras, pour produire des amines estérifiées. Ce type de réaction se réalise souvent en présence d'un agent déshydratant et d'un catalyseur acide. On parle de S-acylation si l'atome d'oxygène de l'alcool est remplacé par un atome de soufre d'un thiol. Lors de la N-acylation qui est présentée voie 4, les acides aminés acylés sont obtenus par la réaction du groupe amine de l'acide aminé avec un acide gras créant une liaison amide (Guo et al., 2022). De plus, en fonction des différentes fonctions chimiques présentes au niveau des chaines latérales des acides aminés, ces derniers peuvent réagir aussi avec les halogénures d'alkyle, les chlorures d'acyles gras et les amines grasses, conduisant à une grande diversité de dérivés (Bordes et al., 2015).

La réaction d'acylation qui est étudiée plus spécifiquement dans ces travaux conduit à l'ajout d'un groupement acyle provenant d'un agent acylant sur une entité moléculaire considérée comme un accepteur d'acyle. Cette réaction permet d'obtenir des dérivés acylés d'acides aminés et de peptides dont les propriétés et le potentiel applicatif ont été présentés précédemment. L'acylation des acides aminés peut être menée selon différentes voies de nature chimique, enzymatique ou chimio-enzymatique. Elle conduit à la fixation du fragment acyle hydrophobe au niveau de la fonction amine α ou au niveau de la chaine latérale de l'acide aminé lorsque celle-ci comporte un groupement accepteur d'acyle (alcool, thiol ou amine).

D'après la littérature, les dérivés N-acylés d'acides aminés peuvent être produits selon plusieurs méthodes de synthèse impliquant l'utilisation des chlorures d'acyle, des anhydrides gras ou des esters méthyliques d'acide gras, ou se basant sur l'hydrolyse du nitrile gras, la carbonylation de l'amide ou encore la méthode à l'huile (Takehara et al., 1972 ; Franceschi et al., 1999 ; Joondan et al., 2016 ; Wang et al., 2019 ; Yang et al., 2020 ; Guo et al., 2022).

La réaction de Schotten Baumann est la voie chimique d'acylation des acides aminés et des peptides la plus utilisée à l'échelle industrielle (Figure I.4). Les matières premières et l'équipement sont faciles d'accès et le procédé est maîtrisé. La réaction est menée en milieu aqueux et implique un chlorure d'acyle et une molécule avec une fonction amine ou hydroxyle libre. L'utilisation d'un chlorure d'acyle à la place d'un acide carboxylique se justifie par sa plus grande réactivité chimique. Il est parfois ajouté en goutte à goutte pour contrôler la vitesse de libération des protons et de réaction, en évitant une acidification du milieu réactionnel. L'ajout de solvants organiques peut faciliter la solubilisation des réactifs. Le milieu réactionnel est maintenu à un pH très basique grâce à l'ajout de soude ou de carbonate de sodium. Les atomes de sodium issus de la soude ou du carbonate de sodium et les atomes de chlore libérés par le chlorure d'acyle forment des sels. La réaction est arrêtée avec l'ajout d'un acide fort, comme l'acide sulfurique, provoquant une précipitation des dérivés acylés, des sels et des substrats en excès. L'abaissement du pH peut provoquer un changement de l'équilibre thermodynamique et une réaction d'hydrolyse des dérivés acylés, libérant des acides gras et complexifiant la composition du milieu. Les dérivés acylés produits sont isolés grâce à des étapes de filtration et de lavages, générant un gros volume d'effluents organiques salins à traiter. Cette réaction a un rendement élevé compris entre 70 et 90 %, ce qui explique son utilisation intensive à l'échelle industrielle. Toutefois elle n'est pas régiosélective et nécessite des étapes de protection et de déprotection des fonctions amines, ce qui augmente les coûts de production (Gerova et al., 2018 ; Bordes et al., 2009 ; Sreenu et al., 2014a ; Sreenu et al., 2014b ; Joodan et al., 2018). Encore actuellement, ce procédé est en cours d'amélioration comme le montre

l'étude de Zhang. L'utilisation d'un algorithme d'optimisation bayésien a permis d'étudier l'influence du débit d'alimentation des substrats et du choix du solvant sur l'efficacité de la réaction et d'optimiser celle-ci en éliminant notamment la réaction d'hydrolyse non souhaitée (Zhang et al., 2024).

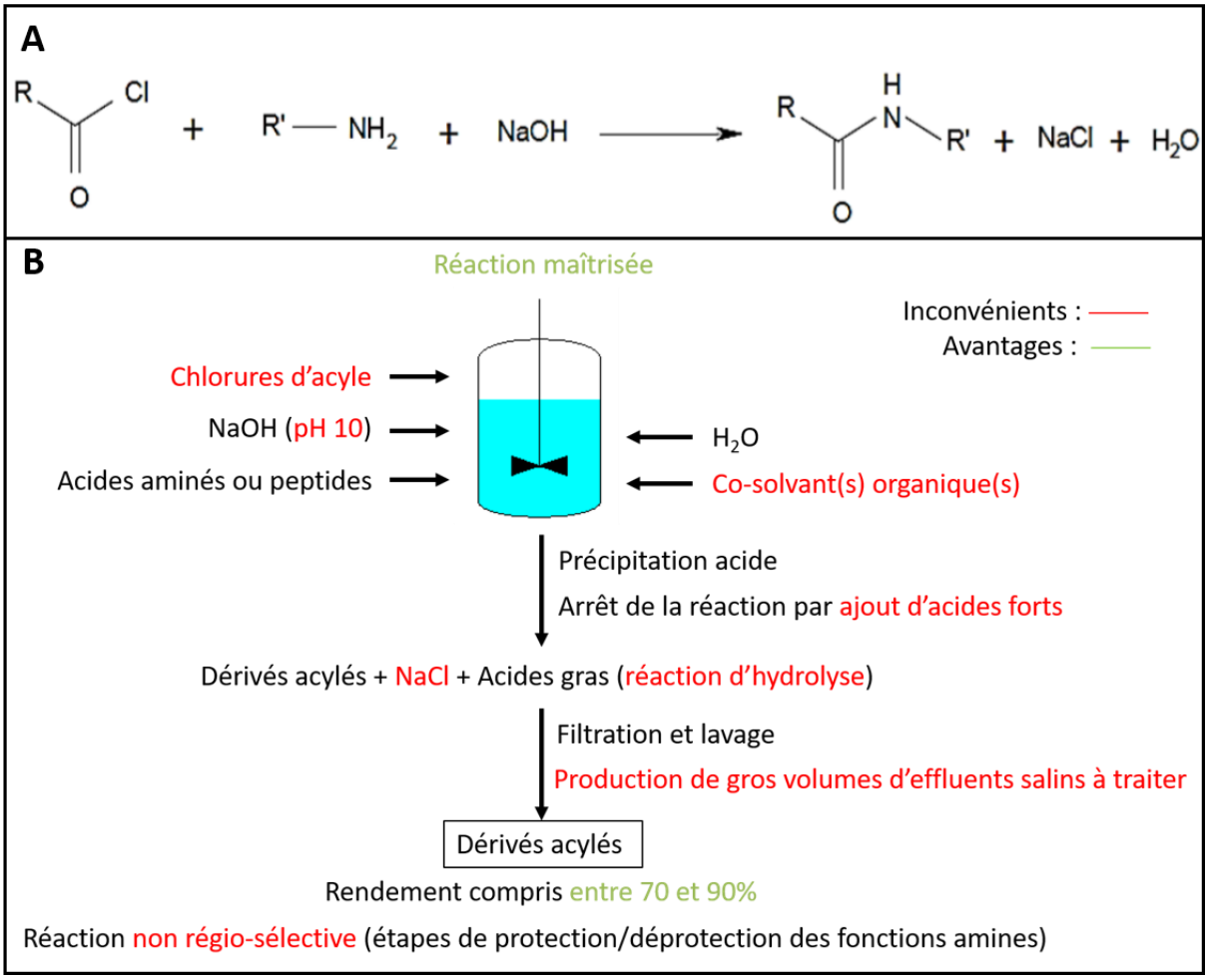


Figure I.4 : A : Réaction chimique de Schotten-Baumann ; B : Production des dérivés N-acylés selon la réaction chimique de Schotten-Baumann à l'échelle industrielle.

1.1.4.2. Synthèse chimio-enzymatique

Une ou plusieurs étapes de biocatalyse peuvent être incluses dans une voie de synthèse chimique. Les différentes étapes peuvent avoir lieu de façon successive ou en one pot. Cette association d'approches permet de combiner les avantages de chacune d'elle tout en limitant leurs inconvénients respectifs. Les deux principaux avantages de cette stratégie de synthèse sont la réduction du nombre d'étapes et du temps de production, car aucun produit intermédiaire ne doit être purifié. Le coût du procédé et la demande énergétique sont alors diminués pour des rendements globaux généralement augmentés. Ce type de procédé peut générer des coûts de

développement élevés et des délais longs, car les conditions des étapes chimiques et enzymatiques doivent être compatibles et optimisées. L'activité et la stabilité des catalyseurs mis en jeu doivent être conservées pendant tout le processus. Le recyclage des cofacteurs enzymatiques est à prévoir. L'absence de procédure de purification intermédiaire augmente le risque de formation de produits secondaires. En effet, des sous-produits peuvent se former si une réaction croisée a lieu entre des réactifs impliqués dans des étapes catalytiques différentes (Rudroff et al., 2018).

L'équipe de Goujard a synthétisé la N-arachidonoyl glycine en utilisant la voie chimio-enzymatique. La glycine est modifiée par l'ajout d'un groupement *tert*-butyle. La réaction d'acylation est catalysée par la lipase B de *Candida antarctica,* puis le groupement *tert*-butyle est clivé par voie chimique. La présence du groupement *tert*-butyle conduit à un taux d'acylation par l'acide arachidonique de 52 % tandis qu'un taux de moins de 6 % est obtenu pour l'acylation de la glycine en son absence (Goujard et al., 2004). Une autre équipe a étudié l'impact de l'ajout d'un groupement carbobenzyloxy sur un acide aminé avant son acylation : ce groupement améliore la solubilité de l'acide aminé et permet d'obtenir un rendement de 92 % lors de l'acylation de la N-Cbz-L-sérine catalysée par les lipases immobilisées de *Candida antarctica* et de *Mucor miehei*. Le rendement d'acylation de la L-sérine est inférieur à 10 % dans les mêmes conditions opératoires (Valivety et al., 1997).

La synthèse des tensio-actifs géminés à base d'arginine est un exemple de synthèse chimio-enzymatique. Ils sont constitués de deux entités N-α-lauroyl-L-arginine éthyl ester monohydrochloride reliées entre elles par une chaine α-ω-diaminoalcane au niveau des groupes α-carboxyliques des résidus arginine (Figure I.5). Pour éviter l'inhibition de l'enzyme par l'agent espaceur, la première étape de synthèse de ces tensioactifs anioniques est une étape chimique où l'ajout de la chaine alcane sur la première entité se fait par acylation spontanée, dans un milieu solvant, à une température correspondant au point de fusion de la chaine alcane. La deuxième étape est la réaction enzymatique catalysée par la papaïne permettant l'ajout de la seconde entité sur le groupe amino aliphatique libre du dérivé formé lors de la première étape (Piera et al., 2000).

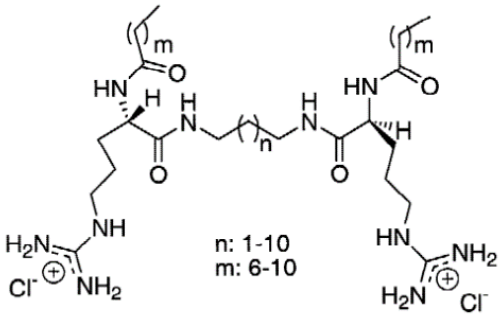


Figure I.5 : Exemple de structure des tensio-actifs cationiques gémini à base d'arginine (Piera et al., 2000).

1.1.4.3. Synthèse enzymatique en milieu solvant

Les dérivés acylés d'acides aminés peuvent également être produits en utilisant uniquement des enzymes pour catalyser la réaction. Ces biocatalyseurs sont naturellement présents pour l'activité métabolique chez les organismes vivants, allant du micro-organisme aux êtres vivants supérieurs. Aux débuts de la catalyse enzymatique, les enzymes étaient parfois présentées comme des catalyseurs coûteux, au potentiel d'action limité à la conversion de leur substrat naturel et nécessitant de respecter leurs conditions physiologiques. Toutefois les avancées scientifiques et technologiques ont permis en quelques années un élargissement du champ d'action des enzymes et leur utilisation intensive à l'échelle industrielle (Sheldon et al., 2018). Les enzymes présentent des propriétés intéressantes de spécificité de substrat et de sélectivité. Lors d'une réaction enzymatique, le plus souvent, il n'est donc pas nécessaire de protéger/déprotéger les groupements non ciblés ni d'utiliser des substrats activés. Ces propriétés évitent également la formation de sous-produits non-désirables.

Les enzymes les plus utilisées à l'échelle industrielle sont les hydrolases, parmi lesquelles les lipases ou les triacylglycérol acylhydrolases (EC 3.1.1.3) constituent la famille la plus représentée. Facilement accessibles et peu coûteuses, les lipases sont très connues pour catalyser la formation ou l'hydrolyse des triglycérides *in vivo.* Elles sont par ailleurs décrites comme étant capables de catalyser d'autres types de réactions dans des solvants non-aqueux (Fatima et al., 2020). Ainsi, de nombreuses études ont montré leur capacité à catalyser la synthèse ou l'hydrolyse d'une grande diversité de molécules, dont les dérivés acylés d'acides aminés font partie (van Rantwijk et al., 2004). Dans cette famille d'enzymes, la lipase B de *Candida antarctica* est la plus étudiée (Zhao et al., 2008).

Les conditions réactionnelles ont un impact important sur les performances catalytiques des enzymes. Le choix du milieu réactionnel doit tenir compte de la solubilité des substrats et des produits, mais aussi éviter la dénaturation de l'enzyme. Récemment, la prise en compte de l'impact environnemental des solvants a favorisé l'émergence de solvants verts tels que le CO_2 supercritique, les liquides ioniques ou les solvants eutectiques profonds dans des procédés catalysés par des lipases. La quantité d'eau dans le milieu de synthèse doit également être maîtrisée. En effet, la réaction d'estérification ou d'amidation catalysée par ces enzymes est réversible. Ainsi, une quantité d'eau trop importante dans le milieu réactionnel favorise les réactions hydrolytiques au détriment de la synthèse d'ester ou d'amide (Dorr et al., 2018). Par ailleurs, l'eau présente dans le milieu peut limiter la diffusion des substrats vers le site actif à cause de la formation de plusieurs couches d'hydratation à la surface de l'enzyme (Mora-Pale et al., 2007). D'un autre côté, une quantité minimale d'eau est nécessaire au maintien de la flexibilité conformationnelle de l'enzyme et de son activité (Clark et al., 2004). Tout ceci justifie les travaux entrepris pour contrôler la quantité d'eau libre dans les milieux réactionnels. On peut citer l'utilisation de tamis moléculaire ou de polymères super-absorbants permettant l'absorption et la rétention des molécules d'eau produites en cours de réaction ou encore le développement de procédés sous pression réduite permettant l'élimination de l'eau en continu (Tufvesson et al., 2007 ; Husson et al., 2009 ; Alves et al., 2014 ; Paludo et al., 2015 ; Nguyen et al., 2018). L'immobilisation des lipases sur des supports très hydrophobes réduit le contact entre l'eau et les enzymes et améliorent les performances enzymatiques (Graebin et al., 2012).
Dans la littérature, au cours des dix dernières années, de nombreux articles rapportent l'utilisation des lipases pour catalyser la synthèse d'esters (1900 articles scientifiques répertoriés sur Web of Science) ; le nombre de travaux portant sur la formation de liaisons amides catalysée par ces enzymes est quant à lui beaucoup plus limité (une cinquantaine d'articles scientifiques répertoriés). Quelques exemples seront cités ici pour illustrer la diversité de conditions mises en œuvre pour ces réactions et les performances obtenues. Lors de l'étude de Soo portant sur la synthèse de N-ε-palmitoyl-L-lysine et de N-ε-oléoyl-L-lysine catalysée par la lipase de *Rhizomucor miehei*, l'effet de la température, du solvant, du rapport molaire acide/amine, de la quantité d'enzyme et de la quantité de tamis moléculaire ajouté ont été étudiés dans le cadre d'un plan d'expérience. Les conditions optimales déterminées pour les deux réactions sont similaires en ce qui concerne la température (aux alentours de 62 °C), le solvant (hexane), le rapport molaire des substrats (3/1) et la quantité d'enzyme (62 mg/mL). Toutefois, la quantité optimale de tamis moléculaire varie du simple au double, d'une réaction à l'autre : respectivement 26,7 mg/mL et 53,3 mg/mL pour la synthèse de N-ε-oléoyl-L-lysine

et de N-ε-palmitoyl-L-lysine. Les auteurs expliquent ce résultat par les caractéristiques physiques et chimiques des deux donneurs d'acyle qui influencent l'état d'hydratation de la lipase et par la suite son activité (Soo et al., 2004).

Des travaux menés au LRGP ont montré une sélectivité identique de la lipase B de *Candida antarctica* lors de l'acylation de la lysine. Cette sélectivité stricte vis-à-vis de la fonction amine ε a pu être démontrée grâce à des analyses par spectrométrie de masse par fragmentation. La réaction de N-acylation de la lysine par l'acide oléique en excès est menée à 55 °C sous agitation (250 rpm), dans du 2-méthyl-2-butanol (2M2B) préalablement déshydraté par des tamis moléculaires, en présence de 10 g/L de lipase immobilisée. Le milieu réactionnel est par ailleurs supplémenté en triéthylamine en large excès afin de favoriser la forme neutre des fonctions amines de la lysine (Dettori et al., 2018a). Des conditions similaires ont été appliquées pour l'acylation des tripeptides Ser-Tyr-Lys, Lys-Tyr-Ser et Leu-Gln-Lys-Trp par l'acide oléique, catalysée par la lipase B de *C. antarctica*. Seul le groupement amine ε de la lysine est acylé au cours de la réaction (Ferrari et al., 2014). L'acylation du tripeptide Lys-His-Ala par l'acide laurique en présence de la lipase immobilisée de *C. antarctica* a été étudiée dans le but de produire un composé multifonctionnel aux propriétés tensioactives, antibactériennes et antioxydantes. Après un travail d'optimisation, les conditions suivantes ont été retenues : 2M2B comme solvant, température réactionnelle de 55 °C, rapport molaire acide/tripeptide de 4 et 72 h de temps de réaction. Ces conditions réactionnelles permettent un rendement de conversion de 75 % (Yu et al., 2020).

D'autres milieux comme le dioxyde de carbone supercritique, les liquides ioniques ou encore les solvants eutectiques profonds, ont été explorés pour mener ce type de réaction. Le CO_2 est non toxique, non inflammable et peu coûteux. Il est supercritique à 31 °C et 7,4 MPa, qui sont des conditions propices à l'activité et à la stabilité des enzymes. Le principal inconvénient est la formation d'acide carbonique à partir du CO_2 et d'eau, entraînant une baisse du pH. De plus, le CO_2 supercritique ne favorise pas la solubilisation des molécules hydrophiles, comme les acides aminés, limitant son utilisation pour la synthèse de dérivés acylés d'acides aminés (Stahl et al., 1978). Lors de réactions impliquant la lysine, il a été démontré que ses groupements amines libres produisent des carbamates en présence de CO_2, faisant ainsi baisser le rendement de la réaction (Kamat et al., 1992). L'équipe de Liu a étudié l'amidation du chlorhydrate de vanillylamine (dérivée de la phénylalanine) par l'anhydride palmitique pour la synthèse du palmitoyl vanillylamide utilisé en industrie alimentaire et déterminé les conditions de synthèse optimales : 50 °C, 17 mPa, pH 8, pendant 23 h, avec un rapport molaire de 1/3 chlorhydrate de

vanillylamine/anydride palmitique et catalysée par la lipase Lipozyme IM de *Mucro miehei* à la concentration de 0,5 % (Liu et al., 2007).

L'activité des enzymes peut être augmentée par lyophilisation en présence de sels en quantité importante (Khmelnitsky et al., 1994). Cette observation a conduit les chercheurs à utiliser les liquides ioniques comme solvants pour les synthèses enzymatiques. Même si une amélioration des performances réactionnelles n'est pas forcément constatée, ils peuvent être privilégiés aux solvants organiques en raison de leur caractère non volatil (Kaar et al., 2003 ; Martin et al., 2008). Ainsi l'acylation du dipeptide Lys-Ser catalysée par Novozym 435® (lipase B de *C. antarctica* immobilisée sur une résine acrylique) a été étudiée dans trois milieux : le 2-méthyl-2-butanol, un milieu sans solvant constitué d'acide oléique en excès, et l'hexafluorophosphate de 1-butyle-3-méthyllimidazolium ($[Bmin]^+$ $[PF_6]^-$). Quel que soit le milieu, la sélectivité de la réaction favorise toujours l'acylation de l'amine ε de la lysine. L'utilisation du liquide ionique conduit à l'amélioration de la solubilité du substrat Lys-Ser, à l'augmentation de la vitesse initiale de synthèse et du taux de conversion du dipeptide (Husson et al, 2009).

Les solvants eutectiques profonds sont issus d'un mélange légèrement chauffé de sels d'ammonium ou de phosphonium avec un donneur de liaison hydrogène comme un alcool, un polyol, un acide carboxylique ou un amide. Ils contiennent des fractions non chargées, ce qui les différencie des liquides ioniques, mais ils possèdent des propriétés similaires telles que la non toxicité, la biodégradabilité et une faible volatilité. La synthèse du peptide N-Ac-Phe-Gly-NH_2 à partir de l'ester éthylique de N-acétylphénylalanine et du chlorhydrate de glycinamide en présence de la chymotrypsine, a été étudiée dans différents solvants eutectiques profonds : une conversion presque complète des substrats est atteinte, en utilisant le glycérol ou l'isosorbide comme donneur de liaison hydrogène (Maugeri et al., 2013). Depuis ces travaux, d'autres synthèses de petits peptides ont été étudiées dans des solvants eutectiques profonds ; une seule étude portant sur la synthèse de dérivés acylés est disponible. Différentes conditions réactionnelles ont été étudiées pour l'acylation de la phénylalanine par l'acide undécenoïque en présence de lipases libres ou immobilisées (lipase B de *C. antarctica* libre ou immobilisée, lipase de *Thermomyces lanuginosus* et lipase de M. *miehei*), dans un mélange choline chloride / urée (1/2) ou choline chloride / glycérol (1/2). La production de phénylalanine N-acylée est infime, voire indétectable. De plus, l'utilisation du mélange choline / urée (1/2) entraîne l'apparition de plusieurs composés secondaires (Bourkaib, 2020).

Enfin, il est aussi possible d'utiliser un milieu exempt de solvant à proprement parler. Dans ce cas, l'un des substrats (généralement le donneur d'acyle) est introduit en excès dans le réacteur et solubilise partiellement ou totalement le second substrat. Lors des travaux menés par Montet

et al., la synthèse de N-ε-acyl-L-lysine à partir de lysine et d'huile de soja, catalysée par Lipozyme, dans différents solvants organiques et dans un milieu sans solvant a été étudiée. Un rendement d'environ 35 % a été obtenu dans le pentane, l'hexane ou l'heptane, en 24 h à 90 °C, tandis que le milieu sans solvant a permis d'atteindre 73 % de conversion du substrat limitant, à 90 °C en 7 jours (Montet et al., 1990).

1.1.5. Conclusions

En conclusion, les dérivés N-acylés d'acides aminés sont constitués de deux entités moléculaires (un acide aminé et la plupart du temps un acide gras), liées de façon covalente *via* une liaison amide. Ce sont des molécules appartenant à la famille des tensio-actifs. En fonction de leur composition atomique et de leur structure, ils présentent de nombreuses propriétés justifiant leur production et leur utilisation intensive dans divers secteurs industriels comme l'hygiène, l'agriculture, l'agroalimentaire, l'énergie, etc. A l'échelle industrielle, ces composés sont synthétisés par voie chimique, *via* la réaction de Schotten Baumann. La voie enzymatique est également étudiée depuis de nombreuses années. Sélective, plus ou moins simple à mettre en œuvre dans des milieux organiques ou aqueux, elle semble surtout performante pour l'acylation des chaines latérales des acides aminés comme la lysine, ce qui limite considérablement son champ d'application. Pour contourner cela, des travaux de recherche visant la découverte d'autres enzymes capables de catalyser la réaction de N-acylation en milieu aqueux, sur un éventail élargi de substrats, ont été menés. Ces derniers mettent en avant l'intérêt d'une catégorie particulière d'amidohydrolases qui sont les aminoacylases I. Les connaissances acquises à ce jour sur ces enzymes font l'objet de la partie suivante.

1.2. Enzymes catalysant l'hydrolyse ou la synthèse de liaisons amides en milieu aqueux

Différentes familles d'enzymes incluant les aminoacyl-transférases (EC 2.3.2), les dipeptidases (EC 3.4.13) et les amidohydrolases (EC 3.5.1) peuvent catalyser l'hydrolyse ou la synthèse de liaisons amides en milieu aqueux. Ces enzymes agissent sur la synthèse ou l'hydrolyse des liaisons de type C-N, pouvant inclure les liaisons peptidiques ou d'autres types de liaisons amides. Ces biocatalyseurs ont leur propre régio-, énantio-, et chimio-sélectivité dans des conditions données, ce qui permet d'éviter les étapes de protection/déprotection des groupements non ciblés et de réduire le nombre d'étapes de synthèse et le coût du procédé. Ces

enzymes sont étudiées pour la production de molécules utilisées dans divers secteurs industriels. Les réactions d'acylation ont lieu dans des milieux aqueux et tamponnés (Sheldon et al., 2022).

1.2.1. Enzymes catalysant l'hydrolyse ou la synthèse de liaisons peptidiques en milieu aqueux

Les enzymes agissant sur les liaisons peptidiques sont très variées et constituent la famille EC 3.4. Parmi elles, les dipeptidases (EC 3.4.13.-) ont un intérêt particulier dans le cadre de ce travail, car elles présentent une homologie avec les amidohydrolases étudiées. Ce sont des enzymes capables de catalyser la N-acylation. Quelques-unes d'entre elles sont présentées ici à titre d'exemples.

Une dipeptidase cytosolique non-spécifique (EC 3.4.13.18), codée par le gène *CNDP2*, a été identifiée chez l'homme et chez d'autres mammifères. Elle appartient à la famille des peptidases M20 et joue un rôle dans la régulation de l'arrêt du cycle cellulaire et de l'apoptose. Elle est active en présence de manganèse, hydrolysant les dipeptides, avec une préférence pour les dipeptides hydrophobes comme la carnosine (β-Ala-His), Cys-Gly ou Cys-Ala (Kaur et al., 2009 ; Pandya et al., 2011 ; Zhang et al., 2014). La structure de l'enzyme humaine a été élucidée et enregistrée sous le code PDB 4RUH (Pandya et al., 2014). En plus de l'action principale évoquée juste avant, elle participe également au métabolisme des acides N-lactoyl-aminés en catalysant leur hydrolyse ou leur formation par protéolyse inverse. Ainsi, il est possible de produire de la N-lactoyl-phénylalanine à partir de lactate (10 mM) et de phénylalanine (25 mM), en présence de $MnCl_2$ (0,1 mM) et de l'enzyme CNDP2 (1 µg), en milieu tampon (Tris-HCl 25 mM, pH 7,4), en une dizaine d'heures et à 37 °C (Jansen et al., 2015).

Dans la même famille, d'autres enzymes ont été caractérisées uniquement pour leur activité hydrolytique, tandis que leur capacité à catalyser la N-acylation n'a jamais été étudiée. Parmi elles, la dipeptidase β-Ala-His (EC 3.4.13.20) catalyse l'hydrolyse préférentielle des liaisons peptidiques de la carnosine, de l'ansérine (β-alanyl-3-méthyl-histidine) et d'autres dipeptides contenant une histidine. Elle est codée par le gène *CNDP1* retrouvé chez de nombreux organismes et son pH optimal est de 8,5 (Teufel et al., 2003 ; Veiga-da-Cunha et al., 2014). Sa structure humaine a été élucidée et son code PDB est 3DLJ (Dong et al., 2008).

L'aminoacylhistidine dipeptidase ou PepD trouvée chez *Vibrio alginolyticus*, dont la structure élucidée est enregistrée sous le code PDB 3MRU, montre une large spécificité de substrats pour les dipeptides contenant une histidine, la plus forte activité ayant été relevée avec le dipeptide His-His (Wang et al., 2008 ; Chang et al., 2010).

Les travaux de Kaur ont permis de découvrir une autre dipeptidase spécifique de l'hydrolyse de Cys-Gly (EC 3.4.13.-), appelée Dug1, présente chez *Saccharomyces cerevisiae*. Cette enzyme fait partie de la famille des peptidases M20 et participe à l'une des dernières étapes d'une voie alternative de la dégradation du glutathion (γ-L-glutamyl-L-cysteinylglycine) (Kaur et al., 2011). Sa structure élucidée est disponible dans la PDB sous le code 4G1P (Singh et al., 2014). Enfin, la dipeptidase β-Ala-Xaa (EC 3.4.13.-) obtenue à partir du gène *pepV* de *Lactobacillus delbrueckii* subsp. *lactis* est peu spécifique et clive une grande variété de dipeptides possédant un résidu β-Ala ou D-Ala terminal, comme la carnosine. Elle peut présenter une légère activité aminopeptidase, car elle catalyse l'hydrolyse de l'acide aminé N-terminal de certains tripeptides (Vongerichten et al., 1994). Sa structure élucidée est disponible sous le code PDB 1LFW (Jozic et al., 2002)

1.2.2. Enzymes catalysant la N-acylation/déacylation en milieu aqueux

1.2.2.1. Les aminoacyltransférases

Parmi les enzymes décrites dans la littérature, les aminoacyltransférases (EC 2.3.2) sont des transférases catalysant le transfert d'une entité possédant une fonction amine, comme un acide aminé par exemple. Ainsi l'acyltransférase de *Mycobacterium smegmatis* catalyse la N-acylation de la (E)-cinnamylamine et de la 2-phénéthylamine par l'acide propanoïque, dans l'eau (Contente et al., 2018). De la même façon, l'acyltransférase de *Candida parapsilosis* permet la N-acylation de la carnosine, par l'acide oléique, dans un tampon phosphate (Husson et al., 2009).

1.2.2.2. Les amidohydrolases

Les amidohydrolases qui sont appelées communément les aminoacylases (EC 3.5.1.-), sont quant à elles, des enzymes hydrolytiques agissant sur les liaisons amides non peptidiques. Leurs substrats sont d'une part les acides aminés N-acylés dont elles catalysent l'hydrolyse, et d'autre part, les amines linéaires et les acides gras dans le cadre des réactions de N-acylation. Quelques travaux montrent le potentiel de ces enzymes catalysent la réaction de N-acylation en milieu aqueux. L'une des premières études rapportant la N-acylation d'acides aminés catalysée par une amidohydrolase a été présentée par l'équipe de Kimura. La synthèse de N-valéryl-L-méthionine a pu être réalisée en milieu aqueux, en présence de l'aminoacylase issue d'une

espèce d'*Aspergillus* (Sigma, USA) (Kimura et al., 2001). Depuis, d'autres travaux ont permis de mettre en évidence l'intérêt de ces enzymes pour catalyser la N-acylation. Les amino-acylases peuvent être regroupées selon leur spécificité de substrat ; les quatre types d'amidohydrolases étudiés lors de ce travail sont décrits ci-après.

1.2.2.2.1. Les N-succinyl-L,L-diaminopimelates désuccinylases (EC 3.5.1.18)

La désuccinylase de l'acide N-succinyl-L,L-diaminopimélique (EC 3.5.1.18), aussi appelée DapE, est une métalloenzyme impliquée dans la voie métabolique du méso-diaminopimélate (DAP) nécessaire à la biosynthèse de la L-lysine et du peptidoglycane constitutif de la paroi cellulaire des bactéries. Elle est présente chez de nombreux micro-organismes comme *Neisseria meningitidis*, *Haemophilus influenzae* ou *Shigella flexneri*. Elle catalyse l'hydrolyse de l'acide N-succinyl-L,L-diaminopimelique (SDAP) en acide succinique et en L,L-2,6-diaminoheptanedioate (DAP) (Figure I.6). L'hydrolyse porte uniquement sur la forme LL du substrat, car elle est inactive envers les formes DD, LD et DL (Born et al., 1998 ; Badger et al., 2005 ; Osipiuk et al., 2021).

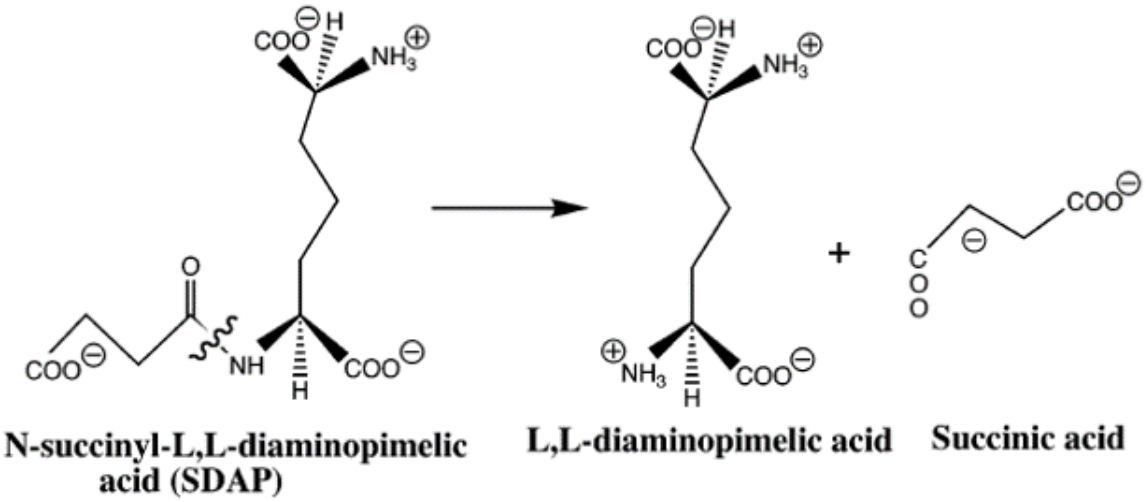


Figure I.6 : Réaction catalysée par la désuccinylase de l'acide N-succinyl-L,L-diaminopimélique (Bienvenue et al., 2003).

L'enzyme produite par *H. influenzae* présente la particularité d'être inactive pour l'hydrolyse des acides aminés acétylés, comme le N-acétyl glutamate, la N-acétyl glutamine, la N-acétyl arginine, la N-acétyl lysine et la N-acétyl ornithine, sur un temps court (30 min). Elle appartient à la famille des peptidases M20 et son pH optimal est de 7. Cette enzyme essentielle à la vie bactérienne est très étudiée en tant que cible pour le développement de nouveaux antibiotiques

face à l'émergence de souches bactériennes résistances aux composés classiquement utilisés (Bienvenue et al., 2003). La structure de cette enzyme a été élucidée à plusieurs reprises et enregistrée sous les codes PDB 3IC1, 3ISZ, 4H2K et 5VO3 (Nocek et al., 2010 ; Nocek et al., 2014 ; Nocek et al., 2018).

1.2.2.2.2. Les pénicillines acylases (ou pénicillines amidases) (EC 3.5.1.11)

Les dérivés de pénicilline sont largement produits par des réactions d'hydrolyse et de synthèse de la liaison amide entre l'acide 6-aminopénicillanique et un donneur d'acyle, en présence de ces enzymes.

L'équipe de Koreishi a mis en évidence la présence d'une enzyme hétérodimérique de 19 et 61 kDa chez *Streptomyces mobaraensis* (Sm-PVA). Elle est activée par le cobalt et son pH optimal est d'environ 8. Elle possède une forte activité hydrolytique vis-à-vis de la capsaïcine (8-méthyl-N-vanillyl-6-nonenamide). Dans un système biphasique eau/n-hexane, les octanoyl-, décanoyl- et lauroyl-vanillylamides qui sont des analogues de la capsaïcine sont synthétisés à partir de la vanillylamine et des acides gras correspondants, avec des rendements supérieurs à 50 %. L'enzyme catalyse l'hydrolyse des acides aminés et des dipeptides N-lauroylés mais aussi la synthèse de dérivés N-lauroylés d'acides aminés et de peptides, en milieu aqueux, en absence et en présence de glycérol. Dans le cas de la lysine, les auteurs mentionnent l'acylation des fonctions amines α et ε (Figure I.7) (Koreishi et al., 2006). Son activité d'acyle transférase permet par ailleurs la production d'amides à partir d'esters méthyliques d'acide carboxylique et de dipeptides (Koreishi et al., 2007). La pénicilline V, la pénicilline G et l'acide 2-nitro-5-phénoxyacétamide benzoïque sont hydrolysés par cette enzyme alors qu'une absence d'activité est relevée vis-à-vis de l'ampicilline et de la céphalosporine. Cette spécificité a amené sa classification dans la famille des pénicilline V acylases (Zhang et al., 2007). Elle catalyse également efficacement la synthèse des antibiotiques de la famille des β-lactamines (Koreishi et al., 2007).

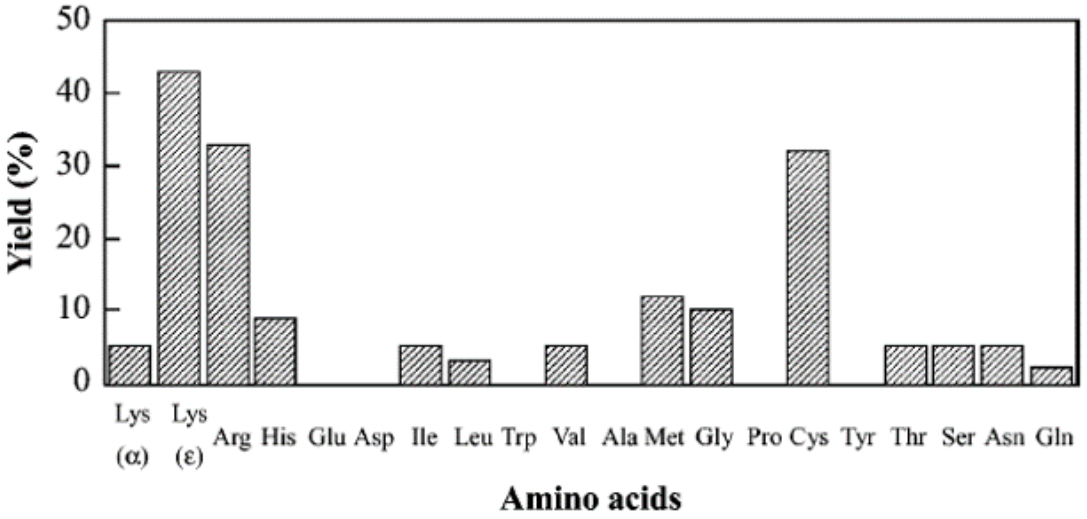


Figure I.7 : Taux de conversion des acides aminés lors de la N-acylation par l'acide laurique. Conditions réactionnelles : acide laurique (7,5 mM), acides aminés (200 mM), 10 unités/mL de l'extrait de *S. mobaraensis* dans du tampon Tris-HCl 100 mM pH 7,5 contenant 0,5 mM de $CoCl_2$ et 78 % de glycérol, à 37 °C, pendant 2 jours (Koreishi et al., 2006).

1.2.2.2.3. Les ε-lysine acylases (EC 3.5.1.17)

Les ε-lysine acylases ou acyl-lysine déacylases sont des enzymes présentes chez de nombreux organismes vivants. Elles catalysent l'acylation de la fonction amine ε de la L-lysine, présentant ainsi la même régiosélectivité que les lipases en milieu solvant. Toutefois, pour ces enzymes, la réaction est menée en milieu aqueux, ce qui représente un avantage certain d'un point de vue environnemental.

Une ε-lysine acylase a été découverte dans le bouillon de culture de *S. mobaraensis* par l'équipe de Koreishi. Cette enzyme (Sm-ELA) d'une taille d'environ 60 kDa est activée par du cobalt ou par du zinc et inhibée par la 1,10-phénanthroline. Son pH optimal est de 8. Elle hydrolyse spécifiquement la liaison amide de divers composés de type N-ε-acyl-L-lysine. Elle catalyse également la réaction réverse de synthèse, en milieu aqueux, avec une préférence pour les donneurs d'acyle de type acides gras aliphatiques comparativement à des donneurs aromatiques. Le rendement est de plus de 90 % dans les synthèses de N-ε-décanoyl-L-lysine, de N-ε-lauroyl-L-lysine et de N-ε-myristoyl-L-lysine, en présence de 10 mM d'acide gras et 0,5 M de lysine (Koreishi et al., 2005a). D'après une recherche d'homologie, l'enzyme appartient à la famille des amidohydrolases métallo-dépendantes de type YtcJ et à la superfamille des hydrolases métallo-dépendantes. La N-ε-lauroyl-L-lysine a été synthétisée dans un tampon Tris-HCl 25 mM à pH 7 et à 37 °C, avec 500 mM de chlorohydrate de L-lysine.

Il a été montré que l'augmentation de la concentration en acide laurique entraîne une diminution du rendement et de la vitesse initiale (Koreishi et al., 2009a). Les dérivés de type N-ε-acyl-L-lysine sont des molécules présentant des propriétés émulsifiantes, hydratantes et antimicrobiennes justifiant leur utilisation dans les produits cosmétiques et les lubrifiants. L'enzyme Sm-ELA permettant la production industrielle de ce type de composés *via* l'acylation sélective de la lysine en position ε en milieu aqueux est protégée par plusieurs brevets déposés par Ajinomoto (Nakanishi et al., 2011 ; Takakura et al., 2019).

Des travaux récents de recherche d'homologie de séquences ont permis d'identifier un gène codant pour une ε-lysine acylase chez *Streptomyces griseus*. Cette enzyme (SgELA) d'une taille approximative de 60 kDa a une activité spécifique de 11,1 µmol/min/mg en présence de 15 mM de N-ε-acétyl-L-lysine à pH 7 et à 30 °C. Toutefois, son instabilité lors de son stockage a limité la poursuite des travaux relatifs à sa caractérisation (Haeger et al., 2023a).

1.2.2.2.4. Les aminoacylases I (EC 3.5.1.14)

Toutes les enzymes de cette classe sont décrites pour leur activité lors de l'acylation d'acides aminés, avec une régiosélectivité et une chimio-spécificité en faveur de leur fonction amine α.

L'aminoacylase 1 est la plus abondante des aminoacylases humaines. Elle a été largement étudiée dans le cadre de travaux portant sur l'influence de son niveau d'expression et de la présence de mutations entraînant le développement de diverses pathologies (van Coster et al., 2005 ; Shi et al., 2013 ; Wei et al., 2014 ; Chen et al., 2021 ; Wang et al. 2022). Cette enzyme cytosolique de 65,11 kDa appartient à la famille des peptidases M20. Sa structure cristallisée a été obtenue partiellement sous la forme d'un mutant T347G. Elle catalyse l'hydrolyse de la N-acétyl-L-méthionine à pH 7,4 dans un tampon borate/phosphate (Lindner et al., 2003). La spécificité de cette enzyme envers divers N-α-acyl-L-méthionine a été également étudiée : l'enzyme a une préférence pour les fragments acyles à chaine droite comparativement aux fragments acyles ramifiés ou benzoylés (Lindner et al., 2008).

Dans le même temps, deux aminoacylases I provenant de *S. mobaraensis* ont été décrites par l'équipe de Koreishi. La première enzyme d'une taille approximative de 100 kDa et au nom inconnu est activée par une addition de cobalt et est inhibée par la 1,10-phénanthroline. Son pH optimal est compris entre 5,5 et 7,5 et sa température optimale est de 60 °C. Elle présente une large spécificité de substrat pour les acides aminés N-acétylés, à l'exception de la N-ε-acétyl-L-lysine et de la N-ε-acétyl-L-proline (Tableau I.1). Cette enzyme catalyse la désacétylation de l'acide 7-aminocéphalosporanique (7-ACA) et de la céphalosporine C ainsi que la synthèse en

milieu aqueux d'acides aminés féruloylés, à l'exception de la proline et de la tyrosine. La meilleure conversion d'acylation entre la lysine et un dérivé d'acide férulique est obtenue avec l'acide hydro-cinnamique avec 43,2 % de conversion (Koreishi et al., 2005b).

Tableau I.1 : Activité hydrolytique de l'aminoacylase au nom inconnu de *S. mobaraensis*. a : une unité correspond à la quantité d'enzyme nécessaire pour hydrolyser 1 µmol de substrat dans du tampon Tris-HCl 50 mM, pH 7,5 en une minute à 37 °C. Concentration en substrats de 15 mM, ou b : 1 mM ou c : 2 mM, en présence d'un aliquot de 20 µL d'enzyme purifiée pendant 20 min. (Koreishi et al., 2005b).

Substrate	Specific activity (units[a]/mg)	Substrate	Specific activity (units[a]/mg)
N-Acetyl-L-arginine	181	*N*-Acetyl-L-proline	0
N-Acetyl-L-histidine	238	*N*-Acetyl-L-tryptophan	200
Nα-Acetyl-L-lysine	41	*N*-Acetyl-L-tyrosine	158
Nε-Acetyl-L-lysine[b]	0	*N*-Acetyl-L-valine	128
N-Acetyl-L-aspartic acid	38	*N*-Acetyl-D-methionine	0
N-Acetyl-L-glutamic acid	110	*N*-Lauroyl-L-methionine[c]	12
N-Acetyl-L-alanine	124	*N*-Chloroacetyl-L-phenylalanine	90
N-Acetyl-L-asparagine	172	*N*-Chloroacetyl-L-tryptophan	118
N-Acetyl-L-cysteine	15	L-Methionyl-L-phenylalanine	<1
N-Acetyl-L-glutamine	201	Glycyl-L-phenylalanine	65
N-Acetyl-L-glycine	24	Z-L-Methionine	1.1
N-Acetyl-L-leucine	188	Z-L-Histidine	<1
N-Acetyl-L-methionine	209	7-ACA[c]	35
N-Acetyl-L-phenylalanine	198	Cephalosporin C[c]	8

La seconde aminoacylase de *S. mobaraensis* (Sm-AA) est une enzyme de masse moléculaire 55 kDa appartenant à la famille des peptidases M20 d'après une étude bio-informatique de recherche de similarité. Elle a un pH optimal de 7,5 et une température optimale de 50 °C. L'enzyme catalyse l'hydrolyse des acides aminés acylés à courte, moyenne ou longue chaine, le meilleur taux de conversion étant obtenu avec la méthionine et l'alanine (Tableau I.2). L'ajout de chlorure de zinc à faible concentration permet d'augmenter l'activité hydrolytique (Koreishi et al., 2009b).

Tableau I.2 : A : Spécificité de substrat de Sm-AA recombinante envers différents acides aminés N-acétylés et N-lauroylés ; B : Spécificité de substrats de Sm-AA recombinante envers différents dérivés de type N-acyl-L-méthionine (Koreishi et al., 2009b).

A

Amino acid	Specific activity (U[a]/mg)	
	N-Acetyl-amino acid[b]	*N*-Lauroyl-amino acid[c]
L-Arg	66	4
L-His	59	6
L-Lys	0	14
L-Asp	8	—
L-Glu	0	—
L-Ala	194	96
L-Asn	31	6
L-Cys	412	17
L-Gln	14	—
Gly	8	34
L-Ile	—	0
L-Leu	60	0
L-Met	530 (173[c])	72
L-Phe	19	0
L-Pro	0	0
L-Ser	—	36
L-Trp	0	0
L-Tyr	0	0
L-Val	24	0

B

Substrate[a]	Specific activity (U[b]/mg)
N-Acetyl-L-Met	173
N-Butyryl-L-Met	169
N-Hexanoyl-L-Met	175
N-Octanoyl-L-Met	251
N-Decanoyl-L-Met	89
N-Lauroyl-L-Met	72
N-Myristoyl-L-Met	106
N-Parmitoyl-L-Met	105

Une seconde amidohydrolase active vis-à-vis des dérivés acylés d'acides aminés a été découverte chez *Streptomyces griseus* grâce une étude bio-informatique. Cette enzyme (SgAA) présente une taille de 50 kDa. Sa spécificité hydrolytique a été étudiée en présence de divers substrats à 15 mM, à pH 7 et 50 °C, pendant 5 min. Il a été montré que l'hydrolyse de nombreux acides aminés acétylés est possible, avec une préférence pour la N-acétyl-L-méthionine, la N-acétyl-L-alanine et la N-acétyl-L-arginine. L'activité hydrolytique diminue lorsque la chaine du donneur d'acyle augmente. Aucune activité dipeptidase n'a été détectée avec les tripeptides Ala-Phe-Ala et Phe-Ala-Ala. L'ajout de cobalt ou de zinc à 1 mM entraîne une diminution de l'activité spécifique de l'enzyme : la concentration en ions métalliques est certainement trop élevée, car l'enzyme est identifiée comme une métalloenzyme appartenant à la famille des peptidases M20 d'après une recherche d'homologie de séquences. La capacité de l'enzyme à catalyser la synthèse des acides aminés lauroylés a été étudiée dans un tampon aqueux Tris-HCl 100 mM, à pH 7 et dans les conditions suivantes : l'acide aminé et l'acide laurique à 100 mM, à 40 °C, pendant 24 h. Seul le produit lauroyl-L-méthionine a été obtenu (Haeger et al., 2023a).

Deux aminoacylases ont été décrites chez *Mycobacterium smegmatis*, sans précision sur leurs séquences protéiques. La première présente une activité hydrolytique envers les dérivés N-acylés d'acides aminés à courte chaine tandis que la seconde catalyse l'hydrolyse de dérivés N-acylés à longue chaine (Matsumoto et al., 1972). Récemment, une recherche d'homologie avec NCBI a permis d'identifier une séquence protéique codant pour une aminoacylase produite par *M. smegmatis* MKD 8. Les études de caractérisation de cette enzyme (MsAA) ont été menées avec des substrats à 15 mM dans un tampon de Tris-HCl 100 mM, à pH 7, à 50 °C et pendant 3 jours. Elle présente une taille d'environ 50 kDa et est activée par les ions zinc. Cette enzyme présente une activité préférentielle d'hydrolyse et de synthèse pour les acides aminés acétylés comparativement aux acides aminés lauroylés, à l'exception de la lauroyl-L-méthionine, où l'activité de l'enzyme envers la lauroyl-L-méthionine est similaire ou supérieure à celles mesurées envers les acides aminés acétylés. Elle possède également une spécificité pour les substrats acylés en position α, ce qui permet de classer cette enzyme dans les aminoacylases I. Ainsi l'hydrolyse de la N-α-acétyl-L-lysine est possible, contrairement à celle de la N-ε-acétyl-L-lysine (Haeger et al., 2023b).

Par ailleurs, une nouvelle aminoacylase de *Paraburkholderia monticola* a été décrite. Elle présente une grande stabilité à la température et au pH. En effet, elle reste active dans une large gamme de pH allant de 4 à 12 et se dénature à partir de 80 °C. Cette enzyme (PmAcy) d'une taille de 47 kDa est activée par l'ajout de Zn, Co ou Mn. L'étude de sa spécificité de substrat a

été menée en présence de 50 µM d'ions métalliques, avec des concentrations en substrat de 15 mM, dans un tampon borate Na-borate 100 mM, à pH 9 et à 50 °C. Les résultats montrent que la meilleure activité spécifique hydrolytique est obtenue avec la N-lauroyl-L-alanine, en présence de cobalt. L'activité de l'enzyme pour la réaction inverse de synthèse de cette molécule est quant à elle presque nulle. L'activité catalytique de l'enzyme pour la synthèse de dérivés acylés d'acides aminés a été étudiée de façon plus approfondie. Une préférence pour les donneurs d'acyle à longue chaine comme les acides gras C16 ou C18, et pour les acides aminés tels que la L-histidine, la L-arginine, la L-leucine ou la L-phénylalanine a été observée, démontrant ainsi une affinité particulière pour les substrats hydrophobes (Figure I.8) (Haeger et al., 2024a).

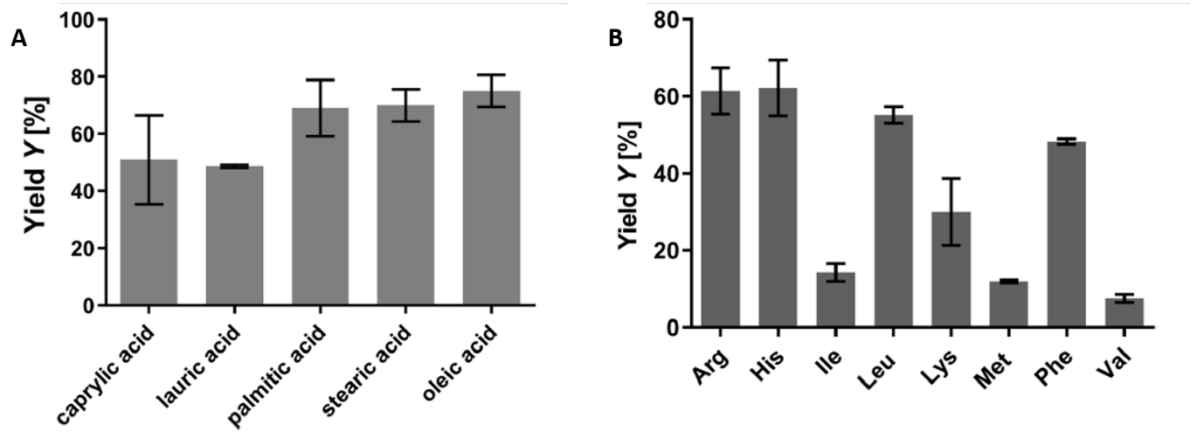


Figure I.8 : Rendements de synthèse de dérivés N-acylés d'acides aminés catalysée par l'aminoacylase de *P. monticola*. A : N-acylation de la L-phénylalanine avec différents acides gras ; B : N-acylation de divers acides aminés par l'acide laurique. Conditions réactionnelles : 200 mM d'acide aminé, 100 mM d'acide gras, milieu tampon Tris 50 mM, pH 8, 50 °C et 24 h (Haeger et al., 2024a).

1.2.2.3. Les aminoacylases de *Streptomyces ambofaciens*

Au sein de l'équipe BioProMo du LRGP, une activité de type aminoacylase a été détectée dans le surnageant de culture de *Streptomyces ambofaciens*. Des activités spécifiques d'hydrolyse des substrats N-α- et N-ε-acétyl-L-lysine et de synthèse de lauroyl-L-lysine ont été mises en évidence (Figure I.9). L'acylation de petits peptides a par ailleurs été étudiée. Les résultats obtenus pour l'acylation des peptides Lys-Ser, Lys-Tyr-Ser, Ser-Tyr-Lys et Leu-Gln-Lys-Tyr ont permis de mettre en évidence une sélectivité originale différente de celle observée en présence de lipases avec une acylation du groupement amine α en position N-terminale des peptides, indépendamment de la nature de l'acide aminé portant ce groupement amine. Cette

régiosélectivité orientée vers les groupements N-terminaux des acides aminés ou des peptides est actuellement peu étudiée et peu décrite dans la littérature. Elle est pourtant d'un grand intérêt car elle permet d'envisager l'acylation de peptides sans modification de leurs chaines latérales ce qui limite le risque de perte de fonctionnalité. Cette voie de synthèse permet en outre de mener les synthèses en milieu aqueux, ce qui constitue un avantage certain d'un point de vue économique et environnemental (Ferrari, 2014).

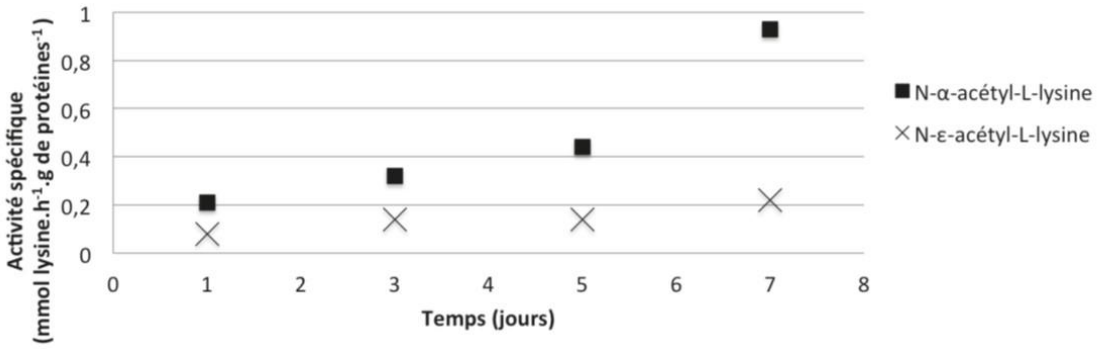


Figure I.9 : Activité spécifique du surnageant de culture de *S. ambofaciens* pour l'hydrolyse de la N-α- ou N-ε-acétyl-lysine, au cours du temps (Ferrari, 2014).

Une purification partielle de l'extrait brut de culture de *S. ambofaciens* par chromatographie ionique a été effectuée pour rechercher les activités d'hydrolyse de la N-α- et de la N-ε-acétyl-L-lysine dans les différentes fractions obtenues. Les résultats ont montré la présence de plusieurs enzymes car l'hydrolyse de la N-α-acétyl-L-lysine et de la N-ε-acétyl-L-lysine sont obtenues en présence de fractions différentes (Tableau I.3). Le génome de *S. ambofaciens* a été comparé aux séquences peptidiques des trois aminoacylases décrites chez *S. mobaraensis* (Koreishi et al., 2007 ; Koreishi et al., 2009a ; Koreishi et al., 2009b), ce qui a permis d'identifier trois séquences génétiques :

- SAM23877_RS06825 (ex-SAM23877_1485) (SamELA) partageant 85 % de similitude avec Sm-ELA.
- SAM23877_RS08060 (ex-SAM23877_1734) (SamAA) partageant 93 % de similitude avec Sm-AA.
- SAM23877_RS04365 (ex-SAM877_0977) (SamPVA) partageant 78 % de similitude avec Sm-PVA.

Une quatrième séquence génétique (SAM23877_6191 ; ex-SAM23877_6191) partageant 65 % de similitude avec Sm-AA a été identifiée suggérant l'existence d'une quatrième enzyme de type aminoacylase chez *S. ambofaciens* (SamAA like). Cette dernière pourrait être homologue à l'aminoacylase non séquencée de *S. mobaraensis* hydrolysant la liaison amide des acides aminés acétylés (Koreishi et al., 2005b).

Des mutants de délétion où l'un des gènes précédemment identifiés chez *S. ambofaciens* est supprimé, ont été produits pour attribuer la spécificité enzymatique à chacun de ces gènes. Cette étude a permis de démontrer que l'activité hydrolytique est majoritairement due aux enzymes SamELA et SamAA. Par ailleurs, les activités hydrolytiques spécifiques de la liaison amide en position α de la lysine et de la liaison amide en position ε de la lysine sont attribuées respectivement aux gènes SAM23877_RS08060 (SamAA) ou SAM23877_RS06825 (SamELA) (Tableau I.3) (Dettori et al., 2018a).

Tableau I.3 : Activités spécifiques d'hydrolyse de la N-α-acétyl-L-lysine et de la N-ε-acétyl-L-lysine relevées dans des fractions issues de la séparation des protéines des extraits bruts des souches sauvage ou mutées de *S. ambofaciens*. ΔSAM23877_1485::scar est la souche mutante sans le gène de SamELA. ΔSAM23877_1734::scar_1734 est la souche mutante sans le gène de SamAA (Dettori et al., 2018a).

Bacterial strain	Aminoacylase activity with N-ε-acetyl-L-lysine as substrate ($mmol_{lys}/(h\ g_{protein})$)		Aminoacylase activity with N-α-acetyl-L-lysine as substrate ($mmol_{lys}/(h\ g_{protein})$)	
	Fraction 3	Fraction 4	Fraction 5	Fraction 6
S. ambofaciens ATCC23877	0.49	0.45	2.9	1.8
Δ *SAM23877_1485::scar*	0.02	0.02	1.1	1.8
Δ*SAM23877_1734::scar*	0.68	0.54	0.04	0

Le potentiel catalytique des aminoacylases de *S. ambofaciens* en termes de sélectivité et de spécificité de substrat en milieu aqueux a été également évalué. Les réactions enzymatiques ont été menées dans un tampon Tris-HCl 25 mM, NaCl 50 mM, à 45 °C et pH 8, avec les différents substrats à hauteur de 100 mM. Les enzymes ont été préalablement purifiées partiellement par précipitation au sulfate d'ammonium et les milieux réactionnels ont été soniqués pour améliorer la solubilité du donneur d'acyle. Les résultats obtenus ont montré qu'il est possible d'acyler la majorité des acides aminés protéinogènes, à l'exception de l'acide aspartique, l'acide glutamique, la proline, la tyrosine et l'histidine, en utilisant l'acide 10-undécénoique comme donneur d'acyle (Figure I.10). Ces travaux ont mis en évidence que les enzymes montrent une énantiosélectivité envers les acides aminés protéinogènes sous forme L. Elles catalysent la réaction d'acylation en utilisant des donneurs d'acyles allant du C2 au C18 en présence et en absence d'insaturation, avec une préférence pour les acides undecenoïque et laurique. La régiosélectivité est en faveur de la position α lors de l'utilisation de la lysine comme accepteur d'acyle, même si l'utilisation d'un donneur d'acyle de longueur de chaine inférieure à 8 atomes de carbone provoque l'apparition minoritaire du produit acylé en position ε. Lors de la synthèse de lauroyl-L-lysine, l'augmentation de la concentration en lysine jusqu'à 150 mM augmente la vitesse initiale de la réaction alors que des concentrations en lysine supérieures à 175 mM et en acide laurique supérieures à 75 mM provoquent une diminution de la vitesse initiale par un

phénomène d'inhibition de substrat. Enfin, l'utilisation d'ions métalliques comme le zinc et plus particulièrement le cobalt permet d'accélérer d'un facteur 7 la vitesse initiale de la réaction de synthèse de lauroyl-L-lysine (Bourkaib et al., 2020a).

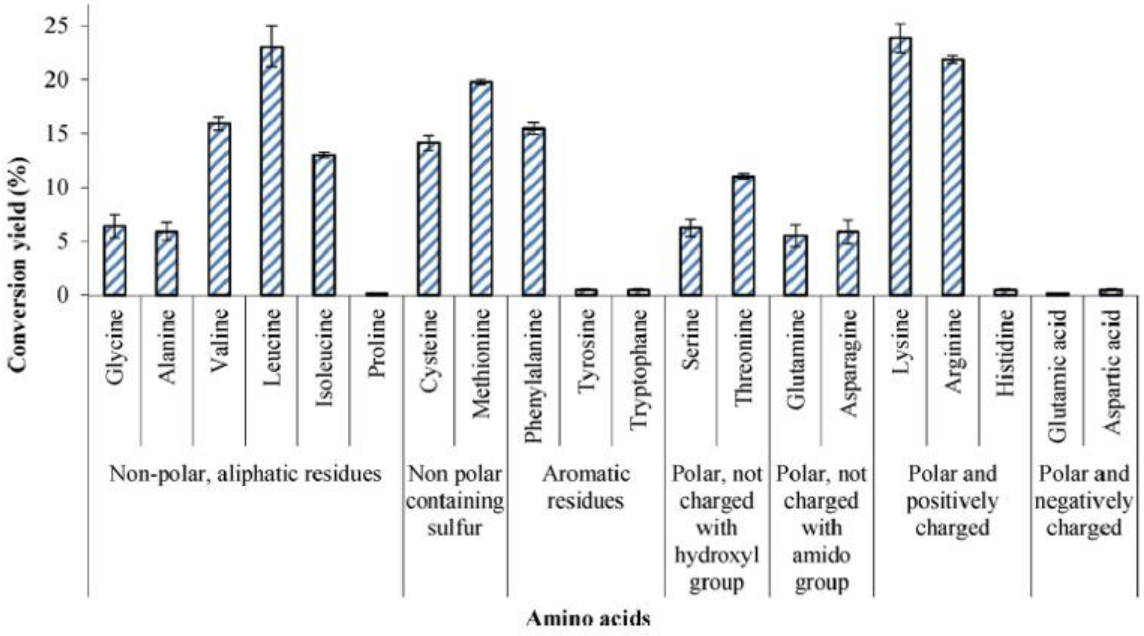


Figure I.10 : Rendements de conversion pour l'acylation de différents acides aminés par l'acide undécénoïque. Conditions réactionnelles : 3 jours, à 45 °C, agitation 250 rpm, milieu tampon Tris-HCl 25 mM, NaCl 50 mM, à pH 8, en présence d'1 g/L d'enzymes (Bourkaib et al., 2020a).

Ce mélange d'enzymes a également été utilisé pour la synthèse de la N-10-undecenoyl-L-phénylalanine dont les conditions opératoires ont été optimisées. La meilleure activité a été obtenue à 45 °C et des valeurs de pH comprises entre 7 et 8. La vitesse initiale de réaction peut être améliorée en augmentant la concentration en phénylalanine ou par l'ajout de cobalt. Une inhibition par le produit a également été observée (Bourkaib et al., 2020b). Des réactions de N-acylation de la lysine ont été menées avec de l'huile de colza ou de l'huile microbienne (issue d'une culture de levure *Cryptococcus curvatus* ATCC 20509) pré-hydrolysée ou non, en milieu aqueux. La meilleure activité des aminoacylases issues de *S. ambofaciens* sauvage est obtenue avec l'huile microbienne hydrolysée dans laquelle les donneurs d'acyle sont sous forme d'acide gras libres. Ces derniers sont majoritairement l'acide palmitique, l'acide oléique et l'acide linoléique. L'analyse des milieux de synthèse a montré une préférence des aminocylases pour l'acide palmitique. La régiosélectivité en faveur du groupement amine α de la lysine est conservée quel que soit le substrat donneur d'acyle (Karayannis et al., 2023).

Une étude portant sur la N-acylation de la lysine par l'acide laurique a été menée en vue de comparer l'activité enzymatique des extraits bruts issus de la souche sauvage à celle d'un mutant issu d'une souche n'exprimant que SamAA, obtenue en délétant les gènes de SamELA

et de SamPVA. Les réactions enzymatiques ont été conduites dans les mêmes conditions opératoires qu'au paragraphe précédent. L'activité spécifique de l'extrait protéique issu de la souche mutante est légèrement inférieure à celle de l'extrait protéique issu de la souche sauvage, ce qui peut s'expliquer par l'absence de SamPVA dont l'enzyme homologue Sm-PVA est capable de catalyser l'acylation des acides aminés avec l'acide laurique (Koreishi et al., 2006). La sélectivité de la N-acylation au niveau de la position α des acides aminés est confirmée par analyse LC-MS/MS. Le profil de spécificité vis-à-vis du substrat accepteur d'acyle est conservé avec la souche mutante. Seuls les taux de conversion obtenus avec la valine, la leucine, l'isoleucine et la thréonine sont légèrement inférieurs à ceux observés avec la souche sauvage en raison de l'absence du gène de SamPVA. Concernant les substrats donneurs d'acyle, le profil de spécificité observé avec l'extrait de la souche mutante suit la même tendance que celle obtenue avec l'extrait de la souche sauvage. Une conversion quasi-nulle de l'acide octanoïque en présence de l'extrait de la souche mutante est observée. L'analyse LC-MS/MS des milieux réactionnels a mis en évidence la présence de deux produits identifiés comme étant les formes α et ε de la N-lauroyl-L-lysine dans le cas de la souche sauvage tandis que seule la N-α-lauroyl-L-lysine est obtenue dans le cas de la souche mutante (Figure I.11) (Bourkaib, 2020). Ce résultat confirme les propriétés de régiosélectivité de SamAA et de SamELA vis-à-vis des positions α et ε de la lysine découvertes lors des études précédentes (Ferrari, 2014 ; Dettori et al. 2018a).

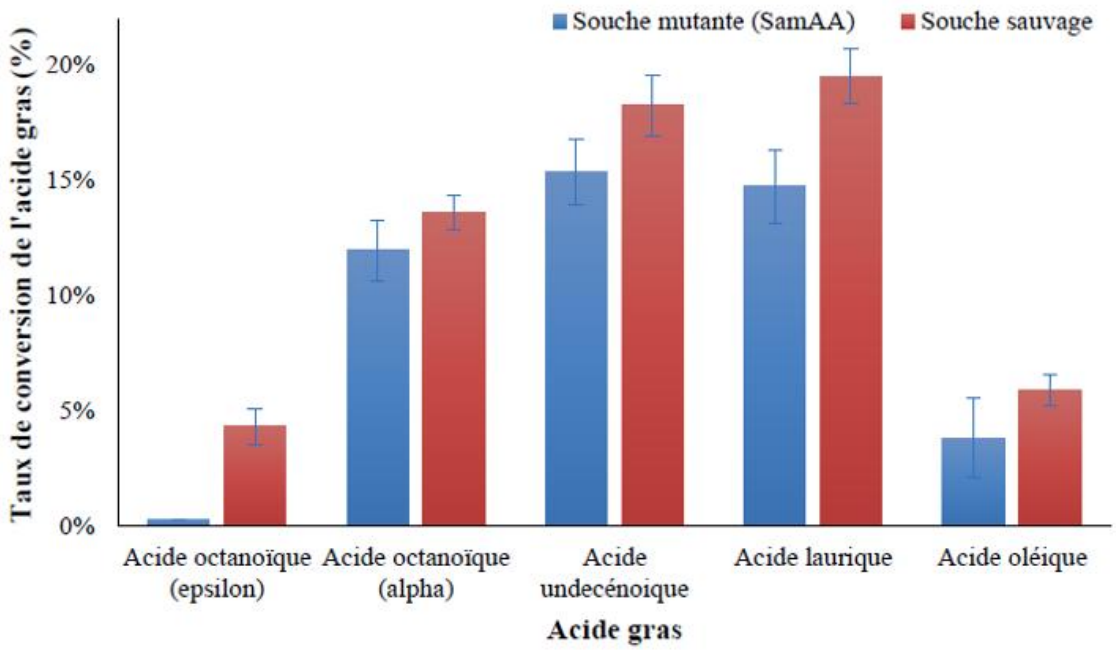


Figure I.11 : Pourcentage de conversion des acides gras lors de la N-acylation de la lysine. Conditions réactionnelles : Substrats à 0,1 M dans du tampon Tris-HCl 25 mM, NaCl 50 mM,

pH 8, à 45 °C, sous une agitation de 500 rpm et en présence de 1 g/L de protéines (Bourkaib, 2020).

Des enzymes solubles, obtenues à partir d'un lysat de culture bactérienne ont été utilisées pour les premières caractérisations enzymatiques des aminoacylases de *S. ambofaciens*. Sous forme soluble, ces enzymes ne peuvent être ni récupérées et ni recyclées. Leur immobilisation a donc été étudiée.

Il existe une grande diversité de configurations d'immobilisation (Figure I.12). Les enzymes peuvent être encapsulées (A), piégées dans une matrice en polymères (B), fixées à un support préfabriqué à l'aide d'un liant chimique (C) ou sous la forme de particules d'enzymes auto-immobilisées (D), comme par les agrégats d'enzymes réticulés (cross-linked enzyme aggregates ou CLEAs). Il est également possible d'immobiliser les cellules entières contenant les enzymes dans des gels ou des matrices de polymères. Quand les enzymes sont fixées sur un support, cette liaison peut réduire l'activité enzymatique si elle a lieu au niveau du site actif ou oriente le site actif vers le support, le rendant ainsi indisponible (E). Cela peut être contourné en utilisant de longues attaches (F) ou en intégrant à la surface de l'enzyme des acides aminés de liaison (G) ou des étiquettes d'affinité positionnant le site actif vers l'extérieur du support (H). L'immobilisation peut se réaliser également *in vivo*, avec la co-expression d'une protéine fusion permettant l'autoaggrégation de l'enzyme à l'intérieur de l'hôte. De plus, différentes enzymes peuvent être co-immobilisées ensemble (I) (Sheldon et al., 2021).

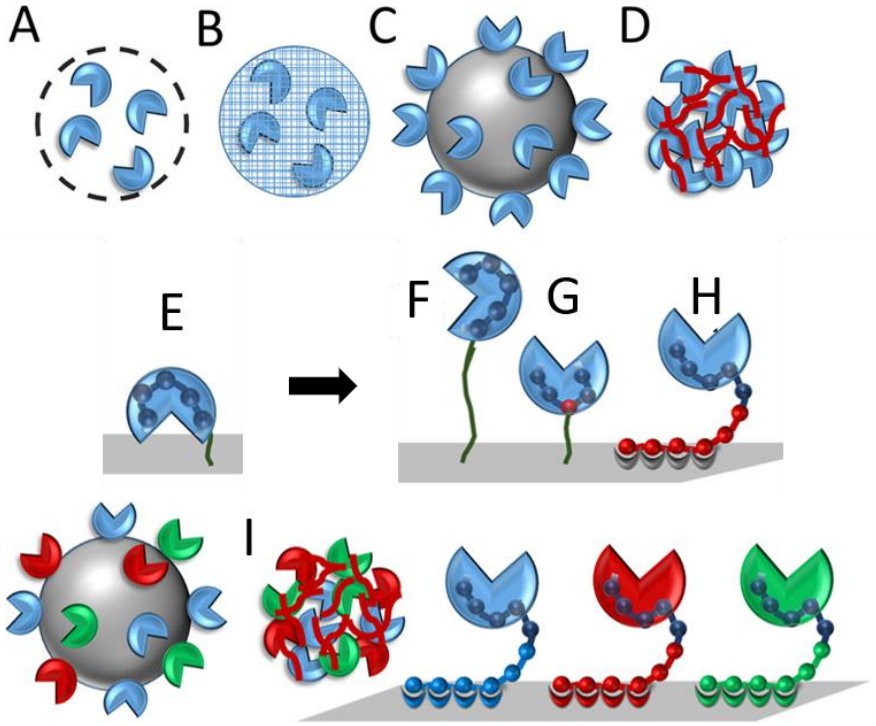


Figure I.12 : Exemples de stratégies d'immobilisation d'enzymes (inspirée de Sheldon et al., 2021).

Une première étude d'immobilisation des aminoacylases de *S. ambofaciens* par chimisorption a été menée impliquant des supports silicatés méso-poreux (SBA-15) fonctionnalisés par deux molécules (glutaraldéhyde ou 3-glycidyloxypropyl-triméthoxysilane). Les résultats ont été comparés à ceux obtenus à l'issue de la simple physisorption des enzymes. L'immobilisation par physisorption a permis d'obtenir un taux d'immobilisation et des activités enzymatiques supérieurs à ceux obtenus avec une immobilisation des enzymes par chimisorption, malgré l'observation d'un phénomène de désorption allant jusqu'à 25 % par cycle, lors du recyclage des enzymes (Dettori et al., 2018b). L'équipe de Bourkaib a poursuivi ces travaux dans une seconde étude comparative portant sur des supports silicatés méso-poreux (SBA-15) et méso-macroporeux. Lorsque les aminoacylases sont immobilisées par physisorption, le taux d'immobilisation et les performances catalytiques sont similaires quel que soit le support. Toutefois la stabilité des enzymes lors des cycles de recyclage est meilleure avec les supports méso-poreux comparativement aux supports méso-macroporeux. Une fonctionnalisation des supports avec différents pourcentages de (3-aminopropyl)triethoxysilane (APTES) conduit à une légère augmentation du taux d'immobilisation pour les deux types de supports ainsi qu'à une augmentation de l'activité enzymatique dans le cas du support méso-poreux. Des tentatives de formation de CLEAs (cross-linked enzyme aggregates) à partir d'enzymes libres ou préalablement immobilisées se sont révélées infructueuses. L'utilisation de support SiHIPE (High Internal Phase Emulsion) ou SBA-si-HIPE (support hybride présentant des méso-pores de type SBA-15 et des macropores de type SiHIPE avec une surface intermédiaire entre les deux) a permis d'augmenter la quantité d'enzyme immobilisée mais l'activité enzymatique mesurée est inférieure à celle obtenue avec les supports méso-poreux et méso-macroporeux (Bourkaib, 2020). Tous ces travaux ont permis de conclure que la meilleure activité enzymatique est obtenue avec l'enzyme immobilisée sur des supports de type SBA-15 fonctionnalisés avec de l'APTES à 50 %. Toutefois, les performances restent inférieures à celles obtenues avec les enzymes libres (Bourkaib et al., 2021).

1.2.2.4. Structure des aminoacylases

La caractérisation des enzymes passe par l'étude de leur potentiel catalytique et de leurs conditions de mise en œuvre mais également par l'élucidation de leur structure tridimensionnelle. Ceci permet en effet d'établir des corrélations entre l'organisation spatiale des enzymes et certaines de leurs caractéristiques comme la sélectivité et la spécificité de substrat, ou encore le mécanisme catalytique. La structure tridimensionnelle peut être élucidée

expérimentalement par cristallographie, RMN ou prédite par des simulations de modélisation moléculaire. A notre connaissance, il existe très peu de structures d'aminoacylases natives disponibles dans la base de données de la PDB. Par ailleurs, la plupart des structures élucidées correspondent à des mutants. Pourtant plus de 80 000 séquences protéiques ont été identifiées comme étant des aminoacylases dans la base de données de séquences protéiques NCBI. Ceci peut s'expliquer par les difficultés rencontrées lors de la production, de la purification de ces enzymes ou lors de l'élucidation de leur structure (Lindner et al., 2003).

Certaines enzymes sont parfois découvertes suite à une recherche d'homologie des séquences génétiques. Les gènes sont insérés dans un système d'expression hétérologue dans le but d'élucider la structure tridimensionnelle de la protéine recombinante produite. Les données obtenues sont ensuite publiées dans les bases de données, comme la Protein Data Bank (PDB), sans qu'aucune étude de caractérisation enzymatique de l'activité ne soit réalisée. Des hypothèses peuvent cependant être émises en étudiant la présence d'éléments conservés caractéristiques de familles d'enzymes ou de domaines. Cette approche a permis de caractériser les enzymes annotées 3ICJ et 3IGH dans la PDB. Ces dernières présentent des séquences conservées relevant de la famille des amidohydrolases 3 ou de la famille des YtcJ-like présentes chez les archées hyperthermophiles *Pyrococcus furiosus* ou *Pyrococcus horikoshii* (Bonanno et al., 2009 ; Patskovsky et al., 2009).

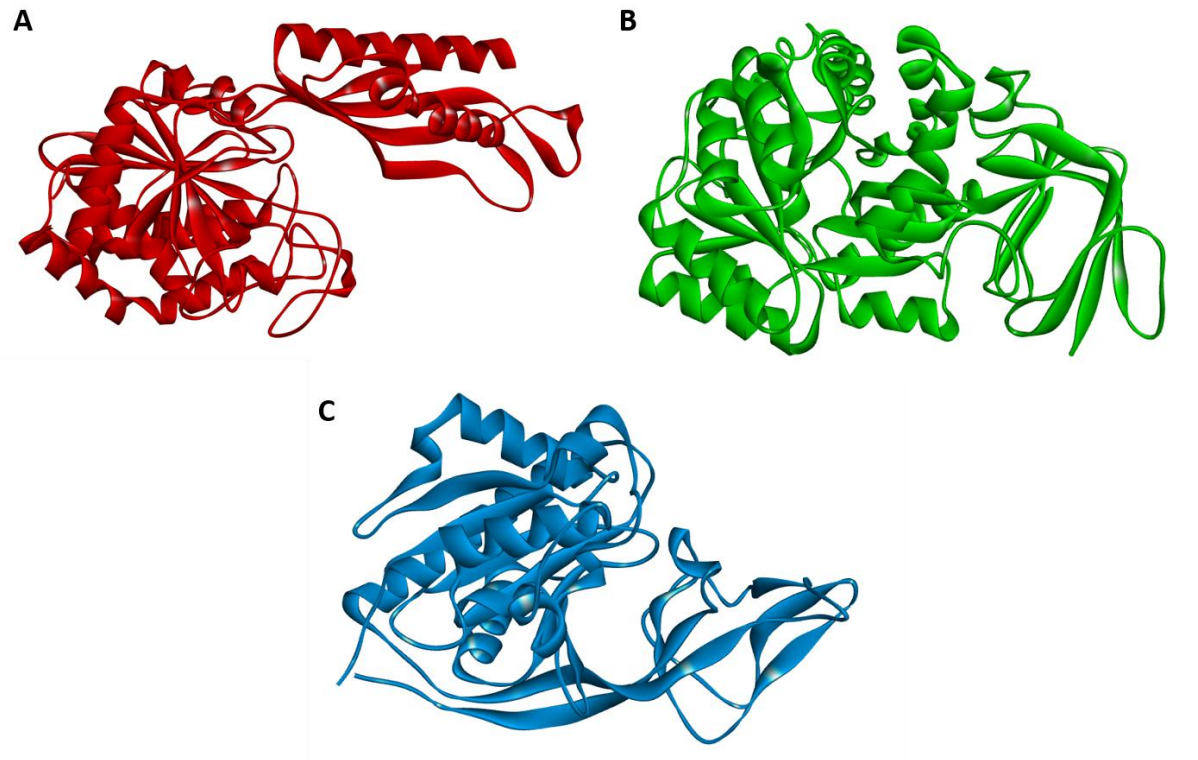


Figure I.13 : Structures tridimensionnelles des monomères constitutifs des aminoacylases. A : Structure du monomère de la N-α-acylglutamine aminoacylase de *Corynebacterium sp.*

appartenant à la famille des peptidases M20 (en rouge) (entrée PDB : 6SLF) (Natsch et al., 2020) ; B : Structure du monomère de l'amidohydrolase (prolidase eah89906) de *Burkholderia lata* (en vert) (entrée PDB : 3N2C) (Xiang et al., 2010) ; C : Structure du monomère de l'aminoacylase III de la souris (en bleu) (entrée PDB : 3NH4) (Hsieh et al., 2010).

Trois structures différentes du monomère constitutif des aminoacylases ont été élucidées, toutes ayant une forme de sifflet (Figure I.13). Le premier monomère (A) est caractéristique de la famille des peptidases M20. Sa structure comporte une première région de forme globulaire et une seconde région de forme cylindrique, les deux étant reliées par une zone de charnière. La première partie est constituée principalement d'un feuillet β entouré de nombreuses hélices α et forme le domaine catalytique. La seconde partie est constituée d'un feuillet β à quatre brins entouré de deux hélices α et forme le domaine de dimérisation. Le site actif est à la limite des deux domaines, sous la zone charnière (Nocek et al., 2010 ; Botelho et al., 2011 ; Natsch et al., 2020). Le deuxième monomère (B) est représentatif de la famille des peptidases M38. Il présente également deux domaines : un petit domaine cylindrique constitué d'un feuillet β et un domaine sphérique plus étendu constitué d'un tonneau $(\beta/\alpha)_8$. Ce dernier est responsable de l'activité catalytique ; il est par ailleurs impliqué dans la dimérisation (Xiang et al., 2010). Le troisième monomère (C) est caractéristique de la famille des peptidases M14. Il présente deux domaines distincts participant à la dimérisation : le premier domaine circulaire N-terminal a une architecture en sandwich αβα dans laquelle le feuillet β est entouré par les hélices α et contient le site d'actif ; le deuxième domaine C-terminal est constitué de brins β et recouvre le site actif de façon à limiter son accessibilité (Bitto et al., 2007 ; Hsieh et al., 2010).

Une autre caractéristique structurale de ces enzymes est la présence ou non d'un ou plusieurs ions métalliques dans le site actif. Ce sont des métalloenzymes quand elles sont capables de fixer des ions métalliques au sein de leur site actif, sinon elles sont considérées comme des N-terminal nucléophile hydrolases (ou Ntn hydrolases). Cette superfamille regroupe entre autres les céphalosporine acylases (EC 3.5.1.93), les homosérine lactone acylases (EC 3.5.1.97) et les pénicilline acylases comme SamPVA (EC 3.5.1.11). Même si elles présentent une faible homologie de séquences, ces enzymes ont une organisation structurelle identique avec un repli caractéristique αββα (Avinash et al., 2016). Les métalloenzymes regroupent principalement les aminoacylases I (EC 3.5.1.14), II (EC 3.5.1.15) et III (EC 3.5.1.114) et nécessitent la présence d'un ou deux ions métalliques divalents et parfois d'une molécule d'eau (Lindner et al., 2003 ; Nocek et al., 2018).

L'organisation quaternaire des aminoacylases, quant à elle, est caractérisée par une grande diversité structurale. En effet, ces enzymes peuvent être monomériques, homo-multimériques

ou hétéro-multimériques. Les structures multimériques ont aussi un nombre variable de sous-unités : deux, quatre ou huit (Haeger et al., 2024b). La structure quaternaire d'une enzyme peut remplir divers rôles, comme l'augmentation de sa stabilité, une aide à son bon repliement, à son activation/inhibition et/ou l'augmentation de son activité catalytique (Chen et al., 2023). Chez les peptidases de la famille M20, la dimérisation de deux sous-unités permet de constituer le site actif car des résidus présents dans les deux chaines peptidiques sont impliqués dans la formation de la totalité du site actif. Des études structurales suggèrent que l'arrivée du substrat au sein du site actif entraîne un changement de conformation au cours duquel les deux domaines de l'enzyme se rapprochent, formant ainsi le site actif (Lindner et al., 2005 ; Nocek et al., 2018).

1.2.2.5. Mécanismes catalytiques

Les aminoacylases sont regroupées en plusieurs catégories selon leur mécanisme catalytique qui dépend de la présence et du nombre d'ions métalliques dans leur site actif.

1.2.2.5.1. Mécanisme catalytique sans ion métallique

Les enzymes appartenant à la superfamille des Ntn hydrolases deviennent actives après le clivage du peptide N-terminal. L'acide aminé catalytique qui peut être une sérine, une thréonine ou une cystéine, est alors accessible pour l'hydrolyse de la liaison amide du susbtrat (Lodola et al., 2012). Toutes les enzymes de cette famille ont un mécanisme catalytique similaire (Figure I.14). L'exemple ci-dessous décrit la proposition de mécanisme de l'hydrolyse de la pénicilline G catalysée par la pénicilline G acylase. Le mécanisme débute par l'attaque nucléophile de la liaison amide par la fonction hydroxyle ou thiole du résidu catalytique et conduit à un intermédiaire stabilisé par des interactions de type liaison hydrogène avec différents acides aminés tels que le couple asparagine-alanine présent dans un trou oxyanionique. Cet état de transition tétraédrique est suivi de la libération du premier produit et la formation d'un acyle enzyme. Puis une attaque nucléophile par une molécule d'eau a lieu pour libérer l'enzyme et le deuxième produit. La réaction inverse de l'hydrolyse (synthèse de liaison amide) est théoriquement possible car toutes les étapes décrites précédemment sont réversibles, en adaptant les conditions réactionnelles (Srirangan et al., 2013 ; Hassan, 2016).

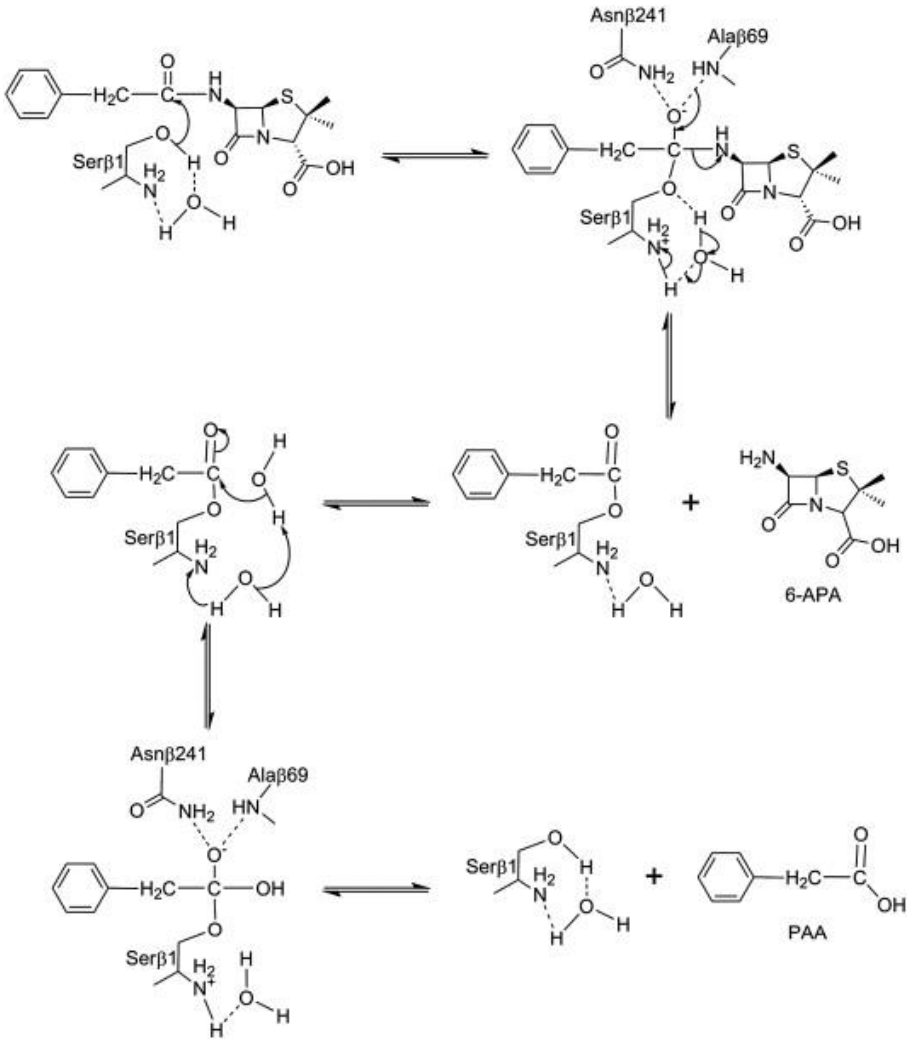


Figure I.14 : Proposition de mécanisme de l'hydrolyse de la pénicilline G en acide phénylacétique et en acide 6-aminopenicillianique, catalysée par la pénicilline G acylase (Srirangan et al., 2013).

1.2.2.5.2. Mécanismes catalytiques avec ion(s) métallique(s)

Les métalloenzymes nécessitent la présence d'un ou deux ions métalliques divalents dans leur site actif. Ce sont principalement des Zn^{2+}, mais ils peuvent être remplacés par du Co^{2+}, du Mn^{2+} ou du Ca^{2+}, selon le type d'enzyme. Ces ions peuvent avoir un rôle dans la stabilité et/ou le mécanisme catalytique de l'enzyme (Henseling et al., 1988 ; Tang et al., 1995 ; Lindner et al., 2003). Les ions métalliques présentent des géométries de coordination bipyramidales trigonales ou tétraédriques déformées et sont liés à quatre ou cinq acides aminés qui sont principalement de type histidine, acide glutamique, acide aspartique, lysine carboxylée ou cystéine, à une distance moyenne de 2 Å. Un acide aminé, le plus souvent un acide aspartique ou une cystéine, fait le pont entre les deux ions métalliques, quand ils sont au nombre de deux.

Ces résidus sont hautement conservés chez les aminoacylases (Haeger et al., 2024b). Une molécule d'eau impliquée dans la catalyse est liée aux ions divalents et à un résidu acide-base essentiel de type glutamate (Davis et al., 2006 ; Nocek et al., 2010).

Les peptidases M20 sont les enzymes appartenant à la sous-sous-classe des aminoacylases les plus étudiées du point de vue de leur caractérisation mécanistique. Ce sont principalement des enzymes homodimériques et binucléaires. Les structures tridimensionnelles de la carboxypeptidase G2 de *Pseudomonas sp.* et de la leucyl aminopeptidase de *Vibrio proteolyticus* (ex-*Aeromonas proteolytica*) ont été élucidées pour étudier leurs domaines de catalyse et de dimérisation ainsi que leur site actif contenant les ions divalents et faire des premières propositions de mécanisme catalytique (Rowsell et al., 1997 ; Desmarais et al., 2007). Dans le même temps, l'aminoacylase 1 humaine et la succinyl diaminopimelate desuccinylase de *Haemophilus influenzae* ont été étudiée par mutagénèse dirigée et par bioinformatique, en utilisant des modèles de relation structure-activité quantitatifs tridimensionnels (QSAR), par une méthode hybride QM/MM ou par une approche d'ingénierie inversée (modélisation de l'état de transition basée sur le produit (PBTSM)). Le but de ces travaux était d'identifier les résidus impliqués dans la fixation des ions métalliques, la catalyse et la dimérisation et d'élucider le mécanisme catalytique des enzymes (Lindner et al., 2005 ; Lindner et al., 2008 ; Dutta et al., 2016 ; Nocek et al., 2018).

Dans les paragraphes suivants, les mécanismes catalytiques d'hydrolyse et de synthèse relevant de la famille des peptidases M20 homodimériques et binucléaires sont présentés. Puis un paragraphe abordera les différences observées entre ce mécanisme et celui des peptidases M20 présentant un seul ion divalent dans leur site actif. Enfin les derniers paragraphes traiteront des différences mécanistiques entre les peptidases M20, M38 et M14.

La première étape du mécanisme catalytique hydrolytique observé dans la famille des peptidases M20 est l'arrivée du substrat dans le site actif. Cette arrivée provoque un important changement de conformation de l'enzyme, qui passe d'une forme ouverte à une forme fermée. Cette hypothèse avait déjà été formulée lors de l'étude de l'aminoacylase 1 humaine en 2003 (Lindner et al., 2003). Ce changement conformationnel s'accompagne du déplacement des résidus conservés Asp274 et Arg276 de la sous-unité A et His206 et Asn263 de la sous-unité B à proximité du centre accueillant les ions métalliques de la sous-unité A (Lindner et al., 2005). Le comportement dynamique de l'auxine amidohydrolase de *Brassica rapa* lors de la liaison à son substrat a été étudié grâce à des simulations de dynamique moléculaire de 10 ns. Les trajectoires obtenues suggèrent un changement conformationnel de l'enzyme tel que celui évoqué précédemment (Simunovic et al., 2010). Cette hypothèse a ensuite été reprise pour

expliquer les résultats obtenus lors de la comparaison des structures cristallographiques de la succinyl diaminopimelate desuccinylase de *H. influenzae* obtenues en présence et en absence des produits issus de l'hydrolyse du substrat naturel de l'enzyme. Les auteurs ont noté une taille plus importante de l'enzyme en l'absence de produits, comparativement à la structure avec ligand (Figure I.15). Lors de la fixation du substrat, le domaine catalytique tourne d'environ 50 °, avec un mouvement de torsion et de flexion au niveau de la zone charnière. Un déplacement de 10 Å au niveau du site actif et pouvant aller jusqu'à 29 Å à l'extérieur de la protéine est observé. Ceci provoque la fermeture du site binucléaire et la formation d'une poche autour du substrat avec l'apparition de nouvelles interactions impliquant les deux chaines protéiques. La figure I.15 montre la conformation ouverte et la conformation fermée de l'enzyme. Le déplacement du résidu His194 de la chaine B vers le site actif du monomère A pour former un trou d'oxyanion stabilisant le fragment succinate est visible (Nocek et al., 2018).

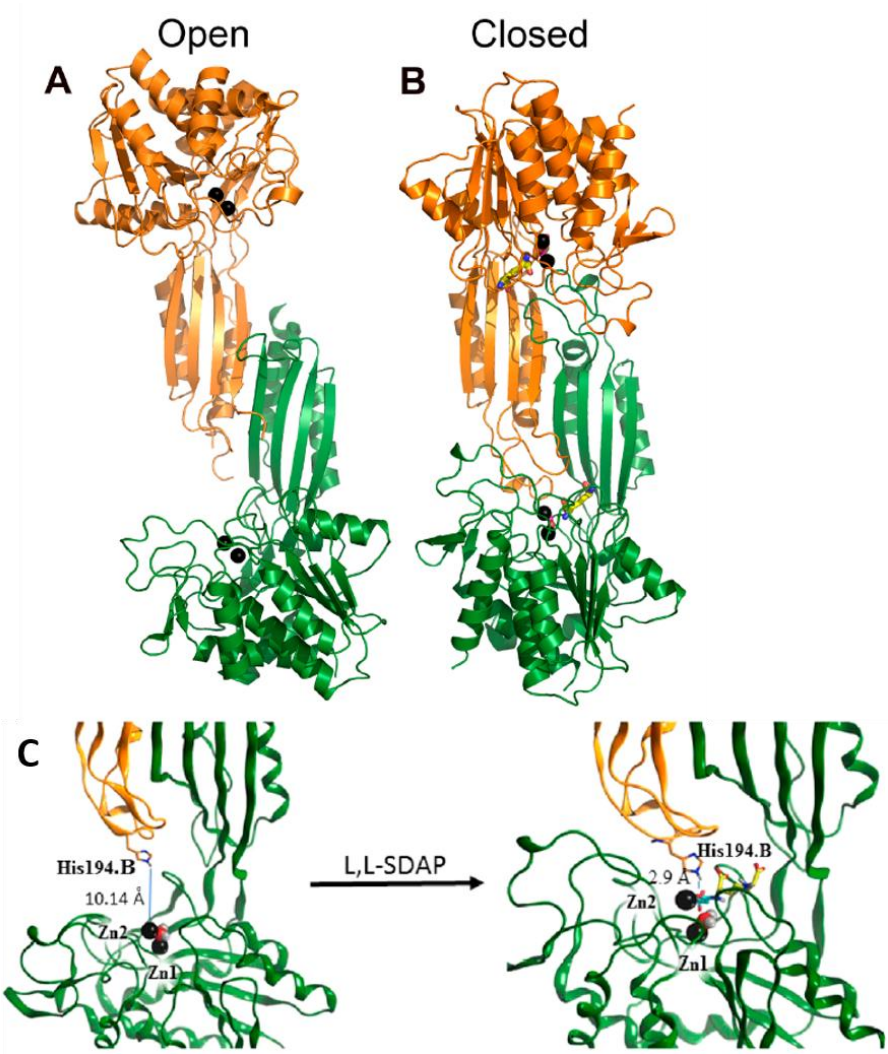


Figure I.15 : Changement conformationnel observé chez l'enzyme HiDapE (A) Diagramme en ruban de la structure dans sa conformation « ouverte » (entrée PDB 3IC1). (B) Structure du complexe enzyme-substrat avec les produits acide succinique (magenta) et acide L-

diaminopimélique (jaune), dans la conformation fermée. Les sous-unités individuelles de l'homodimère sont colorées en vert et en orange. (C) Conformation ouverte et fermée montrant le déplacement du résidu His194.B (Nocek et al., 2018).

En ce qui concerne les interactions avec le substrat, ce dernier est stabilisé par plusieurs interactions entre ses groupements fonctionnels et des résidus constituant le site actif. Le résidu Arg276 de l'aminoacylase 1 est impliqué dans la fixation du substrat (Lindner et al., 2005). Le diaminoheptanedioate et le succinate sont les produits de l'hydrolyse du N-succinyl-(2S,6S)-2,6-diaminoheptanedioate (L,L-SDAP) catalysée par la succinyl diaminopimelate desuccinylase (Figure I.16). Le premier produit/fragment est stabilisé par des interactions entre sa fonction acide terminale et le résidu Asn244.B et entre sa fonction acide proximale et les résidus Arg258.A et Thr325.A. Le groupement acide terminal du second produit/fragment interagit avec les résidus Arg178.A et Tyr197.B. Le substrat est également stabilisé par plusieurs interactions au niveau de sa liaison amide. Ainsi son groupement NH est stabilisé par le résidu Thr325.A appartenant à la boucle de spécificité de l'enzyme (Nocek et al., 2014). Son groupe carbonyle est en interaction avec le résidu His194.B et les deux ions métalliques (Nocek et al., 2018).

L'exemple ci-dessous décrit la proposition de mécanisme catalytique pour l'hydrolyse du L,L-SDAP par la succinyl diaminopimelate desuccinylase d'*H. influenzae* (Figure I.16). La liaison entre Zn2 et le substrat perturbe le pontage de la molécule d'eau liée aux deux ions, qui se déplace en direction du Zn1. Le glutamate catalytique agit comme une base et déprotone la molécule d'eau catalytique. L'ion hydroxyle ainsi formé se situant entre 3 et 4 Å de la liaison amide peut effectuer l'attaque nucléophile responsable de l'hydrolyse. Une supposition est faite que l'arrivée du résidu His194.B dans le site actif de l'enzyme et la formation d'un trou d'oxyanion active la liaison amide et stabilise l'intermédiaire issu de l'hydrolyse.

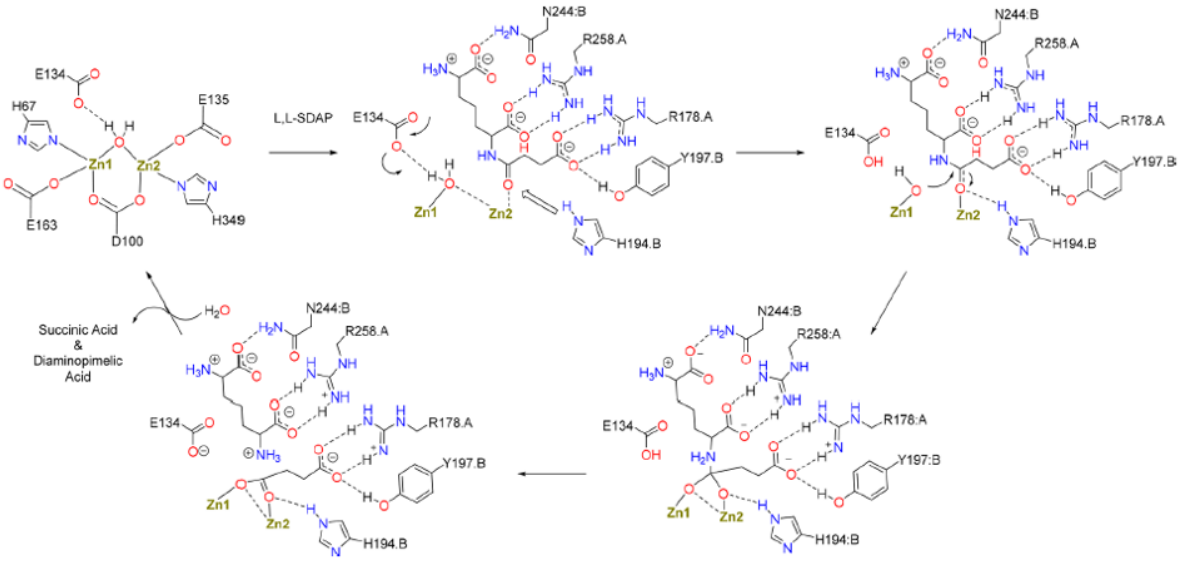


Figure I.16 : Proposition de mécanisme catalytique pour l'hydrolyse du L,L-SDAP par la succinyl diaminopimelate desuccinylase d'*H. influenzae* (Nocek et al., 2018).

Au niveau de cette étape particulière du mécanisme, un angle particulier a retenu notre attention. Il s'agit de l'angle de Bürgi-Dunitz caractérisant le positionnement des substrats lors d'une attaque nucléophile. Des études ont montré que cet angle prend une valeur de 107 ° dans le cas d'une attaque nucléophile impliquant des petites molécules (Bürgi et al., 1972). Cet angle a été également mesuré dans des systèmes moléculaires de grande taille comme les enzymes. Dans ce cas, il peut atteindre des valeurs voisines de 90 °, en raison d'interactions électrostatiques et de Van der Waals et de l'encombrement stérique de l'enzyme provoquant un changement directionnel de l'arrivée du substrat nucléophile (Lodge et al., 1987 ; Radisky et al., 2002 ; Lubkowski et al., 2019).

Lors de l'attaque de la molécule d'eau catalytique, la chaine latérale du résidu Asp100 subit un changement de conformation au cours duquel son groupe carboxylate fait un mouvement de rotation entraînant la rupture de la liaison de coordination avec Zn2 : le pont entre les deux centres métalliques se trouve alors perturbé, ce qui facilite l'activité enzymatique en abaissant l'énergie d'activation (Dutta et al., 2014). Il y a alors formation d'un intermédiaire tétraédrique. Le proton provenant de la molécule d'eau et lié au Glu134.A est transféré vers la fonction amine nouvellement formée, rompant ainsi la liaison amide fragilisée. Il y a alors libération des deux produits et retour de l'enzyme à sa conformation ouverte. Une nouvelle molécule d'eau catalytique peut alors se lier aux atomes de zinc (Dutta et al., 2016 ; Nocek et al., 2018).

Les aminoacylases sont des enzymes majoritairement hydrolytiques, mais elles sont capables de catalyser la réaction reverse de N-acylation. Lors de cette réaction, les deux substrats sont supposés se positionner dans le site actif de la même manière que les deux fragments générés

lors de la réaction d'hydrolyse. Les deux atomes d'oxygène du groupement carboxylate du donneur d'acyle se lient aux deux ions métalliques. Le substrat est stabilisé/activé dans le trou oxyanion. Les fonctions réactives carboxylate et amine des substrats sont chargées et forment une liaison hydrogène. Le mécanisme catalytique supposé est similaire à celui de l'hydrolyse mais se déroule en sens inverse. L'hypothèse d'une stabilisation/activation par le résidu Glu134.A de la fonction amine du substrat accepteur d'acyle, qui joue alors le rôle de nucléophile vis-à-vis du substrat donneur d'acyle, est formulée. Lors de la formation de l'intermédiaire tétraédrique, la fonction amine chargée donne deux protons, un premier au résidu acide-base Glu134.A et un second à l'oxygène du groupement carboxyle qui est lié à Zn1. L'intermédiaire tétraédrique disparaît avec la libération de l'hydroxyle lié à Zn1. Le résidu Glu134.A transfère son proton à l'hydroxyle pour former une molécule d'eau. Le changement de conformation permet la libération du produit (Nocek et al., 2018).

En présence d'un seul ion métallique, le mécanisme catalytique change peu. C'est le résidu His349 qui remplit les fonctions du second ion métallique, pour l'orientation du substrat et l'activation de la fonction carboxyle du donneur d'acyle au niveau du trou oxyanion (Nocek et al., 2010 ; Nocek et al., 2018).

Les peptidases appartenant à la famille M38 présentent un mécanisme catalytique hydrolytique similaire celui proposé pour les peptidases de type M20 (Figure I.17). Le site actif de ces enzymes contient deux ions divalents liés à un hydroxyle. Ce dernier joue le rôle de nucléophile attaquant la liaison amide dont le carbonyle est préalablement activé pour améliorer son caractère électrophile. Un résidu histidine conservé interagissant avec le second ion métallique est à l'origine de cette activation (Xiang et al., 2010). Le résidu jouant le rôle d'acide/base est un aspartate (Xiang et al., 2009).

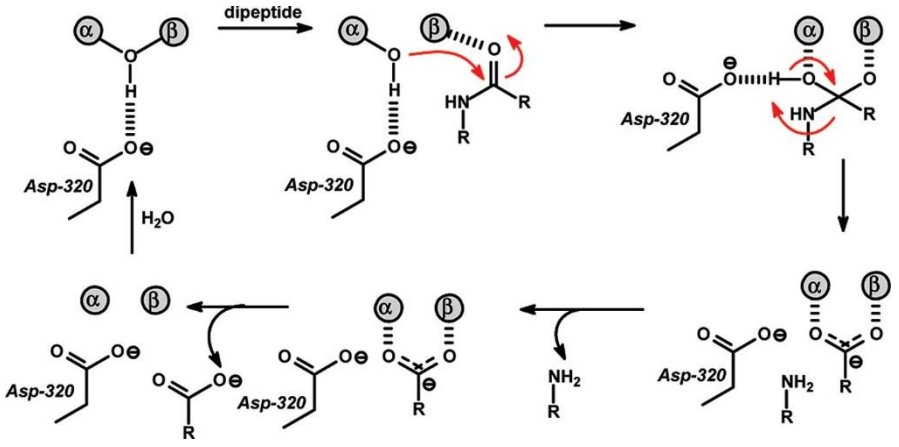


Figure I.17 : Proposition du mécanisme catalytique des peptidases M38 (Cummings et al., 2010).

Les peptidases de la famille M14 sont des enzymes mononucléaires présentant une molécule d'eau dans leur site actif. L'aspartoacylase humaine est parmi les enzymes les plus étudiées de cette famille, car elle est impliquée dans le système métabolique de l'acide N-acétylaspartique dans les mitochondries neuronales. Son absence provoque la maladie neurodégénérative autosomique récessive de Canavan. La présence d'un seul ion métallique implique un mécanisme légèrement différent de celui proposé pour les peptidases M20 (Figure I.18). La sphère de coordination de l'ion Zn^{2+} est composée de trois résidus conservés : His21, Glu24 et His116. Le N-acétyl-L-aspartate est stabilisé par différentes interactions entre ses groupements chargés et les résidus Arg63, Asn70, Arg71, Tyr164, Arg168 et Tyr288. La base catalytique jouant le rôle d'accepteur de proton de la molécule d'eau catalytique est l'acide aminé Glu178. L'ion hydroxyle formé peut alors attaquer la liaison amide du substrat, qui est activée par la présence de l'ion divalent et du résidu conservé Arg63. Il y a formation d'un intermédiaire tétraédrique qui disparaît lors du clivage de la liaison amide, lors d'un transfert de proton au glutamate catalytique. Le résidu Glu178 est impliqué dans ce transfert, jouant un rôle d'acide (Herga et al., 2006 ; Kots et al., 2019). Le résidu Asn117 est glycosylé et joue ainsi un rôle dans le maintien de la stabilité et de l'activité catalytique de l'enzyme (Le Coq et al., 2008).

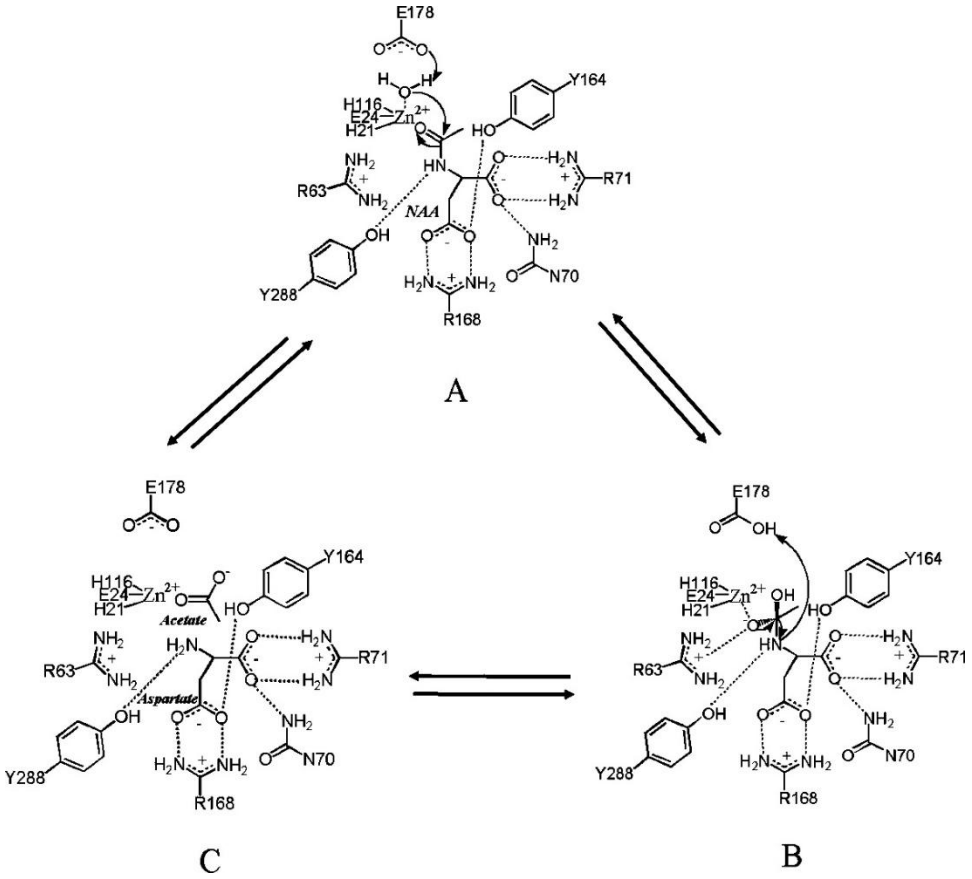


Figure I.18 : Proposition du mécanisme catalytique d'hydrolyse de l'aspartoacylase humaine appartenant à la famille des peptidases M14 (Le Coq et al., 2008).

1.2.3. Conclusions

En conclusion, il existe plusieurs familles d'enzymes catalysant l'hydrolyse ou la synthèse de liaisons amides en milieu aqueux. Les dipeptidases catalysent l'hydrolyse et la synthèse de liaisons peptidiques et parfois la N-acylation en milieu aqueux. La N-acylation/déacylation est aussi catalysée en milieu aqueux par d'autres familles d'enzymes parmi lesquelles sont retrouvées les aminoacyltransférases et les amidohydrolases regroupant les N-succinyl-L,L-diaminopimelates désuccinylases, les pénicillines acylases, les ε-lysine acylases et les aminoacylases I.

Les quatre aminoacylases (SamAA, SamELA, SamPVA et SamAA like) découvertes chez *S. ambofaciens* sont particulièrement intéressantes. En effet, l'une d'elle, SamAA présente une sélectivité en faveur de la N-acylation de la fonction amine α de la L-lysine et d'autres acides aminés et peptides. Une deuxième enzyme, SamELA, catalyse quant à elle la N-acylation de la lysine en position ε. Au vu du potentiel catalytique de ces enzymes, des travaux d'immobilisation ont été entrepris mais ont conduit à une perte partielle de l'activité enzymatique.

La structure tridimensionnelle de quelques aminoacylases a été élucidée. Par exemple, les aminoacylases de la famille des peptidases M20 ont une forme de sifflet, avec la présence d'un domaine globulaire catalytique et d'un domaine cylindrique de dimérisation. L'organisation quaternaire de ces enzymes est variée. Le site actif peut contenir ou non un ou plusieurs ions métalliques divalents, donnant lieu à différents mécanismes catalytiques possibles. Des premiers éléments relatifs à ces mécanismes et aux résidus clé associés ont été identifiés. Le processus global dans le cas de l'hydrolyse repose sur l'attaque nucléophile de la liaison amide par une molécule d'eau activée par un résidu acide-base. Lors de la synthèse, l'attaque nucléophile serait menée par la fonction amine du substrat accepteur d'acyle sur la fonction carboxyle du substrat donneur d'acyle, conduisant à la formation d'un intermédiaire tétraédrique puis à la formation du dérivé acylé.

Ces enzymes n'étant pas disponibles commercialement, la poursuite des travaux visant leur caractérisation nécessite de les produire dans des quantités suffisantes permettant leur mise en œuvre dans des réactions, leur purification et l'élucidation de leur structure. Au vu de la faible quantité basale produite par l'organisme vivant dont est naturellement issue l'enzyme, d'autres méthodes de production doivent être envisagées comme l'expression d'enzymes recombinantes au sein d'organismes vivants génétiquement modifiés.

1.3. Production hétérologue des aminoacylases

Les enzymes peuvent être produites au sein de l'organisme vivant dont elles sont issues ou en utilisant un organisme génétiquement modifié. Il s'agit d'expression dite « homologue » dans le premier cas et d'expression dite « hétérologue » dans le second cas. Les enzymes ainsi produites sont qualifiées de recombinantes.

L'expression de protéines recombinantes est une technique de biotechnologie très répandue, qui s'est développée depuis les premières manipulations d'insertion de nouvelles informations génétiques (Jackson et al., 1972). Elle permet la production et la purification d'une protéine d'intérêt en grande quantité dans le but de la caractériser et éventuellement de la commercialiser. Il y a alors surexpression des protéines recombinantes. La production hétérologue repose sur des techniques de biologie moléculaire, comme le clonage du gène codant la protéine d'intérêt dans un vecteur plasmidique puis l'insertion de ce dernier dans un organisme hôte qui va produire la protéine recherchée. Elle a lieu chez des organismes hôtes procaryotes tels que les bactéries *Escherichia coli*, *Vibrio natriegens*, *Bacillus substilis* et des organismes hôtes eucaryotes tels que les levures *Picha pastoris* et *Saccharomyces cerevisiae*, des champignons filamenteux des genres *Aspergillus*, *Trichoderma*, *Penicillium* et *Rhizopus*, des cellules d'insectes via l'usage de baculovirus, et enfin des cellules végétales ou animales. Le choix d'un hôte hétérologue est une étape importante car chacun possède ses propres avantages et inconvénients qui doivent être pris en compte selon les caractéristiques de la protéine d'intérêt à produire (Pouresmaeil et al., 2023).

Les aminoacylases de *S. ambofaciens* ont déjà été étudiées par Bourkaib (Bourkaib et al., 2020a). Les enzymes obtenues à partir de cultures de *S. ambofaciens* sont produites en petite quantité et ne peuvent qu'être purifiées partiellement, ce qui limite grandement les études portant sur leur caractérisation enzymatique, structurale et mécanistique. La production des aminoacylases par une stratégie d'expression hétérologue permettrait l'obtention d'une quantité plus importante d'enzymes *via* une surexpression protéique et une purification facilitée. La littérature rapporte différents travaux sur la production hétérologue des aminoacylases, dont plusieurs ont déjà été obtenues sous la forme de protéines recombinantes. Ce chapitre décrit les différents hôtes de production utilisés ainsi que les principaux résultats associés.

1.3.1. *Escherichia coli*

Les systèmes bactériens d'expression sont économiques et leur utilisation nécessite peu d'expertise, ce qui fait que ce sont les systèmes d'expression les plus utilisés. Parmi eux, *E. coli* est le micro-organisme utilisé prioritairement pour l'expression et la production de protéines recombinantes, car il est l'hôte le plus abordable et le plus simple à mettre en œuvre (Koehn et al., 2009). C'est un bacille à Gram négatif appartenant à la famille des entérobactéries et possédant un temps de génération rapide d'environ 20 minutes. Il existe une importante diversité génétique et métabolique résultant de la modification de cette bactérie (Rosano et al., 2014).

Dans le cas d'un système d'expression hétérologue procaryote, le gène d'intérêt est inséré dans un plasmide d'expression. Les plasmides sont des molécules d'ADN de petite taille, généralement circulaires, pouvant se répliquer indépendamment à l'intérieur de la cellule hôte. Il existe une grande diversité de plasmides d'expression qui comportent les éléments nécessaires à la production protéique hétérologue : le réplicon, le promoteur, le site de fixation des ribosomes et un système de sélection de souches contenant le plasmide. Ce dernier est généralement un marqueur de résistance à un antibiotique (Singha et al., 2017).

Le réplicon est un segment d'ADN contrôlant le nombre de copies du plasmide à l'intérieur de la cellule hôte. Le nombre de copies est un facteur de sélection pour le choix du plasmide. Augmenter le nombre de copies d'un plasmide permet la plupart du temps d'augmenter l'efficacité de la production de protéines recombinantes, mais ce n'est pas toujours le cas (Lozano Terol et al., 2021). L'équilibre entre la croissance de l'hôte et l'expression en protéines recombinantes est nécessaire, car chaque plasmide supplémentaire entraine une charge métabolique additionnelle qui ne doit pas être néfaste au rendement en protéines recombinantes (Rouches et al., 2022).

Un promoteur fort ayant une forte affinité pour l'ARN polymérase est nécessaire pour augmenter la transcription du gène d'intérêt et produire la protéine à hauteur 10 à 30 % des protéines cellulaires totales. Les promoteurs peuvent être régulés à l'aide d'inhibiteurs ou d'activateurs (Makrides et al., 1996). L'un des promoteurs forts et inductifs les plus utilisés est le promoteur T7 lac présent sur les plasmides de la famille pET. La production de la protéine cible avec ce promoteur peut représenter jusqu'à 50 % des protéines cellulaires totales grâce à la T7 ARN polymérase. Cette enzyme synthétise des ARN cinq fois plus rapidement que la polymérase native d'*E. coli*. En absence d'induction, la bactérie exprime un répresseur Lac, codé par un gène *lacI* présent dans le génome de la bactérie et parfois aussi sur le plasmide. Ce

répresseur se place sur la région opératrice *lacO* présente entre le promoteur et le gène de la T7 ARN polymérase dans le génome de la bactérie, inhibant la synthèse de cette polymérase. La région *lacO* peut être également présente entre le promoteur T7 et le gène de la protéine d'intérêt sur le plasmide, où Lac inhibe l'expression du gène d'intérêt. L'inducteur peut être le lactose ou un analogue du lactose, comme l'IPTG (isopropyl β-D-1-thiogalactopyranoside). Une fois l'inducteur ajouté, ce dernier se fixe au répresseur Lac, empêchant sa liaison à la région *lacO*. L'ARN polymérase d'*E. coli* permet la synthèse de l'ARNm de la T7 ARN polymérase. Cette dernière peut ensuite synthétiser l'ARNm du gène cible (Jia et al., 2016) (Figure I.19).

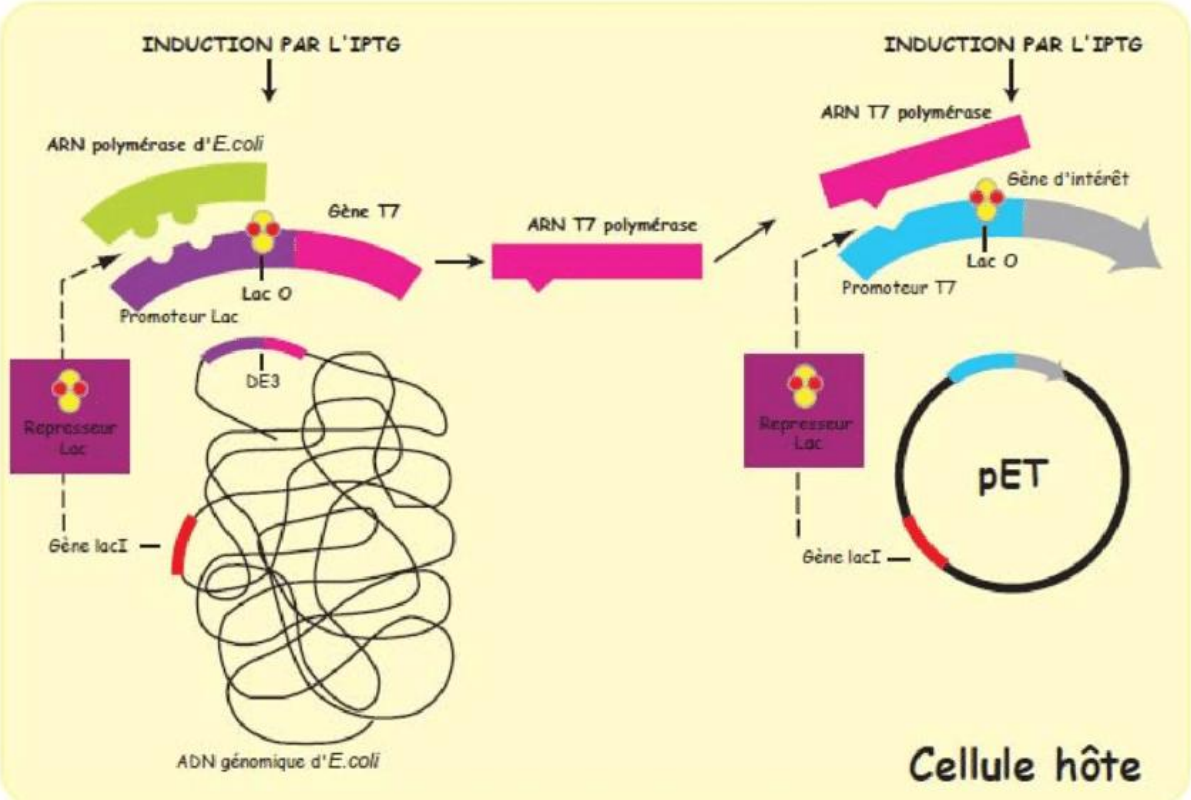


Figure I.19 : Système de production de protéine hétérologue en utilisant *E. coli* et un plasmide pET (Bourgeois, 2011).

Les tags ou étiquettes d'affinité sont composés d'une petite chaine d'acides aminés qui est ajoutée à l'extrémité N ou C des protéines recombinantes. Ils peuvent avoir plusieurs rôles dans l'amélioration de l'expression, l'augmentation de la solubilité et de la stabilité, la diminution de la toxicité sur les cellules hôtes et/ou l'aide à la purification des protéines cibles. Toutefois, ils ne doivent pas troubler la fonction finale des protéines produites (Rosano et al., 2014 ; Costa et al., 2014). Les tags sont très souvent de petits peptides facilitant la purification des protéines cibles, car ils sont accessibles en surface. Il en existe une grande diversité, comme les tags poly-His, poly-Arg, Streptavidin-binding tag (Strep-tag), FLAG. L'étiquette poly-His permettant une purification de la protéine par une unique étape de chromatographie d'affinité au nickel est très

utilisée (Kosobakova et al., 2016). Il est aussi possible d'ajouter un tag sous la forme d'une protéine plus grosse qu'une étiquette. Ce sont des protéines de fusion, qui sont par exemple la protéine de liaison au maltose (Maltose Binding Protein ; MBP) et la Glutathion S transférase (GST). Leur présence peut améliorer la solubilité des protéines recombinantes, tout en facilitant à leur purification. Les tags sont généralement supprimés par clivage enzymatique ou chimique pour obtenir des protéines recombinantes intactes et actives. Pour que cette étape se déroule correctement, un site de clivage reconnu par une protéase (la protéase du virus de la gravure du tabac ou la thrombine par exemple) est ajouté entre la séquence du tag et celle de la protéine recombinante (Jenny et al., 2003 ; Blommel et al., 2007). Comme son nom l'indique, le clivage chimique repose sur l'utilisation de produits chimiques comme le bromure de cyanogène ou des métaux (nickel ou palladium) (Andreev et al., 2010 ; Hwang et al., 2014).

La culture d'*E. coli* est menée principalement dans le milieu de culture LB. Il est également possible d'utiliser des milieux plus complexes et d'ajouter des additifs chimiques qui ont une action positive sur l'expression des protéines recombinantes solubles. Ces additifs peuvent être des déclencheurs de stress osmotique augmentant la stabilité des protéines et favorisant le repliement des polypeptides. Par exemple, la présence de sel facilite l'expression des protéines cibles et stimule l'absorption d'osmolytes comme la bétaïne (Blackwell et al., 1991 ; Papp et al., 2006). Un effet similaire peut être obtenu en réalisant un choc thermique (passage à 47 °C pendant une heure puis poursuite de la culture à 20 °C) avant l'ajout de l'inducteur (Oganesyan et al., 2007 ; Bhatwa et al., 2021).

Les hôtes les plus utilisés pour la production de protéines recombinantes sont les souches *E. coli* K12 et BL21 (DE3) et d'autres qui en dérivent. Le tableau I.4 récapitule les principales caractéristiques des souches *E. coli* disponibles commercialement (Pouresmaeli et al., 2023).

Tableau I.4 : Principales caractéristiques des souches commerciales *E. coli*.

Souches	Caractéristiques	Références
BL21 (DE3)	Absence des protéases Lon et OmpT Possibilité d'induction à l'IPTG	Phillips et al., 1984 ; Studier et al., 1986 ; Grodberg et al., 1988 ; Grossman et al., 1998
Origami B (DE3)	Absence de la thiorédoxine réductase (trxB) et de la glutathion réductase (gor) Cytoplasme oxydant favorisant la formation des ponts disulfures	Stewart et al., 1998 ; Berkmen et al., 2012
Rosetta (2) (DE3)	Présence du plasmide pRARE(2) exprimant les ARNt pour les codons rares d'*E. coli* Neutralisation du biais d'utilisation des codons	Brinkmann et al., 1989 ; Seidel et al., 1992 ; Rosenberg et al., 1993 ; Del Tito Jr et al., 1995 ; Kane, 1995 ; Nakamura et al., 2000
BL21 Gold (DE3)	Présence de la méthylase Dcm et du phénotype Hte augmentant l'efficacité de la transformation	Marinus et al., 1973 ; May et al., 1975 ; Greener et al., 2004
Arctic Express (DE3)	Co-expression des protéines chaperonnes Cpn10 et Cpn60 de *Oleispira antarctica* permettant la croissance bactérienne entre 4 et 12 °C	Ferrer et al., 2003 ; Ferrer et al., 2004a ; Ferrer et al., 2004b

1.3.1.1. *E. coli* K12

E. coli K12 est un hôte d'expression choisi pour l'expression d'une aminoacylase thermostable de *Bacillus stearothermophilus*. Diverses modifications de cette souche et de plasmides ont été mises en œuvre. Les bactéries sont cultivées à 37 °C sur milieu LB. L'induction a lieu quand la DO_{600} est égale à 3, avec l'ajout d'IPTG à 0,1 mM. La concentration de l'inducteur chez *E. coli* est souvent comprise entre 0,1 et 1 mM, pour éviter une expression trop faible dans le cas d'une concentration trop faible en inducteur ou la formation de corps d'inclusion dans le cas d'une concentration trop élevée en inducteur. Les corps d'inclusions sont formés lors du dépassement du taux de traduction par les capacités de repliement protéiques des cellules. La vitesse de croissance d'*E. coli* est également réduite avec une forte concentration en IPTG (De León et

al., 2004 ; Marchall et al., 2016 ; Incir et al., 2024). La culture se poursuit 4 h après l'induction. Les enzymes sont purifiées grâce à deux étapes de chromatographie d'échange d'anions et une chromatographie par filtration sur gel. Le système *E. coli* K12 XS1D2R (F- Δ(*ppc-argE*) *nalA rpoB λ- hsdR recA*) avec le plasmide pAVK66 donne une enzyme recombinante dont l'activité de désacétylation de l'acétylornithine est 30 fois plus élevée que celle des enzymes issues de la souche sauvage *B. stearothermophilus* (Sakanyan et al., 1993).

1.3.1.2. *E. coli* BL21 (DE3)

La souche *E. coli* BL21 (DE3) (F– *ompT hsdS*$_B$ (*r*$_B$–, *m*$_B$–) *gal dcm* (DE3) ; Novagen, Merck, Germany) est très utilisée dans le cadre de l'expression hétérologue. Elle dérive de la souche *E. coli* B834 et présente plusieurs modifications génétiques : deux délétions des gènes de protéases (*Lon* et *OmpT*) et une insertion du gène de l'ARN polymérase T7. L'absence des deux protéases réduit la dégradation des protéines hétérologues exprimées par les bactéries. La présence du gène de l'ARN polymérase T7 permet l'induction de l'expression avec de l'IPTG. Cette enzyme présente chez les souches lysogènes DE3 est fournie par un dérivé du phage Lambda et a généralement un promoteur lacUV5 (Phillips et al., 1984 ; Studier et al., 1986 ; Grodberg et al., 1988 ; Grossman et al., 1998 ; Pouresmaeil et al., 2023).

Une première tentative d'expression de l'aminoacylase 1 porcine (EC 3.5.1.14) a été réalisée avec *E. coli* BL21 et le plasmide pJC20, donnant une protéine purifiée quasiment exempte d'activité. Cette enzyme est produite après une culture de 4 h à 37 °C sous agitation et induite à une DO_{550} de 0,5 par une addition d'IPTG à 0,5 mM (Pittelkow et al., 1998). Un temps de culture court limite l'action des protéases cytoplasmiques endogènes de l'hôte et l'épuisement en nutriments et en oxygène de l'environnement de culture (Gottesman, 1996 ; Weuster-Botz et al., 2001).

L'aminoacylase d'*Achromobacter xylosoxidans* subsp. *denitrificans* ATCC 15173 est exprimée dans *E. coli* BL21 (DE3) avec le plasmide pET-28a. Ce plasmide d'expression est le plus populaire sur le marché ; son utilisation est d'ailleurs citée dans plus de 40000 articles publiés (Shilling et al., 2020). Les bactéries poussent sur milieu LB, pendant 20 h et à 20 °C. La température d'incubation de 20 °C permet une diminution de la formation des corps d'inclusion et une augmentation de l'expression protéique pouvant s'expliquer par un bon repliement des enzymes recombinantes, malgré une baisse de la croissance bactérienne ou un allongement du temps de culture (Vera et al., 2006 ; San-Miguel et al., 2013). L'induction se fait en présence de 0,1 mM d'IPTG. Les aminoacylases recombinantes sont purifiées par chromatographie

d'affinité avec le nickel ; les résultats sont présentés dans le tableau I.5. Les enzymes sont produites majoritairement sous la forme de corps d'inclusion, toutefois une bande de surexpression à la taille attendue de la protéine est visible dans le puits de la fraction soluble après analyse par SDS-PAGE. Après purification, 18 mg d'aminoacylases recombinantes sont obtenues pour un 1 L de culture bactérienne ayant une activité spécifique de 204 UI/mg. La purification permet la récupération de 27 % des protéines de l'extrait cellulaire (Wang et al., 2013).

Tableau I.5 : Résultats relatifs à la purification de l'aminoacylase recombinante d'*A. xylosoxidans* (Wang et al., 2013).

Step	Protein (mg)	Total activity (U)	Specific activity (U/mg)	Recovery (%)	Purification (fold)
Crude extract	560	13455	24	100	1
Ni-NTA	45	6997	155.5	52	6.5
Ultra-filtration	17.8	3632	204	27	8.5

Une aminoacylase a été découverte au sein de la souche LP5_18B de *Burkholderia* sp. Son gène est cloné dans le vecteur pET-22b et le tout est transformé dans *E. coli* BL21 (DE3). Les cultures bactériennes sont réalisées pendant 16 h à 30 °C dans le milieu Overnight Express™ Instant TB Medium (Novagen, Merck KGaA). Ce milieu de culture est un milieu d'auto-induction au lactose, enrichi en glycérol. Il est connu pour augmenter le niveau d'expression protéique. Il n'est pas nécessaire de surveiller la croissance des bactéries ou d'ajouter un inducteur pendant la culture. L'induction se fait par utilisation différentielle des composants métaboliques du milieu, favorisant également une croissance à haute densité. L'hôte d'expression doit être inductible au lactose, c'est-à-dire qu'il doit produire une lactose perméase (codée par le gène *lacY*) pour le transport efficace du lactose et une β-galactosidase (codée par le gène *lacZ*) pour la transformation du lactose en allolactose (Studier et al., 2005 ; Grabski et al., 2005). L'enzyme recombinante n'est pas détectée sur le gel SDS-PAGE obtenu. Cette aminoacylase partage certaines caractéristiques avec plusieurs autres métalloenzymes connues, à zinc ou à magnésium. Son analyse ICP-MS indique la présence de 2 moles de zinc et de 3 moles de magnésium par mole de sous-unité. L'étude de la triazine hydrolase à large spécificité d'*Arthrobacter* et de *Nocardioides* spp. (TrzN) a montré que l'ion métallique Zn présent dans sa sous-unité est nécessaire lors de la synthèse de l'enzyme car il joue un rôle structural et catalytique (Shapir et al., 2006). Sur la base de ces éléments, les cultures sont menées en absence et en présence d'ions métalliques : sulfate de zinc, chlorure de cobalt et sulfate de magnésium à des concentrations de 50, 100 et 200 mg/L. L'activité enzymatique de l'extrait acellulaire est faible en absence d'ions métalliques et en présence de $MgSO_4$ 7 H_2O dans le milieu de culture (0,078 U/mL) ; elle est augmentée d'un facteur 3,8 en présence de 100 mg/L de $CoCl_2$ 6 H_2O, et d'un facteur 12,8 en présence de $ZnSO_4$ 7 H_2O (Takakura et al., 2019).

La souche *E. coli* BL21 (DE3) et le plasmide pET-28b sont utilisés pour la production d'une nouvelle aminoacylase de *Stenotrophomonas maltophilia* (SmAcy). La séquence de son gène a été optimisée pour une expression chez *E. coli*. La culture se fait sur milieu LB pendant 12 h, à 37 °C. Elle est induite par ajout d'IPTG à 0,1 mM à une DO_{600} de 0,6 puis la température est abaissée à 28 °C. Les enzymes recombinantes produites sont purifiées par chromatographie d'affinité avec le nickel. Les auteurs observent une production importante de l'aminoacylase dans la fraction soluble purifiée (plus de 80 %) et une quantité infime d'enzymes dans la fraction insoluble (Figure I.20) (Cao et al., 2020).

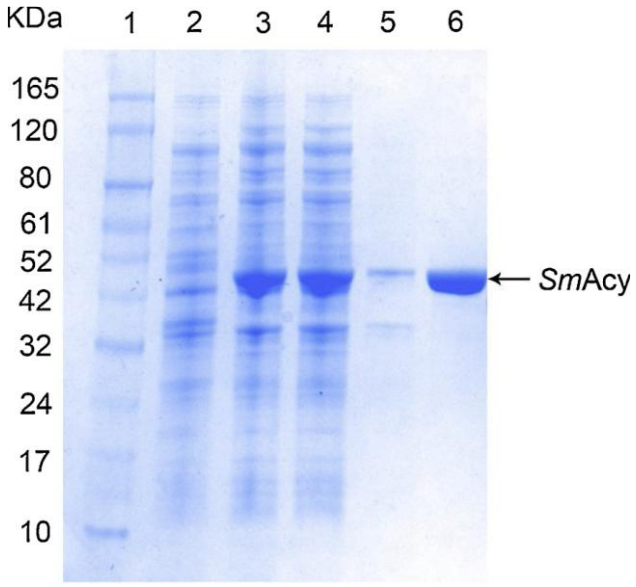


Figure I.20 : Analyse par SDS-PAGE de l'aminoacylase recombinante de *S. maltophilia*, après purification. Puits 1, marqueur de poids moléculaire ; puits 2, *E. coli* BL21 (DE3)/pet-28b vide ; puits 3, *E. coli* BL21 (DE3)/pET-28b/SmAcy induit ; puits 4, fraction soluble *E. coli* BL21 (DE3)/pET-28b/SmAcy ; puits 5, fraction insoluble *E. coli* BL21 (DE3)/pET-28b/SmAcy ; puits 6, fraction purifiée *E. coli* BL21 (DE3)/pET-28b/SmAcy (Cao et al., 2020).

1.3.1.3. *E. coli* Origami B (DE3)

La souche bactérienne *E. coli* Tuner™ (DE3) (F– *ompT hsdS*$_B$ (*r*$_B$–, *m*$_B$–) *gal dcm lacY1* (DE3) ; Novagen, Merck, Germany) dérive de la souche *E. coli* BL21 (DE3) dont elle conserve les modifications génétiques, auxquelles s'ajoute l'élimination du gène *lacY* codant pour une perméase au lactose. Cette perméase fait entrer l'IPTG et le lactose dans les cellules bactériennes qui peuvent ensuite induire l'expression de la protéine recombinante et de cette perméase, créant ainsi une boucle de rétroaction positive (Fernández-Castané et al., 2012). En présence de lactose, le comportement d'induction devient non-progressif et non homogène dans la population bactérienne. Ce phénomène est particulièrement important en présence d'une

faible quantité d'inducteur (Marbach et al., 2012). La délétion de ce gène aide à l'entrée diffusive et uniforme de l'IPTG dans les cellules de la population bactérienne et favorise un niveau d'induction homogène et dépendant de la concentration en inducteur. La solubilité de certaines protéines peut ainsi être améliorée lorsqu'il y a un niveau inférieur d'expression (Binder et al., 2014 ; Binder et al., 2016).

La souche bactérienne *E. coli* Origami B (DE3) (F– *ompT hsdS*$_B$ (*r*$_B$–, *m*$_B$–) *gal dcm lacZY ahpC* (DE3) *gor522::TnT0 trxB* (Kan^R, Tet^R) ; Novagen, Merck, Germany) dérive quant à elle de la souche *E. coli* Tuner (DE3). Elle dérive d'un mutant lacYZ d'*E. coli* BL21, dans lequel la mutation du gène *lacZ* codant pour une β-galactosidase fait disparaître l'activité catalytique de l'enzyme. Elle peut également dérivée des bactéries *E. coli* K12. Cette souche résulte en outre des mutations au niveau des gènes de la thiorédoxine réductase (*trxB*) et de la glutathion réductase (*gor*) ; l'absence de ces deux enzymes améliore la formation des ponts disulfures dans le cytoplasme devenu oxydant et limite le risque de mauvais repliement protéique et la formation de corps d'inclusion (Stewart et al., 1998 ; Berkmen, 2012). Les premiers mutants porteurs de ces deux mutations nécessitaient l'ajout d'un agent réducteur dans le milieu pour obtenir une croissance bactérienne normale (Prinz et al., 1997). Les souches *E. coli* Origami^{TM} sont également porteuses d'une mutation supplémentaire (*ahpC*) provoquant une perte de l'activité peroxydase et permettant aux bactéries un taux normal de croissance sans ajout d'un agent réducteur (Bessette et al., 1999 ; Ritz et al., 2001 ; Hatahet et al., 2014). Les mutations caractéristiques de cette bactérie entraînent une diminution de sa vitesse de croissance et du nombre final de cellules, ce qui conduit à une manipulation plus difficile (Wardenga et al., 2008).

Deux études sur la production hétérologue des aminoacylases de *Streptomyces ambofaciens* ont été menées il y a quelques années. Tout d'abord, les gènes de SamELA et de SamPVA sont clonés dans le plasmide pET-15b. Bien que *E. coli* BL21 (DE3) et *E. coli* Origami B (DE3) soient transformées par les deux types de plasmides, seule l'expression de SamELA a été réalisée. L'étude de la littérature a montré que l'expression de Sm-PVA issues de *S. mobaraensis* et de *S. lavendulae* chez *E. coli* conduit à des formes inactives. Les auteurs expliquent ce résultat par la nature hétérodimérique des enzymes nécessitant une maturation de la structure peptidique primaire impossible à réaliser pour *E. coli* (Zhang et al., 2007 ; Torres-Bacete et al., 2015). La culture d'*E. coli* BL21 transformée par SamELA a lieu à 28 °C pendant 24 h ; elle est induite par un ajout de 1 mM d'IPTG, à une DO_{600} de 0,6. Dans ces conditions, les enzymes sont produites uniquement sous forme insoluble. Pour contourner ce problème, *E. coli* Origami B (DE3) est utilisée comme hôte d'expression. La température de culture après

l'induction est abaissée à 16 °C. La concentration en inducteur est également réduite à 0,1 mM. Une bande pouvant correspondre à SamELA est visible dans la fraction soluble sur le gel SDS-PAGE, toutefois cette bande est également présente avant l'induction ; elle pourrait donc correspondre à une protéine endogène d'*E. coli* de la même taille (Figure I.21). Une légère activité hydrolytique de la N-ε-acétyl-L-lysine est détectée par CCM. Aucun comparatif n'est fait avec un témoin ayant un plasmide vide, ce qui empêche de tirer des conclusions plus précises (George, 2017).

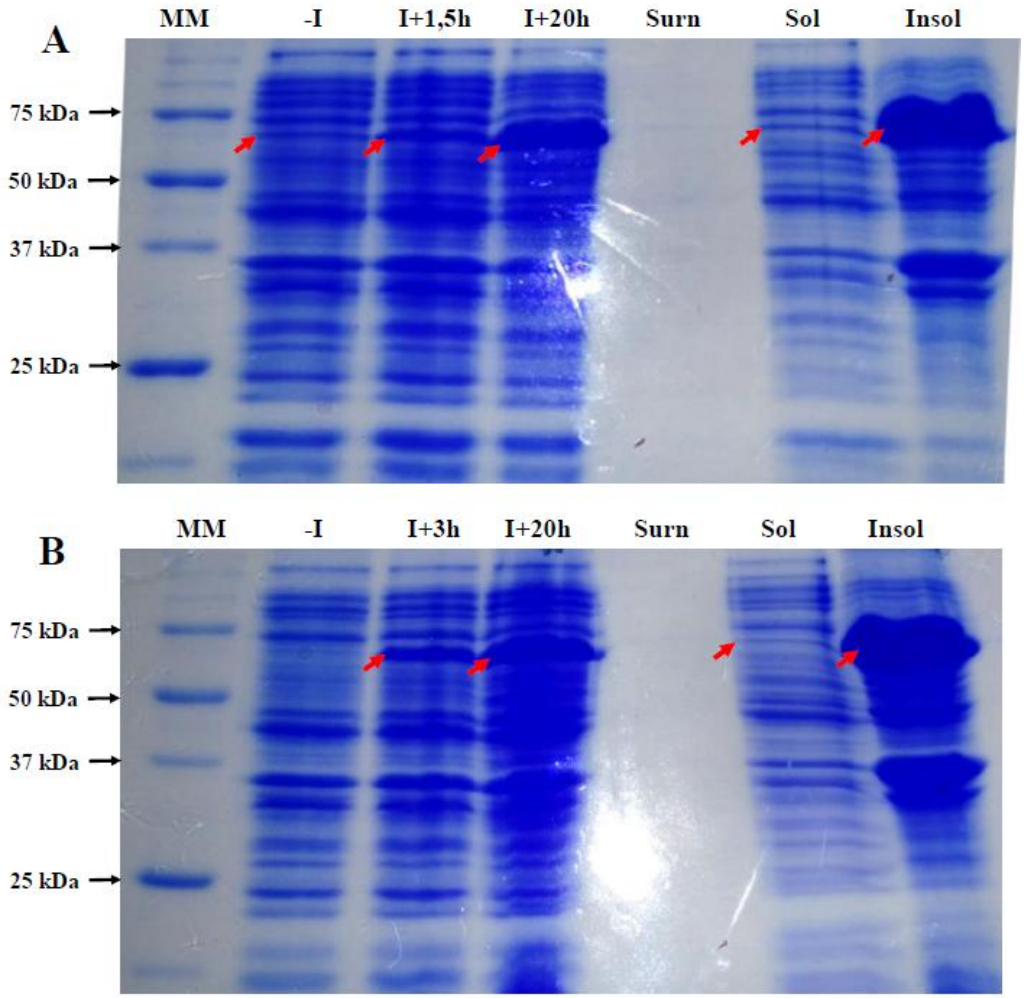


Figure I.21 : Analyse par SDS-PAGE des protéines issues des souches *E. coli* Origami B (DE3) (A) et *E. coli* BL21 (DE3) (B). MM, marqueur de poids moléculaire (unstained Precision Plus Protein standards, BioRad) ; -I, culot cellulaire avant induction ; I+Xh, culot cellulaire X heures après induction à 16 °C par 0,1 mM d'IPTG ; Surn, surnageant de culture après récolte (20 h après induction) ; Sol et Insol, respectivement, la fraction soluble et la fraction insoluble obtenues 20 h après l'induction. Les flèches rouges correspondent à la position supposée de SamELA (George, 2017).

Ces travaux ont été répétés et complétés avec le système *E. coli* Origami B (DE3) transformé par les plasmides pET-15b contenant les gènes de SamPVA et de SamAA. Les conditions de cultures sont similaires à celles optimisées de l'étude précédente. Aucune surexpression

protéique ni aucune activité hydrolytique ou synthétique n'est détectée concernant l'enzyme recombinante SamPVA pour les mêmes raisons que précédemment. La surexpression de SamELA et de SamAA recombinantes est visible sur les gels SDS-PAGE (Figure I.22). Les activités spécifiques des extraits bruts de ces cultures ont été étudiées sur les substrats N-α- et N-ε-acétyl-L-lysine et sont visibles dans le tableau I.6. Une tentative de purification de SamAA par chromatographie d'affinité au nickel est réalisée mais aucune bande de taille correspondant à l'enzyme n'est visible sur le gel SDS-PAGE et l'activité hydrolytique envers la N-α-acétyl-L-lysine est retrouvée uniquement dans les fractions de lavage de la colonne (à 80 mM d'imidazole) (Nili, 2018).

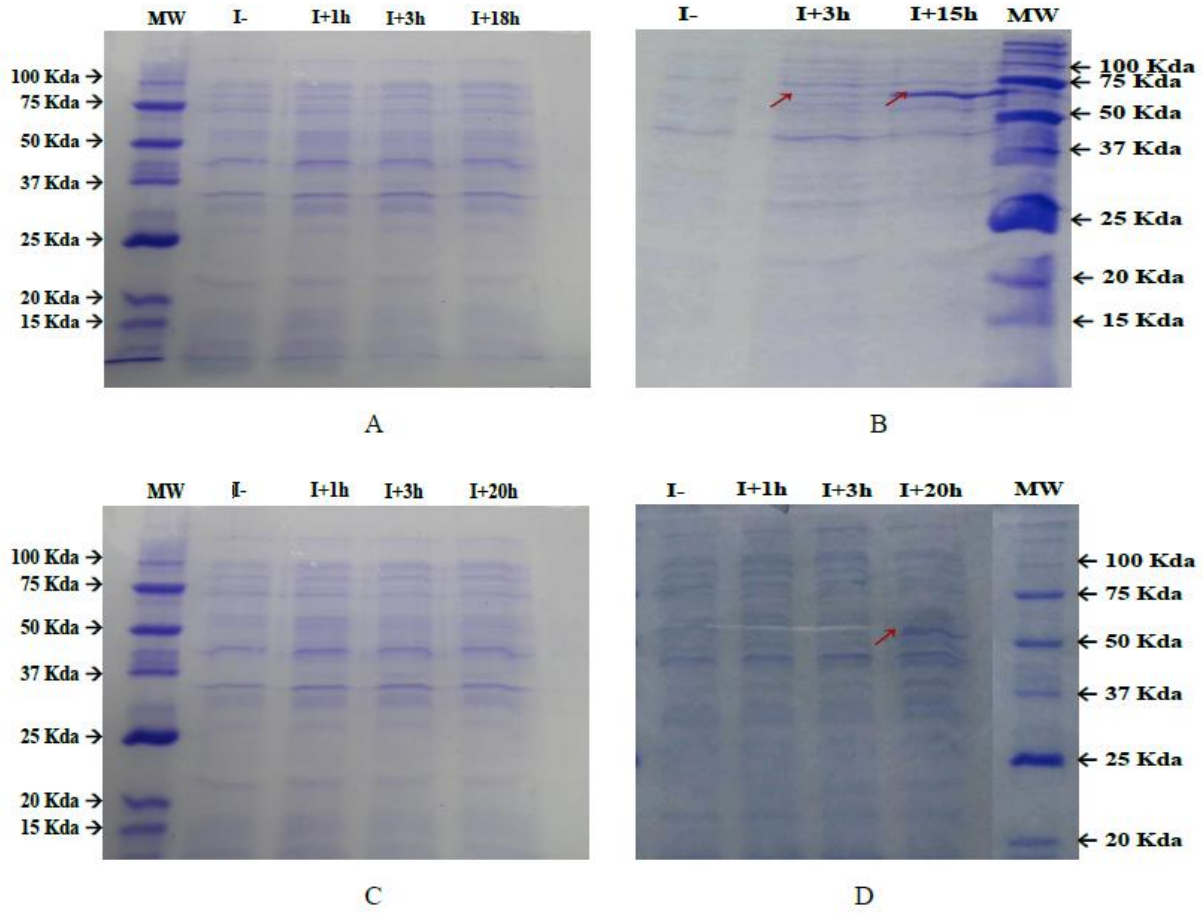


Figure I.22 : Analyse par SDS-PAGE des protéines issues des cultures d'*E.coli* Origami B (DE3) non transformée (A), transformée avec le plasmide pET-15b/SamELA (B), transformée avec le plasmide pET-15b/SamPVA (C), et transformée avec le plasmide pET-15b/SamAA (D). MW : Marqueur de poids moléculaire (Unstained Precision Plus Protein Standards ; BioRad) ; I- : avant induction ; I+Xh : X heures après induction à 0,1 mM IPTG. Les flèches rouges indiquent la position des protéines exprimées (Nili, 2018).

Tableau I.6 : Activités spécifiques apparentes des vitesses de réactions hydrolytiques en présence des fractions solubles des cultures d'*E. coli* Origami B (DE3) transformées (Nili, 2018).

Activité spécifique apparente (nmole Lys/min/mg)	En présence de N-α-acétyl-L-lysine	En présence de N-ε-acétyl-L-lysine
E. coli Origami B (DE3)/SamELA	0,26	3,71
E. coli Origami B (DE3)/SamAA	33,24	0

1.3.1.4. *E. coli* Rosetta (DE3) et Rosetta 2 (DE3)

Les différents ARNt pour un acide aminé sont exprimés de façon variable d'une espèce vivante à l'autre, selon les préférences spécifiques en codons de chaque espèce. La figure I.23 indique les préférences moyennes de codons des génomes de huit organismes fréquemment étudiés. *Streptomyces coelicolor* a le profil d'utilisation des codons le plus extrême parmi les espèces étudiées. Ceci montre que *E. coli* n'est pas l'hôte optimal pour l'expression de protéines codées avec un profil d'utilisation des codons du genre *Streptomyces* (Gustafsson et al., 2004). L'expression de protéines non endogènes à *E. coli* dans cet hôte est donc affectée par un biais d'utilisation des codons provoquant une pénurie d'ARNt rares. Cela provoque une multitude d'erreurs lors de la traduction, comme une terminaison prématurée par un blocage terminal, une mauvaise incorporation d'acides aminés, des décalages de cadre de traduction et/ou une diminution de la qualité et du rendement des protéines recombinantes. Deux stratégies peuvent être mises en place pour éviter ce problème : l'utilisation d'une souche capable de coexprimer les gènes des ARNt rares ou l'utilisation d'un gène dont la séquence est optimisée pour une expression chez *E. coli* (Kim et al., 2020).

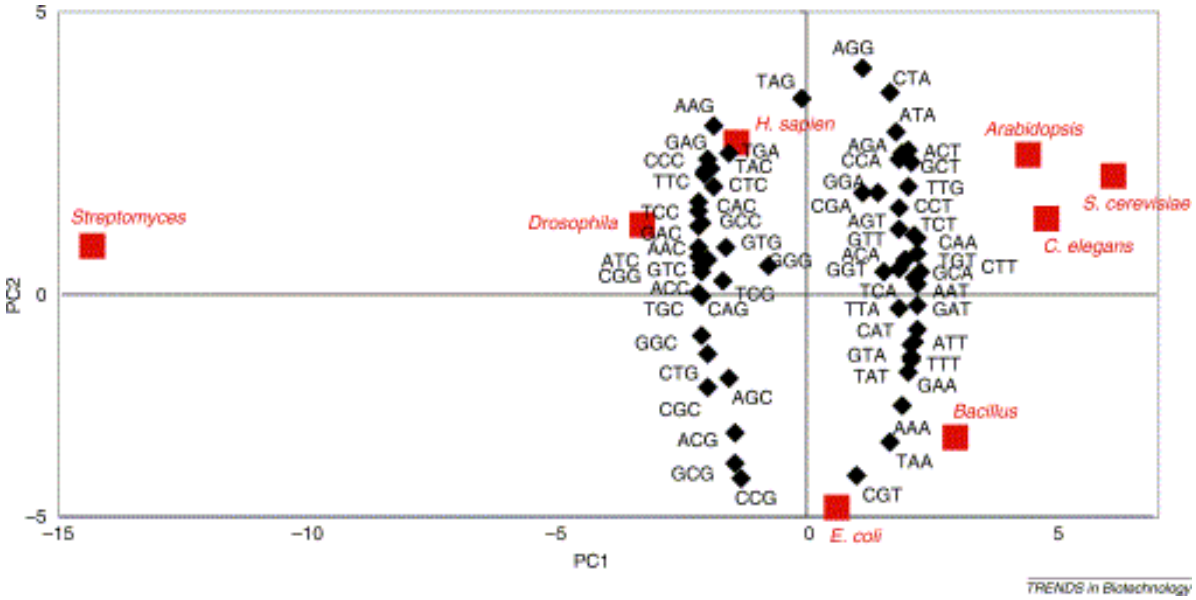


Figure I.23 : Représentation graphique de « l'espace d'utilisation des codons ». L'analyse en composantes principales implique une procédure mathématique qui transforme plusieurs variables corrélées (ici les fréquences des codons) en un nombre plus petit de variables non corrélées appelées composantes principales. La première composante principale représente autant que possible la variabilité des données, et chaque composante suivante représente autant que possible la variabilité restante. Les fréquences auxquelles chaque codon est utilisé dans toutes les protéines de huit organismes fréquemment étudiés ont été tabulées et soumises à l'analyse en composantes principales pour produire une carte de « l'espace d'utilisation des codons ». Les deux codons ATG et TGG qui codent de manière unique Met et Trp, respectivement, ont été omis. Deux dimensions ont été identifiées qui représentaient respectivement 70 % (PC1) et 12 % (PC2) de l'information totale sur la variabilité des codons. Les losanges noirs représentent les charges (c'est-à-dire la contribution de chaque codon aux deux dimensions principales des composants ; par exemple, les codons GAT et CAG ne contribuent pas à PC2 mais ont des contributions négatives et positives à peu près égales à PC1). Les valeurs des charges des codons ont été normalisées par rapport à celles de la distribution des organismes. Les carrés rouges montrent les préférences de chaque organisme tracées dans cet espace. Le tracé a été réalisé à l'aide de MatLab (Mathworks) (Gustafsson et al., 2004).

La possibilité d'obtenir un gène synthétique à moindre coût augmente les utilisations d'un gène d'intérêt dont les codons sont optimisés pour sa synthèse au sein de l'organisme hôte choisi. Des algorithmes automatisés d'optimisation des codons sont développés dans ce sens. Ils modifient les codons de façon à maximiser l'indice d'adaptation des codons du gène à celui de l'hôte d'expression, par randomisation des codons, tout en optimisant les caractéristiques de la séquence comme la teneur en GC et limitant les répétitions ou motifs nuisibles à la transcription et à la traduction (Quax et al., 2015 ; Al-Hawash et al., 2017).

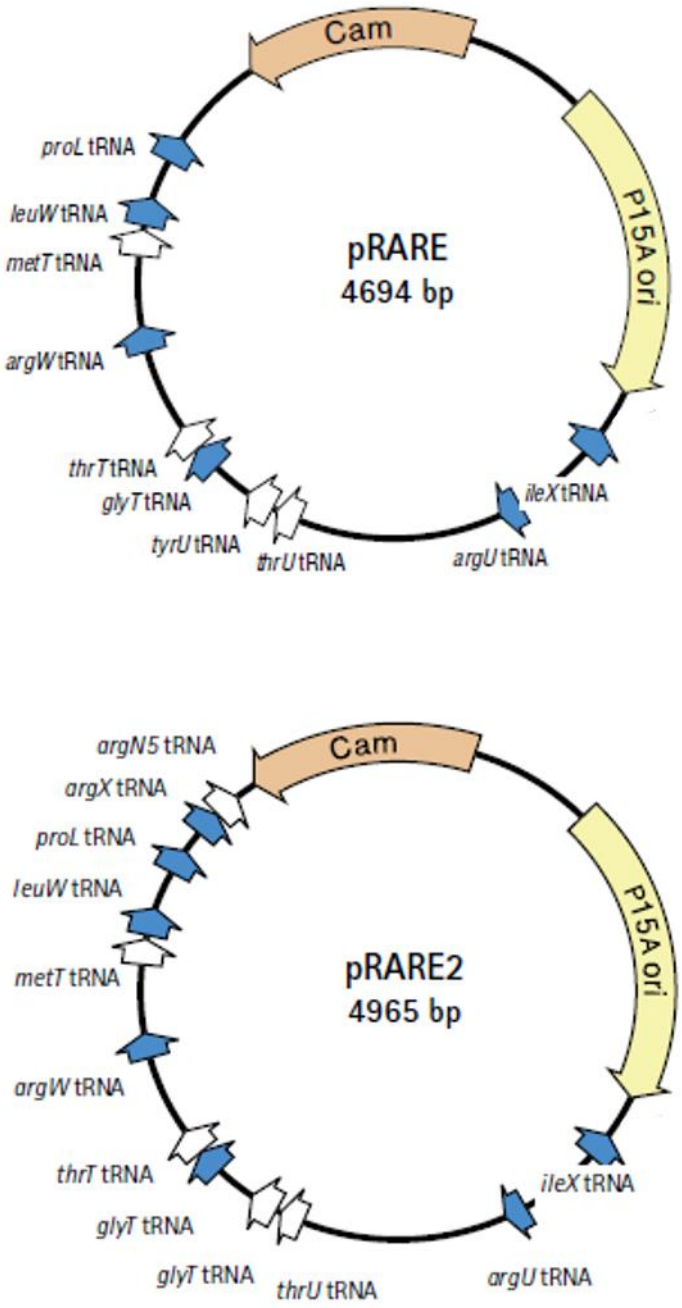


Figure I.24 : Schémas des plasmides pRARE et pRARE 2 permettant la surexpression des ARNt pour les codons AUA (*IleX tRNA*), AGG (*ArgW tRNA*), AGA (*ArgU tRNA*), CUA (*LeuW tRNA*), CCC (*ProL tRNA*), GGA (*GlyT RNAt*) et CGG (*ArgX tRNA*). Les flèches bleues indiquent les gènes surexprimés des ARNt. Les flèches blanches indiquent les gènes des ARNt insérés lors de la construction des deux plasmides (Novagen).

La souche *E. coli* Rosetta^{TM} (DE3) (F– *ompT hsdS*$_B$ (*r*$_B$–, *m*$_B$–) *gal dcm* (DE3) pRARE (Cam^R) ; Novagen, Merck, Germany) a été développée à partir d'*E. coli* BL21 dans le but d'améliorer l'expression des protéines recombinantes d'origine eucaryote avec la surexpression des ARNt des codons rares chez *E. coli*. Cette souche possède le plasmide pRARE portant une résistance au chloramphénicol et fournissant les ARNt pour les codons AUA, AGG, AGA, CUA, CCC et

GGA (Brinkmann et al., 1989 ; Seidel et al., 1992 ; Rosenberg et al., 1993 ; Del Tito Jr et al., 1995 ; Kane, 1995). La souche *E. coli* Rosetta 2 (DE3) (F– *ompT hsd*S_B (r_B–, m_B–) *gal dcm* (DE3) pRARE2 (Cam^R) ; Novagen, Merck, Germany) a le plasmide pRARE2 fournissant les six ARNt mentionnés précédemment et un septième ARNt pour le codon rare CGG (Nakamura et al., 2000) (Figure I.24).

L'aminoacylase 1 porcine exprimée chez *E. coli* Rosetta (DE3) a une activité spécifique presque aussi élevée que l'enzyme native. Le gène codant pour cette aminoacylase est inséré dans le plasmide T7 pET28a. Les cellules bactériennes sont cultivées à 37 °C, pendant 3 h. L'induction est réalisée à une DO_{600} de 0,8 par l'ajout d'IPTG à 0,2 mM (Liu et al., 2006).

Le gène de l'aminoacylase de l'archée hyperthermophile *Pyrococcus horikoshii* est cloné dans le vecteur pET-11a (Merck kGaA, Darmstadt, Allemagne), lui-même transformé dans les cellules d'*E. coli* Rosetta (DE3) pLysS. Le plasmide pLysS est un plasmide supplémentaire résistant au chloramphénicol qui est construit par l'insertion du gène du lysozyme T7 dans le site BamHI du plasmide pACYC184 (Chang et al., 1978 ; Dunn et al., 1983). Ce plasmide est utilisé pour supprimer l'expression basale du promoteur T7, en produisant du lysozyme T7. Cette enzyme peut détruire une liaison spécifique dans la couche de peptidoglycane de la paroi d'*E. coli*, provoquant une lyse cellulaire, ou se lier à l'ARN polymérase T7, inhibant la transcription du gène d'intérêt (Inouye et al., 1973 ; Moffatt et al., 1987 ; Zhang et al., 1997 ; Huang et al., 1999). Dans pLysS, la séquence codante du lysozyme T7 est dans une orientation antisens par rapport au promoteur *tet*, entraînant le contrôle du gène par le promoteur faible *T7 φ3.8* et la production du lysozyme en faible quantité, alors que de grandes quantités de lysozymes sont produites à partir de l'orientation sens du gène sous le contrôle du promoteur *tet* dans pLysE. La présence d'un de ces plasmides augmente la stabilité, le maintien et l'expression des plasmides avec des protéines recombinantes toxiques. Le lysozyme produit par pLysS ou pLysE peut être également utilisé pour lyser les cellules lors de la préparation des extraits cellulaires (O'Mahony et al., 1990 ; Studier, 1991 ; Christodoulou et al., 2002). La culture a lieu dans du milieu LB contenant 1 mM de $ZnCl_2$ à 37 °C, pendant 6 h. L'IPTG ayant le rôle d'inducteur est ajouté au milieu à une concentration de 1 mM, quand la DO_{600} est égale à 0,6. L'enzyme est purifiée par chromatographie d'échange d'anions puis par chromatographie d'interaction hydrophobe. L'aminoacylase est surexprimée avec succès, permettant sa caractérisation structurale et mécanistique grâce à l'introduction de mutations au niveau des acides aminés impliqués dans la fixation des ions métalliques ou dans la catalyse (Tanimoto et al., 2008).

Le gène codant pour une aminoacylase chez *Alcaligens xylosoxydans* (Alcaligenes A-6) dans une version du gène optimisé ou non est inséré dans le plasmide pET-32a. Les plasmides avec le gène optimisé ou non sont placés chez *E. coli* BL21 (DE3). *E. coli* Rosetta 2 (DE3) est transformée uniquement par le plasmide contenant le gène non-optimisé. Toutes les cultures sont menées en milieu LB à 37 °C pendant 3 h et sont induites par ajout d'IPTG à 1 mM à 0,6 de DO_{600}. Les aminoacylases recombinantes sont purifiées par plusieurs étapes de chromatographie ionique et une chromatographie de gel-filtration. Les résultats portant sur l'expression et l'activité spécifique de l'aminoacylase recombinante sont présentés dans le tableau I.7. Le taux d'expression le plus élevé est obtenu avec le système *E. coli* Rosetta 2 (DE3), mais la plus haute activité spécifique est obtenue avec le système *E. coli* BL21 (DE3) contenant le gène optimisé (Hou et al., 2016).

Tableau I.7 : Expression et activité spécifique de l'aminoacylase d'*A. xylosoxydans* selon le système d'expression (Hou et al., 2016).

	Taux d'expression (% d'aminoacylases sur les protéines totales)	Activité spécifique (U/mL)
E. coli BL21 (DE3) avec le gène non optimisé	56	70
E. coli BL21 (DE3) avec le gène optimisé	64	98
E. coli Rosetta 2 (DE3)	78	83

1.3.1.5. *E. coli* BL21 Gold (DE3)

E. coli BL21 Gold (DE3) (F– *ompT hsdS*(r_B– m_B–) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* Hte ; Agilent Technologies, USA) est une souche bactérienne développée par Agilent Technologies et dérivant d'*E. coli* BL21 avec ses deux protéases déficientes (Lon et OmpT). La méthylase Dcm (aussi appelée méthylase Mec), qui méthyle les résidus de cytosine internes dans les séquences CCAGG et CCTGG en position C5, est insérée pour contrer l'action des endonucléases de restriction via la méthylation des sites de restriction (Marinus et al., 1973 ; May et al., 1975). Cette souche présente le phénotype Hte, aussi présent dans la souche XL10-Gold, augmentant l'efficacité de transformation des cellules à plus de 1.10^8 cfu/µg de pUC18 DNA (plasmide de contrôle positif de transformation fourni par Agilent). Le gène codant l'endonucléase 1 (*endA*),

connue pour être responsable de la dégradation rapide de l'ADN plasmidique, est également inactivé (Greener et al., 2004).

Cette souche semble intéressante pour la production des aminoacylases, toutefois, aucune étude ne semble avoir été réalisée jusqu'ici. Elle constitue une alternative possible en cas de production d'une enzyme récalcitrante chez *E. coli*.

1.3.1.6. *E. coli* Arctic Express (DE3)

Les protéines chaperonnes sont des molécules présentes dans la cellule ayant un rôle dans le repliement des autres protéines et dans le maintien de la qualité et de la stabilité du protéome. Elles sont divisées en trois catégories selon leurs fonctions : les protéines chaperonnes de repliement, de maintien et/ou de désagrégation. Elles interviennent pendant ou après l'étape de traduction. Le facteur déclencheur (FT) tout comme GroEL, GroES, DnaK, DnaJ et GrpE sont des protéines chaperonnes identifiées chez *E. coli* (Bhatwa et al., 2021 ; Zhang et al., 2022). Il est possible de les coexprimer lors de la surexpression d'une protéine cible recombinante pour éviter la formation de corps d'inclusion résultant de l'absence de modifications post-traductionnelles, du mauvais repliement protéique, de la faible capacité de solubilité et/ou de la surcharge métabolique de l'hôte. Plusieurs études ont montré que la co-expression d'une ou plusieurs de ces protéines chaperonnes augmente la quantité de protéines cibles solubles et fonctionnelles (Lee et al., 2004 ; Sahdev et al., 2008 ; Wang et al., 2018).

La réduction de la température de la culture bactérienne est une stratégie pour augmenter la quantité de protéines recombinantes solubles. Mais elle est préjudiciable à la capacité des protéines chaperonnes à aider au bon repliement des protéines dans les cellules mésophiles d'*E. coli* (Mendosa et al., 2000). La souche *E. coli* Arctic Express (DE3) (F– *ompT hsdS*(r_B– m_B–) *dcm*$^+$ Tet^r *gal* λ(DE3) *endA* Hte [*cpn10 cpn60* Gent^r] ; Agilent Technologies, USA) a été développée à partir de la souche *E. coli* BL21 Gold (DE3) pour contourner ce problème : elle co- et sur-exprime les protéines chaperonnes Cpn10 et Cpn60, qui sont issues de la bactérie psychrophile *Oleispira antarctica*. Les deux protéines Cpn10 et Cpn partagent 74 et 54 % d'identité respective avec les protéines chaperonnes GroEL et GroES ; elles exercent leur activité de repliement protéiques à des températures comprises entre 4 et 12 °C (Ferrer et al., 2003). Lors de la surexpression des protéines chaperonnes Cpn10 et Cpn60, *E. coli* est capable de se développer à température basse (jusqu'à 4 °C) grâce à l'action de ces protéines sur le repliement des protéines endogènes et recombinantes (Ferrer et al., 2004a ; Ferrer et al., 2004b).

L'aminoacylase 1 porcine a également été exprimée dans un système *E. coli* BL21 possédant les plasmides chaperons de Takara Bio (Otsu, Shiga, Japon) qui sont les plasmides pG-KJE8 (DnaKJE + GroELS), pGro7 (GroELS), pKJE7 (DnaKJE), pG-Tf2 (TF + GroELS) et pTf16 (FT). Le gène inséré dans le plasmide pET-52b(+) est optimisé pour une expression chez *E. coli*. Un milieu complexe (tryptone à 17 g/L, peptone de soja à 3 g/L, NaCl à 5 g/L, K_2HPO_4 à 2,5 g/L et glucose à 2,5 g/L) a été utilisé. L'induction faite à une DO_{600} de 0,7 par ajout d'IPTG à 0,1 mM est suivie d'une baisse de la température à 20 °C pendant 24 h. Du $ZnCl_2$ est également ajouté au milieu de culture à une concentration finale de 5 µM. La purification est facilitée par la présence d'un tag Strep-tag II en position N-terminale de l'enzyme. Le rendement global d'aminoacylase 1 soluble purifiée est de 1 à 2 mg sans protéine chaperonne et de 3 à 4 mg avec TF ou GroEL-GroES pour une culture de 50 mL. L'analyse PAGE native non dénaturante confirme que l'enzyme est un homodimère. L'activité catalytique de l'enzyme pour l'hydrolyse de la N-acétyl-méthionine est similaire à celle de l'enzyme commerciale (Wardenga et al., 2008).

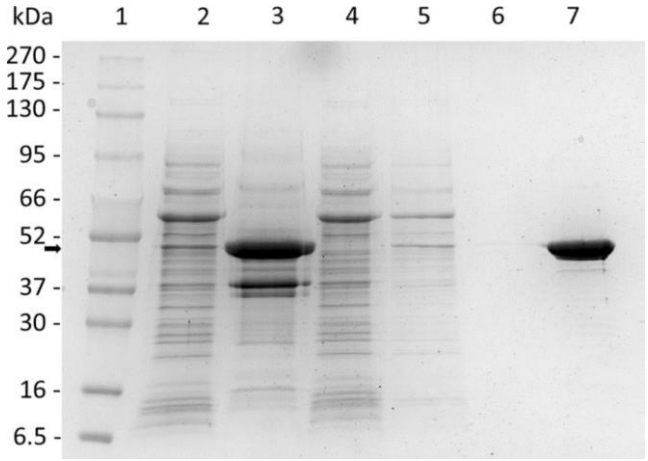


Figure I.25 : Analyse SDS-PAGE des différentes fractions obtenues lors de la purifiation de PmAcy. L'enzyme a été purifiée à partir d'une culture d'*E. coli* BL21 pGro7 avec le plasmide pET-28a-PmAcy (milieu TB-AIM). Piste 1 : marqueur protéique (BlueEasy Prestained Protein Marker, Nippon Genetics) ; piste 2 : extrait brut ; piste 3 : fraction insoluble ; piste 4 : écoulement provenant de la chromatographie d'affinité Strep-tag ; piste 5 : première fraction de lavage ; piste 6 : deuxième fraction de lavage ; piste 7 : élution de PmAcy. La flèche noire indique la hauteur de migration de PmAcy (Haeger et al, 2024a).

Le système *E. coli* BL21 avec les plasmides pGro7 et pET28a est utilisé pour l'expression de l'aminoacylase de *Paraburkholderia monticola*. Le gène est optimisé pour une expression chez cet hôte. Les cellules sont placées dans un milieu d'auto-induction (TB-AIM) de composition complexe (2 % de tryptone de caséine, 2,4 % d'extrait de levure, 25 mM de NaH_2PO_4, 25 mM de KH_2PO_4, 50 mM de NH_4Cl, 2 mM de $MgSO_4$, 5 mM de Na_2SO_4, 0,5 % de glycérol (v/v), 0,5 % de lactose, 0,05 % de glucose et 0,2 % lactose). La culture dure 24 h à 20 °C. Les enzymes

recombinantes sont purifiées par chromatographie d'affinité avec le motif Strep-Tag (Figure I.25). Elles sont majoritairement produites sous la forme de corps d'inclusion, mais l'utilisation du plasmide pGro7 synthétisant une protéine chaperonne permet d'obtenir une forme soluble : 15 mg d'aminoacylase purifiée est obtenue à partir de 100 mL de culture bactérienne (Haeger et al, 2024a).

L'expression de l'aminoacylase de *Mycolicibacterium smegmatis* MKD 8 chez *E. coli* BL21 (DE3) se fait sous la forme de corps d'inclusion. Une approche de co-expression de protéines chaperonnes chez *E. coli* BL21 (DE3) et *Vibrio natriegens* Vmax^{TM} est étudiée pour résoudre ce problème de solubilité. Le plasmide chaperon de Takara Bio contenant pGKJE8 (GroEL/S, DnaK/J/GrpE), pKJE7 (DnaK/J/GrpE), pGTf2 (GroEL/S, Tf), pTf16 (Tf) ou pGro7 (GroEL/S) est transformé chez ces deux bactéries. Le gène est cloné dans le plasmide pET28a avec un Strep Tag et le promoteur fort T7. L'induction peut se faire par ajout d'IPTG (1 mM à DO_{600} de 0,5) ou par auto-induction avec du lactose (0,2 %) en utilisant ce promoteur. Les cultures durent 4 h après induction avec l'IPTG, et 24 h en cas d'auto-induction. Les milieux LB et Terrific Broth (TB : 2 % de tryptone de caséine, 2,4 % d'extrait de levure, 25 mM NaH_2PO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 2 mM $MgSO_4$, 5 mM Na_2SO_4, 0,5 % de glycérol (v/v) et 0,05 % de glucose) sont employés pour les cultures avec *E. coli*. Les milieux TBv2 et BHIv2 (37 g/l Brain-Heart-Infusion Broth, Carl Roth, Allemagne), les deux supplémentés en sels (v2 : 204 mM NaCl, 4,2 mM KCl, 23,4 mM $MgCl_2$), sont utilisés pour les cultures avec *V. natriegens*. Les cultures ont lieu à température réduite : à 20 °C pour *E. coli* et à 30 °C pour *V. natriegens*. L'activité aminoacylase est mesurée par l'hydrolyse de la N-acétyl-L-alanine. La co-expression de GroEL/S chez *E. coli* et *V. natriegens* a la meilleure influence en augmentant l'activité de l'aminoacylase dans l'extrait acellulaire d'un facteur de 1,8 (Haeger et al., 2023b). *E. coli* ArcticExpress^{TM} (DE3) a également été utilisé pour l'expression de l'aminoacylase de *M. smegmatis*. La culture se déroule avec le milieu d'auto-induction TB, sans ajout d'antibiotique, pendant 24 h à 12 °C, avec une croissance préalable de 6 h à 30 °C. Une unité de cette enzyme est définie comme la quantité d'enzyme qui hydrolyse une µmol de N-acétyl-L-alanine par minute dans les conditions données. L'activité aminoacylase et l'activité spécifique de l'extrait acellulaire sont égales à 139,1 (± 6,9) U/mL et à 9,2 (± 0,5) U/mg respectivement. L'activité spécifique de l'enzyme purifiée est quant à elle égale à 127,2 (± 4,8) U/mg (Haeger et al., 2023b).

1.3.1.7. Refolding

E. coli est un hôte d'expression privilégié pour la production de protéines recombinantes, car il a une vitesse de croissance rapide et permet des niveaux d'expression élevés. Ces deux caractéristiques entraînent souvent la formation des protéines souhaitées sous la forme de corps d'inclusion insolubles. Ceci constitue la principale problématique liée à la production d'aminoacylases recombinantes (Haeger et al., 2024b). A notre connaissance, aucune étude de refolding des corps d'inclusions d'une aminoacylase n'a été menée.

Pourtant, transformer ces corps d'inclusion en protéines solubles et actives peut présenter plusieurs avantages. Les corps d'inclusion sont produits en grande quantité et sont résistants à l'action des protéases bactériennes. Ils sont faciles à isoler, en réalisant une simple centrifugation. Le traitement des corps d'inclusion se déroule ensuite en deux étapes critiques : la solubilisation des agrégats de protéines recombinantes puis le repliement sous une forme bioactive (Kaur et al., 2018).

Les corps d'inclusion sont solubilisés en utilisant des réactifs chaotropiques à forte concentration (urée ou chlorhydrate de guanidine), des détergents (SDS ou sarkosyl), un pH alcalin, des agents réducteurs (DTT ou β-mercaptoéthanol) et/ou des agents chélateurs (EDTA) (Puri et al., 1993 ; Khan et al., 1998 ; Patra et al., 2001 ; Jekabsons et al., 2002 ; Yen et al., 2019 ; De Lima et al., 2022). L'emploi de solvants doux, comme les alcools alkyliques, couplé à une forte concentration d'agent chaotrope, permet de restaurer la structure secondaire de type native améliorant la récupération d'une protéine bioactive (Singh et al., 2012).

Après l'étape de solubilisation des corps d'inclusion, une étape de repliement de la protéine est nécessaire et comprend des techniques de dilution, de dialyse ou de chromatographie sur colonne permettant de passer du tampon dénaturant au tampon de repliement (Liu et al., 2015). Ce dernier est composé d'additifs agissant sur la stabilisation de l'état natif, l'accélération de la réaction de repliement et la suppression de l'agrégation non spécifique de la protéine dépliée ou de ces intermédiaires de repliement. Chaque protéine présente des conditions de repliement qui lui sont propres et qu'il est nécessaire d'identifier (Kaur et al., 2018). Divers détergents, agents chaotropes, agents cosmotropes, acides aminés et dérivés d'acides aminés, sucres et osmolytes et alcools sont utilisés seul ou en mélange pour cette étape de repliement (Leibly et al., 2012 ; Lebendiker et al., 2014 ; Chatzileontiadou et al., 2021).

1.3.2. *Streptomyces lividans*

Quand l'expression d'une protéine soluble n'est pas possible chez *E. coli*, malgré une optimisation des codons de la séquence génique, un abaissement de la température de la culture bactérienne, une optimisation de l'induction et/ou la co-expression de protéines chaperonnes, il est possible de changer d'hôte d'expression.

Les micro-organismes du genre *Streptomyces* peuvent être utilisés comme système d'expression alternatif pour la production de protéines recombinantes car ils présentent plusieurs avantages. Ces bactéries forment peu de corps d'inclusion et présente une faible activité protéolytique endogène. Elles ont une capacité de sécrétion élevée, utilisée naturellement pour la sécrétion d'enzymes dégradant les polymères naturels nourriciers. Ce sont des bactéries bien adaptées à l'expression des gènes riches en GC. Cependant, ces cultures bactériennes sont hétéromorphes et visqueuses créant des conditions défavorables à une utilisation industrielle en raison du stress mécanique et oxydatif associé et des problèmes de transfert de chaleur. La morphologie des mycéliums responsable de ce défaut peut être modifié par des facteurs externes comme les milieux et conditions de culture, et par des facteurs génétiques (Berini et al., 2020). La surexpression de SsgA, des protéines impliquées dans le maintien du peptidoglycane et le contrôle des processus nécessitant un remodelage de la paroi cellulaire, provoque une fragmentation des mycéliums et une augmentation de la vitesse de croissance bactérienne ; ceci permet de diminuer les temps de fermentation et d'améliorer la productivité (van Wezel et al., 2006). La sélection des promoteurs forts, la modification des peptides signal ou l'utilisation de différents hôtes de *S. lividans* sont autant d'éléments permettant d'améliorer la production des protéines recombinantes (Sevillano et al., 2016). Grâce à cela, *S. lividans* est devenu l'hôte d'expression préféré pour les gènes du genre *Streptomyces* (Haeger et al., 2024b).

L'expression de l'α-aminoacylase (SgAA) et de l'ε-lysine acylase (SgELA) de *Streptomyces griseus* DSM 40236T a lieu uniquement sous la forme de corps d'inclusion chez *E. coli*, malgré une optimisation des conditions opératoires d'expression. Une stratégie de remplacement de l'hôte d'expression par *Streptomyces lividans*, qui est plus proche génétiquement de la bactérie productrice naturelle, a été mise en place. Les enzymes sont produites au sein de la souche *S. lividans* TK23 avec le plasmide pGH01, qui est un plasmide dérivé du pEM4 porteur du promoteur ermE* de *Saccharopolyspora erythraea* pour l'expression hétérologue (Bibb et al., 1985 ; Quiros et al., 1998). Les cultures en erlenmeyers ont lieu à 30 °C pendant 3 jours dans le milieu YEME (0,3 % d'extrait de levure, 0,5 % de peptone, 0,3 % d'extrait de malt, 1 % de

glucose, 34 % de saccharose, 0,2 % de chlorure de magnésium hexahydraté, le tout (p/v) et 25 ml d'une solution de glycérol à 20 %) supplémenté par 2 µL de polyéthylène glycol (PEG 2000). A l'issue des cultures, de faibles quantités des deux enzymes visées ont pu être purifiées par chromatographie d'affinité. Pour augmenter les quantités produites, une culture en bioréacteur est réalisée. L'expression a lieu dans un bioréacteur de 1,3 L contenant 1 L de milieu de fermentation (0,5 % d'extrait de levure, 3 % de TSB, 0,75 % de glucose, 0,3 % d'extrait de malt, 10 % de saccharose 0,1 % de chlorure de magnésium hexahydraté). La culture est supplémentée avec une solution de glucose à 1,5 g/mL : cette alimentation débute après 15 h d'incubation à un débit de 1 mL/h pendant 7 h, puis le débit est augmenté à 1,5 mL/h jusqu'à la fin de la fermentation. La culture dure 2 et 5 jours pour l'expression respective de l'α-aminoacylase et de l'ε-lysine acylase dans les conditions régulées suivantes : 30 °C, 30 % d'oxygène dissous et pH 7 (Figure I.26). L'α-aminoacylase recombinante purifiée représente 0,56 % des protéines totales dans l'extrait acellulaire. 123 µg d'ε-lysine acylases purifiées sont récupérées, présentant une activité spécifique d'hydrolyse de la N-ε-acétyl-L-lysine (15 mM) de 11,1 U/mg (Haeger et al., 2023a).

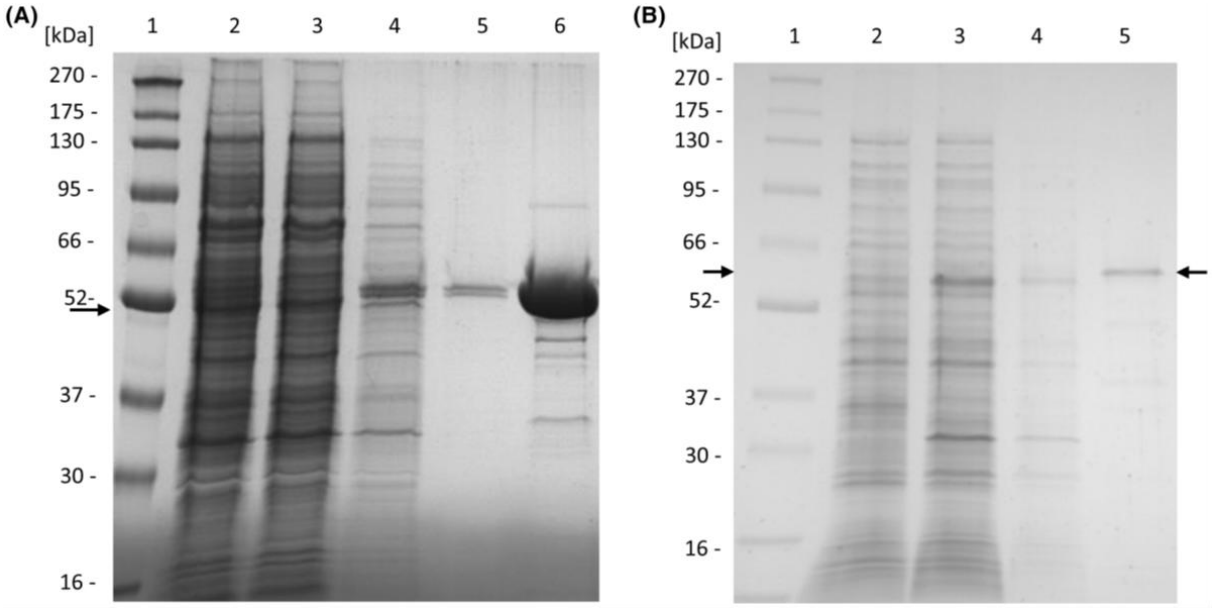


Figure I.26 : Analyse SDS/PAGE des fractions obtenues lors de la purification des aminoacylases recombinantes de *S. griseus*. (A) Chromatographie d'affinité de SgAA. Piste 1 : Marqueur protéique (marqueur protéique précoloré BlueEasy, Nippon Genetics) ; piste 2 : extrait acellulaire ; piste 3 : flux continu ; pistes 4 et 5 : fractions de lavage ; piste 6 : élution de SgAA (B) Chromatographie d'affinité de SgELA. Piste 1 : Marqueur protéique (marqueur protéique précoloré BlueEasy, Nippon Genetics) ; piste 2 : extrait acellulaire ; piste 3 : flux continu ; piste 4 : fraction de lavage ; piste 5 : élution de SgELA. Les flèches noires indiquent la position des protéines exprimées (Haeger et al., 2023a).

Plusieurs aminoacylases issues des génomes de *Streptomyces spp.* sont exprimées chez *E. coli* et *S. lividans*. L'expression a lieu chez *E. coli* BL21 (DE3) avec le plasmide pET-22b, à 20 °C, pendant 24 h et dans le milieu Overnight Express^{TM} Instant TB (Novagen Inc.). Elle a lieu chez la souche *S. lividans* 1326 avec le plasmide pTONA5, construite par conjugaison avec *E. coli* S17-1. Les spores germent sur une boite de pétri à l'agar de soja (2 % de farine de soja, 2 % de mannitol et 2 % d'agar) pendant 5 jours à 30 °C. S'en suit une culture en milieu liquide (0,8 % de K_2HPO_4, 2 % de glucose ou de glycérol, 0,05 % $MgSO_4$ 7 H_2O, 0,5 % de polypeptone et 0,5 % d'extrait de levure) pendant 7 jours à 30 °C. Les enzymes recombinantes sont purifiées partiellement par précipitation différentielle au sulfate d'ammonium, suivie d'une chromatographie ionique. Les taux d'activités les plus élevées sont obtenues avec les enzymes recombinantes de *S. lividans* car il y a formation de corps d'inclusion chez *E. coli*. L'aminoacylase de *S. sp.* 64E6 permet d'obtenir l'activité enzymatique la plus élevée. Il y a 1,1 mg d'aminoacylases purifiées ayant une activité spécifique de 36,1 µmol/min/mg, pour 112 mg de protéines ayant une activité spécifique apparente de 24,7 µmol/min/mg dans l'extrait brut. La production des enzymes recombinantes est augmentée d'un facteur 4,2 en remplaçant le glucose par du glycérol, dans le milieu de culture (Arima et al., 2013).

L'α-aminoacylase (Sm-AA) et l'ε-lysine acylase (Sm-ELA) de *S. mobaraensis* sont deux enzymes qui ont été exprimées avec succès chez *S. lividans*. Le gène de la première enzyme est inséré dans le plasmide pSH19 ayant un promoteur fort et inductible (Herai et al., 2004). Le plasmide résultant est introduit chez *S. lividans* TK24. La culture a lieu dans du milieu TSB, à 30 °C pendant 3 jours. Elle est induite le deuxième jour de croissance avec l'ajout d'isovaleronitrile ou de l'ε-caprolactame à 0,1 % (v/v). L'enzyme recombinante est purifiée par précipitation au sulfate d'ammonium puis par chromatographie d'interaction hydrophobe (Fukatsu et al., 2005). Sm-AA sauvage représente 0,02 % des protéines totales de l'extrait acellulaire de *S. mobaraensis* alors que Sm-AA recombinante représente 3,24 % des protéines totales de l'extrait acellulaire de *S. lividans*. L'activité de Sm-AA obtenue avec l'induction à l'isovaleronitrile est 42 fois supérieur à l'activité de Sm-AA sauvage issue du surnageant de culture (Tableau I.8) (Koreishi et al., 2009b). Le gène de la seconde enzyme sous le contrôle d'un promoteur fort est inséré dans le plasmide pUC702, construit à partir des plasmides pIJ702 et pUC19 (Katz et al., 1983 ; Obata et al., 1989a ; Obata et al., 1989b). *S. lividans* est transformé avec le plasmide obtenu. Les conditions de culture sont similaires à celles de la première enzyme, sauf que la culture dure 6 jours et se fait sans ajout d'inducteur. L'enzyme recombinante est purifiée par chromatographie d'interaction hydrophobe. L'activité totale de

l'enzyme recombinante est 300 fois supérieure à celle des cellules sauvages de *S. mobaraensis* (Koreishi et al., 2009a).

Tableau I.8 : Résumé de la purification de l'aminoacylase SmAA recombinante issue de *S. lividans* TK24 (Koreishi et al., 2009b).

Purification step	Protein (mg)	Total activity (U)	Yield (%)	Specific activity (U/mg)	Purification (fold)
Cell-free extract	142	13600	100	96	1.0
Ammonium sulfate precipitation	55	7088	52	129	1.3
Octyl Sepharose 4 Fast Flow	4.6	2438	18	530	5.5

La penicilline V acylase de *S. mobaraensis* (Sm-PVA) a été exprimée chez *E. coli*, mais l'enzyme recombinante n'est probablement pas exprimée avec ce système à cause de sa structure hétérodimérique. Les équipes de Zhang et Koreishi mentionnent malgré tout un succès d'expression avec une souche de *S. lividans* dont les résultats sont en attente de publication (Zhang et al., 2007 ; Koreishi et al., 2007 ; Torres-Bacete et al., 2015).

1.3.3. *Vibrio natriegens*

Vibrio natriegens est une bactérie marine anaérobie facultative isolée dans la boue des marais salants. Elle a été identifiée comme l'organisme non pathogène à croissance le plus rapide, avec un temps de génération minimal de 9,8 min en milieu complexe (Eagon, 1962). Récemment, elle est devenue un sujet d'étude pour des applications en biologie synthétique, en remplacement d'*E. coli*. Outre sa vitesse de croissance rapide, elle présente plusieurs avantages. Elle peut utiliser diverses sources de carbone grâce à son métabolisme diversifié et peut croître sur des milieux complexes et minimaux, comme ceux utilisés habituellement pour *E. coli*. Son nombre de ribosomes est élevé soutenant sa croissance rapide. Un certain nombre d'éléments génétiques d'*E. coli*, comme les promoteurs, les cassettes de résistances aux antibiotiques, les plasmides, et autres, sont fonctionnels chez *V. natriegens*. La plupart des outils ou techniques de biologie moléculaire établis sont aussi compatibles, comme les méthodes de transformation de l'ADN et les techniques d'ingénierie du génome (Hoffart et al., 2017 ; Hoff et al., 2020 ; Xu et al., 2021).

La souche *V. natriegens* Vmax™ est construite avec un système d'expression T7 fonctionnel et permet d'exprimer des protéines « difficiles » qui sont insolubles ou produites en faibles quantités dans d'autres systèmes d'expression (Weinstock et al., 2016 ; Weinstock et al., 2019 ;

Weinstock et al., 2022). L'α-aminoacylase de *M. smegmatis* est exprimée par cette souche avec succès (Haeger et al., 2023b).

1.3.4. *Komagataella phaffii* (syn. *Pichia pastoris*)

Au cours des dernières années, les levures sont devenues des hôtes eucaryotes attractifs pour la production de protéines hétérologues car elles présentent des avantages significatifs comme les capacités d'effectuer des modifications post-traductionnelles et de former les liaisons disulfures, contrairement aux hôtes procaryotes. Les levures nécessitent des cultures simples à mettre en œuvre, donnant une croissance cellulaire rapide et adaptable à la production à grande échelle. Les gènes hétérologues peuvent être intégrer dans le chromosome de la levure ou dans des plasmides (Unver et al., 2024).

Komagataella phaffii est une levure méthylotrophique, c'est-à-dire qu'elle utilise le méthanol comme source de carbone pour sa production d'énergie dans la voie MUT, avec la présence des deux alcool oxydases (AOX1 et AOX2). Le promoteur de AOX1 peut être utilisé comme le promoteur du gène de la protéine cible, nécessitant la présence de méthanol pour l'induction (Karbalaei et al., 2020). C'est également un micro-organisme approprié pour la sécrétion de protéines recombinantes directement dans le milieu de culture, facilitant ainsi leurs purifications (Damasceno et al., 2012 ; Juturu et al., 2018).

L'expression de l'ε-lysine acylase de *S. ambofaciens* (SamELA) est réalisée chez la levure *Pichia pastoris* X33 où son gène optimisé est cloné dans le plasmide pPICZαA. Cette souche possède les deux gènes codant pour des alcool oxydases, nommées AOX1 et AOX2, qui sont sous le contrôle de promoteurs forts inductibles par du méthanol : son phénotype Mut+ signifie qu'elle est capable de se développer avec du méthanol comme seule source de carbone. Toutefois, une activité d'hydrolyse de la N-α-acétyl-L-lysine est observée montrant une altération de spécificité de l'enzyme pour la N-ε-acétyl-L-lysine (Ferrari, 2014).

1.3.5. Cellules d'insectes

Les aminoacylases 1 humaine et porcine étaient isolées à partir de tissus rénaux. Le système d'expression *Baculovirus* dans les cellules d'insectes *Spodoptera frugiperda* a été utilisé pour la production de ces enzymes. Les gènes codant pour ces deux aminoacylases sont clonés dans le vecteur pVL1393 (PharMingen). La première étape est l'obtention des baculovirus recombinants : les virus sont cultivés pendant 24 h dans du sérum de veau fœtal auquel sont

ajoutés le vecteur recombinant, l'ADN du virus de la polyédrose et la lipofectine. Puis le milieu de culture est échangé contre du milieu commercial SF-900 (Gibco-BRL) supplémenté avec 10 % de sérum de veau fœtal et 0,1 % de Pluronic F68 et inoculé par des cellules de *S. frugiperda* (Sf21). Cette culture est menée à 27 °C durant 5 jours. Une autre culture de cellules de *S. frugiperda* menée dans le milieu commercial supplémenté est infectée par le surnageant de la première culture, pour obtenir un stock de virus à titre élevé. L'expression des aminoacylases 1 a lieu avec une culture de cellules *S. frugiperda* infectées pendant 6 jours. Les enzymes sont purifiées par chromatographie d'interaction hydrophobe puis par chromatographie anionique. L'expression des aminoacylases porcine et humaine est à son maximum après 3 et 4-5 jours de production respectivement. L'aminoacylase 1 porcine est également exprimée dans un système d'expression bactérien de type *E. coli* BL21 (DE3). L'activité spécifique obtenue envers la N-acétyl-méthionine est très faible : elle est de 1 U/mg pour l'enzyme recombinante bactérienne contre 200 U/mg pour l'enzyme recombinante obtenue à partir des cellules d'insecte (Pittelkow et al., 1998).

1.3.6. Conclusions

Pour conclure, les conditions d'expression hétérologue des aminoacylases sont diverses et variées, suggérant des difficultés pour produire des enzymes en grande quantité sous une forme soluble et active (Heager et al., 2024b). Dans les travaux menés jusqu'ici, l'hôte classique de production, *E. coli* BL21 (DE3) a été abandonné au profit de versions dérivées ou d'hôtes plus complexes à mettre en œuvre, comme les bactéries *S. lividans* et *V. natrigens,* la levure *K. phaffii* et les cellules d'insectes. Les aminoacylases ont souvent une structure multimérique, ce qui laisse penser que la souche *E. coli* Origami (B) (DE3) présente un intérêt pour leur expression. Le biais des codons peut être contourné en utilisant une version optimisée du gène d'intérêt pour son expression chez l'hôte choisi ou en utilisant la souche *E. coli* Rosetta 2 (DE3). La co-expression de protéines chaperonnes semble aider au bon repliement des aminoacylases. Les travaux antérieurs montrent que les conditions opératoires de la culture de l'hôte peuvent être optimisées : utilisation d'un milieu enrichi, réduction du temps de culture, abaissement de la température post-induction ou ajout d'ions métalliques jouant alors un rôle structural pour l'aminoacylase. Le type d'induction et la concentration en inducteur sont aussi des paramètres à prendre en compte. La réussite de l'expression hétérologue d'une aminoacylase se cache dans la bonne combinaison de ces différents éléments, dont certains peuvent apparaître comme des détails alors qu'ils constituent des points clé de la production. La détermination des conditions

favorables qui conduiront à une production d'aminoacylase en grande quantité, sous une forme soluble et active, nécessite un travail approfondi car aucune tendance n'est ressortie de ce travail bibliographique pour l'élaboration de la démarche à suivre.

Jusqu'ici, malgré des travaux menés sur la recherche de ces conditions, les aminoacylases demeurent produites majoritairement sous la forme de corps d'inclusion. Envisager une stratégie de valorisation ces corps d'inclusion peut être une solution. Au vu de ces difficultés, le délai nécessaire à la production, la purification et la caractérisation structurale de ces enzymes reste incertain. Dans ce contexte et dans le but d'acquérir davantage de connaissances sur ces enzymes, l'utilisation d'approches numériques apparaît comme une stratégie pertinente en attendant de pouvoir élucider leur structure tridimensionnelle.

1.4. Modélisation moléculaire appliquée à la biocatalyse

En modélisation moléculaire, les lois de la biologie, de la chimie et de la physique sont exploitées dans la création de programmes informatiques pour l'étude du comportement des systèmes moléculaires. La modélisation a de plus en plus d'applications comme la prédiction de la structure tridimensionnelle des protéines et la compréhension de la sélectivité des réactions enzymatiques. En 2018, 106 millions de séquences protéiques sont identifiées comme séquences de référence et non redondantes dans la base de données publiée par le National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/), alors que seulement 144000 structures tridimensionnelles de protéines élucidées expérimentalement sont référencées dans la banque mondiale de données sur les protéines (Protein Data Bank, PDB, https://www.rcsb.org/) (Muhammed et al., 2018). L'obtention des structures résolues des enzymes en vue de leur caractérisation peut se révéler complexe, augmentant l'attrait pour les méthodes de prédiction de la structure tridimensionnelle des protéines (Cavasotto et al., 2009). Deux approches majeures sont développées pour la prédiction de la structure d'une protéine d'intérêt :

- La modélisation basée sur un ou plusieurs templates, dont la structure tridimensionnelle est déjà résolue. Celle-ci est utilisée pour modéliser la structure inconnue de la protéine d'intérêt. Cette méthode appelée modélisation par homologie repose sur l'identification d'un ou plusieurs templates et l'alignement de sa (leurs) séquence(s) et de la séquence d'intérêt. C'est cette approche qui sera mise en œuvre dans le cadre de ce travail, c'est pourquoi elle sera détaillée par la suite.

- La modélisation sans template, où la similarité globale avec une structure de la PDB n'est pas utilisée. Cette démarche met en œuvre un échantillonnage à grande échelle et l'application de fonctions énergétiques basées sur les lois de la physique (Kuhlman et al., 2019).

La structure tridimensionnelle des aminoacylases de *S. ambofaciens* est inconnue à ce jour, comme c'est le cas pour de nombreuses enzymes de cette classe. Toutefois, comme dit précédemment, il est possible de prédire la structure d'une protéine par modélisation. Une fois le modèle obtenu, il devient possible d'étudier les interactions enzyme-substrat par des simulations de docking moléculaire, comme dans le cas des enzymes de structure connue (Selvam et al., 2017).

1.4.1. Modélisation par homologie

La structure d'une protéine est déterminée par sa séquence en acides aminés. Des séquences similaires en acides aminés sont supposées se replier en structures identiques. Il est admis que la structure tridimensionnelle des protéines présente un niveau de conservation supérieur à celui des séquences car l'évolution structurale est plus lente que l'évoluation génétique. La modélisation par homologie s'appuie sur ces trois éléments. C'est une méthode informatique de prédiction de la structure tridimensionnelle d'une protéine, à partir de sa séquence en acides aminés. Elle se base sur la séquence et la structure élucidée d'un ou plusieurs modèles appelés « templates ». La construction du modèle sur la base de la similarité de séquences implique que la protéine cible et le ou les template(s) partagent un pourcentage d'homologie le plus élevé possible.

La littérature scientifique décrit une méthode permettant la construction de modèles d'homologie robustes, en suivant sept étapes qui seront détaillées par la suite (Figure I.27) :

- Identification et sélection des templates.
- Alignements de séquences et correction d'alignement.
- Construction de la chaine peptidique.
- Construction des boucles.
- Modélisation des chaines latérales des acides aminés.
- Optimisation du modèle.
- Validation du modèle (Muhammed et al., 2018).

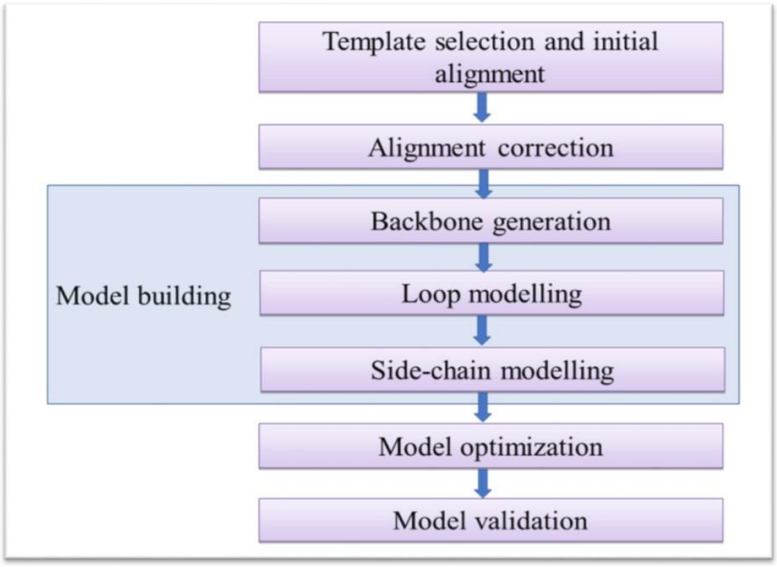


Figure I.27 : Les sept étapes de la modélisation par homologie (Murmu et al., 2022).

Un mauvais choix de template, des erreurs d'alignement, des erreurs de décalage dans des régions quasi-alignées, des imprécisions au niveau des chaînes latérales et/ou un faible pourcentage d'homologie peuvent être des sources d'inexactitude dans la création d'un modèle d'homologie (Haddad et al., 2020).

1.4.1.1. Identification des templates

La première étape dans la modélisation par homologie est l'identification des templates, qui sont des protéines dont la structure tridimensionnelle a déjà été résolue et partageant un fort pourcentage d'homologie avec la protéine cible. Pour trouver ces templates, la séquence en acides aminés de la protéine cible est comparée à celles disponibles dans une base de données comme la PDB.

Plusieurs outils reposant sur des approches différentes sont disponibles pour identifier les templates potentiels. L'outil de recherche d'alignement local de base (BLAST) est disponible sur le site du NCBI. Il produit un alignement séquence-séquence par paires (Altschul et al., 1997; Altschul et al., 2005; Johnson et al., 2008). D'autres outils utilisent une approche différente en réalisant des alignements profil-profil ou des modèles de Markov cachés (Hidden Markov Models ; HMM), parfois en intégrant avec des propriétés structurelles. Ainsi MMseqs2 est un logiciel disponible sur l'interface web ColabFold (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) générant des alignements de séquences multiples en utilisant une approche HMM sur la base de données de séquences UniRef (https://www.uniprot.org/uniref) (Mirdita et al., 2017 ;

Steinegger et al., 2019). L'interface web de Swiss Model (https://swissmodel.expasy.org/interactive) utilise le logiciel HH_suite3 produisant des alignements sous forme de profil HMM avec les bases de données de séquences de protéines en cluster (Uniclust90, Uniclust50 et Uniclust30 ; https://uniclust.mmseqs.com/). Les clusters sont réalisés à partir de la suite logicielle MMseqs2 en utilisant la banque de données UniProtKB (https://www.uniprot.org/uniprotkb). Il est intéressant de travailler avec un représentant des différents sous-ensembles ayant des séquences similaires. Cela permet une économie de ressource de calcul et une meilleure sensibilité dans les résultats de recherche de similarité de séquences (Camacho et al., 2009 ; Mirdita et al., 2017 ; Steinegger et al., 2019).
Le choix du ou des template(s) est primordial. Il est conseillé d'avoir un pourcentage de similarité supérieur à 25% pour générer une structure tridimensionnelle précise et fiable (Rost et al., 1996). Les facteurs environnementaux (pH, type de solvant, présence d'un ligand, etc.) et la résolution de la structure des templates sont également des facteurs à prendre en compte pour le choix définitif des templates, car ils interviennent dans la construction de la structure de la protéine cible (Saxena et al., 2013).

1.4.1.2. Alignements de séquences et corrections d'alignement

Une fois le ou les templates sélectionnés, les alignements de séquences avec la séquence de la protéine cible peuvent avoir lieu. Il est possible de réaliser un alignement entre la protéine cible et un ou plusieurs templates et entre les templates quand plusieurs templates sont utilisés. Les alignements entre la protéine cible et les templates peuvent être corrigés et optimisés pour augmenter la qualité des structures tridimensionnelles modélisées, car l'erreur au niveau d'un acide aminé dans l'alignement provoque le décalage du carbone α du résidu correspondant, entraînant la modification de la rotation d'une hélice α par exemple (Dalton et al., 2007). La correction d'un alignement de séquences prend majoritairement la forme d'une modification des paramètres du calcul d'alignement permettant l'identification et la correction des mésappariements (substitutions d'acides aminés) et des écarts (insertions et délétions d'acides aminés) (Muhammed et al., 2018).
La suite informatique Discovery Studio® propose deux types de calculs pour la réalisation des alignements de séquences « Align Sequences… » et « Align Sequences to Templates… ». La méthode utilisée, Align123, est basée sur le programme ClustalW et permet d'effectuer des alignements multiples. Ce programme aligne les séquences grâce à un algorithme d'alignement progressif par paires (Thompson et al., 1994). Les structures secondaires prédites et les

séquences multimériques peuvent être prise en compte dans la construction et la correction de l'alignement sur Discovery Studio®. Il n'est pas possible de modifier les alignements générés par Swiss Model et ColabFold car ils sont automatisés.

1.4.1.3. Modélisation de la structure tridimensionnelle des protéines

Une fois le(s) template(s) sélectionné(s) et l'alignement de séquences obtenu, les modèles de la structure de la protéine cible peuvent être générés. Les méthodes d'assemblages de corps rigides (rigid-body assembly methods), les méthodes de mise en correspondance de segments (segment matching methods), les méthodes de contraintes spatiales (spatial restraint methods) ou les méthodes d'évolution artificielle (artifical evolution methods) sont des approches mises en œuvre pour la construction des modèles (Muhammed et al., 2019).

La méthode basée sur les contraintes spatiales construit le modèle en tenant compte des contraintes issues de la structure tridimensionnelle du template. Ces contraintes sont stéréochimiques et impliquent les longueurs et les angles de liaison, les angles dièdres et les distances de contact de Van der Waals. Cette méthode est utilisée dans le programme Modeller (Sali et al., 1993).

En ce qui concerne l'approche par assemblage des corps rigides, la structure protéique est décomposée en éléments conservés, appelés les corps rigides. Ces derniers sont assemblés pour générer la structure tridimensionnelle de la protéine cible. Cette méthode est appliquée dans le programme Promod3 implémenté dans le serveur Swiss Model (Guex et al., 2009 ; Bienert et al., 2017 ; Waterhouse et al., 2018). Dans le cadre d'une étude portant sur la modélisation par homologie des structures multimériques, 28 critères ont été analysés montrant la pertinence de l'identité de séquences, de la similarité de la structure secondaire de la zone de contact monomère-monomère et de la courbe des scores de conservation des résidus pour la détermination de la structure quaternaire de la protéine cible. La zone de contact entre les monomères peut être facilement identifiée car c'est un élément conservé. Le score de conservation repose sur trois calculs : un calcul d'énergie (entropie relative) calculé pour chaque résidu d'un alignement, un calcul d'énergie au niveau de l'interface (préalablement déterminée comme la zone accessible aux solvants) pour évaluer la conservation des acides aminés de l'interface et enfin un calcul d'itération tenant compte de l'enfouissement des acides aminés dans le domaine de dimérisation. Ce score est négatif si le résidu appartient à la zone d'interface et égal à zéro pour les autres résidus. Il est alors possible de tracer une courbe sur laquelle les différentes zones d'interface sont clairement identifiées (Bertoni et al., 2017).

AlphaFold est un logiciel d'intelligence artificielle développé par l'entreprise Google DeepMind en 2018. Il repose sur un réseau de neurones résiduel appelé ResNet qui a pour architecture un réseau neuronal convolutionnel très profond. Il utilise des algorithmes statistiques globaux, avec des paramètres de couplage de pseudo-variance et de couplage direct, pour générer les caractéristiques d'entrées des cellules (fields), au nombre de 64 x 64 (Kandathil et al., 2019). Le réseau neuronal convolutif profond d'AlphaFold utilise des informations globales relatives à des paires de résidus avec des scores de coévolution, pour améliorer la prédiction de la distance entre résidus. Autrement dit, il prédit la distribution de la distance entre les résidus puis convertit cette distribution en potentiel de distribution spécifique aux protéines. Ce potentiel est minimisé grâce à un échantillonnage conformationnel et des méthodes basées sur le gradient pour la création des modèles tridimensionnels. AlphaFold n'utilise pas les structures secondaires au moment de l'inférence, mais il modélise les distances et les angles suivant des distributions de probabilité qui mènent indirectement à la structure secondaire de la protéine. L'apprentissage profond intègre des caractéristiques structurales unidimensionnelles, des contacts bidimensionnels et des scores de qualité de la structure tridimensionnelle, ce qui améliore la prédiction des distances entre les résidus et résout presque entièrement le problème de prédiction de la topologie de repli des protéines (Xu et al., 2019 ; Senior et al., 2019 ; Hou et al., 2019).

Deux ans plus tard, Google DeepMind dévoile une nouvelle version du logiciel, AlphaFold2, dont l'architecture est totalement différente de celle d'AlphaFold. Ce nouveau système itératif est basé sur une architecture de réseau neuronal de type Evoformer, plutôt que de simples convolutions comme AlphaFold. AlphaFold2 traite directement les informations des alignements de séquences multiples et les représentations intermédiaires par paires de résidus (y compris les informations issues des templates) (Jumper et al., 2021). Le code source de ce logiciel est publié, rendant son utilisation libre. ColabFold (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb) est une interface web développée par une équipe de recherche d'Havard permettant d'utiliser facilement AlphaFold2 (Mirdita et al., 2022). MMseqs2 est intégré à ColabFold et permet une prédiction accélérée de la structure tridimensionnelle d'une protéine, *via* une recherche d'homologie rapide (Kim et al., 2024).

En 2024, Google DeepMind propose AlphaFold3 (https://alphafoldserver.com/welcome). En plus de prédire la structure tridimensionnelle des protéines, le logiciel peut donner des informations sur les interactions avec l'ADN, l'ARN et certains ligands comme les ions. Il utilise l'architecture Pairformer, qui est similaire mais plus simple que l'architecture Evoformer

introduite avec AlphaFold2. AlphaFold3 est disponible depuis un serveur gratuit pour la recherche non commerciale (Abramson et al., 2024).

La modélisation des boucles dans les structures protéiques constitue un véritable challenge. Les boucles déterminent souvent la spécificité d'une protéine. Or elles correspondent à des lacunes ou des insertions dans les séquences des protéines, entrainant l'apparition de zones non-alignées au sein des alignements de séquences. Ce ne sont pas des éléments conservés. Elles présentent une variabilité structurelle élevée comparativement aux feuillets β ou aux hélices α, rendant la prédiction de leur structure difficile et augmentant la probabilité d'une conformation erronée (Kmiecik et al., 2016). Il existe deux méthodes principales pour la modélisation des boucles : la première repose sur une recherche dans les bases de données tandis que la deuxième est basée sur une recherche conformationnelle, parfois couplée avec des méthodes *de novo* (Fernandez-Fuentes et al., 2006 ; Werner et al., 2012).

En ce qui concerne les chaines latérales des résidus, celles-ci sont modélisées en les ajoutant sur le squelette de la chaine peptidique préalablement modélisée. Leur placement est ajusté grâce à des simulations *ab initio* et/ou en utilisant les coordonnées provenant du template dans le cas d'un partage de haut niveau d'identité de séquences. Les rotamères qui sont les différentes conformations possibles des chaines latérales sont sélectionnés sur la base de critères énergétiques, de la séquence protéique et des coordonnées de la chaine peptidique (backbone). Le positionnement des résidus constituant le cœur hydrophobe de la protéine est plus précis que celui des résidus exposés à la surface (Samudrala et al., 1998).

En fonction des outils choisis, la modélisation du squelette de la chaine peptidique, des boucles et des chaines latérales peut se faire en une à trois étapes. Les outils présentés précédemment modélisent la structure tridimensionnelle en une seule étape parfois modulable en fonction des options. Par exemple, le programme Modeller propose l'option "Refine Loop » qui permet d'affiner la conformation des boucles (Fiser et al., 2000 ; Shen et al., 2006). Il existe des concours pour étudier les performances des différents programmes de modélisation par homologie et aider les utilisateurs à choisir le bon outil selon leurs besoins de recherche. Le projet CAMEO (Continuous Automated Model EvaluatiOn ; www.cameo3d.org) est un projet communautaire dans lequel les performances des différents outils utilisés en modélisation de la structure des protéines sont évaluées par différents scores de façon automatisée. Les résultats sont communiqués de façon hebdomadaire pour deux catégories de serveurs de modélisation (prédiction de structure de protéine (3D) et estimation de la qualité du modèle (QE)), et de façon discontinue pour deux autres catégories d'outils de modélisation (prédiction de contact (CP) et liaison du ligand (LB)). La communauté travaille par ailleurs au développement d'une nouvelle

catégorie disponible en version béta évaluant les complexes protéiques (Haas et al., 2018 ; Haas et al., 2019 ; Robin et al., 2021 ; Leemann et al., 2023). L'initiative CASP (Critical Assessment of Protein Structure Prediction ; predictioncenter.org) est une expérience débutée en 1994 et menée tous les deux ans dont le but est d'étudier la précision et l'amélioration de la modélisation de la structure tridimensionnelle des protéines. Lors des éditions CASP12 et de CASP13, la modélisation par homologie a fait de gros progrès dans quatre domaines : l'alignement séquence-structure, la combinaison des templates servant de structures de référence, la modélisation des boucles et l'assemblage des protéines (Moult et al., 2017 ; Guzenko et al., 2019). En 2018, les programmes Modeller et Swiss Model sont présentés comme deux des meilleurs outils de modélisation de structures tridimensionnelles protéiques par modélisation d'homologie. AlphaFold a fait son entrée avec une première place à l'édition CASP13, mais ce logiciel n'est pas divulgué par DeepMind (Kryshtafovych et al., 2019). Lors de l'édition CASP14 en 2020, AlphaFold2 confirme sa place de leader dans la prédiction de structures protéiques grâce à l'utilisation de l'apprentissage profond (Kryshtafovych et al., 2021). DeepMind publie alors son code source rendant son utilisation possible. Depuis, des logiciels dérivés d'AlphaFold2 ont été développés et ont obtenu les meilleurs scores lors de CASP15 (Kryshtafovych et al., 2023).

1.4.1.4. Validation de la structure tridimensionnelle modélisée

La vérification et la validation des modèles obtenus est nécessaire car leur précision dépend de la similarité des séquences, des outils de modélisation employés, de la qualité des templates. Il est recommandé d'utiliser plusieurs stratégies pour valider la robustesse et la fiabilité des modèles, car chaque méthode d'évaluation étudie la structure sous un angle différent. Les méthodes d'évaluation sont réparties en quatre catégories (Haddad et al., 2020) :

- Les méthodes basées sur les lois de la physique reposant sur des calculs d'énergie grâce à des champs de force et vérifiant la présence d'une conformation optimale.
- Les méthodes reposant sur la connaissance. L'approche de validation applique des scores représentatifs des énergies obtenus statistiquement à partir de structures tridimensionnelles résolues expérimentalement et répertoriées dans une base de données.

Swiss Model a recours à l'analyse énergétique Qualitative Model Energy ANalysis (QMEAN) évaluant l'énergie d'interaction Cβ, l'énergie par paires de tous les atomes, l'énergie d'angle de torsion et l'énergie de solvatation au sein des modèles (Benkert et al., 2011). Disco est, quant à lui, un score de contrainte de distance basé sur le consensus, utilisant la distribution de

distances inter-résidus des structures protéiques issues de la PDB qui est comparée à celle des modèles évalués. Ce score est ajouté au score de QMEAN permettant ainsi l'évaluation d'un ensemble de modèles tridimensionnels par le biais d'un score composite QMEANDisCO. Ce dernier combine la précision d'une méthode de consensus (Disco) et d'une approche de modèle unique (QMEAN) permettant une estimation de la qualité des modèles au niveau global et local. Actuellement, ce score est normalisé sous le nom de QMEANDisCo Global (Studer et al., 2020).

La suite informatique Discovery Studio® met à disposition deux outils d'évaluation pour les structures tridimensionnelles modélisées : le DOPE Score normalisé et le Profiles-3D. Le DOPE Score normalisé est un score d'énergie prenant en compte la taille des protéines. Il permet de comparer les structures modélisées en leur attribuant un score, dont le calcul est basé sur la méthode MODELER DOPE (Discrete Optimized Protein Energy) (Shen et al., 2006 ; Eramian et al., 2009). Profiles-3D vérifie la validité d'une structure tridimensionnelle en calculant un score de compatibilité au niveau de chaque résidu et de la protéine entière dans l'environnement tridimensionnel donné, avec une comparaison à deux références. Les références sont les scores issus de l'analyse statistique des structures à haute résolution recensées dans la PDB et ayant la même composition en acides aminés que la protéine cible. Il y a création de groupes de résidus valides (avec un score positif) et de résidus invalides (avec un score négatif) (Bowie et al., 1991).

- Les méthodes utilisant l'apprentissage automatique.
- Les méthodes expérimentales, où la structure obtenue par modélisation est comparée aux données expérimentales sur la liaison aux ligands ou sur la structure résolue de la protéine cible ou d'un template.

La méthode la plus classique est la détermination de la déviation quadratique moyenne (RMSD) qui donne une moyenne des écarts entre tous les atomes de deux structures tridimensionnelles. Il est recommandé d'appliquer cette méthode en divisant la protéine en fragments, de façon à isoler la contribution des boucles dans la valeur globale de RMSD (Haddad et al., 2016).

Les scores lDDT et PAE quant à eux sont calculés sur les modèles générés par ColabFold. Le lDDT (test de différence de distance locale) évalue la reproduction des interactions locales trouvées dans plusieurs structures protéiques de référence au sein de la structure modélisée (Mariani et al., 2013). Le PAE (Predicted Align Error) est un score sous la forme d'une matrice colorée en deux dimensions indiquant l'erreur de position du résidu x si la structure prédite et la structure de référence sont alignées sur le résidu y (Varadi et al., 2021).

1.4.1.5. Optimisation du modèle

L'optimisation de la structure tridimensionnelle modélisée peut se faire avant ou après l'étape de validation. Elle est conseillée pour diminuer le niveau énergétique du modèle et ajuster sa structure pour obtenir une conformation dite « naturelle ». Le modèle est généralement optimisé par mécanique moléculaire (Haddad et al., 2020).

La mécanique moléculaire modélise l'énergie potentielle d'un système en utilisant la disposition spatiale des atomes constitutifs. Des approximations sont utilisées pour simplifier la description du système modélisé : les atomes sont traités comme des particules sphériques, de charge constante et caractérisées par leur rayon de Van der Waals ; les liaisons covalentes sont considérées comme des ressorts. L'énergie potentielle du système moléculaire peut être exprimée sous la forme d'une fonction correspondant à un champ de force. Un champ de force classique est composé de deux parties décrivant d'une part les interactions entre les atomes reliés par des liaisons covalentes (comme les étirements des liaisons chimiques, les torsions et les flexions des angles dièdres et les variations d'angle impropre) et d'autre part, les interactions non-covalentes regroupant les interactions de Van der Waals et les interactions électrostatiques désignées sous les termes respectifs de Lennard-Jones et de Coulomb (Ponder et al., 2003). Le champ de force est choisi en fonction de la molécule étudiée. Les champs de force CHARMm (Chemistry at Harvard Molecular mechanics) ou AMBER (Assisted Model Building and Energy Refinement) sont principalement utilisés pour les simulations menées sur des structures protéiques (Momany et al., 1992 ; Case et al., 2005 ; Brooks et al., 2009).

Une fois le modèle de structure généré, il est possible d'optimiser celui-ci en effectuant une minimisation d'énergie potentielle qui est une simulation de mécanique moléculaire utilisant un champ de force. Le but est de rechercher la conformation de la protéine possédant la plus faible énergie possible en parcourant sa surface d'énergie potentielle. Cet état de basse énergie est susceptible de correspondre à la conformation la plus probable du système. Le processus est itératif et continue jusqu'à ce que la différence d'énergie calculée entre deux conformations successives soit inférieure à un seuil préalablement établi, indiquant qu'il y a convergence au niveau d'un minimum énergétique. La descente la plus raide (steepest descent), les gradients conjugués et Newton-Raphson sont des algorithmes couramment utilisés pour les minimisations d'énergie (Werner et al., 2012). Parfois, ces outils de minimisation sont directement implémentés dans les programmes de modélisation par homologie. C'est le cas de Swiss Model qui réalise systématiquement une étape de minimisation d'énergie sur les modèles de structure tridimensionnelle qu'il génère (Studer et al., 2021). La suite logicielle Discovery

Studio® propose également plusieurs algorithmes de minimisation d'énergie (Powell, 1977 ; Hilderbrandt, 1977 ; Durmus et al., 2008). L'utilisation de ces algorithmes permet de mettre en place des restrictions de déplacement des atomes et la mise en œuvre de la minimisation d'énergie par étapes progressives. Pour une optimisation plus poussée des modèles, des méthodes comme la dynamique moléculaire peuvent être employées (Muhammed et al., 2019).

1.4.2. Simulations de docking moléculaire

Une fois le modèle tridimensionnel d'une enzyme obtenu grâce à la modélisation par homologie, il devient alors possible d'étudier les interactions avec des substrats, en réalisant des simulations de docking. Celles-ci permettent la prédiction des modes d'interaction préférentiels entre deux molécules et donnent accès à des informations importantes comme la position et l'orientation d'un substrat dans la cavité catalytique d'une enzyme. Pendant une simulation de docking, chaque angle de torsion du substrat est incrémenté de manière aléatoire, générant des conformères issus de son espace conformationnel. Ces derniers sont ensuite placés selon différentes orientations au sein de la cavité catalytique. En fonction, du programme utilisé, il est nécessaire ou pas d'identifier le site actif de l'enzyme en amont des simulations (Agarwal et al., 2016). Certaines cibles protéiques de docking contiennent des ions métalliques et des molécules d'eau ayant des rôles importants pour la catalyse et les liaisons du substrat à l'enzyme. Il convient alors de les prendre en compte lors des simulations de docking (Cheng et al., 2012). Les données produites à l'issue du docking correspondent à un ensemble de poses représentant les différentes conformations de complexes substrat/enzyme favorables sur le plan énergétique et géométrique. Le scoring est le classement de ces poses à l'aide de fonctions de score évaluant la complémentarité du substrat avec l'enzyme selon la forme, l'énergie du complexe ou sa stabilisation par des interactions (Yuriev et al., 2015).

Il existe trois catégories d'algorithmes de docking, selon la prise en compte de la flexibilité du substrat et/ou de l'enzyme : le docking des corps rigides, le docking semi-flexible et le docking flexible. Dans le cas du docking rigide, le substrat et l'enzyme sont considérés comme deux corps rigides, limitant la précision des résultats mais raccourcissant les temps de calcul. Les protocoles de docking semi-flexible et flexible permettent une flexibilité au niveau du substrat ou au niveau de l'enzyme pour les premiers et une flexibilité du substrat et de l'enzyme pour les seconds, augmentant l'espace de recherche conformationnelle et le temps de calcul. La précision des simulations est améliorée lorsque le substrat et l'enzyme sont flexibles (Fan et al., 2019). La suite logicielle Discovery Studio® propose plusieurs outils de docking de

ligands comme CDOCKER, LibDock ou encore un algorithme de docking flexible. CDOCKER est une méthode de docking semi-flexible basé sur CHARMm où le récepteur est maintenu rigide tandis que plusieurs conformations du ligand sont testées (Wu et al., 2003). LibDock quant à lui, utilise des caractéristiques de site protéique appelées HotSpots où les HotSpots polaires et apolaires sont susceptibles d'interagir préférentiellement avec les atomes polaires et apolaires du ligand, respectivement. Le fichier HotSpot de l'enzyme est déterminé avant la simulation des poses du ligand. Puis lors de la simulation, le ligand s'apparie aux HotSpots identifiés. Les atomes d'hydrogène du ligand ne sont pas présents lors de l'étape de génération des poses, ce qui nécessite une étape d'optimisation post-docking pour garantir leur positionnement (Diller et al., 2001).

Les simulations de docking intègrent une étape de scoring dont le rôle est d'évaluer les poses générées selon diverses fonctions de notation. Plus précisément, le score est utilisé pour évaluer l'énergie des poses et la complémentarité géométrique et chimique des éléments constitutifs de ces poses (par exemple une enzyme et son substrat). Les outils de scoring utilisent des fonctions de notation empiriques, basées sur la connaissance ou l'apprentissage automatique ou sur des simulations de dynamique moléculaire (Paggi et al., 2024). Généralement, chaque outil de docking possède sa propre fonction de notation permettant le classement des complexes récepteur-ligand générés sous la forme d'une liste. Par exemple, chaque pose générée avec CDOCKER se voit attribuer un score CDOCKER incluant l'énergie de contrainte interne du ligand et l'énergie d'interaction récepteur-ligand (Wu et al., 2003). LigScore1, LigScore2, PLP1, PLP2 et Jain sont des fonctions de notations empiriques. Elles sont composées de plusieurs termes renvoyant aux interactions protéine-ligand, qui sont reconnues expérimentalement comme importantes pour la détermination de l'énergie de liaison au sein d'un complexe et qui sont additionnées. Par exemple, les atomes se chevauchant et la répulsion électrostatique conduisent à une pénalité de score alors que l'attraction électrostatique et les interactions favorables de type Van der Waals et liaison hydrogène contribuent positivement au score. LigScore1 et LigScore2 sont des fonctions de notation pour la prédiction de l'affinité récepteur-ligand fonctionnant avec les champs de force CCF (Consistent Force Field) ou Dreiding (Krammer et al., 2005). Le score PLP1 (Piecewise Linear Potential) est la somme des valeurs fonctionnelles des interactions par paires comprenant les liaisons hydrogène et les interactions stériques dans un complexe récepteur-ligand. Le score PLP2 est également une somme des valeurs de trois types d'interactions par paires (la liaison hydrogène, la dispersion et la répulsion) en attribuant un rayon atomique à chaque type d'atome au sein du complexe (Gehlhaar et al., 1995 ; Gehlaar et al., 1999). Le score Jain évalue les structures et les affinités

de liaison d'une série de complexes protéine-ligand et donne la somme de cinq termes d'interaction (interactions lipophiles, interactions d'attraction polaire, interactions de répulsion polaire, la solvatation de la protéine et du ligand et un terme d'entropie pour le ligand) (Jain, 1996). PMF et PMF4 sont des fonctions de notation basées sur la connaissance et plus spécifiquement sur l'analyse statistique des structures ligand-récepteur connues et la fréquence d'occurrence d'interactions ligand-récepteur spécifiques, sans tenir compte de l'affinité de liaison. La fonction de score PMF est développée à partir d'une analyse statistique de structures tridimensionnelle de complexes protéine-ligand. Ce score est calculé en additionnant les termes d'interaction par paires sur toutes les paires interatomiques présente au sein du complexe (Muegge et al., 1999). Le score PMF04 est une version actualisée du score PMF, incluant davantage de données ainsi que les potentiels des ions métalliques et des halogènes (Muegge, 2006). Toutes les fonctions de scores présentées évaluent de façon différente les complexes protéine-ligand. C'est pourquoi il est intéressant de les combiner dans un protocole de consensus, permettant d'identifier les meilleurs complexes protéine-ligand en prenant en compte les résultats de plusieurs fonctions de score.

1.4.3. Conclusions

La modélisation par homologie est une méthode informatique de prédiction ayant pour objectif de modéliser la structure tridimensionnelle d'une protéine, dans l'attente de l'élucidation de sa structure par une voie expérimentale. La modélisation se base sur des templates qui sont des protéines dont la structure tridimensionnelle est déjà résolue et partageant un fort pourcentage d'homologie avec la protéine cible. Des résidus clés de l'enzyme étudiée peuvent être identifiés grâce des alignements de séquences, si des résidus impliqués dans la catalyse sont déjà connus chez les templates. L'alignement de séquences et la structure modélisée nécessitent une optimisation avant leur utilisation ou analyse. Une fois le modèle tridimensionnel construit, les simulations de docking moléculaire permettent d'identifier les interactions ligand-protéine et d'étudier par exemple la sélectivité d'une enzyme. Les meilleurs poses du ligand au sein de la protéine sont déterminées par plusieurs fonctions de score dont les résultats peuvent être condensés dans un score consensus. Une fois les complexes les plus probables identifiés, il est possible d'étudier en détail les jeux d'interaction, d'évaluer la proximité des substrats aux résidus catalytiques, de faire des hypothèses sur le rôle de certains résidus dans la reconnaissance et/ou la fixation du substrat ou encore de proposer un mécanisme catalytique de l'enzyme étudiée.

Une seule étude récente portant sur la prédiction de la structure tridimensionnelle d'une aminoacylase de *Stenotrophomonas maltophilia via* une méthode de modélisation par homologie a été publiée à notre connaissance. Les travaux visaient l'étude de la régiosélectivité de l'enzyme en utilisant des simulations de docking moléculaire. La structure tridimensionnelle de l'enzyme a été modélisée sur le serveur de Swiss Model. Grâce au template identifié, les résidus impliqués dans la fixation des ions Mn^{2+} et dans la catalyse ont pu être identifiés. Des simulations de docking ont été menées avec le programme Autodock 4.0 (https://rsd3.scripps.edu/), pour expliquer la synthèse énantiosélective du L-phosphinothricine par cette enzyme (Cao et al., 2020).

Dans le cadre de notre étude, la modélisation présente un intérêt particulier car la structure des aminoacylases de *S. ambofaciens* reste inconnue à ce jour. L'identification des éléments responsables de leurs propriétés de sélectivité nécessite une étude approfondie de leurs interactions avec les substrats impliqués dans la réaction de N-acylation, à l'échelle moléculaire. Dans l'attente de l'élucidation expérimentale de leur structure et grâce aux énormes progrès faits dans la prédiction du repliement des chaines protéiques, des modèles tridimensionnels de ces enzymes peuvent être générés. Une fois optimisés, ces derniers peuvent être utilisés comme cible de docking de différents substrats, pour lesquels les résultats expérimentaux sont connus. Ces résultats, couplés aux observations issues des modèles et aux données disponibles sur les templates, pourraient expliquer la sélectivité particulière de ces enzymes.

2. Chapitre II : Matériels et méthodes

2.1. Etude numérique : Modélisation moléculaire

Les structures tridimensionnelles protéiques de SamAA et de SamELA ont été construites par modélisation par homologie et optimisées. La modélisation par homologie est une méthode informatique de prédiction de la structure tridimensionnelle d'une protéine à partir de sa séquence en acides aminés, en se basant sur la séquence et la structure élucidée d'un ou plusieurs templates. Les séquences des protéines SamAA et SamELA ont été découvertes lors de l'étude de Ferrari (Ferrari, 2014). Plusieurs outils de modélisation par homologie ont été utilisés pour obtenir des modèles tridimensionnels ayant la meilleure qualité possible. Puis l'étude des modes d'interaction enzyme/substrats a été réalisée grâce à des simulations de docking pour mimer la N-acylation de la L-lysine par trois acides carboxyliques différents (acide acétique, acide caproïque, acide laurique), catalysée par SamAA et de SamELA et pour expliquer la sélectivité de ces deux enzymes vis-à-vis à des fonctions amine α et ε de la L-lysine.

2.1.1. Identification des templates

La première étape de l'obtention d'une structure quaternaire d'une protéine par modélisation par homologie est l'identification des templates. Un template est une protéine dont la structure a déjà été élucidée par cristallographie et qui a un pourcentage élevé d'identité de séquence avec la protéine de structure inconnue. Trois outils ont été utilisés pour identifier les templates de SamAA et de SamELA.

Le premier programme est l'interface web BLASTp de NCBI disponible à cette adresse web : https://blast.ncbi.nlm.nih.gov/Blast.cgi (Altschul et al., 1997 ; Altschul et al., 2005 ; Johnson et al., 2008). La recherche de templates de SamAA a été effectuée le 26 novembre 2020. La séquence en acides aminés de l'aminoacylase a été comparée à celles disponibles dans la base de données des protéines déjà cristallographiées « PDB ». Tous les paramètres liés à la recherche de templates sont laissés par défaut. La recherche de templates pour SamELA a été réalisée le 2 novembre 2022 en suivant la même procédure.

Le second programme est l'interface web Swiss Model d'Expasy disponible à cette adresse : https://swissmodel.expasy.org/interactive#sequence (Camacho et al., 2009 ; Mirdita et al.,

2017 ; Steinegger et al., 2019). Les templates de SamAA ont été recherchés le 13 janvier 2021. Ces résultats ont été actualisés le 11 septembre 2021. Les templates de SamELA ont été identifiés le 24 octobre 2024. Ces trois recherches ont été menées avec l'option « Search For Templates ».

Le dernier outil est le logiciel appelé MMseqs2 (https://mmseqs.org) qui est disponible sur l'interface web ColabFold couplé à AlphaFold2. L'alignement multiple de séquences est obtenu à partir des bases de données de séquences UniRef et Environmental (Mirdita et al., 2017 ; Steinegger et al., 2019). Cette recherche a été effectuée le 22 novembre 2021 pour SamAA et le 19 janvier 2023 pour SamELA.

Ces outils permettent d'identifier les templates et donnent par ailleurs également un premier alignement des séquences de la protéine cible et des templates permettant l'identification des résidus conservés.

2.1.2. Construction des modèles par modélisation par homologie

Une fois les templates identifiés, la structure tridimensionnelle de SamAA a été prédite par modélisation par homologie après avoir effectué un alignement des séquences de l'enzyme et des templates (sur la base de la similarité des séquences). Les programmes « Modeller » de Discovery Studio® (version 4.1 à version 22.1), « ProMod3 » de Swiss Model, « AlphaFold2 » de ColabFold et « AlphaFold3 » de Google Deepmind ont été mis en œuvre. Les trois premiers outils ont été choisis sur la base des résultats publiés par CASP13 et CASP14 (Kryshtafovych et al., 2019 ; Kryshtafovych et al., 2021). AlphaFold3 est la version la plus récente d'AlphaFold au moment où les travaux numériques ont été finalisés.

2.1.2.1. Modeller (Discovery Studio®)

Une stratégie d'optimisation de la qualité du modèle de la structure tridimensionnelle de SamAA a été suivie en combinant différentes options relatives aux calculs d'alignement et à la construction des modèles par Modeller dans la suite logicielle Discovery Studio®.

Discovery Studio® propose deux types de calculs d'alignement de séquences, pour identifier le ou les templates. Le calcul « Align Sequences… » permet de prendre en compte les informations provenant des structures secondaires des templates et de SamAA et/ou d'aligner des séquences de protéines multimériques, en activant les options respectives « Use Secondary Structures… » (nommée « DSC ») et « Align by Chain… ». Le second calcul, appelé « Align

Sequences to Templates… » peut utiliser les informations de la structure tridimensionnelle des templates avec l'option « Create Sequences Profiles… » et/ou utiliser l'alignement obtenu à partir du Blastp (Thompson et al., 1994). Les séquences des templates identifiés lors de la recherche d'homologie avec SamAA ont été utilisées pour réaliser les alignements, en combinant les deux types de calculs et leurs multiples options. Chaque nouvel alignement produit a été soumis à Modeller paramétré par défaut, pour générer un nouveau modèle de la structure tridimensionnelle de SamAA dont la qualité a été vérifiée (Sali et al., 1993). Les quatre alignements conduisant aux modèles tridimensionnels ayant la meilleure qualité ont été sélectionnés pour continuer l'amélioration de la qualité du modèle de la structure tridimensionnelle.

Modeller est disponible dans la suite logicielle Discovery Studio®, avec plusieurs options disponibles. La structure tridimensionnelle peut être modélisée avec les ligands du template (par exemple des ions métalliques) présents dans le site actif (« Ligands ») et avec des molécules d'eau de solvatation (« Water »). La position des boucles peut être optimisée (« Refine Loop »). Les quatre alignements sélectionnés précédemment ont été réutilisés pour étudier l'impact du choix des options de Modeller sur la qualité des modèles générés.

2.1.2.2. Swiss Model

La séquence monomérique de SamAA a été soumise au serveur Swiss Model disponible à l'adresse https://swissmodel.expasy.org/interactive. Cet outil permet de prédire la structure quaternaire de la proteine cible et ne nécessite pas de paramétrage particulier en utilisant l'option « Build Model » (Guex et al., 2009 ; Bienert et al., 2017 ; Bertoni et al., 2017 ; Waterhouse et al., 2018 ; Studer et al., 2021). Il est également possible d'imposer un template. Dans le cas de SamAA, la structure tridimensionnelle monomérique et dimérique de la succinyl-diaminopimelate desuccinylase de *Neisseria meningitidis* (entrée PDB 4PPZ) a été utilisée comme template.

2.1.2.3. ColabFold (AlphaFold 2)

L'interface web GitHub ColabFold permet l'utilisation couplée de MMseq2 et d'AlphaFold2 disponible à https://colab.research.google.com/github/sokrypton/ColabFold/blob/v1.2.0/AlphaFold2.ipynb (Jumper et al., 2021 ; Mirdita et al., 2022). Les modélisations ont été réalisées sur trois cycles.

Les modèles tridimensionnels ont été générés à partir de la séquence dimérique de SamAA sur la version v1.2 du logiciel et à partir de la séquence monomérique de SamELA sur la version v.1.4.

2.1.2.4. AlphaFold3

AlphaFold3 est disponible sur le serveur https://alphafoldserver.com/ (Abramson et al., 2024). Les séquences de SamAA et de SamELA ont été soumises en précisant deux copies de la chaine peptidique et un ion en deux copies sous la forme de Zn^{2+} pour la première enzyme et une copie de la chaine peptidique et un ion en deux copies sous la forme de Zn^{2+} pour la seconde enzyme.

2.1.3. Validation des modèles d'homologie

La qualité des modèles d'homologie a été évaluée par différentes fonctions de score et en vérifiant la présence de plusieurs éléments spécifiques de la famille des peptidases M20 dans le but de sélectionner les meilleurs modèles.

2.1.3.1. Scoring

Chaque outil de modélisation par homologie a sa propre méthode d'évaluation de la qualité des modèles tridimensionnels obtenus sous la forme d'un score.
Les scores GMQE (Global Model Quality Estimate) et QMEANDisCo ont été appliqués de façon automatique aux modèles générés par Swiss Model. Le GMQE est une estimation de la précision attendue de la structure tertiaire du modèle sous la forme d'un score normalisé. Le QMEANDisCo est une estimation locale et globale de la qualité absolue du modèle par comparaison avec des structures tertiaires de même taille déjà résolues. Ce score est actuellement remplacé par le QMEANDisCo Global donnant un score normalisé entre 0 et 1 (Waterhouse et al., 2018 ; Studer et al., 2020).
Les modèles de ColabFold ont été automatiquement analysés par lDDT (local Distance Difference Test) et PAE (Predicted Align Error). Le lDDT est un score de superposition qui évalue les différences locales de la position de tous les acides aminés d'un modèle par rapport à une référence ; il doit être supérieur à 70 % pour que toutes les régions d'un modèle soient considérées correctement modélisées. Le PAE est un score indiquant l'erreur de position au

résidu x montrant la différence d'alignement du modèle et de la référence sur le résidu y, sous la forme d'une matrice colorée en 2D (Mariani et al., 2013).
Les modèles générés par AlphaFold3 ont aussi été automatiquement analysés par lDDT et PAE. Deux fonctions supplémentaires de score (Scores pTM et ipTM) ont été ajoutées au serveur pour l'analyse de la qualité des complexes macromoléculaires. Le pTM évalue la précision de la structure globale au niveau des chaines et du complexe. L'ipTM mesure la précision des positions relatives prédites des sous-unités au sein du complexe (Zhang et al., 2004 ; Xu et al., 2010).
Tous les modèles ont été analysés par le score DOPE normalisé (disponible sur Discovery Studio®) permettant de comparer la qualité de tous les modèles, indépendamment du logiciel utilisé pour les construire. Ce score utilise la méthode Modeller DOPE (Discrete Optimized Protein Energy) qui est un score d'énergie, et est un score Z dérivé des statistiques des scores de DOPE brut permettant la prise en compte de la taille des protéines. Les modèles sont jugés de bonne qualité pour un score DOPE négatif. Les scores inférieurs à -1 indiquent un modèle de type natif. Une valeur positive de score indique un modèle de mauvaise qualité (Shen et al., 2006 ; Eramian et al., 2008).

2.1.3.2. Etude de la conservation des éléments de spécificité

Le site de liaison des ions métalliques des templates de SamAA présente une géométrie de coordination à base trigonale ou carrée déformée (Bienvenue et al., 2003 ; Davis et al., 2006). Cette caractéristique structurale a été supposée conservée chez SamAA. Ce critère a été utilisé pour sélectionner le modèle final de la structure de SamAA, en superposant les structures tridimensionnelles de la succinyl-diaminopimelate desuccinylase de *Haemophilus influenzae* (entrée PDB 5VO3) ou de la carnosinase 2 humaine (entrée PDB 4RUH) et le modèle tridimensionnel de SamAA, à l'aide de l'outil « Align and Superimpose Proteins… » de Discovery Studio® (Thompson et al., 1994). Le respect de la géométrie mentionnée ci-dessus a été évalué par un calcul de RMSD (écart quadratique moyen).

2.1.4. Optimisation des modèles d'homologie

L'optimisation des modèles tridimensionnels de SamAA et de SamELA a été réalisée sur Discovery Studio. Elle a consisté à ajouter les ions métalliques lorsque ceux-ci étaient absents lors de la construction des modèles. Une molécule d'eau catalytique a par ailleurs été ajoutée

au niveau des sites actifs des enzymes en s'inspirant de la structure élucidée des templates. Les structures ont été solvatées puis minimisées sur le plan énergétique.

2.1.4.1. Ajout des ions métalliques et de la molécule d'eau catalytique

Deux ions Zn^{2+} ont été ajoutés manuellement dans chacun des deux sites actifs du modèle dimérique de SamAA et dans le site actif du modèle monomérique de SamELA, en s'inspirant de la géométrie des sites actifs des templates (succinyl-diaminopimelate desuccinylase de *H. influenzae* (entrées PDB 3IC1 et 5VO3) pour SamAA (Figure II.1) et hydrolase métal-dépendante non caractérisée de *Pyrococcus furiosus* (entrée PDB 3ICJ) pour SamELA).

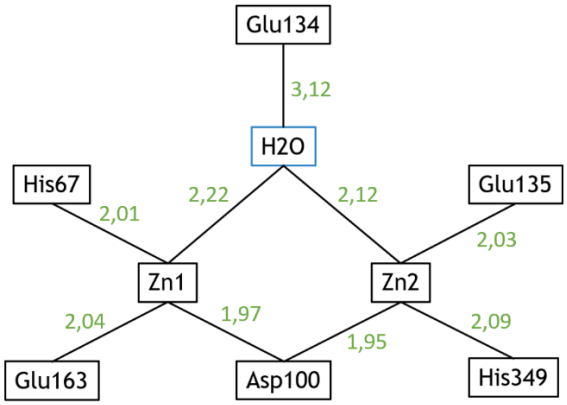


Figure II.1 : Représentation schématique du site actif de la succinyl-diaminopimelate desuccinylase de *H. influenzae* mentionnant les distances entre les différents éléments (Nocek et al., 2010 ; Nocek et al., 2018).

Une molécule d'eau interagissant avec ces deux ions métalliques a été ajoutée manuellement dans chaque site actif pour que l'architecture caractéristique des métalloenzymes soit respectée.

2.1.4.2. Solvatation et minimisation d'énergie progressive

Le modèle a été soumis au champ de force Momany et Rone CHARMm (Momany et al., 1992). La structure a été solvatée dans une boite d'eau de forme orthorhombique. Le système a été neutralisé avec des contre-ions (Na^+ et Cl^-).

Une minimisation de l'énergie a été appliquée en deux étapes à l'aide de la méthode Smart Minimizer (RMS gradient à 0,1 Å et sous contrainte fixe appliquée à l'ensemble du système sauf aux atomes d'hydrogène et aux molécules d'eau) suivie de la méthode du Conjugate Gradient (RMS gradient à 0,01 Å et sous contrainte fixe appliquée à la chaine peptidique et une

contrainte restreinte appliquée aux chaines latérales des acides aminés, aux ions métalliques et à la molécule d'eau catalytique) (Powell, 1977 ; Tronrud, 1992).

2.1.5. Simulations de docking et scoring

Les simulations de docking et les analyses de score ont été réalisées sur la suite logicielle Discovery Studio®.

2.1.5.1. Structure des ligands

Les outils du module « Sketch Molecules » de Discovery Studio® ont été utilisés pour construire les structures tridimensionnelles des ligands. Le champ de force Momany et Rone CHARMm et une correction géométrique (« clean geometry ») ajustant la longueur des liaisons et des angles ont été appliqués sur chaque structure.
Les molécules N-α-, N-ε-acétyl-L-lysine, N-α-, N-ε-caproyl-L-lysine, N-α-, et N-ε-lauroyl-L-lysine ont été utilisées pour mimer la réaction d'hydrolyse. Les molécules L-lysine, acide acétique, acide caproïque, acide caprylique et acide laurique ont été utilisées pour mimer la réaction inverse de synthèse des dérivés acylés de la L-lysine. Les substrats ont leurs fonctions amines protonées et leurs fonctions acides déprotonées.

2.1.5.2. Définition du site actif

Le site actif est la zone dans laquelle le ligand va être amarré. Il a été défini par la méthode dite des cavités, qui identifie les sites probables de liaison au niveau des cavités présentes à la surface de l'enzyme. Cette méthode permet l'identification de plusieurs sites possibles. Le site actif a été choisi sur la base des connaissances relative à la famille d'enzymes auxquelles appartiennent SamAA et SamELA : il doit contenir les ions métalliques et les résidus catalytiques conservés.

2.1.5.3. Simulations de docking

Le docking moléculaire est une méthode de calcul prédisant les modes d'interactions préférentiels entre un ligand et le site actif d'une enzyme cible. Cette méthode est basée sur des

algorithmes explorant l'espace conformationnel du ligand et le positionnement de chacune de ces conformations selon différentes orientations dans le site actif. Il existe plusieurs types d'algorithmes de docking permettant d'appliquer ou non une flexibilité sur le ligand et sur l'enzyme. Le programme de docking semi-rigide CDOCKER a été utilisé pour l'ensemble des simulations de docking concernant SamAA. Le programme LibDock a également été utilisé lors des simulations concernant SamELA.

Les meilleurs modèles tridimensionnels optimisés de SamAA et de SamELA ont été utilisés comme cible de docking, en présence ou en absence de la molécule d'eau catalytique lors de l'étude de la réaction d'hydrolyse ou de synthèse des dérivés α- et ε-acylés (Wu et al., 2003). 20 ou 50 complexes enzyme-ligand, également appelés poses, ont été retenus à l'issue de chaque simulation. Le seuil RMSD, le nombre de conformations aléatoires générées et le nombre d'orientations à affiner ont été fixés respectivement à 0,1, 20 et 20.

2.1.5.4. Analyses des poses de docking

Les premières analyses des complexes enzyme/ligand ont consisté à appliquer des filtres géométriques. Ainsi, les poses sont considérées comme similaires si le RMSD entre elles est inférieur à 2 Å. Sur cette base, les poses ont été réparties dans des groupes d'orientation préférentielle (Cole et al., 2015). Ensuite, la mesure d'angles ou distances spécifiques impliquées dans le mécanisme catalytique des enzymes de la même famille que SamAA et SamELA, a été effectuée pour évaluer la faisabilité de la réaction d'hydrolyse ou de synthèse :

- La distance comprise de 2 à 2,5 Å, entre l'atome Zn2 et l'atome d'oxygène de la liaison amide ou de la fonction carboxylate du donneur d'acyle.
- La distance d'environ 4 Å entre l'atome de carbone de la liaison amide ou de la fonction carboxylate du donneur d'acyle et l'atome d'oxygène de la molécule d'eau catalytique ou l'atome d'azote de la fonction amine de l'accepteur d'acyle.
- La distance comprise de 1 à 4,7 Å entre l'atome d'oxygène terminal du résidu Glu154 et l'atome d'hydrogène de la liaison amide ou de la fonction amine de l'accepteur d'acyle (Dutta et al., 2016 ; Nocek et al., 2018).
- L'angle d'attaque nucléophile vis-à-vis du centre électrophile, appelé angle de Bürgi-Dunitz, égale à 90 ° dans le cas d'une attaque nucléophile enzymatique (Bürgi et al., 1974 ; Lodge et al., 1986 ; Radisky et al., 2002 ; Lubkowski et al., 2019).

En parallèle, le classement des poses issues des simulations de docking a été réalisé en appliquant des fonctions de score. Les poses générées au cours de chaque simulation de docking ont été classées automatiquement selon le score négatif « CDOCKER ENERGY » qui prend en compte l'énergie contrainte interne du ligand et l'énergie d'interaction enzyme-ligand. Des analyses de score supplémentaires ont été appliquées en utilisant plusieurs fonctions de score : LigScore1_Dreiding, LigScore2_Dreiding, PLP1, PLP2, Jain, PMF et PMF4 (Gehlhaar et al., 1995 ; Jain, 1996 ; Gehlhaar et al., 1999 ; Muegge et al., 1999 ; Mueggue, 2005 ; Krammer et al., 2005). Afin d'améliorer la fiabilité du classement des poses, un consensus des résultats des sept fonctions de score citées précédemment a été réalisé. Le consensus score attribue une valeur de 1 à chaque pose lorsque celle-ci fait partie des 20 % de meilleures poses d'une des fonctions de score considérées. Ainsi, un score de 7 est obtenu lorsqu'une pose est considérée comme l'une des meilleures par sept fonctions de score. Si une égalité de consensus de score est observée entre plusieurs poses, le classement des poses par la fonction de score CDOCKER ENERGY est observée. Ce consensus de score permet donc d'identifier les poses les mieux classées en tenant compte de l'ensemble des fonctions de score appliquées.
Enfin, les interactions entre les substrats et les résidus constituant la cavité catalytique de SamAA ou de SamELA ont été étudiées avec l'outil « Analyze Ligands Poses » de Discovery Studio®.

2.2. Etude expérimentale : Production par expression hétérologue

Les enzymes SamAA et SamELA ont été précédemment étudiées pour identifier et confirmer leur spécificité respective de N-acylation envers les fonctions α et ε de la L-lysine (Dettori et al., 2018a ; Bourkaib et al., 2020a). Pour poursuivre les travaux de caractérisation, en particulier l'élucidation de leurs structures tridimensionnelles et la mise en œuvre d'un procédé enzymatique de N-acylation, il est nécessaire d'augmenter la quantité d'enzymes produites naturellement par *Streptomyces ambofaciens*, en utilisant un hôte d'expression hétérologue.

2.2.1. Systèmes d'expression choisis et milieux de culture

La production de SamAA et de SamELA a été effectuée en utilisant plusieurs systèmes d'expression combinant les souches *Escherichia coli* et *Streptomyces lividans* et les plasmides pET-15b, pET-28a, pIB139, pIJ86, pIJ8600 et pIJ8660 pouvant contenir ou non le gène des enzymes optimisé à l'expression chez *E. coli* ou non (Tableaux II.1, II.2 et II.3).

Le plasmide pET-15b contient une version du gène de SamAA non-optimisé pour l'expression chez *E. coli*. Les systèmes contenant le plasmide pET-15b ont été construits lors de travaux antérieurs (Nili, 2018) (Figure II.2). Au début de ces travaux, le gène extrait du plasmide pET-15b a été séquencé pour vérifier l'absence de mutation chez Genewiz (https://www.genewiz.com/en-gb/). Les plasmides pET-28a avec ou sans le gène de SamAA et avec le gène de SamELA ont été commandés chez GenScript (https://www.genscript.com/) et contiennent une version du gène optimisé pour son expression chez *E. coli* (Figure II.3). Les systèmes avec un plasmide pET-28a ont été construits lors de ces travaux en suivant la procédure présentée au point suivant. Les cultures bactériennes d'*E. coli* ont été réalisées dans le milieu LB liquide (pour 1 L, 20 g de LB both (Sigma Aldrich) ou solide (pour 1 L, 1 L de milieu LB liquide et 15 g d'agar)), dans le milieu Overnight Express™ Instant TB Medium (pour 1 L, un sachet de EasyPak et 10 mL de glycérol) (Novagen, Merck KGaA) ou dans le milieu TB-AIM (pour 1 L, 20 g de tryptone, 24 g d'extrait de levure, 3 g de NaH_2PO_4, 3,4 g de KH_2PO_4, 2,68 g de NH_4Cl, 492 mg de $MgSO_4$, 710 mg de Na_2SO_4, 500 mg de glucose, 2 g de lactose et 5 mL de glycérol). Les solutions plasmidiques ont été conservées dans une solution aqueuse stérile à -20 °C. L'ensemble de ces souches ont été conservés à -80 °C dans du milieu LB supplémenté à 10 % en glycérol.

Les systèmes de *S. lividans* ont été construits chez l'UMR 1128 DynAMic (Dynamique des Génomes et Adaptation Microbienne) (Tableau II.2). Le plasmide pIJ8660 possède également une séquence *vsi* codant pour un peptide signal pour la sécrétion de la protéine recombinante. Les spores ont été obtenues sur un milieu solide MSF (Mannitol Soya Flour) (Pour 1 L, 20 g de mannitol, 20 g de farine de soja et 20 g d'agar). Les cultures bactériennes de *S. lividans* ont eu lieu dans le milieu de fermentation (pour 1 L, 5 g d'extrait de levure, 30 g de Tryptic Soy Broth, 8,25 g de glucose monohydraté, 3 g d'extrait de malt, 200 mL de solution de saccharose à 50 % (p/v) et 50 mL de solution de $MgCl_2$ à 20,4 g/L). Ces systèmes ont été conservés sous forme de spores dans une solution aqueuse de NaCl 0,9 % supplémentée à 25 % en glycérol à -20 °C.

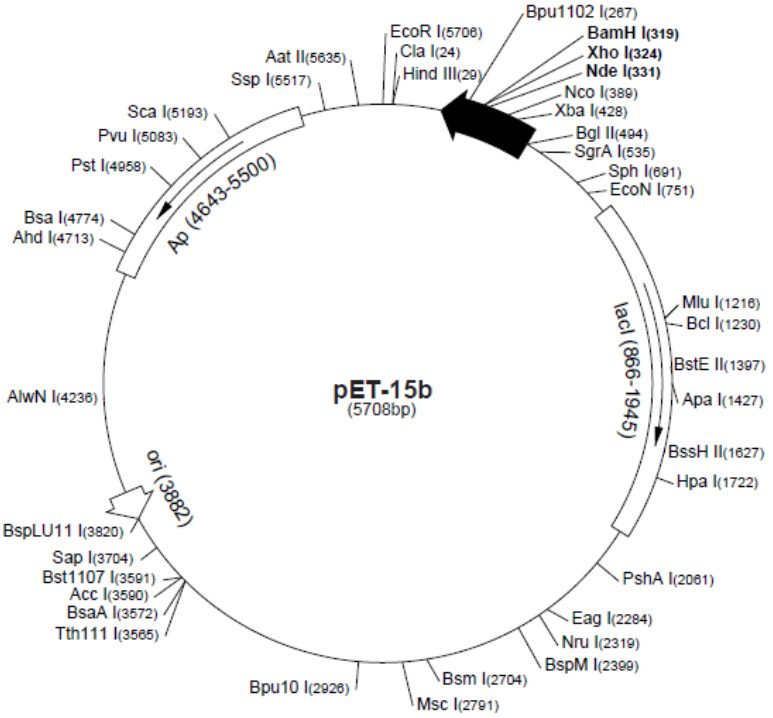
Le procédé de purification a été anticipé par la présence de tag facilitant la purification des protéines recombinante taguée. Le tag 6-His a été choisi pour réaliser une purification des aminoacylases par chromatographie d'affinité au nickel (IMAC). Dans le plasmide pET-15b, le tag 6-His est présent à l'extrémité N-terminale du gène de SamAA. Dans le plasmide pET-28a, le tag 6-His est placé en position N- ou C-terminale pour contourner une éventuelle gêne au bon repliement des enzymes recombinantes par la présence du tag à une de leurs extrémités. Le tag StrepTag (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) a également été étudié, en remplacement du tag 6-His dans l'hypothèse où son affinité aux ions métalliques divalents pourrait être un

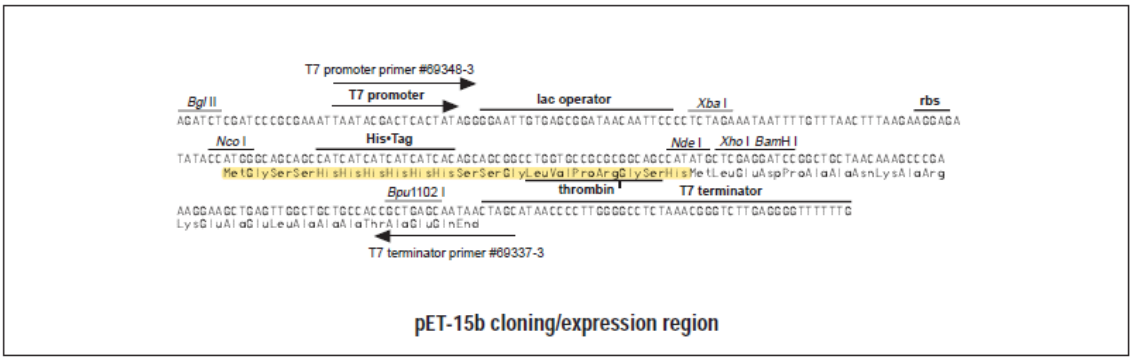
problème au bon repliement des deux métalloenzymes et au processus de purification par chromatographie d'affinité au nickel. Le strepTag, pouvant être positionné aussi bien en C qu'en N terminal présente une affinité particulière et non métallo dépendante pour une résine à base de streptavidine modifiée appelée commercialement Strep-Tactin®. Ce second tag est présent dans le plasmide pET-28a et dans les plasmides des systèmes *S. lividans*, en position N-terminale du gène de SamAA.

Tableau II.1 : Récapitulatif des caractéristiques des systèmes *E. coli* étudiés pour l'expression de SamAA.

Souche	Plasmide	Type de tag	Position du tag
Origami B (DE3)	pET-15b vide	6-His	N-ter
	pET-15b-SamAA		
Rosetta 2 (DE3)	pET-28a-SamAA	6-His	N-ter
			C-ter
Rosetta 2 pLysS (DE3)	pET-15b vide	6-His	N-ter
	pET-15b-SamAA		
BL21 Gold (DE3)	pET-28a vide	6-His	N-ter
	pET-28a-SamAA	6-His	N-ter
			C-ter
BL21 Gold pLysS (DE3)	pET-28a-SamAA	6-His	N-ter
			C-ter
Arctic Express (DE3)	pET-28a-SamAA	6-His	N-ter
		Strep-Tag	

pET-15b sequence landmarks	
T7 promoter	463-479
T7 transcription start	452
His•Tag coding sequence	362-380
Multiple cloning sites	
(*Nde* I - *Bam*H I)	319-335
T7 terminator	213-259
lacI coding sequence	(866-1945)
pBR322 origin	3882
bla coding sequence	4643-5500



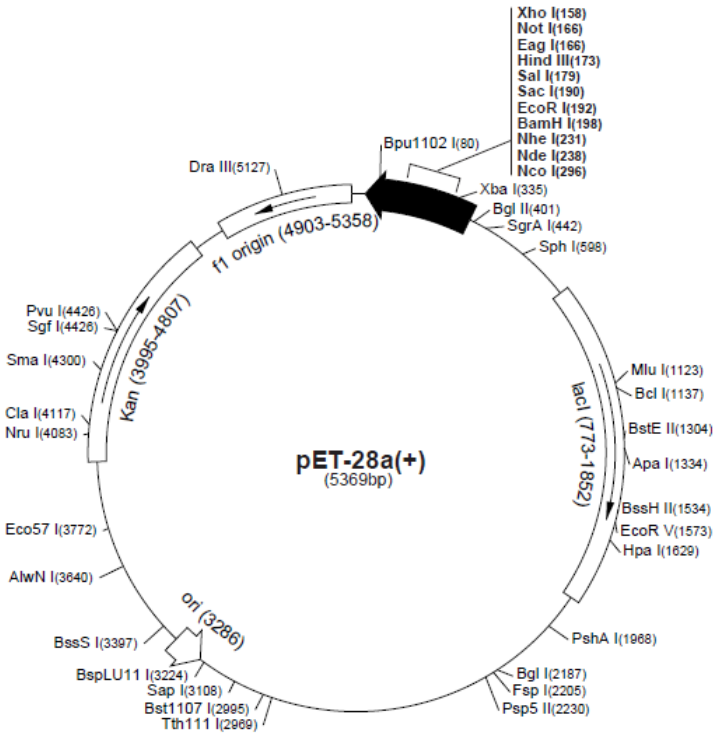


pET-15b cloning/expression region

Figure II.2 : Schéma du plasmide pET-15b (Novagen).

pET-28a(+) sequence landmarks	
T7 promoter	370-386
T7 transcription start	369
His•Tag coding sequence	270-287
T7•Tag coding sequence	207-239
Multiple cloning sites (*Bam*H I - *Xho* I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	773-1852
pBR322 origin	3286
Kan coding sequence	3995-4807
f1 origin	4903-5358

The maps for pET-28b(+) and pET-28c(+) are the same as pET-28a(+) (shown) with the following exceptions: pET-28b(+) is a 5368bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-28c(+) is a 5367bp plasmid; subtract 2bp from each site beyond *Bam*H I at 198.



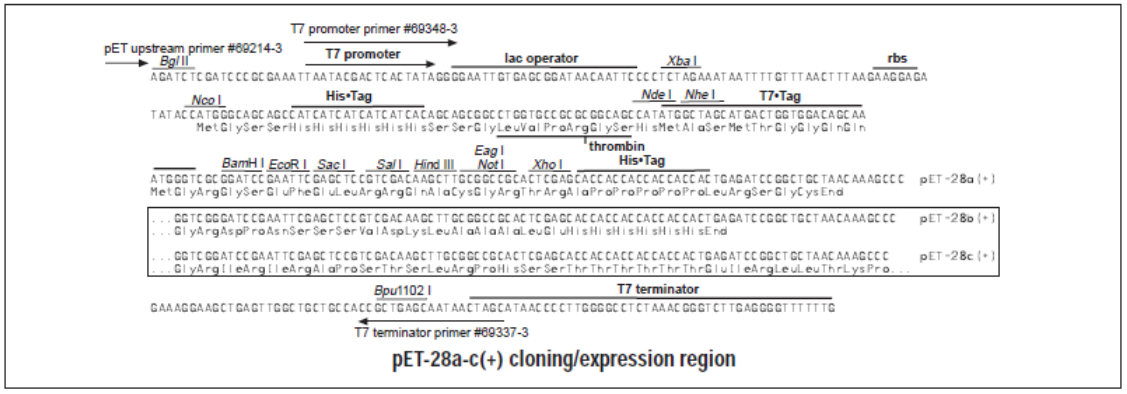


Figure II.3 : Schéma du plasmide pET-28a (Novagen).

Tableau II.2 : Récapitulatif des caractéristiques des systèmes *S. lividans* étudiés pour l'expression de SamAA.

Souche	Plasmide	Type de tag	Position du tag	Type de vecteur	Promoteur
STK24	pIB139 vide	StrepTag	N-ter	Vecteur intégratif	Fort et constitutif (*ermE*p*)
	pIB139-SamAA				
	pIJ8600 vide			Vecteur intégratif	Inductible au thiostrepton
	pIJ8600-SamAA				
	pIJ86 vide			Vecteur réplicatif	Fort et constitutif (*ermE*p*)
	pIJ86-SamAA				
	pIJ8660 vide			Vecteur intégratif	Constitutif
	pIJ8660-SamAA				

Tableau II.3 : Récapitulatif des caractéristiques des systèmes *E. coli* étudiés pour l'expression de SamELA.

Souche	Plasmide	Type de tag	Position de tag
Rosetta 2 (DE3)	pET-28a-SamELA	6-His	N-ter
			C-ter
BL21 Gold (DE3)			N-ter
			C-ter
BL21 Gold pLysS (DE3)			N-ter
			C-ter
Arctic Express (DE3)			N-ter
			C-ter

2.2.2. Transformations bactériennes d'*E. coli*

La transformation bactérienne consiste en l'insertion d'un vecteur (ici un plasmide) dans une cellule hôte. Les cellules d'*E. coli* ont été achetées compétentes : les souches *E. coli* BL21 Gold (DE3) avec et sans pLysS et *E. coli* Arctic Express (DE3) auprès d'Agilent Technologies

(https://www.agilent.com/) et les souches *E. coli* Rosetta 2 (DE3) auprès de Novagen (https://www.sigmaaldrich.com/FR/fr). L'ensemble des cellules compétentes a été stocké à -80 °C.

Les cellules compétentes ont été lentement décongelées dans la glace. 1 µL de solution plasmidique ayant une concentration comprise entre 1 et 50 ng d'ADN plasmidique a été ajouté dans chaque tube, qui a été replacé dans la glace entre 5 à 30 minutes. Le choc thermique a été ensuite réalisé dans un bain d'eau à 42 °C : il a duré exactement 20 s pour les souches *E. coli* BL21 Gold (DE3) et *E. coli* Arctic Express (DE3) et 30 s pour les souches *E. coli* Rosetta 2 (DE3). Puis les tubes de cellules compétentes ont été de nouveau plongés dans la glace pendant 2 min. Un volume de milieu SOC Medium stérile (pour 100 mL, 2 mL de solution de glucose à 20 % (p/v) et 98 mL de SOC Medium (pour 1L, 20 g de tryptone, 5 g d'extrait de levure, 500 mg de NaCl avant stérilisation et 10 mL de solution stérile de $MgCl_2$ à 1 M)) a été ajouté dans les tubes qui ont été placés dans un incubateur à 37 °C pendant 1 h sous agitation. Le volume de cellules compétentes, la concentration de la solution plasmidique et le volume du SOC Medium à ajouter après le choc thermique sont différents selon le fournisseur des cellules compétentes. Des boites de pétri contenant du milieu LB-agar supplémenté en antibiotiques appropriés permettant la sélection des clones transformés ont été inoculées par la technique de cadran avec 50 µL de solution de cellules transformées. Les boites de pétri ont été placées dans une étuve à 37 °C pendant 24 à 48 h pour la croissance des clones transformés. Une fois les colonies formées, les boites de pétri ont été emballées de paraffine et ont été stockées à 4 °C. Une colonie isolée a été choisie pour le passage en culture liquide et pour la mise en collection des nouveaux systèmes obtenus.

2.2.3. Cultures bactériennes

Les cultures bactériennes des systèmes ont lieu en présence d'antibiotiques qui assurent le maintien de la conservation des plasmides présents chez les bactéries. Le tableau II.4 récapitule les antibiotiques et leurs concentrations utilisées selon la souche et le plasmide étudiés.

Tableau II.4 : Antibiotiques et leurs concentrations utilisées selon les souches bactériennes et les plasmides étudiés.

Plasmide/Souche	Antibiotiques	Concentrations (µg/mL)
Plasmide pET-15b	Ampicilline	100
Plasmide pET-28a	Kanamycine	50
Plasmide pLysS	Chloramphénicol	50
Plasmides de *S. lividans*	Apramycine	50
E. coli Origami (DE3)	Tétracycline	12,5
	Kanamycine	50
E. coli Rosetta 2 (DE3)	Chloramphénicol	25
E. coli BL21 Gold (DE3)	Tétracycline	12,5
E. coli Arctic Express (DE3)	Gentamycine	20

2.2.3.1. *Escherichia coli*

Pour limiter l'expression des deux aminoacylases sous la forme de corps d'inclusion, de nombreux paramètres et conditions ont été étudiés et sont présentés dans la figure II.4.

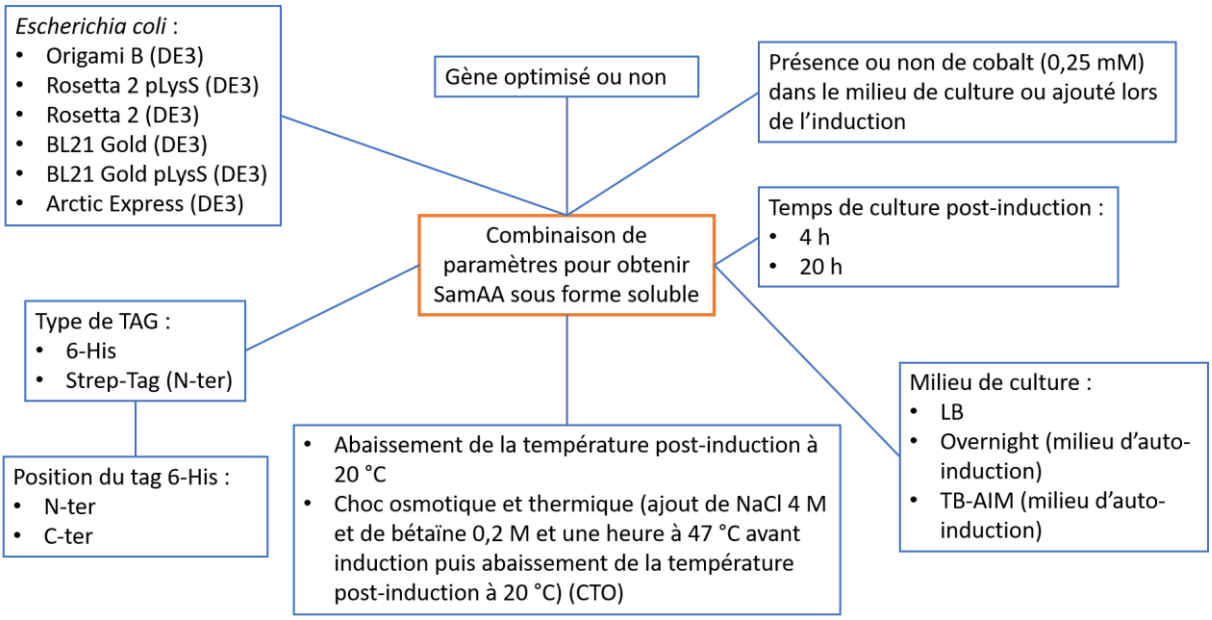


Figure II.4 : Récapitulatif des paramètres et des conditions étudiés pour la surexpression de SamAA sous une forme soluble et active chez *E. coli*.

Tout d'abord, une préculture a été réalisée dans 5 mL de milieu LB additionné des antibiotiques nécessaires au système avec un inoculum de 1 %, soit 50 µL d'une ancienne pré-culture ou d'un tube de la collection. Elle a lieu à 37 °C sous agitation (200 rpm). La DO_{600} a été mesurée au

bout de 24 h pour vérifier la croissance bactérienne. L'ensemble des pré-cultures a été conservé à 4 °C pendant quelques semaines.

Les cultures bactériennes ont été réalisées dans des erlenmeyers de 250 mL contenant 30 mL de milieu stérile avec un inoculum frais de pré-culture à 1 % dans les milieux LB et TB-AIM et à 2,5 % dans le milieu Overnight. Les cultures dans le milieu LB ont été faites à 37 °C, sous agitation (200 à 235 rpm), en présence et en absence de cobalt à 0,25 mM et pendant 8 h ou 24 h. Un temps de culture plus court est étudié pour limiter le temps d'action des protéases. L'expression protéique a été initiée par ajout de 0,1 mM d'IPTG à une DO_{600} de 0,6 en absence de cobalt ou à une DO_{600} de 0,4 en présence de cobalt. La température peut être abaissée à 20 °C et un choc thermique et osmotique (ajout de NaCl à 4 M et de bétaïne à 0,2 M et une heure à 47 °C) peut être réalisé après l'induction. Les cultures dans le milieu Overnight ont été faites à 37 °C, sous agitation (300 rpm), pendant 16 h et en présence et en absence de cobalt à 0,25 mM. Il est recommandé d'ajouter une concentration de 50 µg/mL d'ampicilline et de 100 µg/mL de kanamycine en cas d'utilisation du milieu Overnight. Le suivi de la croissance bactérienne a été réalisé en mesurant la DO_{600} du milieu de culture, de l'inoculation à l'induction et à l'arrêt des cultures.

La précédente étude de caractérisation enzymatique de SamAA conclut que l'ajout de cobalt augmente l'activité de l'enzyme, suggérant que SamAA serait une métalloenzyme (Bourkaib et al., 2020a). L'ajout de $CoCl_2$ au milieu de culture a été étudié car les ions cobalt seraient susceptibles de jouer également un rôle sur le plan structural, permettant la limitation de la production de l'aminoacylase sous la forme de corps d'inclusion. Une concentration de 0, 25 mM a été choisie pour tenir compte de la toxicité du cobalt pour *E. coli* (toxique à partir d'une concentration de 0,1 mM) et de l'aide à la solubilisation de la protéine recombinante (jusqu'à une concentration de 0,8 mM) (Sari et al., 2007 ; Matjan et al., 2011 ; Takakura et al., 2019).

La souche *E. coli* Arctic Express (DE3) est une souche possédant les gènes des protéines chaperonnes Cpn10 et Cpn60 de la bactérie *Oleispira antarctica*, qui sont des protéines actives à des températures entre 4 et 12 °C (Ferrer et al., 2004a ; Ferrer et al., 2004b). Leurs présences nécessitent un abaissement de la température de croissance et une adaptation du temps de culture. Les cultures dans le milieu LB ont eu leur température abaissée à 11 °C après l'induction pendant 24 h. La température des cultures dans le milieu Overnight ou le milieu TB-AIM a été de 30 °C pendant 6 h puis de 11 °C pendant 16 h. Les cultures d'*E. coli* Arctic Express sont réalisées en absence d'antibiotique.

2.2.3.2. Streptomyces lividans

2.2.3.2.1. Obtention des spores

Les spores ont été produites sur le milieu MSF solide (pour 1 L, 20 g de mannitol, 20 g de farine de soja et 20 g d'agar) à partir d'une suspension aqueuse de spores. Les boites de pétri ont été incubées à 28 °C pendant une quinzaine de jours : le mycélium blanc doit devenir gris foncé ou noir. Les spores ont été décollées à l'aide d'un râteau dans une solution stérile aqueuse de NaCl 0,9 % et de Tween 20 0,1 %, puis lavées dans une solution aqueuse de NaCl 0,9 % et stockées dans une solution aqueuse de NaCl 0,9 % additionnée de glycérol 25 %.

2.2.3.2.2. En erlenmeyer

Les pré-cultures ont été réalisées dans des erlenmeyers à baffles de 250 mL contenant 25 mL de milieu de fermentation et quelques gouttes d'anti-mousse (polypropylène glycol 2000). Elles ont été inoculées à 0,1 %, soit 25 µL de solution de spores fraiches. Elles ont duré deux jours sous agitation (150 rpm) à 28 °C. Puis les cultures ont été réalisées dans des erlenmeyers à baffles de 500 mL avec 50 mL de milieu de fermentation additionné d'anti-mousse. Elles ont été inoculées à 5 %, soit 2,5 mL de pré-culture fraiche. Elles ont duré trois jours sous agitation (140 rpm) à 28 °C.
Si le plasmide pIJ86 est présent, il est nécessaire d'ajouter de l'apramycine à 50 µL/mg pour le maintien du plasmide. Les autres plasmides étudiés sont intégratifs et ne nécessitent pas l'ajout de l'antibiotique. Le plasmide pIJ8600 contient un promoteur inductif : l'induction a été réalisée 24 h après l'inoculation en ajoutant du thiostrepton à 10 µg/mL.

2.2.3.2.3. En bioréacteur

Le système *S. lividans* STK24 avec le plasmide pIJ86-SamAA a été aussi étudié en bioréacteur de 1 L.
Une première pré-culture a été réalisée dans un erlenmeyer à baffles de 500 mL contenant 50 mL de milieu de fermentation, quelques gouttes d'anti-mousse (polypropylène glycol 2000), de l'apramycine à 50 µg/mL et un inoculum à 0,1 %, soit 50 µL de solution de spores fraiches. Elle a duré deux jours sous agitation (180 rpm) à 28° C. Puis une seconde pré-culture d'un volume plus important a été réalisée dans un erlenmeyer à baffles de 1 L, contenant 100 mL de

milieu de fermentation, quelques gouttes d'anti-mousse, de l'apramycine à 50 µg/mL et un inoculum à 5 %, soit 5 mL de la pré-culture précédente. Elle a duré trois jours sous agitation (180 rpm) à 28 °C.

Le bioréacteur contient 1 L de milieu de fermentation additionné avec de l'apramycine à 50 µg/mL et de l'anti-mousse. L'inoculum à 2,5 % provenant de la seconde pré-culture a été ajouté. Un prélèvement de 30 mL a été réalisé au moment de l'inoculation puis toutes les 24 heures pendant 7 jours, pour suivre la production de SamAA recombinante.

La régulation de l'oxygène dissous entre 30 et 100 % a été faite en modifiant manuellement l'aération (débit de l'arrivée d'air) et l'agitation tout au long de la culture bactérienne. Le pH a été régulé automatiquement à 7,2 à l'aide d'une solution de H_2SO_4 à 20 % (v/v) et d'une solution de NaOH à 2 M. Suite à un dysfonctionnement des pompes des solutions régulatrices de pH, ayant eu lieu 29 heures après l'inoculation, le pH a été ensuite régulé manuellement plusieurs fois par jour avec les solutions acide et base. Une supplémentation en glucose à 1 g/L devait débuter après 15 h de culture à un débit de 1,333 mL/h pendant 7 h puis à un débit de 2 mL/h jusqu'à la fin de la culture (Haeger et al., 2023a). Cependant, la solution de glucose a rencontré des problèmes de solubilisation dans la bouteille et dans le tubing au moment de débuter la supplémentation. Malgré un chauffage de la bouteille et du tubing, l'ajout correct du glucose n'a pas pu avoir lieu.

2.2.4. Extraction et lyse cellulaire

A la fin des cultures bactériennes en erlenmeyers et après les prélèvements issus du bioréacteur, l'extraction des cellules contenues dans les milieux de culture a été menée en réalisant deux centrifugations (8000 ou 10000 rpm, 4 °C, 10 min ; Centrifuge 5804 R, Eppendorf). La vitesse de rotation est augmentée quand la croissance bactérienne est faible. Chaque culot cellulaire a été lavé une fois par 25 mL de tampon (25 mM Tris HCl, 50 mM NaCl, pH 8) puis a été resuspendu dans 15 mL de tampon pour être stocké à -20 °C dans l'attente de l'analyse des extraits bruts (EBNL). Le tampon est supplémenté en cobalt (0,25 mM) quand la culture bactérienne a eu lieu en présence de cobalt.

Les cellules bactériennes de chaque extrait ont été ensuite lysées pour libérer les protéines intracellulaires à l'aide d'une presse de French. Les bactéries *E. coli* ont été lysées par un cycle de pression-décompression à 2000 kbar tandis que les bactéries *S. lividans* ont été lysées par trois cycles de pression-décompression à 2000 kbar. Les débris cellulaires ont été séparés des protéines solubles par centrifugation (11000 rpm, 4 °C, 20 min). La fraction de protéines

solubles (FS) a été récupérée et stockée à -20 °C. Le culot des débris cellulaires (FI) a été resuspendu dans 15 mL de tampon et a été également stocké à -20 °C.

2.2.5. Protocole de refolding

Les fractions insolubles issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamAA 6-His C-ter ont été utilisées pour l'étude d'un protocole de refolding visant à exploiter l'importante production de SamAA sous la forme de corps d'inclusion insolubles.
Les fractions insolubles ont été centrifugées (11000 rpm ; 4 °C ; 20 min) pour permettre un changement de tampon. Le processus de refolding a eu lieu dans un tampon Tris base 100 mM, $CoCl_2$ 6 H_2O 0,5 mM, pH 11,5 et s'est déroulé pendant 24 h à 6 °C sous agitation. Puis, une nouvelle centrifugation a eu lieu pour séparer la fraction soluble refoldée de la fraction insoluble refoldée. Cette dernière a été resuspendue dans du tampon Tris base. Les deux fractions refoldées ont été stockées à -20 °C dans l'attente de leurs analyses.

2.2.6. Dosages protéiques

La concentration protéique des fractions solubles a été déterminée par le dosage de Bradford et le dosage BCA pour la réalisation des analyses SDS-PAGE et enzymatiques. La concentration protéique des fractions solubles concentrées, des fractions refoldées et des fractions purifiées a également été déterminée par ces deux dosages.

2.2.6.1. Dosage de Bradford

La méthode de Bradford est basée sur le dosage colorimétrique de l'interaction entre le bleu de Coomassie G-250 et les résidus basiques et hydrophobes des protéines. Le réactif de Bradford est une solution contenant 200 mg de bleu de Coomassie G-250 dissout dans 100 mL d'éthanol, 200 mL d'acide orthophosphorique et 200 mL d'eau (Bradford, 1976). 20 µL d'échantillon ou de standard a été mélangé à 980 µL du réactif de Bradford (Sigma Aldrich). Le mélange réactif est laissé pendant 15 min à l'obscurité, puis le changement de l'absorbance est mesuré à 595 nm. La concentration en protéines a été calculée à partir d'une droite de calibration d'une solution étalon (0,25 à 1,4 g/L) d'albumine de sérum bovin (BSA).

2.2.6.2. Dosage BCA

Le dosage BCA a été effectué pour calculer la concentration protéique des fractions purifiées dont la concentration protéique n'a pas pu être déterminée par un dosage de Bradford car la concentration protéique de ces fractions était inférieure à la limite basse de détermination par la méthode de Bradford fixée à 0,1 g/L. C'est pourquoi la méthode classique du dosage BCA a été adaptée dans le but de calculer la concentration protéique des fractions purifiées inférieure à 0,1 g/L.

Le principe du dosage colorimétrique reste inchangé. Il est basé sur la formation d'un complexe entre le Cu^{2+} et les protéines en conditions alcalines suivie d'une réduction du Cu^{2+} en Cu^{+} catalysée par les résidus cystéine, tryptophane et tyrosine et les liaisons peptidiques des protéines. Le taux de réduction est donc proportionnel à la quantité de protéines présentes. La couleur pourpre apparait grâce à la présence de l'acide bicinchoninique lors de la formation des complexes avec le Cu^{+}.

Le réactif BCA est un mélange de deux réactifs A et B, préparé en mélangeant 50 unités du réactif A (mélange d'acide bicinchoninique, de carbonate de sodium, de tartrate de sodium et de bicarbonate de sodium dans de la soude 0,1 M) avec 1 unité du réactif B (solution de sulfate penta-hydrate de cuivre à 4 % (p/v)). 1 mL de réactif BCA a été mélangé à 50 µL d'échantillons ou de standard. Le mélange a été incubé à 60 °C pendant 15 min. Après avoir laissé refroidir, l'absorbance à 562 nm a été mesurée. La concentration en protéines a été calculée à l'aide d'une droite de calibration d'une gamme d'étalonnage réalisées avec du BSA (0,005 à 0,25 g/L).

2.2.7. Concentration par ultrafiltration

Quand la concentration protéique de la fraction soluble est faible (inférieure à 1 g/L), une concentration d'un facteur 10 a été effectuée par ultrafiltration dans des systèmes de filtration Amicon Ultra-15 10 KDa (Millipore) en réalisant des séries de centrifugation (4000 g, 4 °C, 15 min). Les fractions solubles concentrées (FSc) ont été récupérées et sont stockées à -20 °C.

2.2.8. Electrophorèse SDS-PAGE

L'analyse des protéines de l'ensemble des fractions obtenues est effectuée par électrophorèse sur gel de polyacrylamide en présence de dodécyl-sulfate de sodium (SDS-PAGE), par

migration verticale. Cette technique permet de séparer des molécules en fonction de leur poids moléculaire. Les protéines sont dénaturées et chargées négativement par le SDS.
Des gels commandés préalablement coulés (4-12 % Criterion^{TM} XT Bis-Tris Protein Gel ; 18 puits de 30 µL ; BioRad) ont été utilisés. Chaque échantillon, ayant un volume de 18 µL, a été préparé en mélangeant 6 µL de tampon de charge (XT Sample Buffer 4X ; BioRad) et 1 µL d'agent réducteur (XT Reducing Agent 20X ; BioRad). Une concentration de 1 g/L par puit est conseillée par BioRad : si la concentration protéique de la fraction était supérieure à 1,5 g/L, l'échantillon était dilué dans du tampon (Tris-HCl 25 mM, NaCl 50 mM, pH 8). Le mélange a été ensuite dénaturé à 95 °C pendant 5 min. 10 µL de marqueur de poids moléculaire (Precision Plus Protein Unstrained Standards ; BioRad) et 20 µL d'échantillons préparés ont été déposés dans les puits du gel. La migration a eu lieu dans le tampon XT MES Running Buffer 1X (BioRad) à 100 V. Une fois la migration terminée, les gels ont été lavés trois fois pendant 5 min dans l'eau ultrapure sous agitation. Ils ont été colorés dans la solution de coloration au bleu de Coomassie (Bio-Safe Coomassie ; BioRad) sous agitation pendant une heure, puis ils ont été décolorés par plusieurs bains d'eau ultrapure de 30 min sous agitation, avant d'être photographiés.

2.2.9. Synthèse enzymatique

La réaction de synthèse de la N-α-lauroyl-L-lysine est utilisée pour vérifier la présence des enzymes recombinantes et comparer la quantité d'enzymes produites sous une forme soluble et active dans les différentes conditions opératoires.
La réaction de N-acylation a été réalisée dans des tubes à vis équipés de barreaux aimantés et chauffés dans un bain d'huile par une plaque chauffante thermorégulée (IKA). Les deux substrats (acide laurique et L-lysine) étaient en condition équimolaire à 100 mM. Le tampon Tris HCl 25 mM, NaCl 50 mM, pH8 a été supplémenté en cobalt à 0,5 mM. La solubilisation du donneur d'acyle (acide laurique) a été améliorée en réalisant une émulsion à l'aide d'une sonication pendant 2 min à 40 % de la puissance maximum. Les réactions enzymatiques ont débuté par l'ajout des fractions solubles à 0,5 g/L, pour un volume réactionnel final de 2 mL. Elles ont été réalisées à 45 °C pendant 48 h sous agitation (500 rpm). Des échantillons ont été prélevés en début et en fin de réaction et ont été dilués 5x dans un mélange méthanol/eau (80/20) puis filtrés avant d'être analysés par CCM (Chromatographie sur Couche Mince) et/ou par HPLC (High Performance Liquid Chromatography).

2.2.10. Hydrolyse enzymatique

Les réactions d'hydrolyse de la N-α- et de la N-ε-acétyl-L-lysine ont été réalisées pour évaluer l'activité des fractions solubles de SamELA. L'hydrolyse a été réalisée à 37 °C dans des tubes à vis contenant du tampon Tris-HCl 25 mM, NaCl 50 mM, pH 8 additionné en cobalt à 0,5 mM et en substrat à 4 mM pour un volume réactionnel final de 2 mL. Les réactions enzymatiques ont débuté par l'ajout des fractions solubles à 0,5 g/L et ont duré pendant 24 h sous agitation (300 rpm). La présence de L-lysine libérée lors de cette réaction a été vérifiée par CCM.

2.2.11. Méthodes analytiques

2.2.11.1. Chromatographie sur couche mince (CCM)

Les prélèvements des réactions d'hydrolyse et de synthèse réalisées lors de l'étude de la production hétérologue de SamELA ont été analysées par chromatographie sur couche mince (CCM). Des plaques de silice de 20 x 20 cm (Silica gel 60, Merck) sur lesquelles ont été déposées 5 µL de prélèvement ou de standard à 2 cm de la marge de la plaque, ont été utilisées. Les standards de la réaction de synthèse étaient la L-lysine et la N-ε-lauroyl-L-lysine (Sigma-Aldrich) tandis que les standards de la réaction d'hydrolyse étaient la L-lysine et la N-ε-acétyl-L-lysine. Une fois les dépôts faits et séchés, les plaques ont été placées dans une enceinte fermée hermétiquement contenant l'éluant qui était un mélange de butanol (60 %), d'acide acétique (20 %) et d'eau (20 %). Après la migration, les plaques ont été séchées à l'air libre. Une solution éthanolique de ninhydrine à 0,1 % a été utilisée lors de la révélation, car la ninhydrine est un composé réagissant avec les fonctions amines libres. Les plaques ont été chauffées à 100 °C dans une étuve ou avec un sèche-cheveux, entrainant l'apparition d'un chromophore violet.

2.2.11.2. Analyse par HPLC

La quantification de la N-α- et de la N-ε-lauroyl-L-lysine produites lors des réactions de synthèse a été réalisée par HPLC sur une chaine Shimadzu® LC 20 puis LC 40. Elle est équipée d'une colonne Alltima HP C18 « amide » (150 x 2,1 mm, 5 µm, Alltech Grace®) avec une pré-colonne (7,5 x 2,1 mm, 5 µm, Alltech Grace®) maintenue à 25 °C dans un four colonne, d'un injecteur automatique d'échantillons, d'un détecteur UV à barrette de diode et d'un détecteur évaporatif à diffusion de lumière à faible température (DEDL). La composition et le gradient

des solvants d'élution sont présentés dans le tableau II.5. Le débit des solvants était de 0,2 mL/min. Le volume d'injection était de 10 µL. Le détecteur DEDL réglé à une pression d'air comprimé de 350 kPa et une température de 40 °C permet de détecter et de quantifier les substrats et les produits de la réaction de synthèse par spectrométrie UV à 214 nm. La concentration en N-α-lauroyl-L-lysine a été déterminée en se référant à la droite d'étalonnage du standard (Genscript).

Tableau II.5 : Composition et gradient des solvants d'élution utilisés lors des analyses HPLC.

Temps (min)	Phase A (méthanol/eau/acide trifluoroacétique (TFA), 60/40/0,1)	Phase B (méthanol/TFA, 100/0,07)
0 – 1	95 %	5 %
1 – 7	5 %	95 %
7 – 20	5 %	95 %
20 – 21	95 %	5 %
21 – 35	95 %	5 %

2.2.12. Purification de la protéine recombinante SamAA

La purification de la protéine recombinante SamAA a été réalisée à partir d'une fraction soluble issu d'une culture d'*E. coli* BL21 Gold (DE3) transformée avec le plasmide pET-28a-SamAA N-ter 6-His sur le milieu LB-Co en utilisant la chromatographie d'affinité aux ions métalliques immobilisés (IMAC). La colonne HisTrap^{TM} Excel (Cytiva) contenant du nickel (Ni^{2+}) immobilisé sur la phase stationnaire permet la rétention de l'étiquette poly-histidine (6-His) présente en position N-terminale de la protéine recombinante.

2.2.12.1. Système seringue

Une première purification de SamAA recombinante a été réalisée en utilisant un système seringue manuelle, avec un débit d'environ 1 mL/min. La colonne a été lavée à l'éthanol puis équilibrée avec le tampon Tris-HCl 25 mM, NaCl 50 mM, pH 8. Le lavage et l'élution ont été réalisés avec le tampon en présence d'imidazole dont la concentration est croissante (20 mM, 80 mM, 160 mM et 250 mM). Toutes les fractions ont été récupérées et ont été stockées à -20 °C.

2.2.12.2. ÄKTA Pure

Une seconde purification de SamAA recombinante a été réalisée sur l'automate ÄKTA Pure (Cytiva). Ce système permet l'automatisation de la purification par chromatographie d'affinité et le suivi de l'élution par la présence d'un spectrophotomètre réglé à 214 nm. Les tampons utilisés étaient les tampons Tris-HCl 25 mM, NaCl 50 mM, pH8 supplémentés en imidazole à 20 mM ou à 250 mM. Les paliers d'élution à 80 mM et à 160 mM ont été gérés par l'automate. Toutes les fractions ont été récupérées et ont été stockées à -20 °C.

2.2.12.3. Dialyse – Concentration protéique

Chaque fraction récupérée au cours des deux purifications a été concentrée par ultrafiltration à l'aide des systèmes de filtration Amicon Ultra-15 10 KDa (Millipore) pour les analyses SDS-PAGE et enzymatiques. Cette technique permet également d'éliminer l'imidazole.

2.2.13. Dot-Blot

Un dot-blot repose sur une méthode d'immuno-détection permettant de visualiser la présence de molécules porteuses de l'antigène déposées sur une membrane. Il a été réalisé pour vérifier l'accessibilité du tag 6-His en position N-terminale de la protéine recombinante SamAA dans le cadre de sa purification par chromatographie d'affinité au nickel. L'échantillon à analyser a été transféré sur une membrane par diffusion simple et mis en contact avec des anticorps primaires anti-tag 6-His (anticorps penta-His, QIAGEN ; solution diluée au 1000ème dans du tampon bloquant (Tris-HCl 10 mM, NaCl 150 mM, BSA 3 % (p/v), pH 7,5)) dans un premier temps puis avec des anticorps secondaires anti-anticorps primaires (Anti-mouse IgG HPR-linked antibody, Ozyme ; solution diluée au 10000ème dans du tampon TBS/lait (Tris-HCl 10 mM, NaCl 150 mM, poudre de lait 5 % (p/v))) dans un second temps. Les anticorps secondaires sont couplés à une peroxydase (HPR) permettant une détection par chimioluminescence.
Certaines fractions solubles concentrées et obtenues après purification issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA ont été testées. Elles ont été préparées par dilution au 5ème dans un tampon dénaturant (Urée 8 M, NaH_2PO_4 100 mM, Tris-HCl 10 mM, pH 8) ou natif (NaH_2PO_4 50 mM, NaCl 300 mM, pH 8). 1 µL de chaque échantillon a été déposé sur une membrane PVDF (polyfluorure de vinylidène), préalablement activée dans un bain d'éthanol absolu de 5 min. La membrane a été séchée à l'air libre puis a été plongée dans

un second bain d'éthanol absolu. Elle a été ensuite lavée deux fois pendant 10 min avec du tampon TBS (Tris-HCl 10 mM, NaCl 150 mM, pH 7,5) et a été incubée pendant 1 h dans du tampon TBS/lait. La membrane a été lavée deux fois 10 min dans du tampon TBS/Tween (Tris-HCl 20 mM, NaCl 150 mM, Tween 20 0,05 % (v/v), pH 7,5). Ensuite, elle a été placée une heure dans la solution d'anticorps primaires et a été de nouveau lavée deux fois pendant 10 min dans du tampon TBS/Tween et une fois pendant 10 min dans du tampon TBS. La membrane a été ensuite placée dans la solution d'anticorps secondaires pendant 1 h avant d'être lavée deux fois pendant 10 min dans du tampon TBS/Tween et une fois pendant 10 min dans du tampon TBS. La réaction de détection chimioluminescente peut alors être réalisée dans le laboratoire d'IMOPA au BioPôle avec le kit de révélation Immobilon^{TM} Western (Millipore) et l'automate Fusion Solo (Vilber).

3. Chapitre III : Modélisation de la synthèse et de l'hydrolyse de dérivés N-acylés de la L-lysine catalysées par SamAA

3.1. Introduction

Les dérivés N-acylés d'acides aminés sont des molécules obtenues par acylation de leur fonction amine. Les acides aminés les plus abondants sont α-aminés ; parmi eux, certains comme la lysine, présentent en outre une fonction amine primaire au niveau de leur chaîne latérale, permettant d'envisager différents sites de N-acylation. Leur diversité structurale leur confère des propriétés fonctionnelles différentes menant à un large champ d'applications, dans les secteurs des détergents et de l'hygiène domestique, de l'alimentation, de l'énergie et de la médecine (Le Guenic et al., 2018 ; Tripathy et al., 2018 ; Guo et al., 2022). Ils sont produits actuellement par la réaction chimique de Schotten-Baumann, qui est une réaction conduisant à des rendements compris entre 70 et 90 %. Cependant, cette réaction a lieu dans un environnement très basique ; elle fait intervenir des chlorures d'acyles et nécessite la protection des groupements ne devant pas être modifiés car elle n'est pas régiosélective. Les produits de la réaction sont isolés par précipitation acide, filtration et lavage, créant ainsi un volume important d'effluents salins à traiter (Gerova et al., 2008 ; Bordes et al., 2009 ; Sreenu et al., 2014a ; Sreenu et al., 2014b ; Joondan et al., 2018).

Les acides aminés N-acylés peuvent être également produits par voie enzymatique. Les enzymes les plus décrites et utilisées pour catalyser la réaction de N-acylation sont les lipases. Elles catalysent le greffage d'un groupe acyle (donneur d'acyle) sur une fonction amine (accepteur d'acyle) dans des milieux pauvres en eau tels que les solvants organiques anhydres, les systèmes sans solvant ou encore les liquides ioniques (Montet et al., 1990 ; Soo et al., 2004 ; Husson et al., 2008 ; Husson et al., 2009 ; Kua et al., 2023). La réaction de N-acylation peut aussi être catalysée par des enzymes de la famille des aminoacylases. La réaction enzymatique est alors menée en milieu aqueux, ce qui favorise la solubilité du substrat et réduit l'impact environnemental du procédé comparativement aux réactions conduites en milieu solvant. Plusieurs aminoacylases (Sm-AA, Sm-ELA et Sm-PVA) catalysant la N-acylation des différents acides aminés ont été identifiées dans le surnageant de culture de *Streptomyces mobaraensis* (Koreishi et al., 2005a ; Koreishi et al., 2007 ; Koreishi et al., 2009a ; Koreishi et al., 2009b). Plus récemment, d'autres aminoacylases produites par *Streptomyces ambofaciens* ont été découvertes : les gènes codant pour les enzymes SamAA, SamELA et SamPVA

partagent 86, 80 et 69 % d'identité de séquences avec les gènes de Sm-AA, Sm-ELA et Sm-PVA, respectivement. L'étude de mutants de délétion a montré une diminution spectaculaire de l'hydrolyse de la N-α- et de la N-ε-acétyl-L-lysine dans les fractions provenant du mutant dépourvu du gène de SamAA et du gène de SamELA (Dettori et al., 2018a). La spécificité de SamAA pour l'acylation du groupe amine α de la L-lysine a été confirmée dans le cadre des travaux de Bourkaib (Bourkaib et al., 2020a ; Bourkaib et al., 2021). Jusqu'à présent, les études sur ces aminoacylases visaient à préciser leur potentiel pour la N-acylation, sans fournir d'explication pour leurs propriétés de sélectivité.

Les travaux de ce chapitre portent plus spécifiquement sur l'aminoacylase SamAA. Cette dernière est d'un grand intérêt, car sa sélectivité favorise l'acylation de la fonction amine α de la L-lysine. Cette régiosélectivité la distingue d'autres enzymes plus connues comme les lipases et permet d'envisager la N-acylation d'une grande diversité d'acides aminés et de peptides. L'objectif principal de cette étude est d'expliquer la sélectivité de SamAA à l'échelle moléculaire, c'est-à-dire sur la base d'éléments structurels. A ce jour, aucune structure d'aminoacylase de *S. ambofaciens* n'a encore été élucidée ; les difficultés rencontrées pour produire ces enzymes sous forme soluble en constituent la principale raison.

En attendant que des données structurales expérimentales soient disponibles, une stratégie de modélisation de la structure tridimensionnelle de l'enzyme a été choisie, en utilisant une approche de modélisation par homologie (Muhammed et al., 2018 ; Haddad et al., 2020). Plusieurs outils de prédiction de structure ont été étudiés de façon à proposer un modèle robuste de la structure tridimensionnelle de SamAA. Ce dernier a été utilisé comme cible de docking pour étudier les modes d'interaction entre l'enzyme et les substrats des réactions de N-acylation et N-déacylation et identifier les éléments responsables de la sélectivité de l'enzyme (Selvam et al., 2017).

3.2. Etude par modélisation moléculaire de la structure de l'aminoacylase SamAA et sa sélectivité envers la N-acylation de la L-lysine

Les travaux présentés sous la forme d'un article ont pour but d'expliquer la sélectivité de SamAA lors de la synthèse et de l'hydrolyse des dérivés acétylé et caproylé de la L-lysine à l'échelle moléculaire. Trois outils de prédiction de la structure tridimensionnelle de protéines, à partir de leur séquence en acides aminés, ont été utilisés, conduisant à différents modèles dont la robustesse a été évaluée. Le modèle le plus robuste a été choisi pour émettre des premières hypothèses sur la structure de SamAA et les éléments impliqués dans la catalyse. Ce modèle a

par ailleurs servi de cible de docking d'une part pour la L-lysine et les donneurs d'acyle acétyl et caproyl dans le cas de la réaction de N-acylation, d'autre part pour les dérivés N-acétyl- et N-caproyl-L-lysine dans le cas de l'hydrolyse. Enfin, sur la base des résultats obtenus, un mécanisme d'hydrolyse et de synthèse de la N-α-caproyl-L-lysine est proposé.
La présentation de ces travaux est suivie de résultats complémentaires, portant plus spécifiquement sur la modélisation de l'hydrolyse et de la synthèse de la N-lauroyl-L-lysine.

Title: Molecular insights in *Streptomyces ambofaciens* aminoacylase structure and its selectivity towards L-lysine N-acylation

Auteurs : Laureline Gennesseaux[1], Stéphane Delaunay[1], Bertrand Aigle[2], Isabelle Chevalot[1], Yann Guiavarc'h[1], Catherine Humeau[1]
1 : Université de Lorraine, CNRS, LRGP, F-54000 Nancy, France
2 : Université de Lorraine, INRAE, DynAMic, F-54000, Nancy, France

Abstract:
Synthesis and hydrolysis of L-lysine acylated derivatives are successfully catalyzed in aqueous medium by the aminoacylase SamAA from *Streptomyces ambofaciens* with strict selectivity towards α-amine function enabling the preparation of products of interest by an environmentally friendly way. To date no explanation was given for this selectivity that distinguishes this enzyme from other better known biocatalysts like lipases. This study focuses on the prediction of SamAA 3D structure and its interaction with L-lysine, acetate and caproate as acyl donors and the corresponding acylated derivatives using homology modelling and docking simulations. After investigating three homology modelling softwares (Modeller, Swiss Model and AlphaFold2), a 3D structure of SamAA was proposed in the form of a binuclear homodimer with a whistle-shaped monomer. Alignment with templates allowed identifying SamAA family and key conserved residues. SamAA belongs to the M20 peptidase family. Residues His88, Asp120, Glu155, Glu182 and His416 are involved in the binding of metal ions whereas residue Glu154 was identified as the essential acid-base residue. N-α/ε-acetyl-L-lysine and N-α/ε-caproyl-L-lysine were used as model substrates for docking simulations to mimic their hydrolysis or synthesis. The results strongly suggested the selective acylation/deacylation of L-lysine α-amine function, while acylation/deacylation of its ε-amine function appeared unlikely to occur which perfectly matches experimental results. Catalytic mechanism for either

hydrolysis or synthesis of N-α-caproyl-L-lysine was proposed. These results constitute the first elements featuring the structure, the selectivity and the mechanism of this aminoacylase.

Keywords: aminoacylase; *Streptomyces ambofaciens*; N-acylation; selectivity; structure prediction; molecular modelling

3.2.1. Introduction

N-acyl-amino acids are bimodular structures resulting from the acylation of amino acid amine function (Figure III.1). Depending on their structure, they exhibit a variety of technofunctional properties and biological activities, giving them a high application potential. They are commonly used in detergents and household hygiene, food, energy and medical sectors (Le Guenic et al., 2018; Tripathy et al., 2018; Guo et al., 2022). Currently, they are produced by the Schotten-Baumann chemical reaction. This reaction is mastered and has a yield of between 70 and 90 %. However, it requires the use of acyl chlorides and amino acids with protected amine functions for the choice of reactive function in a strongly basic medium. Then, the acylated derivatives are recovered by acid precipitation, followed by filtration and washing, generating a large volume of saline organic effluents to be treated (Gerova et al., 2008; Bordes et al., 2009; Sreenu et al., 2014a; Sreenu et al., 2014b; Joodan et al., 2018).

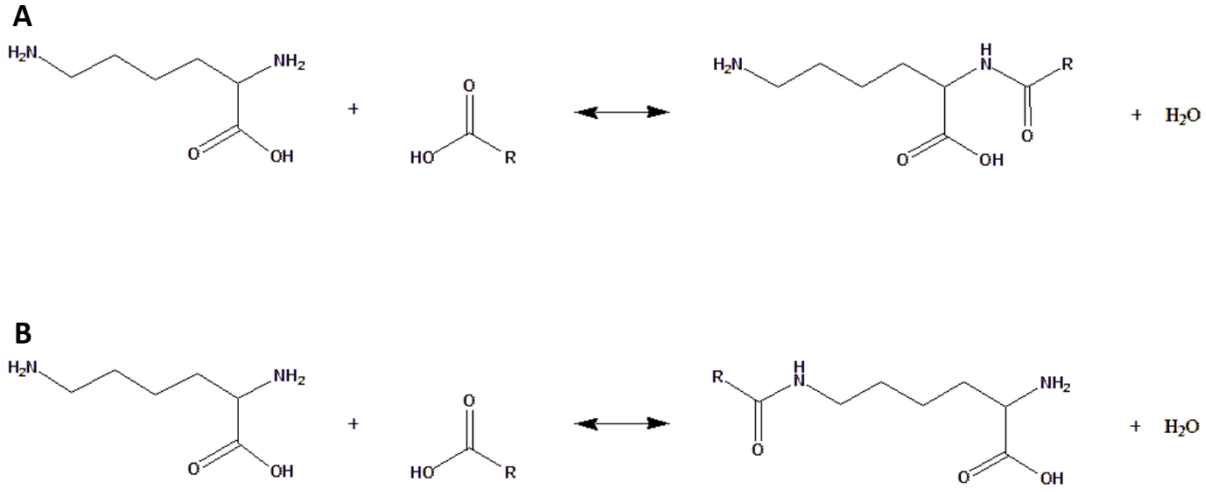


Figure III.1: N-acylation of L-lysine; A: N-α-acylation and B: N-ε-acylation.

N-acyl amino acids can also be produced by the enzymatic pathway. This reaction is chemo-, regio- and enantio-selective and is performed under mild reaction conditions, thus complying with the twelve principles of green chemistry (Sheldon et al., 2018). Lipases are the best known

and most widely described of the enzymes capable of catalysing the acylation reaction. Lipases catalyse the transfer of an acyl group from an acyl donor to an amine function considered as an acyl acceptor in various media such as dry organic solvents, solvent-free systems and ionic liquids (Montet et al., 1990; Husson et al. 2008). For example, acylation of the dipeptide Lys-Ser (carnosine) by oleic acid took place on the lysine ε-amino group in the ionic liquid $[Bmim]^+$ $[PF_6]^-$ (1-butyl-3-methyllimidazolium hexafluorophosphate) which improved the solubility of the dipeptide leading to an increase in the initial rate and the conversion yield of the reaction (Husson et al., 2009). *Rhizomucor miehei* lipase showed enzymatic specificity towards the ε-amine function of lysine during its acylation by palmitic acid and oleic acid in several organic solvents (Soo et al., 2004). More recently, a mutant of this lipase (RML(D156G)) with an eleven-fold higher hydrolytic activity than the wild-type enzyme was selected and used to catalyse the synthesis of N-lauroylglycine in an aqueous system containing glycerol (70 % v/v). A 80 % yield of conversion of L-lysine was obtained that surpassed all other known results observed with similar enzymes (Kua et al., 2023).

Enzymes belonging to the aminoacylase family have been described for their ability to catalyse the N-acylation reaction in an aqueous medium, resulting in improved substrate solubility and less environmental impact compared to processes carried out in solvent media. A first 60 kDa monomeric aminoacylase from *Streptomyces mobaraensis* discovered by Koreishi's team was shown to catalyse the hydrolysis and synthesis of several N-ε-acyl-L-lysine derivatives. This catalysis takes place in an aqueous buffer and its yield can exceed 90 % (Koreishi et al., 2005a). This enzyme, called Sm-ELA, achieves a yield close to 100 % when using 500 mM L-lysine hydrochloride and 50 or 100 mM lauric acid after 6 h or 9 h of reaction respectively (Koreishi et al., 2009a). Another acylase called Sm-PVA is a 61 kDa and 19 kDa heterodimer from *S. mobaraensis* that catalyses the synthesis of β-lactam antibiotics, capsaicin derivatives and N-lauroyl-amino acids. By way of example, this enzyme catalyses the acylation of the lysine side chain by lauric acid (Koreishi et al., 2007). A third monomeric 55 kDa aminoacylase (Sm-AA) is produced by *S. mobaraensis* and catalyses the hydrolysis of short and long acyl derivatives of amino acids (Koreishi et al., 2009b).

Other aminoacylases were discovered in a crude culture extract of *S. ambofaciens ATCC23877*. Specific activities towards N-α-acetyl-L-lysine and N-ε-acetyl-L-lysine were found in different fractions from partial purification of the crude extract, suggesting the presence of several aminoacylases from this bacterium. A sequence similarity search found three genes in *S. ambofaciens* sharing 86, 80 and 69 % identity respectively with the genes of the three aminoacylases Sm-AA, Sm-ELA and Sm-PVA identified in *S. mobaraensis*. Deletion mutants

retaining only the SamAA or SamELA gene were created. A dramatic decrease in N-α-acetyl-L-lysine hydrolysis was observed in fractions from the mutant lacking the SamAA gene. A similar result was obtained upon N-ε-acetyl-L-lysine hydrolysis in fractions from the mutant lacking the SamELA gene. These results led to the identification of the enzymes SamAA and SamELA specifically catalysing α- and ε-acylation of L-lysine respectively (Detorri et al., 2018a). An additional study investigated the acylation of the twenty proteogenic amino acids with 10-undecylenic acid: one of the highest yields was obtained with positively charged amino acids, such as lysine. The enzymatic specificity of SamAA was also confirmed by acylating L-lysine with octanoic acid in the presence of wild-type aminoacylases from *S. ambofaciens* and the mutant containing only the SamAA gene. Only the α-derivative was obtained with SamAA, while the α- and ε-derivatives were detected in the presence of the aminoacylases mixture (Bourkaib et al., 2020b). This study focuses on the aminoacylase SamAA which is particularly interesting due to its selectivity that promotes the N-α-acylation of L-lysine. This regioselectivity differs from that of lipases, making it possible to consider the acylation of any other amino acid or peptide.

To date, studies on *S. ambofaciens* aminoacylases focused on the reaction scale and have not been able to explain the high selectivity of these enzymes. Yet there is a particular interest to explain the selectivity and the specificity of enzymes at a molecular scale but it requires knowing their structure. In the case of *S. ambofaciens* aminoacylases, no structure has been yet elucidated, as is the case for many enzymes of this class. However, recent advances in bioinformatics now allow us to predict the structure of proteins of known sequence. Homology modelling allows to predict the 3D structure of a protein from its amino acid sequence and known 3D structure of one or more templates. The target protein and the template(s) must present the highest possible sequence homology percentage (Muhammed et al., 2018; Haddad et al., 2020). Once the model of the 3D structure has been obtained, it can be used in numerical simulations such as molecular docking simulations, pending experimental elucidation of the 3D structure (Selvam et al., 2017).

The main objective of this study is to explain the selectivity of SamAA during the synthesis and the hydrolysis of N-acyl-L-lysine derivatives based on structural elements, which involves building a 3D model of the enzyme. Three homology modelling tools were used leading to several models among which the most robust was chosen for further docking simulations. Structural elements that are required for catalysis were assumed to be similar to those of the templates. The most probable binding modes between the enzyme and L-lysine, acetic acid and

caproic acid as well as the corresponding N-α-acyl or N-ε-acyl-L-lysine derivatives were studied to explain the selectivity observed experimentally.

3.2.2. Materials and methods

3.2.2.1. Search for templates

The first step to obtain a quaternary structure of a protein by homology modelling is the identification of templates. Three tools were used to identify SamAA templates. The first program is the NCBI BLASTp web interface available at https://blast.ncbi.nlm.nih.gov/Blast.cgi (Altschul et al., 1997; Altschul et al., 2005; Johnson et al., 2008). The amino acid sequence of the aminoacylase was compared to those available in the database for the three-dimensional structural data of proteins "PDB". The second program is the Expasy Swiss Model web interface available at https://swissmodel.expasy.org/interactive#sequence (Camacho et al., 2009; Mirdita et al., 2017; Steinegger et al. 2019). The last tool is the application software called MMseqs2 (mmseqs.org) which is available on the ColabFold web interface coupled with AlphaFold2. Multiple sequence alignment was generated against the UniRef and Environmental sequence databases (Mirdita et al., 2017; Steinegger et al., 2019). The result of multiple protein sequence alignment was visualised using ESPript 3.0 (https://espript.ibcp.fr/ESPript/ESPript/) (Robert et al., 2014).

3.2.2.2. Homology modelling

Once the templates were identified, the 3D structure of SamAA was created by homology modelling after carrying out a sequence alignment of the templates and SamAA. The programs "Modeller" from Discovery Studio® (version 4.1 to version 22.1), "ProMod3" from Swiss Model and "AlphaFold2" from ColabFold were used to predict the 3D structure of SamAA. These tools were chosen based on the results issued from CASP13 and CASP14 (Critical Assessment of protein Structure Prediction) (Kryshtafovych et al., 2019 ; Kryshtafovych et al., 2021).

The use of Modeller, available in the Discovery Studio® suite, requires prior sequence alignment between the protein sequences of HiDapE as a template and SamAA. The "Align Sequences... " tool, also available in Discovery Studio®, was used by selecting the dimeric sequences of the two enzymes and the options "Use Secondary Structures... " (use of

information from secondary structures) and “Align by Chain... ” (alignment of multimeric protein sequences) (Thompson et al., 1994). The sequence alignment and the dimeric 3D structure of HiDapE were used in Modeller with default parameters to generate the model for SamAA 3D structure (Sali et al., 1993).

The monomeric sequence of SamAA was entered on the Swiss Model server available at https://swissmodel.expasy.org/interactive. This tool allows to predict the quaternary structure of the target protein and does not have any particular setting for neither sequence alignment nor 3D structure prediction (Guex et al. 2009; Bienert et al., 2017, Bertoni et al., 2017, Waterhouse et al., 2018).

The ColabFold GitHub web interface allows the coupled use of MMseqs2 and AlphaFold2 available at https://colab.research.google.com/github/sokrypton/ColabFold/blob/v1.2.0/AlphaFold2.ipynb (Jumper et al., 2021 and Mirdita et al. 2022). Version v1.2 was used. 3D models were generated from SamAA dimeric sequence. The predictions were performed over three cycles.

3.2.2.3. Evaluation of homology models

Each homology modelling tool has its own evaluation method and scores for classifying the different models obtained. The GMQE (Global Model Quality Estimate) and QMEANDisCo scores were applied to the models from Swiss Model (Waterhouse et al., 2018; Studer et al., 2020). The models from ColabFold were analyzed by lDDT and PAE (Mariani et al., 2013). All models were analyzed by the normalized DOPE Score (Shen et al., 2006; Eramian et al., 2008). This score, allowing models to be compared, was used for the selection of the best final model. The metal ion binding site of the templates presents a distorted trigonal or square bipyramidal coordination geometry (Bienvenue et al., 2003; Davis et al., 2006). This structural feature was used to select the final model for SamAA, by overlaying HiDapE and SamAA structures using the “AlignandSuperimposeProteins…” tool of Discovery Studio® (Thompson et al., 1994).

3.2.2.4. Optimisation of the model

The optimization of the best model was performed with Discovery Studio suite. Two Zn^{2+} ions were added manually within the two active sites of the dimeric structure. Their position was inspired from HiDapE after superimposition of the structures. A water molecule interacting with the two metal ions was added manually in each active site so that the characteristic

architecture of metalloenzymes was respected (Nocek et al., 2018). Then the model was typed with the Momany and Rone CHARMm force field (Momany and al., 1992). A solvation of the system was applied, followed by a two-step energy minimization using the Smart Minimizer (RMS Gradient at 0.1 Å) and the Conjugate Gradient (RMS Gradient at 0.01 Å) methods (Powell, 1997; Tronrud, 1992). This final model was used as a target for further docking simulations.

3.2.2.5. Docking simulations and scoring

Docking simulations and score analyses were performed on DiscoveryStudio suite. CDOCKER program was used for all docking simulations (Wu et al., 2003). The 50 best poses were requested. The binding site was defined "From receptor cavities... " and chosen so that it contains the metal ions and the catalytic residues. The RMSD threshold, the number of random conformations and the number of orientations to refine were set to 0.1, 20 and 20 respectively. Different ligands were docked according to this protocol in order to mimic either the hydrolysis of N-acyl-L-lysine derivatives or their synthesis. Then, N-α-, N-ε-acetyl-L-lysine, N-α- and N-ε-caproyl-L-lysine were docked to mimic the hydrolysis reaction. L-Lysine, acetic acid and caproic acid were docked aiming to mimic the reverse reaction. It that case, the water molecule interacting with the metal ions was removed from the target. The substrates to be docked have their amine functions protonated and their acid functions deprotonated.
The poses issued from docking simulations were ranked according to the negative score "CDOCKER ENERGY" that includes the internal constrained energy of the ligand and the receptor-ligand interaction energy. Additional scoring analyses were performed using different scoring functions (LigScore1_Dreiding, LigScore2_Dreiding, PLP1, PLP2, Jain, PMF, PMF4) gathered in the "Consensus Score" (Krammer et al., 2005; Gehlhaar et al., 1995; Gehlhaar et al., 1999; Jain, 1996; Muegge et al. ?, 1999; Muege, 2006). The poses were assumed to be similar if RMSD between them was less than 2 Å and clustered in order to determine the main binding modes of the ligands (Cole et al., 2005). Then, the interactions between the ligands and SamAA residues were studied with the "Analyze Ligands Poses" tool of Discovery Studio.

3.2.3. Results and discussion

Templates are structures that have been already elucidated and that are used as a source of inspiration for modelling unknown structures. Here, these templates guided the construction of

SamAA 3D model, and also provided usefull information for making hypotheses about the key residues and the mechanism of this enzyme.

3.2.3.1. Identification of the templates and main structural features

The templates were identified after performing a protein sequence alignment with NCBI and Swiss Model tools, showing the sharing of sequence identity with SamAA. The sequence identity and similarity percentages between SamAA ans its templates are presented in Table III.1.

Table III.1: Percentages of sequence identity and similarity between SamAA and its templates.

Enzyme name	Short name	PDB entries	Identity percentage	Similarity percentage
Mutant of human aminoacylase-I	mhAcy1	1Q7L	33 %	48 %
Succinyl-diaminopimelate desuccinylase from *Neisseria meningitidis* MC85	NmDapE	4PPZ	25 %	49 %
Succinyl-diaminopimelate desuccinylase from *Haemophilus influenzae*	HiDapE	3IC1 5VO3	24 %	31 %
Succinyl-diaminopimelate desuccinylase from *Shigella flexneri*	SfDapE	7LGP	22 %	30 %
Human Carnosinase-2	hCN2	4RUH	17 %	28 %

mhAcy1 is a partial mutant of human aminoacylase 1 responsible for the hydrolysis of N-acyl-L-amino acids (Lindner et al., 2003). NmDapE, NiDapE and SfDapE are succinyl-diaminopimelate desuccinylases from *Neisseria meningitidis* MC85, *Haemophilus influenzae* and *Shigella flexneri*, respectively. These enzymes catalyse the hydrolysis of L,L-N-succinyl-diaminopimelic acid to succinate and LL-2,6-diaminopimelic acid in the lysine and meso-diaminopimelic acid biosynthesis pathway (Nocek et al., 2014; Nocek et al., 2018; Osipiuk et al., 2021). hCN2 is a human carnosinase-2 catalysing the hydrolysis and synthesis of N-lactoyl-amino acids and the hydrolysis of mainly hydrophobic dipeptides (Pandya et al., 2014).

These five enzymes are homodimeric and binuclear functioning with zinc for the first four, cobalt for HiDapE and manganese for hCN2. mhAcy1, NmDapE, HiDapE and SfDapE have a water molecule in interaction with the two metal ions (Nocek et al., 2010). The structure of hCN2 was obtained in the presence of bestatin as an inhibitor close to the manganese ions.
Residues involved in metal ion binding and catalytic mechanism of the enzymes designated as templates have already been identified in previous studies. Electron absorption analyses and EPR (electron paramagnetic resonance) spectra allowed to determine the coordination geometry of HiDapE metal ion sites. The two metal ions are in a distorted trigonal bipyramidal coordination geometry composed of one aspartate residue, two glutamate residues, two histidine residues and one water molecule (Nocek et al., 2018). This geometry produces a characteristic organization of the active site and is necessary for the catalytic activity with the addition of another aspartate and glutamate residue. This architecture was also found in hCN2 where the two manganese ions have a square-based bipyramidal coordination geometry involving one glutamate, two histidine, two aspartate residues and the inhibitor bestatin. The active site is completed with two catalytic residues, one aspartate and one glutamate.

3.2.3.2. Protein sequence analysis

Multiple sequence alignment of SamAA and templates revealed the presence of conserved amino acids. These are shown in Figure III.2 and Table III.2.

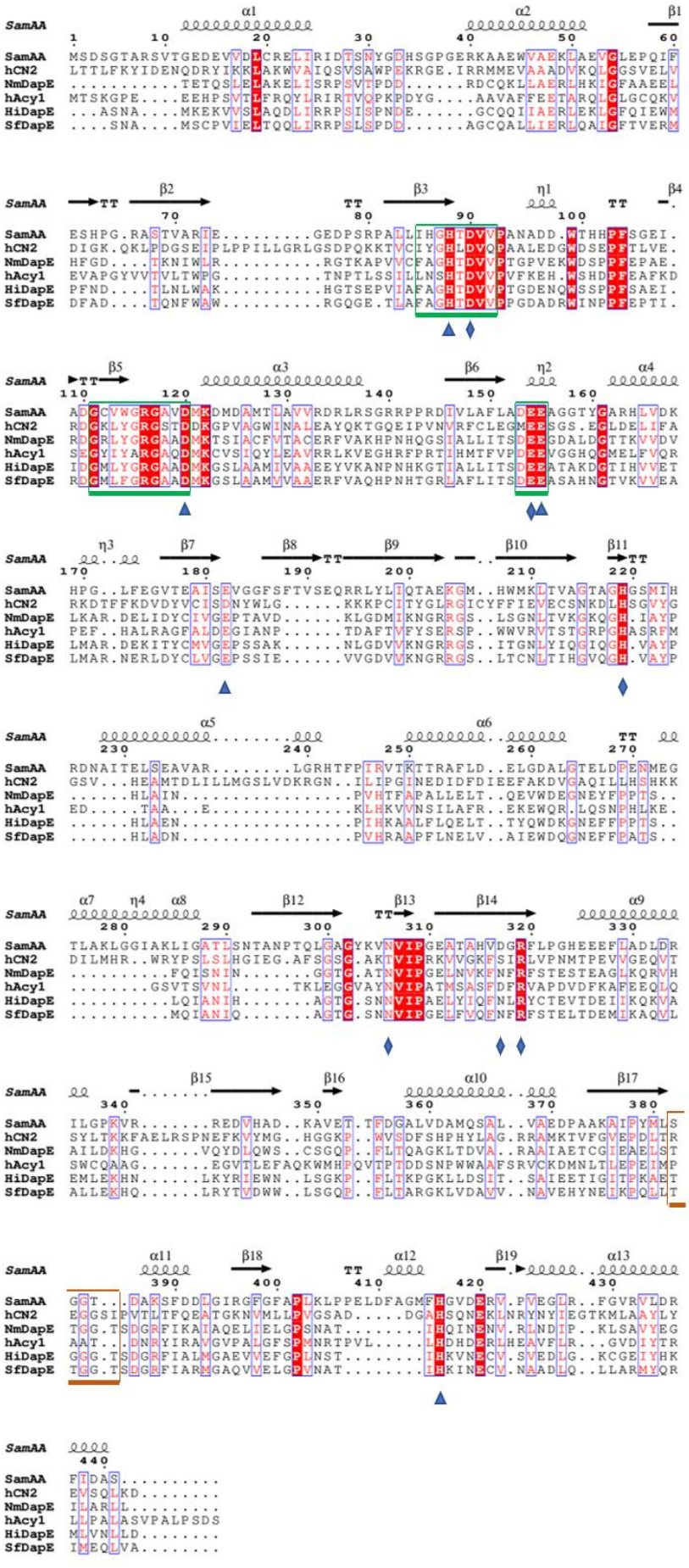

SamAA
hCN2
NmDapE
hAcyl
HiDapE
SfDapE

Figure III.2: Multiple sequence alignment of SamAA and templates. Residues that are likely to bind metal ions are annotated with a triangle. Residues that are supposed to have a role in the catalytic mechanism are annotated with a diamond. The loop that is assumed to be responsible for substrate specificity is identified by a brown frame. The three characteristic domains of the M20 peptidase family are identified by a green frame.

Table III.2: List and role of conserved residues.

Function	SamAA	HiDapE	NmDapE	mhACY1	hCN2	Chain
Metal-binding	His88	His67	His68	His80	His99	A
	Asp120	Asp100	Asp101	Asp113	Asp132	
	Glu155	Glu135	Glu136	Glu148	Glu167	
	Glu182	Glu163	Glu164	Glu175	Asp195	
	His416	His349	His350	His373	His445	
Catalytic	Asp90	Asp69	Asp70	Asp82	Asp101	A
	Glu154	Glu134	Glu135	Glu147	Glu166	
	Asp317	Asn256	Asn257	Asp274	Ser341	
	Arg319	Arg258	Arg259	Arg276	Arg343	
	His219	His194	His195	His206	His228	B
	Asn306	Asn245	Asn246	Asn263	Thr330	
Specificity loop	Ser382	Gly322	Thr323	Pro344	Glu414	A
	Gly383	Gly323	Gly324	Ala345	Gly415	
	Gly384	Gly324	Gly325	Ala346	Gly416	
	Thr385	Thr325	Thr326	Thr347	Ser417	

Site-directed mutagenesis study and partial elucidation of the 3D structure of mhAcy1 confirmed the role of the residues His80, Asp113, Glu148, Glu175 and His373 in the binding of the two Zn^{2+} ions (Lindner et al., 2003). The residue Glu134 was identified as an essential acid/base residue that is involved in the catalytic mechanism of HiDapE (Davis et al., 2006). These results allow us to assume that the conserved residues His88, Asp120, Glu155, Glu182 and His416 are involved in the binding of the metal ions in SamAA while the residue Glu154 participates to the catalytic mechanism.

Other key residues of mhAcy1 were identified as playing a part in catalysis, such as dimerization or substrate binding. An assomption was made that a conformational change occurs in the presence of the substrate which would cause a movement of the residues Asp274 and Arg276 of the chain A and residues His206 and Asn263 of the chain B towards the active

site of the chain A located near the metal ions (Lindner et al., 2005). This was confirmed by Nocek et al.: the crystallographic structure of HiDapE with and without ligand shows a shift of the residue His194 (corresponding to the residue His206 in mhAcy1 structure). The latter moves into the active site of the other peptidic chain forming an oxyanion hole (Nocek et al., 2018). All these amino acids are conserved in SamAA sequence and correspond to the residues Asp317, Arg319 of the chain A and His219, Asn306 of the chain B.

Another interesting feature is a hydrogen bond between an oxygen atom of the residue Asp69 and the Nε proton of the residue His67 that was discovered in HiDapE crystal structure. The presence of this bond allows the formation of an Asp-His-Zn triad regulating the Lewis acidity of the metal ion and facilitating the coordination of a double-bonded oxygen atom of the residue Asp100 and/or Glu135 with the second metal ion (Nocek et al., 2010). This triad is also conserved in SamAA involving the residue Asp90.

Concerning the substrate specificity loop, the residues Gly322 to Thr325 were identified in HiDapE structure after site-directed mutagenesis and bioinformatics studies. This loop takes part in active site construction and substrate stabilisation (Nocek et al., 2014). In addition, it was shown that mutating the conserved residue Thr347 of mhAcy1 to Ser increases the selectivity towards N-benzoylamino acids (Lindner et al., 2008). This loop corresponding to the residues Ser382 to Thr385 seems also conserved in SamAA structure.

Finally, the enzymes belonging to the M20 peptidase family present three conserved domains whose sequences are as follows: [L, F, V] – X – [S, G, A] – H – X – D – X – V; [G, D] – X – X – [Y, W, F] – [G, A] – X – G – X – X – D; X – E – E (Koreishi et al., 2009b). The five templates identified for SamAA belong to this family. Moreover, these sequences are also found in SamAA as follows: from Ile85 to Val92 ([I – H – G – H – T – D – V – V]), from Gly111 to Asp120 ([G – C – V – W – G – R – G – A – Y – D]) and from Asp153 to Glu155 ([D – E – E]). Regarding the conservation of these features coupled with the conservation of the residues responsible for the binding of metal ions or involved in the catalytic mechanism, SamAA is assumed to belong to the M20 peptidase family.

3.2.3.3. Construction of homology models

Three homology modelling tools were used to predict the 3D structure of SamAA. In 2018, at CASP13, Modeller, Swiss Model and AlphaFold were presented as three of the best modelling tools for the prediction of protein 3D structures by homology modelling, but AlphaFold was not disclosed by DeepMind (Kryshtafovych et al., 2019). Two years later, at CASP14,

AlphaFold2 retained its first place and its source code was published by DeepMind and interfaced by ColabFold (Kryshtafovych et al., 2021; Miridta et al., 2022). In this study, the best models issued from these simulations were identified basing on the normalized DOPE Score. Then the conservation of some structural features, especially those concerning the binding of metal ions, was considered to choose the best structure as a target for further docking simulations.

A sequence alignment between the template HiDapE and SamAA was obtained using the "Align Sequences... " tool and was used with the 3D structure of the template to model the dimeric structure of SamAA using Modeller. This model has a normalised DOPE Score value of 0.11.

The Swiss Model server led to seven models of SamAA 3D structure. Each model based either on one template previously identified by the server (hCN2, HiDapE, LdPepV, mSaPepV, VaPepD) or on an imposed template (monomeric and homodimeric NmDapE). The models based on hCN2 and HiDapE structures were ranked as the best according to the GMQE and QMEAN assessments. This result was confirmed by the normalized DOPE Score (values of – 0.47 and – 0.26 for the models based on hCN2 and HiDapE structures, respectively). These models were in the form of homodimers.

ColabFold is a web interface using the AlphaFold2 source code coupled with MMseqs2 for the alignment of target and template sequences. Regarding the results obtained with Swiss Model, an homodimeric sequence was used as input in the simulation. Five homodimeric models with very close normalized DOPE Score values (around – 1.50) were obtained. The best model was identified basing on lDDT and PAE scores.

Four models were therefore selected: one from Modeller, two from Swiss Model and one from ColabFold. All of these were analyzed based on the position of key residues. More specifically, SamAA appears as an enzyme belonging to the M20 peptidase family in which an acid-base residue and the metal ion-fixing residues have an essential role in the catalytic mechanism (Nocek et al., 2018). Sequence alignment analysis allowed to identify these conserved residues as: His88, Asp120, Glu154, Glu155, Glu182 and His416. A distorted trigonal or square bipyramidal coordination geometry was observed for the metal ion binding site of the templates (Bienvenue et al., 2003; Davis et al., 2006). The respect for such an organisation was therefore used to refine the choice of the best homology model by overlaying the models and the structure of the templates HiDapE and hCN2. These results were detailed in the following paragraph.

3.2.3.4. Structural evaluation of the models

According to the sequence alignment between HiDapE and SamAA, five of the six residues forming the active site are conserved. Only the residue Glu163/HiDapE is shifted by one residue from the residue Glu182/SamAA leading to a rotation of almost 90 ° of the acid terminal function of the glutamate in the model obtained with Modeller. The other six conserved residues were well superimposed with RMSD values between residues lower than 2 Å (Figure III.3 A and B).

The six residues of interest are conserved in the alignment between the template hCN2 and SamAA generated by Swiss Model, but only five of them were superimposed. The residue pairs Asp195/hCN2 and Glu182/SamAA, His445/hCN2 and His416/SamAA were not superimposed, leading to a RMSD value greater than 2 Å (Figure III.3 C and D). In the model ranked second by Swiss Model based on HiDapE template, the residue Glu163/HiDapE is shifted by one residue relative to the residue Glu182/SamAA leading to the rotation of the terminal acid function of the residue and a RMSD value around 5.7 Å (Figure III.3 E and F).

MMseq2 performed more than 13,188 alignments between the SamAA sequence and fragmentary or complete sequences from the PDB70 and UniRef databases, allowing the identification of numerous potential templates. The groups of N-acyl-aliphatic-L-amino-acid amidohydrolases and succinyl-diaminopimelate desuccinylases, which include the templates mhACY1, HiDapE, NmDapE and SfDapE, were found in this list, and that is why the pre-selected model from ColabFold was superimposed on the HiDapE structure. The six residues of interest were well superimposed. The highest RMSD value was observed for the residues Glu163/HiDapE and Glu182/SamAA (around 2.3 Å) (Figure III.3 G and H).

To conclude, the model ranked first by Colabfold respects the most the architecture of the templates and was therefore retained as a target for further docking simulations.

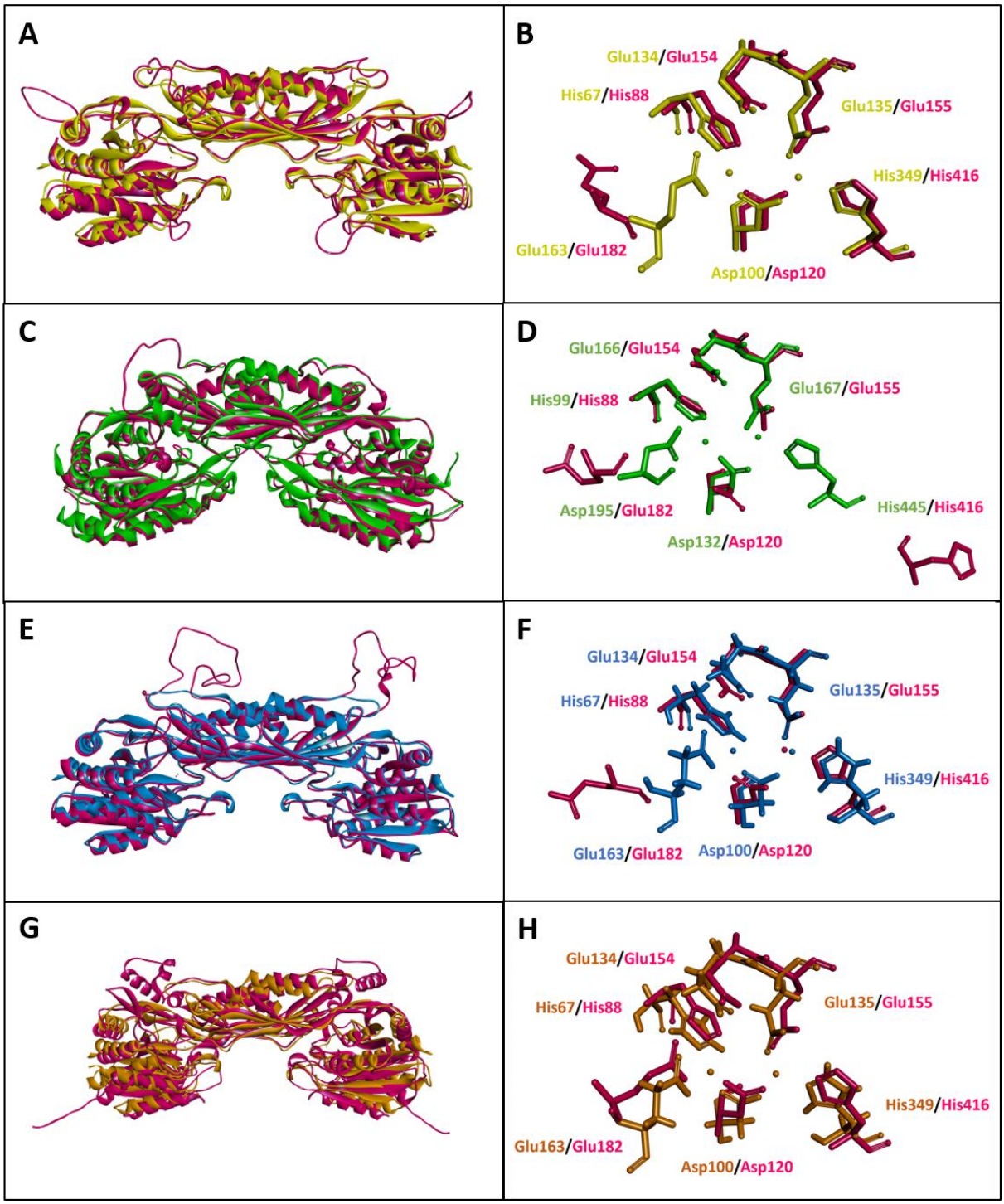


Figure III.3: A and B: molecular overlay of the best model obtained with Modeller (in pink) and HiDapE structure (in yellow), B: focus on the metal ion binding sites, C and D: molecular overlay of the best model obtained with Swiss Model from hCN2 as a template (in pink) and hCN2 structure (in green), D: focus on the metal ion binding sites, E and F: molecular overlay of the best model obtained with Swiss Model from HiDapE as a template (in pink) and HiDapE structure (in blue), F: focus on the metal ion binding sites, G and H: molecular overlay of the best model obtained with ColabFold (in pink) and HiDapE structure (in orange), H: focus on the metal ion binding sites.

3.2.3.5. Description of SamAA predicted 3D structure

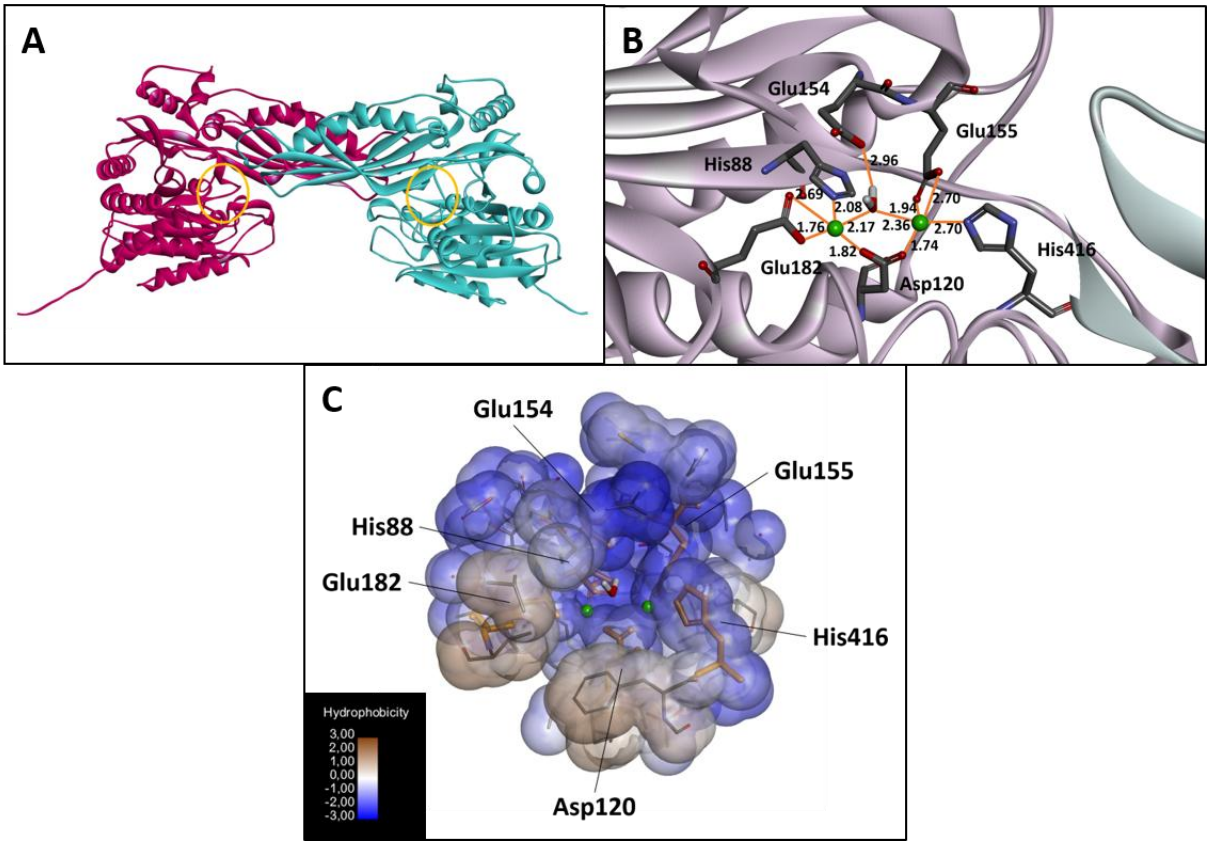


Figure III.4: A: Predicted 3D structure of SamAA in its dimeric form. The orange circles indicate the active site of each monomer; B: Predicted architecture of SamAA active site with the catalytic water molecule, the catalytic residue, the metal ion-binding residues and the two Zn^{2+} ions (in green); C: Van der Walls-type surface of all on atoms present in a 10 Å sphere around the catalytic water molecule showing the catalytic cavity of the monomer. The surface is coloured according to hydrophobicity.

The final model of SamAA is formed by two identical monomers, each 441 residues long. Each monomer has a "whistle" shape composed of two distinct elements: a first spherical part formed by the residues 1-202 and 359-441, and a second cylindrical part formed by the residues 205-352 (Figure III.4A). The hinge zone between these two elements ranging from residues 203-205 to 353-358 is partially conserved. The two N-terminal ends of each monomer are modeled as structureless fragments of about ten residues each, being a consequence of the use of AlphaFold2 to predict the 3D structure (Jumper et al., 2021). SamAA seems to have a neutral to very hydrophilic surface, with the exception of the dimerization zone which is slightly hydrophobic. The core of the catalytic and dimeric domain is neutral to slightly hydrophobic.

The dimerization domain is composed of a first layer formed by a β-sheet with four strands (β10, β12, β14 and β15) in antiparallel orientation and a second layer formed by two α-helices (α5 and α9). The β15 strand is extended by the small β16 strand at the hinge zone. This

organization in a two-layer α+β sandwich fold is characteristic of the M20 peptidase family (Rowsell et al., 1997 ; Nocek et al., 2010). The dimerization domain extends from residue 204 to residue 353 ; it is interrupted by three small fragments (residues 215-222, 244-289, 302-309) from the catalytic domain. The main element of the catalytic domain is the β-sheet consisting of nine strands, the first three of which (β1, β2 and β6) are antiparallel. The next two (β3 and β7) are parallel, then a double succession of two antiparallel strands (β8, β9, β17 and β18) is observed. This β-sheet is surrounded by 14 helices α, on one side the helices α1, α2, α3, α10, α13, on the other side the helices α4, α6, α7, α8, α11, α12. A second small β-sheet located on the surface of the catalytic domain is made up of two antiparallel strands (β5 and β19). Finally a third β-sheet consisting of two small anti-parallel strands (β11 and β13) is found at the end of the dimerization domain (Figure III.5).

The dimerization interface involves the dimerization domain and the catalytic domain, next to the α5 helix and the β12 and β14 strands. Numerous π-alkyl and hydrogen bonds interactions but no disulfide bridges are observed. There are also interactions between the catalytic site of one monomer and the two final loops of the dimerization domain of the second monomer. More specifically, hydrogen bond interactions are observed between the residues Lys304.A-Glu155.B, Asn306.A-His416.B and Asn306.A-Glu155.B (and *vice versa*). The residues His219.A-His416.B and Val307.A-His416.B (and *vice versa*) interacts throught π-π and π-alkyl interactions, respectively.

There is one active site per monomer, which is located in the center of the catalytic domain and above the hinge zone. It is delimited by the end of strands β3, β6 and β7 and surrounded by α-helices. The catalytic cavity of SamAA is a slit-shaped cavity perpendicular to the metal ion axis. It is divided into two distinct zones: a hydrophilic part extending from the upper part of the cavity (between α4 and α11 helices) to the two metal ions and a hydrophobic part extending from the metal ions to the lower part of the cavity (Figure III.4C).

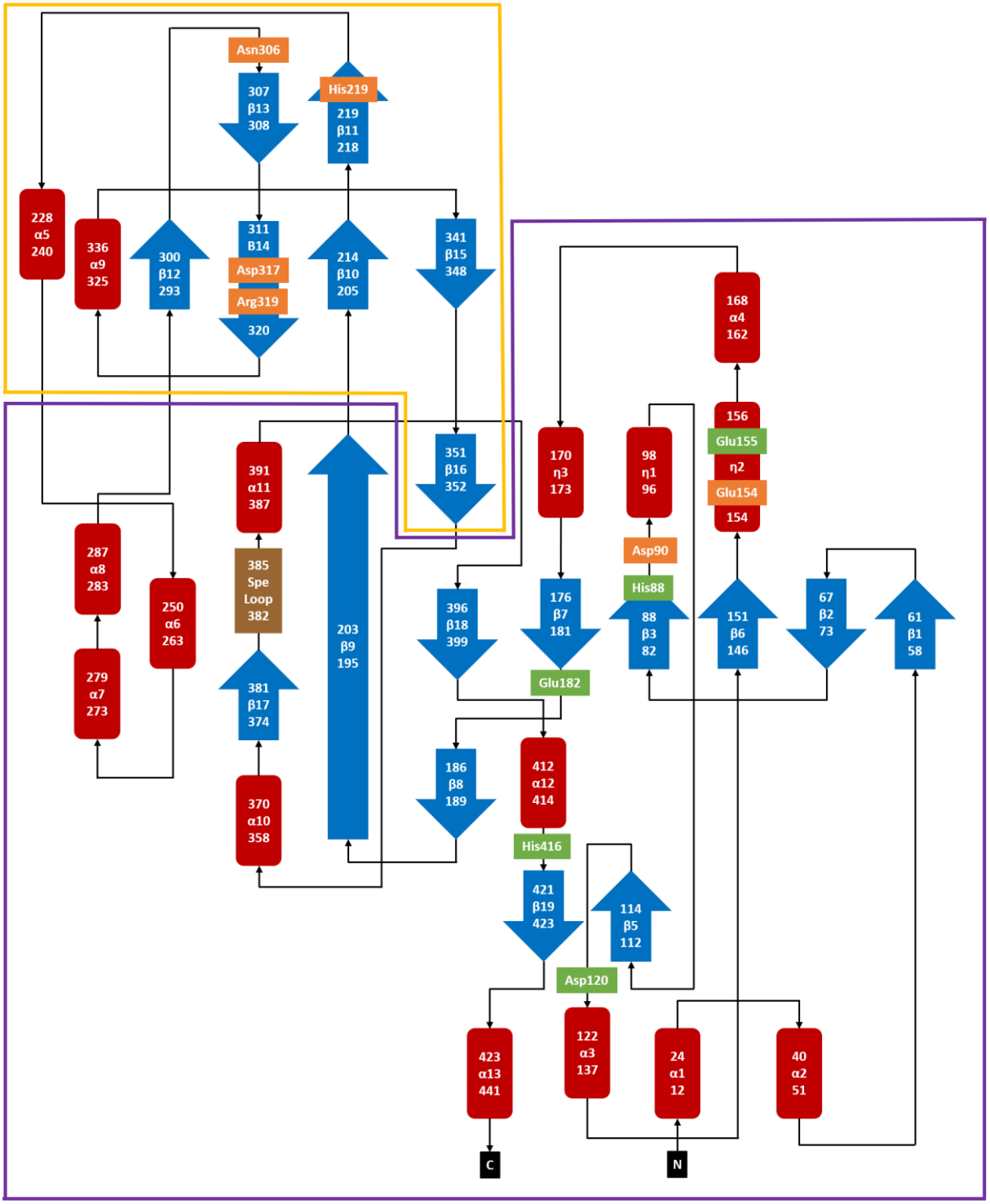


Figure III.5: Graphical representation of the secondary structure of SamAA. The purple and yellow frames delimit the catalytic domain and the dimerisation domain, respectively. The blue arrows represent the sheets β and the red cylinders correspond to helices α and η. Conserved residues are shown in orange (catalytic role) and green (binding of metal ions). The brown rectangle indicates the conserved specificity loop, abbreviated as Spe Loop.

The amino acids forming the active site are located at the extremities of the secondary structures. His88 and Asp120 are located at the end of strand β3 and before the beginning of helix α3 respectively. Glu154 and Glu155 are located at the beginning of helix 3_{10} n°2, Glu182

and His416 are positioned at the end of strand β7 and at the end of helix α12, respectively. Two Zn^{2+} ions and a catalytic water molecule are at the center of the active site. These three elements were placed manually in such a way as to respect the coordination geometry observed in HiDapE structure. The two hydrogen atoms of the water molecule are oriented towards the end of the Glu154 residue at an average distance of 2.96 (+/- 0.55) Å forming hydrogen bonds (Nocek et al., 2010; Nocek et al., 2018). The active site is closed in part by the loops present at the end of the dimerization domain of the second chain.

As previously noted, the two metal ions are in a distorted trigonal bipyramidal coordination geometry where the axial positions correspond to the atoms Asp120(OD2), Glu182(OE1) for Zn1 and the atoms Asp120(OD1) and Glu155(OE2) for Zn2. The equatorial positions are occupied by the atoms His88(NE2), Glu182(OE2) and the catalytic water molecule for Zn1 and the atoms Glu155(OE1), His416(NE2) and the catalytic water molecule for Zn2. The coordination covalent bond of templates has a precise length of 2 Å. The distances between the metal fixing residues and the ions are 2.10 (+/- 0.37) Å for the first metal center and 2.29 (+/- 0.44) Å for the second metal center (Figure III.4B). The two ions are separated by a distance of 3.98 (+/- 0.08) Å. The angle values of the two metal centers are summarized in the table III.3.

SamAA predicted structure as well as the templates HiDapE, NmDapE, SfDapE and hCN2 are characterized by a whistle-shaped monomer that is a structural feature of the M20 peptidase family (Pandya et al., 2014; Starus et al., 2015; Nocek et al., 2018; Osipiuk et al., 2021). It is also found in the following enzymes: the N-α-acylglutamine aminoacylase from *Corynebacterium sp.* (EC 3.5.1.133; PDB entry 6SLF), the acetylornithine deacetylase from *Rhodopseudomonas palustris* (EC 3.5.1. 16; PDB entry 3PFO); the auxin-conjugated amidohydrolase from *Arabidopsis thaliana* (EC 3.5.1.-; PDB entry 1XMB) and the N-acetylcysteine deacetylase from *Bacillus subtilis* (EC 3.5.1.-; PDB entry 1YSJ) (Minasov et al., 2005; Bitto et al., 2008; Joint Center for Structural Genomics 2010; Natsch et al., 2020). Other enzymes showing lower sequences similarity share this structural feature, such as the metallodipeptidase DUG1 from *Saccharomyces cerevisiae* (EC 3.4.13. - PDB entry 4G1P), human carnosine dipeptidase 1 (EC 3.4.13.20; PDB entry 3DLJ), aminoacyl-histidine dipeptidase from *Vibrio alginolyticus* (VaPepD) (EC 3. 4.13.3; PDB entry 3MRU) or the dipeptidase PepV from *Lactobacillus delbrueckii subsp. lactis* (LdPEPV) (EC 3.4.13.-; PDB entry 1LFW) (Jozic et al., 2002; Dong et al., 2008; Chang et al., 2010; Singh et al., 2014).

Table III.3: Angle measurements around each metal ion after energy minimization. The equatorial and axial angles of the coordination geometry are highlighted in green and orange respectively.

	Description	Angles (°)	Standard deviation (°)
Site Zn1	Glu182(OE2)-Zn1-His88(NE2)	121.39	5.66
	H_2O(O)-Zn1-His88(NE2)	79.47	12.27
	H_2O(O)-Zn1-Glu182(OE2)	127.84	4.23
	Glu182(OE1)-Zn1-Glu182(OE2)	52.28	2.02
	Glu182(OE2)-Zn1-Asp120(OD2)	120.56	3.47
	His88(NE2)-Zn1-Asp120(OD2)	116.29	7.01
	Glu182(OE1)-Zn1-His88(NE2)	75.28	0.05
	H_2O(O)-Zn1-Asp120(OD2)	68.32	1.94
	H_2O(O)-Zn1-Glu182(OE1)	99.59	8.53
Site Zn2	Glu155(OE1)-Zn2-His416(NE2)	109.01	11.31
	H_2O(O)-Zn2-His416(NE2)	139.39	4.72
	H_2O(O)-Zn2-Glu155(OE1)	86.72	2.93
	Glu155(OE1)-Zn2-Asp120(OD1)	135.65	8.95
	His416(NE2)-Zn2-Asp120(OD1)	103.50	7.81
	Glu155(OE2)-Zn2-Glu155(OE1)	51.44	3.40
	Glu155(OE2)-Zn2-His416(NE2)	69.79	5.90
	H_2O(O)-Zn2-Asp120(OD1)	85.93	13.76
	H_2O(O)-Zn2-Glu155(OE2)	95.51	2.09

All the templates and the eight above mentionned enzymes have two divalent ions in each of their active sites. These metal ions are Zn^{2+}, Co^{2+}, Mn^{2+} or Ni^{2+}. Regarding the metal ion binding site, SamAA and its templates have two histidines, two glutamates and one aspartate, with the exception of hCN2 where a glutamate is replaced by an aspartate, which modifies the substrate specificity. Similar organisation was observed for acetylornithine deacetylase from *R. palustris* whereas few differences were found in other enzymes. The aspartate binding the metal ions is replaced by a cysteine in aminoacylases from *Corynebacterium*, *A. thaliana* and *B. subtilis*. Human carnosine dipeptidase 1 has an ion-binding site similar to that of hCN2. The ions of DUG1 metallodipeptidase from *S. cerevisiae* and aminoacyl-histidine dipeptidase from *V. alginolyticus* (VaPepD) are partially bound to histidines, glutamates or aspartates. The

dipeptidase PepV from *L. delbrueckii subsp. lactis* (LdPepV) has an inhibitor linked to metal ions, thus modifying the interactions between ions and residues.

3.2.3.6. Docking results

Molecular docking simulations were performed on the best model of SamAA as a target to explain the selectivity of the enzyme observed experimentally (Dettori et al., 2018a; Bourkaib et al., 2020a). The superposition of the two monomers indicated that there is no structural difference between them (RMSD value on main chains: 0.04 Å). Docking results presented below were obtained with SamAA chain A. The model including one water molecule was used to mimic the binding of the acyl-aminoacids to be hydrolyzed whereas the model without water was used to mimic the binding of the substrates during the reverse reaction. More specifically N-α-acetyl-L-lysine, N-ε-acetyl-L-lysine, N-α-caproyl-L-lysine and N-ε-caproyl-L-lysine were docked within SamAA catalytic site in order to compare their binding modes during hydrolysis. L-Lysine, acetic acid and caproic acid were docked within SamAA cavity after the removing of the catalytic water molecule to mimic the synthesis of N-acyl-L-lysine derivatives. The catalytic water molecule was moved manually following the structure of the HiDapE to the arrival of the substrates: it is then located approximately 2.0 Å from the Glu154.A residue and 3.8 Å from Zn2 (Nocek et al., 2018). These substrates were chosen in one hand to explain the selectivity of SamAA towards α-acylated substrates and in the other hand to study the effect of the acyl donor substrate on interaction modes.

Few discriminating criteria were used to analyze docking results and conclude about the feasibility and the selectivity of the reactions under study: the orientation of the substrates regarding the hydrophilic/hydrophobic nature of SamAA, the position and the proximity of the ligand functional groups to the catalytic residues and the angle involved in the nucleophilic attack of either the amide bond by the water molecule in hydrolysis reaction or the acyl donor by the acyl acceptor in synthesis reaction.

These criteria were established based on the 3D structure of HiDapE that was elucidated in the presence of the two products from N-succinyl-(2S,6S)-2,6-diaminoheptanedioate hydrolysis. This crystal provides information on the positioning of the acyl donor and the acyl acceptor, which are opposite the two metal ions and the Glu134.A residue (Nocek et al., 2018). Particular distances were measured that were used as standards to analyze docking results. These include the distances between the Zn2 and the amide oxygen atom (from 2.0 to 2.5 Å), the distance between the catalytic water oxygen atom and the amide carbon atom (close to 4 Å), and the

distance between the amide hydrogen atom and Glu134.A oxygen atom (from 1.0 to 4.7 Å) (Dutta et al., 2016; Nocek et al., 2018). Furthermore the Bürgi-Dunitz angle for nucleophilic attack has also to be considered. It is admitted that this angle between the nucleophile and the carbonyl carbon atom to be attacked for the optimal achievement of a nucleophilic reaction between small molecules is approximately 107 ° (Bürgi et al., 1972). Other studies performed on larger molecules like enzymes showed that the value of this angle would be 90 ° because of electrostatic and Van der Waals interactions and steric hindrance that modifies the trajectory of the nucleophile (Lodge et al., 1986; Radisky et al., 2002; Lubkowski et al., 2019).

3.2.3.6.1. Hydrolysis reaction

3.2.3.6.1.1. Docking of N-α-acetyl-L-lysine

Docking results showed that the substrate N-α-acetyl-L-lysine took two preferred orientations within SamAA catalytic pocket, corresponding to two groups of poses.

In the first group of 23 poses corresponding to the top-ranked poses, the acetyl and lysyl moieties fit into the hydrophobic and the hydrophilic parts of the cavity, respectively. The amide bond comes near the catalytic water molecule (3.2 Å between the carbon atom to be attacked and the water molecule oxygen atom; 2.2 Å between the amide oxygen atom and the Zn2 ion) suggesting that the nucleophilic attack could occur. The value of the corresponding Bürgi-Dunitz angle is 82 °. The atoms forming the amide function interact with the residues Glu154.A and Gly384.A whereas the ε-amine function interacts with the residues Gly157.A, His346.A and Asp386.A. The distance between the hydrogen atom carried by the amide nitrogen and the terminal oxygen atom of Glu154.A is equal to 2.4 Å. The carboxyl group forms an ionic bond with the residue Arg319.A and hydrogen bonds interactions with residues Gly384.A, His416.A, His219.B and Asn306.B (Figure III.6A).

In the second group of 27 poses, the amide bond is around 6 Å from the catalytic water molecule and is therefore too far away for nucleophilic attack. As before the acetyl moiety and the lysyl moiety fit into the hydrophobic and the hydrophilic parts of the cavity, respectively. The α- and the ε-amine functions point towards the ions zinc and the residues Glu154.A and Glu155.A, respectively. The carboxyl group interacts both with the two metal ions and the catalytic water molecule through hydrogen bonding. The ε-amine function is stabilised by ionic and hydrogen bonds interactions with residues Glu154.A and Asp386.A.

3.2.3.6.1.2. Docking of N-ε-acetyl-L-lysine

Docking simulations with N-ε-acetyl-L-lysine were performed in order to compare its binding modes within SamAA catalytic pocket, knowing that its hydrolysis has never been observed experimentally. Three preferential orientations were obtained most of them being far from the catalytic water molecule.

In the first one corresponding to 22 poses, the amide bond comes approximatively at a distance of 6.5 Å from the catalytic water molecule, that is too far to consider the nucleophilic attack. The acyl moiety is placed in the hydrophilic part of the catalytic cavity, the carboxyl group interacting with the zinc ions and the catalytic water molecule. The lysyl moiety of the substrate is stabilised through few interactions with residues Asp120.A, Glu154.A, Glu182.A, Gly384.A. and Trp208.A.

In the second orientation corresponding to the best ranked poses (16 poses of 50) the carboxyl group points towards the cavity entrance, the amide bond being far from the catalytic machinery (more than 7 Å). The substrate is positioned in such a way that the α-amine function interacts with residues Glu154.A and Glu155.A whereas the carboxyle group interacts with residues Gly384.A, Asn306.B and Arg319.A throught salt bridges and hydrogen bond interactions. Alkyl or π-alkyl hydrophobic interactions are also observed between the lysyl moiety and residues Val183.A, His219.B and His416.A.

Finally, the third binding mode (12 poses of 50) corresponding to poses with poor scores looks similar to one orientation observed with the N-α-acetyl-L-lysine, where the acetyl moiety comes near the catalytic water molecule and the metal ions, the lysyl moiety being located in the hydrophilic part of the cavity. The amide bond interacts with the catalytic water molecule and residue Gly384.A that also interacts with the carboxyl group of the substrate. The free α-amine function is stabilised by ionic interaction with residues Glu154.A and Asp386.A. The distance between the amide oxygen atom and the Zn2 ion is 2.2 Å. The hydrogen atom of the amide bond is at an average distance of 2.5 Å from the terminal Glu154.A oxygen atoms, but no interaction was observed between them. The average distance between the amide bond and the catalytic water molecule allows to consider the nucleophilic attack (3.1 Å) however an average angle of attack of 68 ° was found that is far from the 90 ° value identified in Lubkowski's study (Lubkowski et al., 2019). All of the above would explain the lack of catalysis encountered with N-ε-acetyl-L-lysine (Figure III.6B).

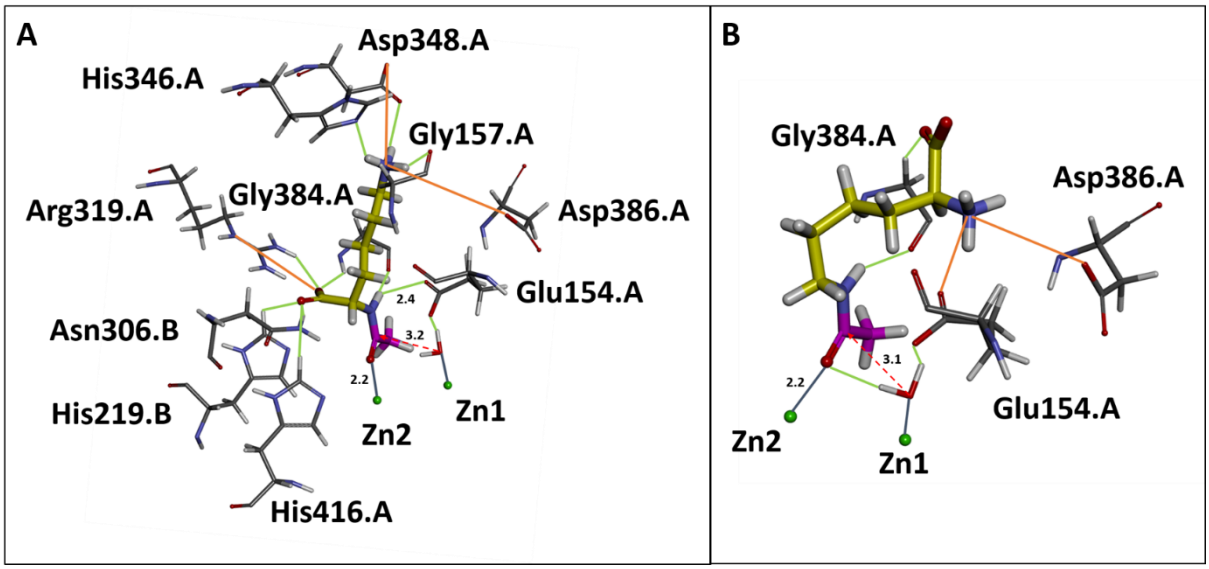


Figure III.6: Preferential orientations of N-α-acetyl-L-lysine (A) and N-ε-acetyl-L-lysine (B) within SamAA catalytic pocket. For clarity, only residues interacting with the substrates are shown. The two metal ions are in green. The lysyl and the acetyl moieties of the substrates are in yellow and pink, respectively. Hydrogen bonds, ionic interactions and dative bonds are shown in green, orange and grey solid lines. The dotted red arrow mimics the nucleophilic attack of the amide bond by the water molecule.

3.2.3.6.1.3. Docking of N-α-caproyl-L-lysine

In contrast to the proven specificity of SamAA towards N-α-acetyl-L-lysine, the hydrolysis of N-α-caproyl-L-lysine has never been demonstrated experimentally. Just as for N-α-acetyl-L-lysine, two preferential orientations of N-α-caproyl-L-lysine were observed within SamAA catalytic pocket.

A first group of 44 poses corresponding to the best-ranked poses was observed. In this orientation, the amide bond comes near the catalytic water molecule (3.2 Å) with an average angle for nucleophilic attack of 83 °. The distance between the amide function and the Glu154.A terminal ends is 2.4 Å. The orientation of the acylated derivative suggests the possibility of a nucleophilic attack on the amide bond. The substrate is stabilised through electrostatic interactions between the amide carbonyl group and the Zn2 ion (distance of 2.2 Å) and hydrophobic alkyl and π-alkyl interactions between the caproyl chain and residues Val183.A, His219.B, Leu381.A and Phe415.A. The amide function interacts with residues Glu154.A, Asn306.B and Gly384.A whereas the carboxyl group is stabilised by residues His416.A and His219.B through hydrogen interactions and residue Arg319.A through ionic interaction. The terminal ε-amine function interacts with residues Gly157.A, His346.A, Asp348.A and Asp386.A through hydrogen bond and ionic interactions respectively (Figure III.7A).

In the second group of 6 poses, the distances between the amide bond and the Zn2 ion and the catalytic water molecule are comprised between 4 and 5 Å. The angle for nucleophile attack is greater than 150 °, making these poses unlikely to react. In this orientation, the substrate is shifted by about 1.5 Å comparing to the first orientation. Above mentionned hydrophobic interactions stabilising the caproyl chain as well as hydrogen bond and ionic interactions stabilising the amine functions are conserved.

3.2.3.6.1.4. Docking of N-ε-caproyl-L-lysine

In the case of N-ε-caproyl-L-lysine, two main orientation were also observed among docking results.

A first group of 11 well ranked poses was obtained where the substrate remains far from the metal ions and the catalytic water molecule (more than 6.5 Å) that is incompatible with the supposed hydrolysis mechanism. The acyl derivative is mainly located in the hydrophilic part of the catalytic cavity, the caproyl moiety being above the residues involved in metal ion binding and catalysis. The amide bond is opposite residue Glu154.A, sharing a hydrogen bond interaction with it. The carbon chain of the lysyl moiety is balanced by hydrophobic interactions with residues His219.B and His416.A. The free carboxyl group interacts with metal ions and the catalytic water molecule whereas the free α-amine function is stabilised by hydrogen bond and ionic interactions with residues Asp120.A, Glu154.A, Glu182.A and Gly384.A.

The second main orientation of the substrate (39 poses) is similar to that observed with N-α-caproyl-L-lysine. Although the amide bond comes near the catalytic water molecule (distance of 3.4 Å) the 62 ° angle between the electrophilic center and the nucleophile does not allow to consider the nucleophilic attack. In such an orientation, the caproyl chain interacts with the hydrophobic part of the cavity, especially with residues Val183.A, Leu381.A, Phe415.A and His219.B. Residue Gly384.A stabilises the amide bond of the ligand near the water molecule and the Zn2 ion at a distance of 2.2 Å. The carboxyl group and the free α-amine function are stabilised by ionic and/or hydrogen bond interactions with residue Lys388.A and residues Glu154.A and Asp386.A respectively (Figure III.7B).

In view of these results, hydrolysis of N-ε-caproyl-L-lysine seems unlikely to occur which correlate to experimental results obtained with various acyl-chain lengths.

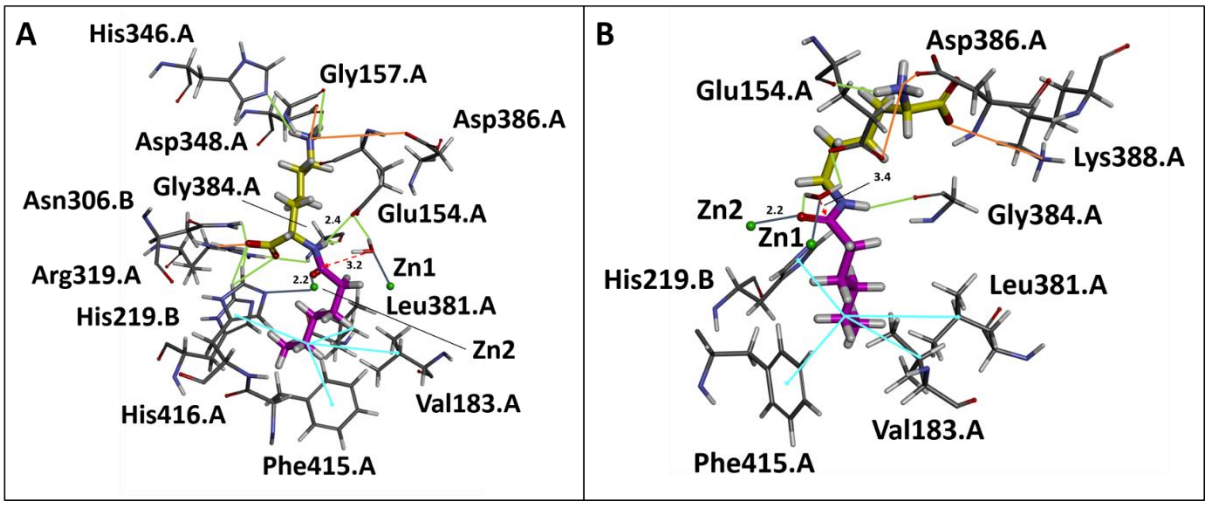


Figure III.7: Some poses from docking of N-α-caproyl-L-lysine (A) and N-ε-caproyl-L-lysine (B) where the amide bond to be hydrolyzed comes near SamAA catalytic machinery. For clarity, only residues interacting with the substrate are shown. The two metal ions are in green. The lysyl and the acetyl moieties are in yellow and pink respectively. Hydrogen bonds, ionic interactions and dative bonds are shown in green, orange and grey solid lines respectively. The dotted red arrow mimics the nucleophilic attack of the amide bond by the water molecule. The black solid lines indicate the name of a residue or a distance value.

3.2.3.6.2. Synthesis reaction

The specificity of SamAA favouring the synthesis of N-α-acyl-L-lysine derivatives has already been demonstrated experimentally (Bourkaib et al., 2020a). In that case the model from which the catalytic water molecule was removed was used for docking simulations, aiming to elucidate substrates binding modes and therefore the selectivity of the enzyme towards lysine α- or ε-amino group. Just as before the effect of the acyl chain length was addressed by comparing acetate and caproate binding modes.

3.2.3.6.2.1. Docking of acyl donor in first

The complexes issued from the docking of acetate showed two main orientations. In the first one that corresponds to 28 well ranked poses, acetate interacts with the hydrophobic part of the cavity. Interactions between the two metal ions and carboxylate are observed, with an average distance of 2.5 Å. The carbon chain is oriented towards the outside of the catalytic cavity (Figure III.8A). In the second group of 22 poses, the acetyl chain points towards the hydrophilic part of the cavity near residues Glu155.A and Glu154.A, the carboxylate interacting with the two metal ions (average distance of 3 Å).

In the case of caproic acid, three main orientations were observed. In the first most realistic one (8 top-ranked poses), the acyl donor is located in the hydrophobic part of the catalytic cavity, the carboxylate group interacting with the two metal ions at a distance of 2.2 Å. Alkyl and π-alkyl hydrophobic interactions between the acyl chain and residues Leu381.A, Phe415.A, His219.B and Met222.B are observed (Figure III.8B). In the second orientation (37 poses), the acyl donor is located in the hydrophilic part of the catalytic cavity, that should rather accommodate the acyl acceptor. Finally in the third group of 5 poses, the caproyl chain faces the metal ions and their fixing residues and the carboxylate group points towards the cavity entrance, at a distance of more than 5 Å from metal ions.

The top-ranked complexes issued from acetate and caproate docking simulations were chosen as targets for further docking of L-lysine.

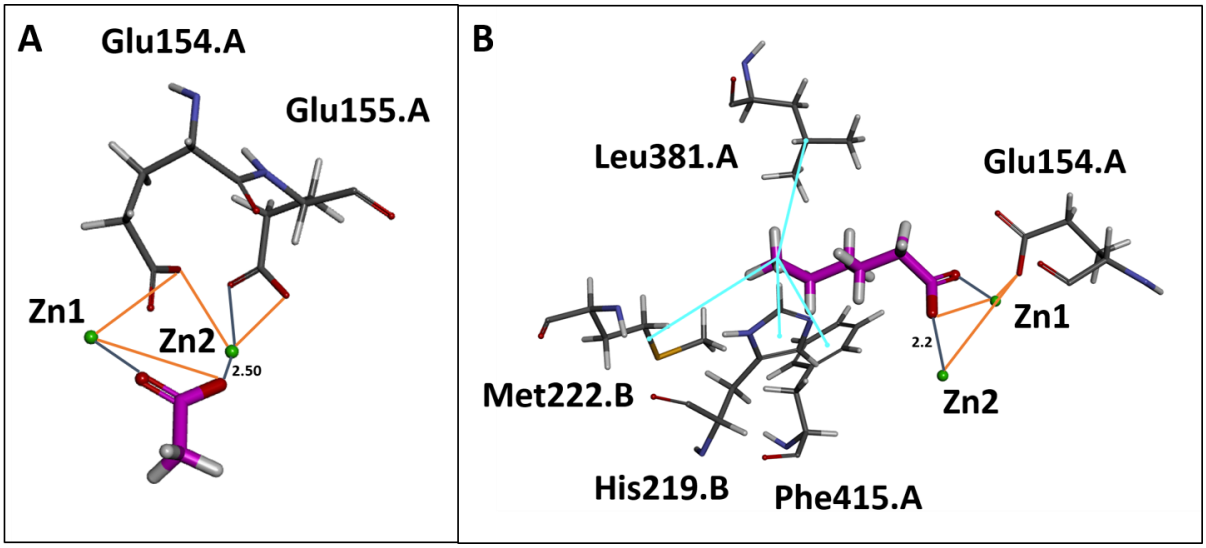


Figure III.8: Main orientations and interactions of acyl donors acetate (A) and caproate (B) within SamAA active site. For clarity, only residues interacting with acyl donors are shown. The two metal ions and the acyl donors are in green and in pink respectively. Ionic interactions, hydrophobic interactions and dative bonds are shown in orange, blue and grey solid lines.

3.2.3.6.2.2. Docking of L-lysine within SamAA in complex with acyl donor substrate

Whatever the acyl donor used, a single preferential binding mode was obtained where L-lysine location is highly similar to that observed in hydrolysis reaction. The α-amine function of L-lysine comes close to the acid-base residue Glu154.A at a distance around 2 Å and interacts with the acyl donor carboxylate group and to residues Glu155.A and Gly384.A. The distance between the nucleophilic nitrogen atom and the acyl donor carboxylate is around 3 Å. The angle

for nucleophilic attack is 100 ° that is consistent with the criteria previously defined by Lubkowski and coworkers (Lubkowski et al., 2019). L-lysine is located in the hydrophilic part of the cavity, just above the acyl donor substrate and close to residues Glu154.A and Glu155.A. The ε-amine function interacts with residues Gly157.A, Asp348.A and Asp386.A through ionic and hydrogen bond interactions whereas the carboxylate group is stabilised by residues Arg319. A, Gly384.A, His219.B and Asn306.B (Figure III.9).

Regardless of the acyl donor, all lysine poses were obtained with the α-amino group oriented towards the catalytic residues whereas the ε-amino group pointed towards the bottom of the hydrophilic cavity.

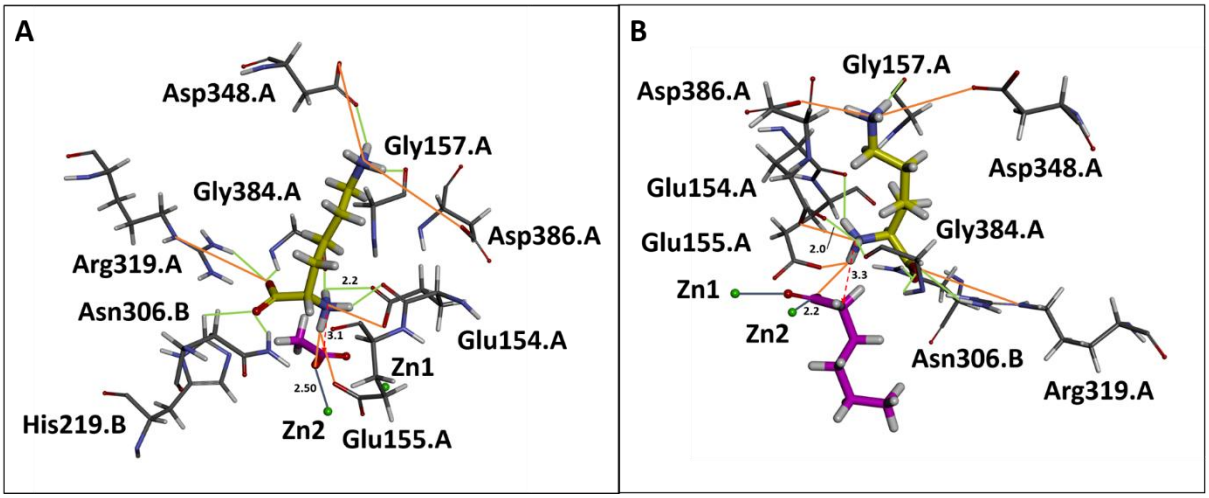


Figure III.9: Main orientations and interactions of L-lysine within SamAA active site in complex with acyl donor substrates acetate (A) or caproate (B). For clarity, only residues interacting with L-lysine are shown. The two metal ions, the acyl donors and L-lysine are in green, pink and yellow respectively. Hydrogen bond, ionic, hydrophobic interactions and dative bonds are shown in green, orange, blue and grey solid lines. The dotted red arrow mimics the nucleophilic attack of the acyl donor by L-lysine α-amine function. The solid black line indicates the distance length.

3.2.3.6.2.3. Docking of L-lysine in first

New docking simulations were performed by reversing the order of arrival of the two substrates to verify its impact on the regioselectivity of the reaction. The docking of L-lysine within SamAA catalytic cavity led to four main binding modes.

In the first group of 21 top-ranked poses, the L-lysine is located in the upper hydrophilic part of the catalytic cavity. The α-end of L-lysine is near the residues Glu154.A, Glu155.A and Gly384.A, forming ionic and hydrogen bonds. The α-amine function is 2 Å from the terminal end of the catalytic residue Gu154.A, which is favourable for its acylation. The carboxylate

function is stabilised by ionic bond with Arg319.A and hydrogen bond interactions with Arg319.A, Gly384.A and Asn306.B. The ε-amine function forms ionic and hydrogen interactions with Gly157.A, Asp348.A and Asp386.A (Figure III.10).

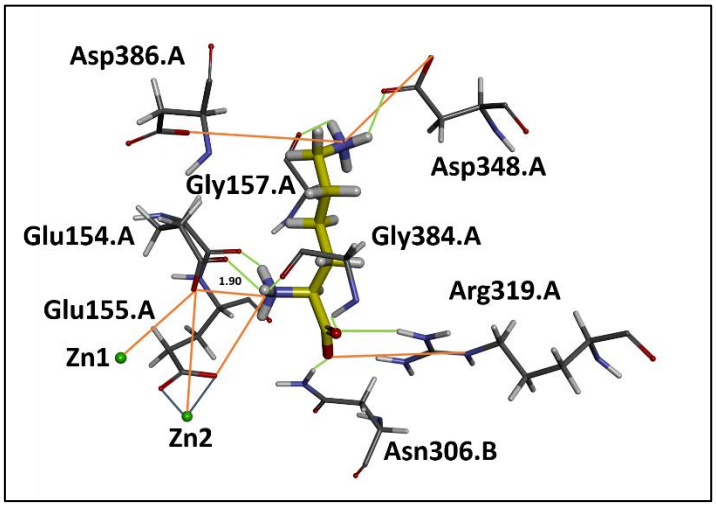


Figure III.10: L-lysine in interaction with SamAA active site. For clarity, only interacting residues are shown. The two metal ions and the L-lysine are in green and yellow respectively. Hydrogen bonds, ionic interactions and dative bonds are shown in green, orange and grey solid lines.

A second group of 5 poses showed the L-lysine carboxyl group pointing towards the two metal ions blocking the accessibility of the two amine functions to an acyl donor.

In the third group of 13 poses, the two amine functions point towards residues Glu154.A and Glu384.A, with the carboxylate function creating steric hindrance in the catalytic cavity and blocking the accessibility of the two amine functions and Zn2 to an acyl donor.

In the fourth group of 11 poses the L-lysine α-amine function comes distant the catalytic acid-base residue Glu154.A (distance around 5 Å) which seems incompatible with a proton transfert. Moreover, L-lysine is just in front of residues Glu154.A, Glu155.A and the metal ions, hindering the acyl-donor arrival.

The top-ranked SamAA-L-lysine complex was selected as a target for further docking of acyl donor substrates and shown in figure III.10.

3.2.3.6.2.4. Docking of acyl donor substrate within SamAA catalytic pocket in complex with L-lysine

Acyl donor substrates were docked into the catalytic cavity of SamAA, which already contains L-lysine to verify whether nucleophilic attack could occur.

Two preferential binding modes were observed when docking acetate. In the first one that corresponds to the 33 poses with highest score, acetic acid carboxylate function interacts with the two metal ions and the L-lysine α-amine function at a distance of 3.8 Å and forming an angle of 100 ° with it. A distance of 2.3 Å is measured between the Zn2 ion and acetic acid carboxylate function (Figure III.11A). In the second preferential binding mode (17 poses), the formed Bürgi-Dunitz angle seemed incompatible with nucleophilic attack.

In the case of caproic acid, a single preferential orientation was obtained where the acyl donor interacts with the hydrophobic area of the active site, especially the residues Leu381.A, Phe415.A, His219.B and Met222.B. The carboxylate function forms ionic bonds with L-lysine and the two metal ions, at a distance of 2.3 Å from Zn2 ion. The L-lysine α-amine function is 3.9 Å from caproic acid electrophilic center and forms an angle of 95 ° that allows considering the nucleophilic attack and the subsequent acylation of L-lysine on the α position (Figure III.11B).

To conclude, the order of arrival of the substrates seems to have no influence on neither the feasibility nor the selectivity of the N-acylation reaction catalysed by SamAA. Main orientations and interactions are almost the same whether L-lysine enters the cavity first or second and support the selective acylation od its α-amino group.

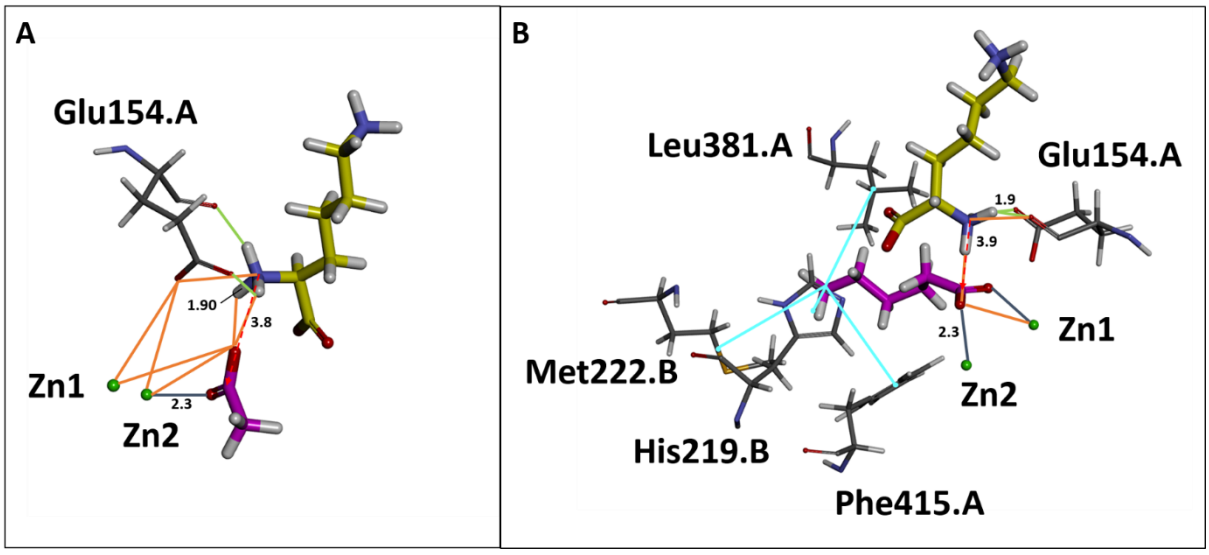


Figure III.11: Main orientations and interactions of the acyl donors acetate (A) and caproate (B) within SamAA active site in complex with L-lysine. For clarity, only residues interacting with acyl donor subtrate are shown. The two metal ions, the acyl donors and the L-lysine are in green, pink and yellow respectively. Hydrogen bonds, ionic, hydrophobic interactions and dative bonds are shown in green solid lines, orange, blue and grey solid lines. The dotted red arrow mimics the nucleophilic attack of the acyl donor by L-lysine α-amine function.

3.2.3.7. Proposed catalytic mechanism of N-α-caproyl-L-lysine hydrolysis or synthesis

Here is a proposal for N-α-caproyl-L-lysine hydrolysis and synthesis inspired from the mechanism elucidated for the HiDapE enzyme as well as docking results (Dutta et al., 2016; Nocek et al., 2018).

Hydrolytic mechanism begins with the arrival of N-α-caproyl-L-lysine in SamAA active site, where several residues stabilise the substrate. The ε-amine function of the lysyl fragment is stabilised by the non-conserved residues Gly157.A and His346.A and the conserved residue Asp386.A. The amide and the carboxyl functions of the substrate interact with the conserved residues Glu154.A, Gly384.A, and Arg319.A, Gly384.A, His416.A, His219.B, Asn306.B, respectively. Residue Gly384.A is supposed to belong to the SamAA specificity loop as identified in HiDapE and stabilises the lysyl α-extremity (Lindner et al., 2008; Nocek et al., 2014). Previous studies on hACY1 showed the role of residues Asp274.A, which is conserved as Asp317.A in SamAA sequence. This residue is involved in catalysis, without being in direct contact with the substrate, as its mutation leads to a decrease in the catalytic constant while the Michaelis constant remains unchanged (Lindner et al., 2005). In the case of SamAA, the terminal end of residue Asp317.A stabilises the positive charges of the residue Arg319.A that interacts with the substrate carboxyl group. The caproyl moiety of the substrate seems to be stabilised through hydrophobic interactions with the non-conserved residues Val183.A, Leu381.A, Phe415.A, and the conserved residues His416.A and His219.B. This latter residue corresponds to the residue His194.B in HiDapE structure which role in the formation of the oxyanion hole was suggested: residue His194.B establishes a hydrogen bond interaction with the oxygen atom of the succinyl fragment bound to Zn2, thus forming the oxyanion hole and activating the amide bond to be hydrolyzed in the active site A. Such an interaction is made possible by the change from an opened to a closed conformation of the enzyme caused by the arrival of the substrate, bringing the residue His194.B closer to the active site of the chain A. This residue formerly located 10.1 Å away from the Zn2 ion in the open conformation is 2.9 Å in the closed one (Nocek et al., 2018). In the case of SamAA, residue His219.B is 5.3 Å from the Zn2, suggesting that the modelled structure presents an intermediate state between the open and the closed conformations.

Once the N-α-caproyl-L-lysine is stabilised in the active site, causing the enzyme to change its conformation, the hydrolytic mechanism is supposed to begin with the deprotonation of the catalytic water molecule linked to the two metal ions, by the acid base residue Glu154.A and

the subsequent nucleophilic attack of the substrate amide bond that is activated by the oxyanion hole. A tetrahedral intermediate is then formed and stabilised by the oxyanion hole, increasing the distance between the two metal ions and modifying their interactions: Zn1 is linked to the deprotonated water molecule and Zn2 is linked to the subtrate. Then a proton is transferred from residue Glu154.A to the α-amine function of the lysyl fragment, leading to the release of L-lysine and caproate and the transition to an open conformation of the enzyme, allowing the binding of a new catalytic water molecule (Figure III.12).

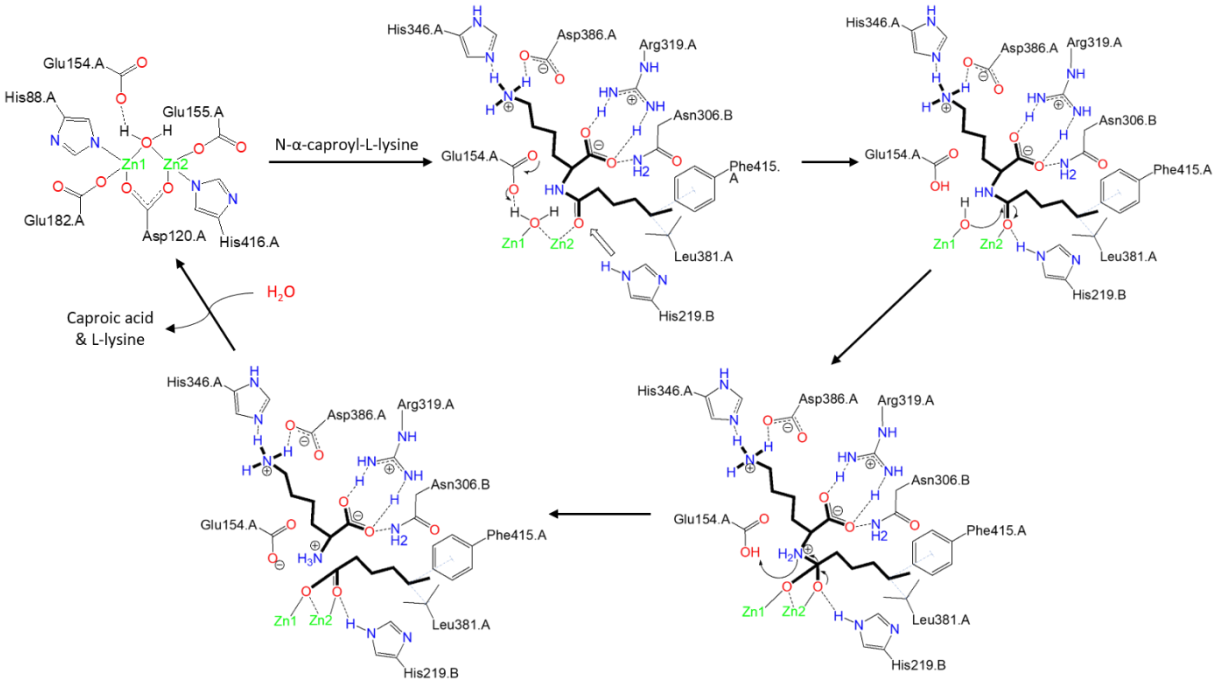


Figure III.12: Proposal for SamAA catalytic mechanism during N-α-caproyl-L-lysine hydrolysis.

The mechanism by which SamAA catalyses the N-acylation of L-lysine is supposed to be similar to that of hydrolysis, but proceeding in the reverse way. Once caproate and L-lysine are in the active site, causing the enzyme to move into its closed conformation, the acid base residue Glu154.A deprotonates the lysine α-amine function that is supposed to attack the caproate electrophilic center, leading to a tetrahedral intermediate that is stabilised by the oxyanion hole. This intermediate disappears with the release of the hydroxyle bound to Zn1, then its protonation by residue Glu154.A leading to one water molecule. Subsequent conformational change allows the release of the product N-α-caproyl-L-lysine (Figure III.13).

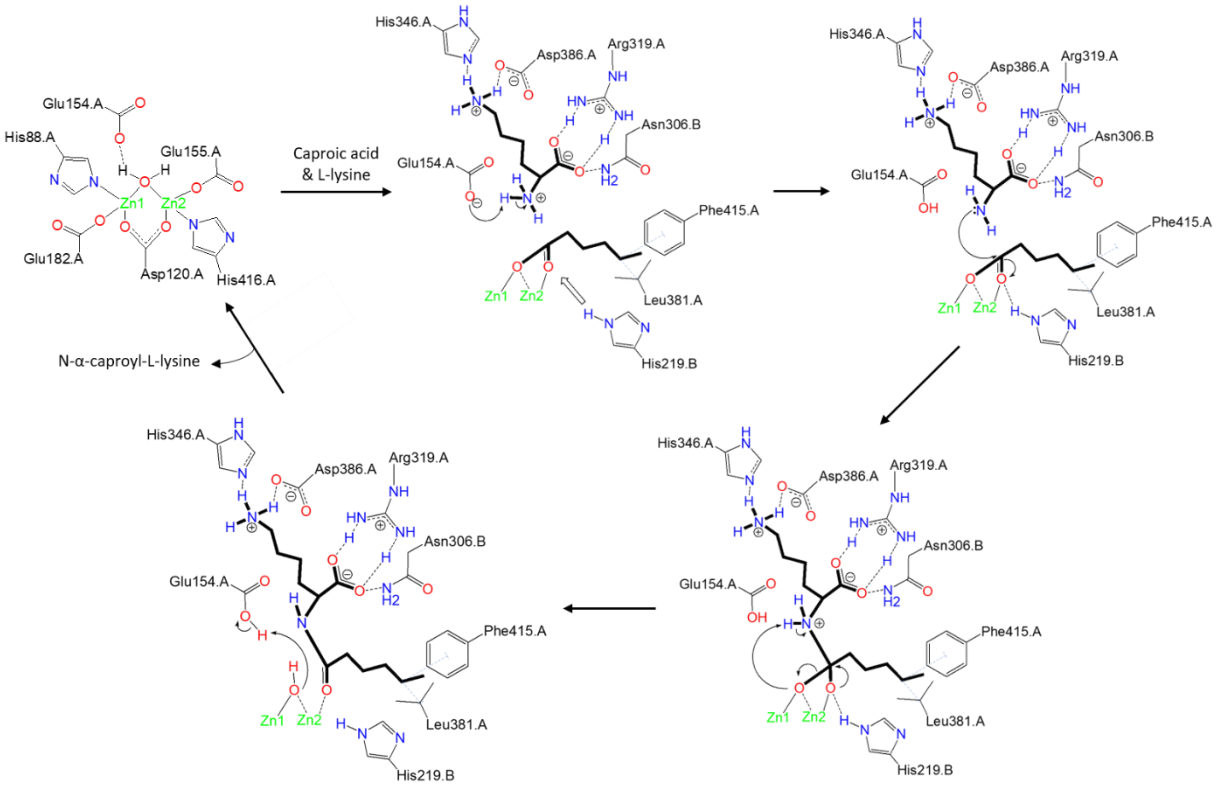


Figure III.13: Proposal for SamAA catalytic mechanism leading to the synthesis of N-α-caproyl-L-lysine.

3.2.4. Conclusions

SamAA from *S. ambofaciens* is an enzyme whose catalytic activity for hydrolysis or synthesis of L-lysine acyl derivatives has already been demonstrated experimentally, but there was no explanation given for its selectivity towards L-lysine α-amine function. This work aimed to provide structural information about this enzyme while awaiting the elucidation of its structure and to explain its selectivity using homology modelling and docking simulations. Conserved residues were identified and a 3D structure was proposed, suggesting that SamAA is a binuclear homodimer with a whistle-shaped monomer. According to the templates, SamAA belongs to the M20 peptidase family. Key residues were identified as His88, Asp120, Glu155, Glu182 and His416 that seem to be involved in the binding of the metal ions, and as and Glu154 which is believed to be the essential acid-base residue. Docking simulations were performed to mimic the hydrolysis and the synthesis of N-acyl-L-lysine. The results strongly suggest the selective acylation/deacylation of the lysine α-amine function whereas the acylation/deacylation of the ε-amine function seems unlikely to occur, what is in perfect agreement with experimental results. Based on previous studies on homologous enzymes a catalytic mechanism is proposed.

These results constitute the first elements featuring the structure, the selectivity and the mechanism of SamAA.

3.3. Contribution de l'article

Ces travaux ont permis de proposer un modèle de la structure tridimensionnelle de SamAA prédit par modélisation d'homologie et d'expliquer la sélectivité de l'enzyme lors de la synthèse/hydrolyse de dérivés N-acylés de la L-lysine, à l'échelle moléculaire. Des templates proches de SamAA du point du vue de leur séquence ont été identifiés ; ceux-ci ont été utilisés pour guider la modélisation de sa structure tridimensionnelle. L'étude des templates a également permis l'identification des résidus clés impliqués dans la fixation des ions métalliques ou dans l'activité catalytique. L'étude des modes d'interaction enzyme-substrat menée grâce à des simulations de docking moléculaire apporte les premiers éléments d'information pour la compréhension de la sélectivité particulière de SamAA.

Pour rappel, la modélisation par homologie permet de prédire la structure tridimensionnelle d'une protéine à partir de sa séquence en acides aminés, en se basant sur la séquence et la structure tridimensionnelle élucidée d'un ou plusieurs templates. Pour obtenir un modèle robuste, la protéine cible et le ou les templates doivent présenter le pourcentage d'homologie de séquence le plus élevé possible (Muhammed et al., 2018 ; Haddad et al., 2020). Cinq templates ont été identifiés pour SamAA *via* une recherche d'homologie de séquences : l'aminoacylase 1 humaine (entrée PDB 1Q7L), les succinyl-diaminopimelate desuccinylases de *Neisseria meningitidis* MC85 (entrée PDB 4PPZ), de *Haemophilus influenzae* (entrée PDB 5VO3) et de *Shigella flexneri* (entrée PDB 7LGP), et la carnosinase 2 humaine (entrée PDB 4RUH). La structure des cinq templates a été étudiée pour obtenir des informations sur l'organisation de leur site actif, en faisant l'hypothèse d'une organisation similaire chez SamAA. Deux ions métalliques (Zn^{2+}) et une molécule d'eau catalytique seraient présents au niveau de chaque site actif de chaque chaine peptidique, dans une géométrie de coordination bipyramidale trigonale déformée. Des résidus clés conservés ont été identifiés :

- His88.A, Asp120.A, Glu155.A, Glu182.A et His416.A impliqués dans la fixation des ions métalliques.
- Asp90.A, Glu154.A, Asp317.A, Arg319.A, His219.B et Asn306.B jouant un rôle dans la catalyse enzymatique.
- Ser382.A, Gly383.A, Gly384.A et Thr385.A formant la boucle de spécificité.

Trois outils de modélisation par homologie ont été utilisés pour prédire la structure 3D de SamAA : Modeller, Swiss Model et ColabFold (interface web de AlphaFold2). Ces outils ont été choisis sur la base du classement des meilleurs logiciels de prédiction de la structure de protéines établi à l'issue des concours CASP13 et CASP14 (Kryshtafovych et al., 2019 ; Kryshtafovych et al., 2021). La qualité des modèles tridimensionnels de SamAA générés a été estimée par une fonction de score appelée DOPE Score normalisé et par l'évaluation de la similarité dans l'organisation des sites actifs de SamAA et de la succinyl-diaminopimelate desuccinylase de *H. influenzae*. D'après ces analyses, le meilleur modèle de la structure tridimensionnelle de SamAA est celui produit par CobalFold. En effet, celui-ci présente la valeur de DOPE Score normalisé la plus faible (égale à -1.50). De plus, dans ce modèle, le site actif prédit de SamAA présente une grande similarité avec celui observé chez le template. SamAA serait un homodimère binucléaire dont le monomère en forme de sifflet présente un domaine catalytique et un domaine de dimérisation.

La structure prédite de SamAA a été utilisée comme cible pour le docking de différents substrats (N-α/ε-acétyl-L-lysine, N-α/ε-caproyl-L-lysine, L-lysine, acide acétique, acide caproïque) pour étudier les modes d'interaction enzyme-substrat et expliquer la sélectivité particulière de SamAA. Les résultats issus des simulations suggèrent que l'acylation et la déacylation sélective de la fonction amine α de la L-lysine est possible, tandis que l'acylation et la déacylation de la fonction amine ε est peu probable, en raison d'un mauvais positionnement des substrats dans la cavité. Ceci est cohérent avec les résultats expérimentaux obtenus précédemment (Dettori et al., 2018a ; Bourkaib et al., 2020a). Dans les complexes supposés conduire à une réaction de N-acylation, la fonction carboxylate du substrat donneur d'acyle interagit avec les deux ions métalliques de l'enzyme et la chaine carbonée est stabilisée par des interactions avec les résidus Leu381.A, Phe415.A, His219.B et Met222.B constituant la région hydrophobe de la cavité. La fonction amine α de la L-lysine se place, quant à elle, à proximité des ions métalliques et du donneur d'acyle. Elle interagit avec les résidus Glu154.A, Glu155.A et Gly384.A alors que la fonction amine ε est stabilisée à l'autre extrémité de la cavité catalytique par les résidus Gly157.A, Asp348.A et Asp386.A. Sa fonction carboxyle interagit avec les résidus Arg319.A, Gly384.A et Asn306.B. Sur la base de ces éléments et des quelques informations disponibles dans la littérature au sujet de cette famille d'enzymes, un mécanisme catalytique faisant intervenir un résidu acide-base essentiel (Glu154) a été proposé pour l'hydrolyse ou la synthèse de la N-α-caproyl-L-lysine.

Les travaux présentés dans l'article sont accompagnés de résultats complémentaires présentés juste après. Ceux-ci portent sur l'optimisation du modèle de SamAA généré par Modeller et la

modélisation de la réaction de N-acylation de la L-lysine par l'acide laurique et de la réaction inverse d'hydrolyse.

3.4. Résultats complémentaires

3.4.1. Construction du modèle de la structure de SamAA avec Modeller

Cette étude complémentaire à l'article présente la démarche suivie pour la construction du modèle par homologie de SamAA avec Modeller, outil implémenté dans la suite logicielle Discovery Studio®.

L'aminoacylase SamAA de *Streptomyces ambofaciens* a une activité enzymatique démontrée pour la réaction de N-acylation de la fonction amine α de la L-lysine et la réaction inverse d'hydrolyse des dérivés α-acylés de L-lysine (Dettori et al., 2018a ; Bourkaib et al., 2020a). Cette sélectivité n'a jamais été expliquée jusqu'ici à l'échelle moléculaire car aucune structure tridimensionnelle des aminoacylases de *S. ambofaciens* n'a été élucidée. La prédiction des structures tridimensionnelles des protéines dont la séquence est connue est possible grâce aux progrès de la bio-informatique. La modélisation par homologie permet de prédire la structure tridimensionnelle d'une protéine à partir de sa séquence en acides aminés et en se basant sur la structure tridimensionnelle connue d'un ou plusieurs templates. Ces derniers sont choisis en fonction de leur homologie de séquence avec la protéine cible, qui doit être la plus élevée possible (Muhammed et al., 2018 ; Haddad et al., 2020).

Une première étape du travail a donc consisté à mener des recherches de similarité de séquences avec les outils NCBI (Blastp) et Swiss Model (Search For Templates). Ces derniers ont permis d'identifier plusieurs templates : le mutant partiel de l'aminoacylase humaine 1 (mhAcy1 ; entrée PDB 1Q7L), les succinyl-diaminopimelate desuccinylases de *Neisseria meningitidis* (NmDapE ; entrées PDB 1VGY et 4PPZ), de *Haemophilus influenzae* (HiDapE ; entrée PDB 5VO3) et de *Shigella flexneri* (SfDapE ; entrée PDB 7LGP) et la carnosinase 2 humaine (hCN2 ; entrée PDB 4RUH). Ces résultats sont décrits dans l'article présenté précédemment.

La deuxième et la troisième étapes de la démarche correspondent respectivement à l'alignement des séquences des templates et de la séquence de SamAA et à la construction de la structure tridimensionnelle de celle-ci par modélisation par homologie. Une stratégie d'optimisation de chacune de ces étapes a été mise en place, *via* l'étude des différentes options et combinaisons de calculs disponibles. Le meilleur modèle issu de cette étude est celui présenté dans l'article précédent.

3.4.1.1. Optimisation de l'alignement de séquences

La suite logicielle de Discovery Studio® propose deux protocoles d'alignement de séquences, permettant de choisir le nombre de template(s) et différentes options. Le calcul « Align Sequences… » permet de prendre en compte les informations provenant des structures secondaires des templates et de SamAA et/ou d'aligner des séquences de protéines multimériques, en activant les options respectives « Use Secondary Structures… » (nommée « DSC ») et « Align by Chain… ». Le calcul « Align Sequences to Templates… » permet d'utiliser les informations de la structure tridimensionnelle des templates (« Create Sequences Profiles… ») et/ou d'utiliser l'alignement (partiel ou entier) obtenu à partir du Blastp.

37 alignements ont été réalisés en combinant les deux types de calcul, en variant les différents paramètres à disposition et les cinq templates. Les caractéristiques des alignements de séquences et les pourcentages d'identité et de similarité obtenus sont récapitulés dans le tableau III.4. Si un nouvel alignement est obtenu, il est utilisé pour générer un nouveau modèle de la structure tridimensionnelle de SamAA avec Modeller, paramétré par défaut. 30 modèles de la structure de SamAA ont été obtenus dont la qualité a été vérifiée par un calcul de DOPE Score normalisé (Tableau III.4). L'amélioration des pourcentages de similarité et d'identité ne suffit pas à elle seule à améliorer la qualité de la structure tridimensionnelle modélisée. Par exemple, un alignement entre SamAA-7LGP ayant un pourcentage d'identité de 40 % permet d'obtenir un modèle avec une valeur de DOPE Score normalisé de 0,88 tandis qu'un alignement entre ces deux séquences ayant un pourcentage d'identité plus faible (égale à 33 %) permet d'obtenir un modèle de meilleure qualité (DOPE Score normalisé égale à 0,48).

Le calcul de DOPE Score normalisé a permis de sélectionner les quatre meilleurs modèles tridimensionnels. Les trois premiers retenus ont été construits sur la base d'un seul template et ont les valeurs de DOPE Score normalisé les plus faibles (0,34 ; 0,42 ; 0,44). Discovery Studio conseille d'utiliser plusieurs templates dans l'alignement de séquences dans le but d'améliorer la qualité du modèle de la protéine cible, c'est pourquoi le quatrième retenu a un alignement avec plusieurs templates et la valeur de DOPE Score normalisé la plus faible (0,47) en respectant ce critère (Tableau III.4). Les quatre alignements à l'origine de ces quatre modèles sont utilisés par la suite pour une amélioration de la qualité de la structure tridimensionnelle modélisée de SamAA.

Tableau III.4 : Caractéristiques et pourcentages de similarité et d'identité des alignements de séquences réalisés et valeurs du DOPE Score normalisé des modèles obtenus. Sur la base de la valeur du DOPE Score normalisé des modèles générés, les alignements de séquences choisis et leurs caractéristiques sont surlignés en vert clair.

Templates	Type de séquences	Commande d'alignement	Use Secondary Structures	Align by Chain	Create Sequence Profile	Alignement Blastp	Pourcentage d'identité	Pourcentage de similarité	DOPE Score normalisé
1VGY	Monomère	Align Sequences	None	False	***	***	20,80%	37,60%	0,51
1Q7L	Chaine A	Align Sequences	None	False	***	***	27,80%	47,20%	0,45
		Align Sequence to Templates	***	***	False	***	27,80%	47,20%	
4PPZ	Monomère	Align Sequences	None	False	***	***	21,50%	37,10%	0,7
			DSC	False	***	***	21,50%	37,10%	
		Align Sequence to Templates	***	***	False	Non	21,50%	37,30%	0,54
			***	***	True	Oui	13,50%	28,60%	1,16
	Homodimère	Align Sequences	None	True	***	***	21,50%	37,10%	0,69
			DSC	True	***	***	21,50%	37,30%	0,73
1Q7L et 4PPZ	Chaine A et monomère	Align Sequences	None	False	***	***	6,10%	11,70%	0,51
		Align Sequence to Templates	***	***	False	Non	6,10%	11,30%	0,52
			***	***	True	Oui	5,40%	9,10%	0,76
7LGP	Monomère	Align Sequences	None	False	***	***	21,80%	40,20%	0,88
			DSC	False	***	***	21,80%	40,20%	
		Align Sequence to Templates	***	***	True	Non	17,90%	33,20%	0,48
			***	***	True	Oui	15,50%	30,30%	0,86
	Homodimère	Align Sequences	None	True	***	***	21,70%	39,90%	0,85
			DSC	True	***	***	21,70%	39,90%	
1Q7L, 4PPZ, 5VO3 et 7LGP	Chaine A et monomères	Align Sequence to Templates	***	***	True	Oui	5,20%	8,60%	0,82
4RUH	Monomère	Align Sequences	None	False	***	***	16,20%	36,90%	0,47
			DSC	False	***	***	16,20%	36,90%	
		Align Sequence to Templates	***	***	True	Oui	13,50%	32,30%	0,45
			***	***	True	Non	13,30%	30,10%	0,83
	Homodimère	Align Sequence	DSC	True	***	***	16,70%	36,90%	0,44
			None	True	***	***	16,70%	36,90%	
5VO3	Monomère	Align Sequences	None	False	***	***	20,90%	40,20%	0,34
			DSC	False	***	***	20,90%	40,20%	
		Align Sequence to Templates	***	***	True	Oui	15,70%	33,30%	0,76
			***	***	True	Non	15,70%	33,30%	0,77
	Homodimère	Align Sequences	DSC	True	***	***	20,90%	40,10%	0,42
			None	True	***	***	21,10%	40,60%	0,47
1Q7L, 4PPZ, 5VO3, 7LGP et 4RUH	Chaine A et monomères	Align Sequences	None	False	***	***	3,20%	7,90%	1,26
		Align Sequence to Templates	***	***	True	Oui	2,90%	5,70%	0,58
			***	***	True	Non	0,10%	0,70%	0,65
	Chaines A et C et homodimères	Align Sequences	DSC	True	***	***	3,50%	7,90%	0,47
			None	True	***	***	3,20%	7,90%	1,26
1Q7L, 7LGP,4RUH	Chaine A et monomères	Align Sequence to Templates	***	***	True	Oui	2,50%	5,30%	0,76

3.4.1.2. Optimisation de la structure du modèle de SamAA

Modeller est disponible dans la suite de logiciels Discovery Studio®, avec plusieurs options possibles. Le modèle de la structure tridimensionnelle peut être construit en présence des ligands présents dans le site actif du template comme les ions métalliques (« Ligands ») et de molécules d'eau solvatant la structure tridimensionnelle (« Water »). La position des boucles peut être optimisée (« Refine Loop »). Dans le but d'améliorer la qualité des modèles obtenus jusque-là, les alignements de séquence des quatre meilleurs modèles obtenus pour SamAA ont été réutilisés pour générer de nouveaux modèles de la structure tridimensionnelle de SamAA en étudiant ces trois options disponibles. La qualité de ces nouveaux modèles a été analysée par le DOPE Score normalisé, dont les résultats sont récapitulés dans le tableau III.5.

Tableau III.5 : Caractéristiques des alignements de séquences choisis et valeurs du DOPE Score normalisé des modèles obtenus.

Template	Structure 3D du modèle	Alignement	Ligands	Water	Refine Loop	Dope SCORE
5VO3	Monomère	Align Sequence Sans DSC	***	***	***	0,34
			2 zincs	***	***	0,44
			***	True	***	0,72
			2 zincs	True	***	0,75
			***	***	True	0,17
	Homodimère	Align Sequence avec DSC	***	***	***	0,42
			4 zincs	***	***	0,43
			***	True	***	0,78
			4 zincs	True	***	0,77
			***	***	True	0,11
			4 zincs	***	True	0,22
4RUH	Homodimère	Align Sequence Sans DSC	***	***	***	0,44
			***	True	***	0,63
			4 manganèses	***	***	0,47
			4 manganèses	True	***	0,66
			***	***	True	0,37
			4 manganèses	***	True	0,35
Chaines A et C de 1Q7L, 4PPZ, 4RUH, 5VO3 et 7LGP	Homodimères	Align Sequence avec DSC	***	***	***	0,47
			Oui	***	***	0,55
			***	True	***	3,9
			Oui	True	***	3,97
			***	***	True	0,45

L'utilisation des options permettant l'insertion des ligands présents dans la structure tridimensionnelle du template et la solvatation du modèle lors de sa modélisation n'ont pas permis d'améliorer la qualité du modèle. L'option « Refine Loop » améliore la qualité du modèle généré, mais le temps de calcul est augmenté. Le modèle de SamAA ayant la valeur de DOPE Score normalisé la plus basse et égale à 0,11 a été obtenu avec l'option « Refine Loop » et la structure homodimérique de 5VO3. L'alignement utilisé lors de cette modélisation a été obtenu avec le calcul « Align Sequence… » accompagnée de l'option « Use Secondary Structures » (ou « DSC ») et utilisant la séquence homodimérique du template (choix de l'option « Align by Chain »).

3.4.1.3. Conclusions

La stratégie d'optimisation de l'alignement de séquence et de construction de la structure tridimensionnelle de SamAA a permis d'obtenir un modèle ayant la valeur la plus basse de DOPE Score normalisé qui est égale à 0,11. Cette valeur qui est positive indique un modèle de mauvaise qualité présentant des problèmes structuraux (Shen et al., 2006 ; Eramian et al., 2009).

Ce modèle a été obtenu en utilisant la succinyl-diaminopimelate desuccinylase de *N. meningitidis* (entrée PDB 5VO3) sous une forme homodimèrique comme template. Le calcul « Align Sequence… » avec les options « DSC » et « Align by Chain » et les séquences homodimériques du template et de SamAA permettent d'obtenir l'alignement utilisé dans la modélisation de la structure de SamAA réalisée par Modeller avec l'option « Refine Loop ». Ce modèle est celui décrit dans l'article et comparé aux modèles issus d'autres logiciels de modélisation par homologie (Swiss Model et AlphaFold2).

3.4.2. Modélisation de la synthèse/hydrolyse de la N-lauroyl-L-lysine

Lors de précédents travaux visant à préciser les performances des aminoacylases de *S. ambofaciens* pour la production de dérivés acylés d'acides aminés, la réaction de N-acylation de la lysine a été menée en présence de divers substrats donneurs d'acyle de longueur de chaîne allant de 8 à 18 atomes de carbone. Les résultats obtenus ont montré une augmentation de l'activité avec l'accroissement de la chaine grasse jusqu'à 12 atomes de carbone puis une diminution de l'activité pour les chaines de taille supérieure (Bourkaib et al., 2020a). Par ailleurs une sélectivité stricte en faveur de la fonction amine α de la lysine avait été observée. Au vu de ces résultats, l'étude des modes d'interaction entre SamAA et les substrats impliqués dans la synthèse ou l'hydrolyse de la N-lauroyl-L-lysine semblait pertinente et complémentaire des travaux décrits dans l'article présenté précédemment et portant sur l'acylation de la lysine par des donneurs d'acyle à chaîne courte. Certains impacts du mode de construction d'un modèle sur sa qualité et les modélisations qui lui sont associées sont également mentionnés.
Comme précédemment décrit dans l'article, le meilleur modèle de SamAA issu de ColabFold a servi de cible de docking pour étudier les modes d'interactions entre SamAA et la L-lysine, l'acide laurique et les dérivés α- et ε-N-acylés correspondants. Selon la réaction étudiée (hydrolyse ou synthèse), le modèle avec ou sans molécule d'eau catalytique a été utilisé comme cible, selon le protocole décrit dans l'article. Pour rappel, des critères ont été appliqués pour analyser les résultats de docking et conclure sur la faisabilité et la sélectivité des réactions : le positionnement des substrats donneur ou accepteur d'acyle dans la cavité catalytique de SamAA en lien avec ses propriétés de surface, la proximité des groupes fonctionnels des substrats par rapport aux résidus catalytiques et les positions relatives des éléments impliqués dans la catalyse évaluées notamment *via* l'angle de Bürgi impliqué dans l'attaque nucléophile soit de la liaison amide par la molécule d'eau dans la réaction d'hydrolyse, soit de l'acide laurique (donneur d'acyle) par la L-lysine (accepteur d'acyle) dans la réaction de synthèse (Bürgi et al.,

1974 ; Lodge et al., 1987 ; Radisky et al., 2002 ; Lubkowski et al., 2019). Ces critères ont été définis sur la base de la structure tridimensionnelle de la succinyl-diaminopimelate desuccinylase de *Haemophilus influenzae* élucidée en présence des deux produits issus de l'hydrolyse du N-succinyl-(2S,6S)-2,6-diaminoheptanedioate. Cette enzyme a été utilisée comme template pour la prédiction de la structure de SamAA et comme référence pour l'architecture du site actif (Dutta et al., 2016 ; Nocek et al., 2018). Ainsi des distances caractéristiques dont le détail est disponible dans le chapitre Matériels et Méthodes et dans l'article, ont été mesurées et utilisées comme références pour analyser les poses issues des simulations de docking. Pour faciliter la compréhension des résultats, ces poses sont toutes orientées de manière identique, selon les indications données en figure III.14.

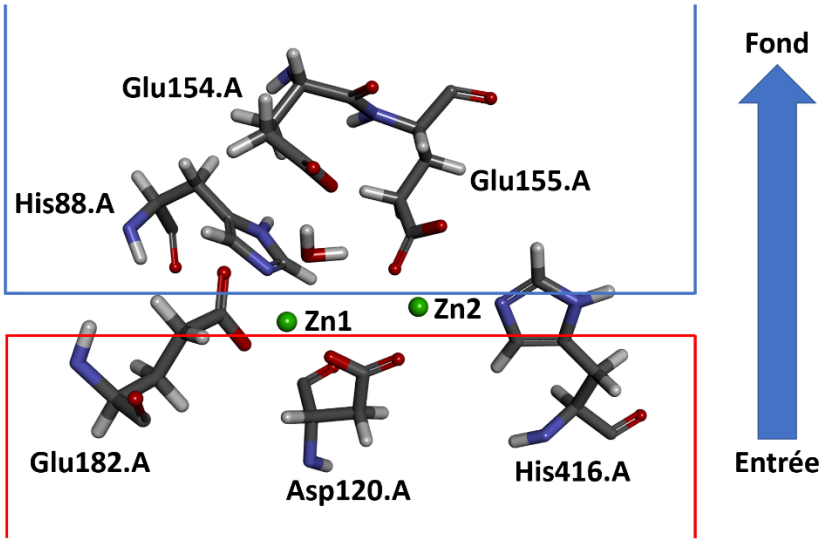


Figure III.14 : Architecture du site actif prédit de SamAA avec la molécule d'eau catalytique, les deux ions métalliques (en vert), le résidu catalytique acide-base et les résidus fixateurs des ions métalliques. Les cadres partiels bleu et rouge indiquent respectivement la région hydrophile et la région hydrophobe de la cavité catalytique. La fléche bleue indique la direction de l'entrée et du fond de la cavité.

3.4.2.1. Etude de l'hydrolyse de la N-lauroyl-L-lysine

Comme rappelé précédemment, des travaux antérieurs ont montré que SamAA présente une spécificité d'hydrolyse pour les dérivés N-α-acylés de la L-lysine quelle que soit la longueur de chaine de l'entité acylé, contrairement aux dérivés N-ε-acylés qui ne sont pas pris en charge par l'enzyme (Dettori et al., 2018a). Des simulations de docking ont été menées pour tenter d'expliquer cette caractéristique à l'échelle moléculaire.

Les résultats de docking montrent que les complexes associant la N-α-lauroyl-L-lysine et SamAA présentent trois orientations préférentielles du substrat dans la cavité catalytique de l'enzyme.

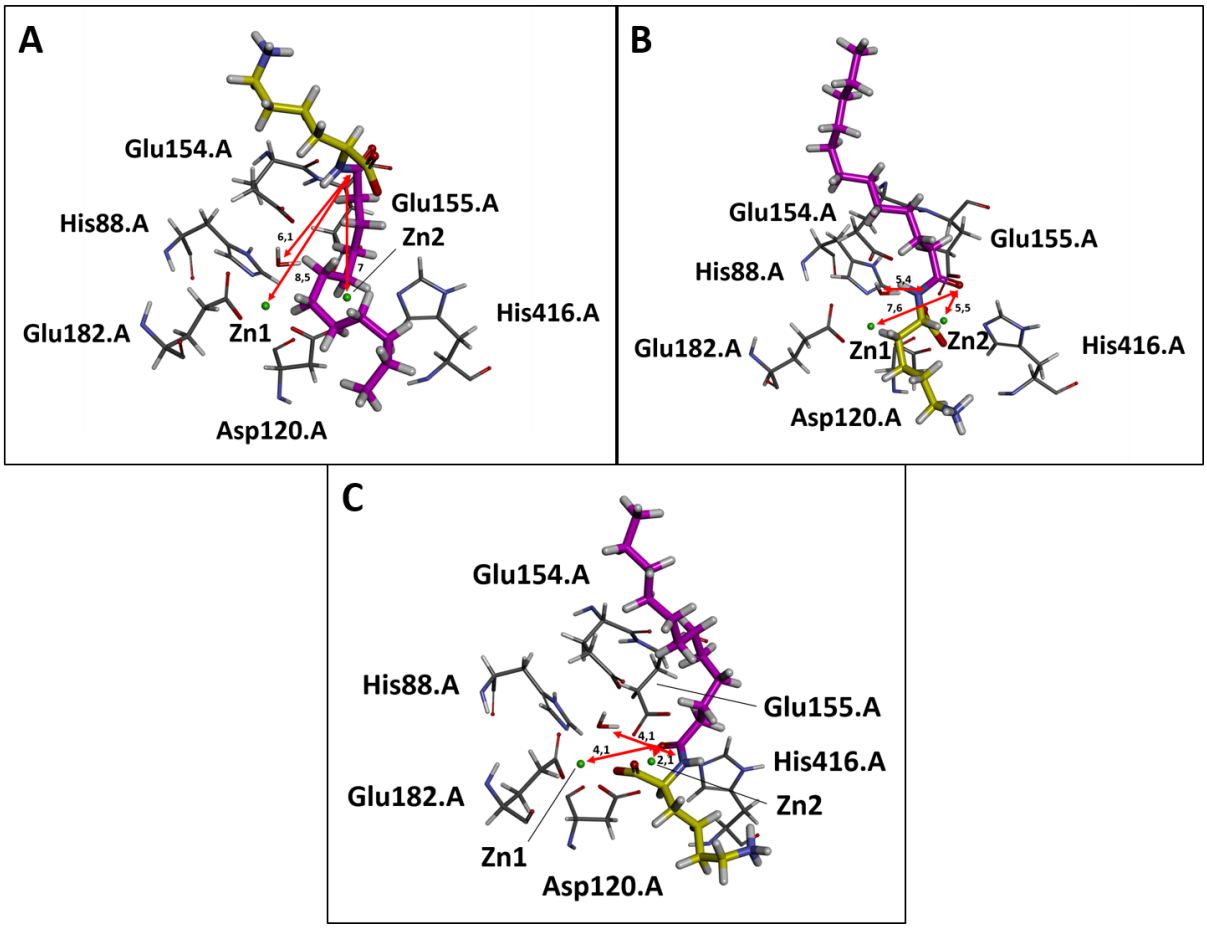


Figure III.15 : Orientations préférentielles de la N-α-lauroyl-L-lysine dans le site actif de SamAA. Les deux ions sont en vert. Les fragments lysyl et lauroyl sont respectivement en jaune et en rose. Les flèches rouges indiquent les distances entre le substrat et les ions métalliques ou la molécule d'eau catalytique. Les traits noirs indiquent le nom des éléments visibles.

Dans le premier groupe des 20 poses les mieux classées, le fragment lauroyl du dérivé acylé se positionne devant les deux ions métalliques et les résidus Asp120.A et Glu155.A. Le fragment lysyl interagit avec la partie hydrophile de la cavité, la fonction amine de la liaison amide pointant en direction du résidu acide-base Glu154.A. Les ions métalliques et la molécule d'eau catalytique sont à une distance comprise entre 6 et 8 Å de la liaison amide (Figure III.15A).

Dans la seconde orientation correspondant à 8 poses, le fragment lauroyl du dérivé acylé est à proximité des résidus Glu154.A et Glu155.A et interagit avec la partie hydrophile de la cavité catalytique tandis que le fragment lysyl est proche de la molécule d'eau catalytique, des ions métalliques et du résidu Asp120.A dans la partie hydrophobe de la cavité. La fonction amine

de la liaison amide pointe vers le résidu Glu154.A, mais demeure éloignée des ions métalliques (5,5 à 8 Å) et de la molécule d'eau catalytique (5 à 6 Å) (Figure III.15B).

Enfin dans le troisième groupe de 22 poses, le dérivé acylé adopte une position similaire à celle du deuxième groupe de poses, où les fragments lauroyl et lysyl interagissent respectivement avec la partie hydrophile et hydrophobe de la cavité catalytique. La liaison amide interagit avec les ions métalliques à une distance comprise entre 2 et 4 Å. La fonction amine libre est orientée vers l'extérieur du site actif. La molécule d'eau est à 4 Å de la liaison amide. Toutefois l'angle d'attaque est évalué à 40 ° (Figure III.15C).

Les résultats issus du docking de la N-ε-lauroyl-L-lysine suggèrent deux orientations préférentielles dans la cavité catalytique de SamAA.

Dans la première orientation correspondant aux 14 poses les mieux classées, le fragment lauroyl de la N-ε-lauroyl-L-lysine est à proximité des ions métalliques, de la molécule d'eau et des résidus catalytiques. Le fragment lysyl est quant à lui à la périphérie hydrophile du site actif près des résidus Glu154.A et Glu155.A. La liaison amide est située à 8 Å des ions métalliques et à 6 Å de la molécule d'eau catalytique (Figure III.16A).

Dans le second groupe de 36 poses, le fragment lysyl de la N-ε-lauroyl-L-lysine est dans la partie hydrophobe de la cavité catalytique, à côté du résidu Asp120.A. Le fragment lauroyl est dans la partie hydrophile du site actif, sa chaine aliphatique étant proche des résidus Glu154.A et Glu155.A et orientée vers l'extérieur du site. Le groupe carbonyle de la liaison amide est en interaction avec les deux ions métalliques (de 2,2 à 4,3 Å). Il est également proche de la molécule d'eau catalytique (3,8 Å), toutefois l'angle d'attaque nucléophile affiche une valeur de 51 ° (Figure III.16B).

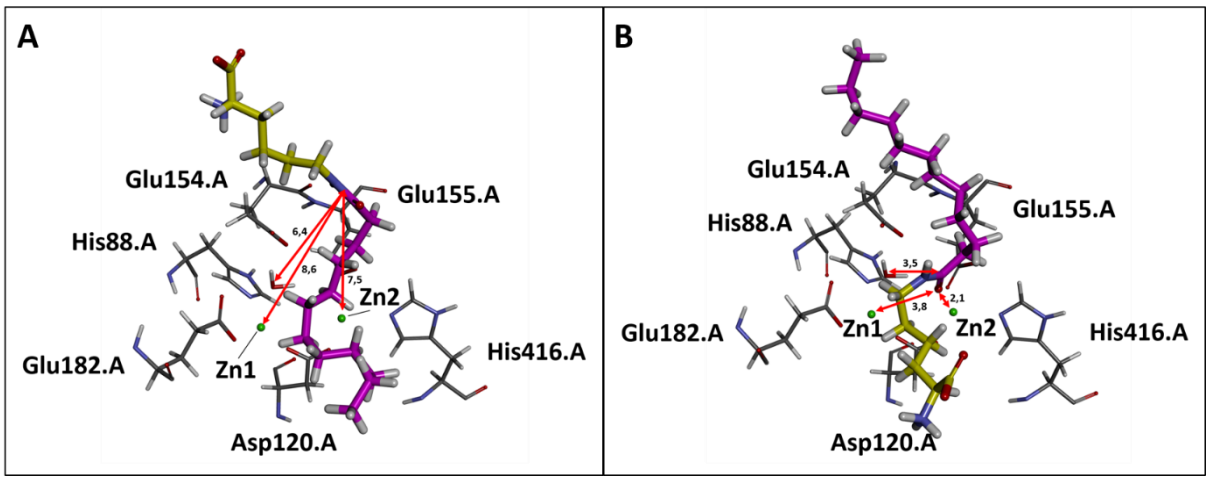


Figure III.16 : Orientations préférentielles de la N-ε-lauroyl-L-lysine dans le site actif de SamAA. Les deux ions sont en vert. Les fragments lysyl et lauroyl sont respectivement en jaune et en rose. Les flèches rouges indiquent les distances entre le substrat et les ions métalliques ou la molécule d'eau catalytique. Les traits noirs indiquent le nom des éléments visibles.

En conclusion, d'après les résultats des simulations, aucune des orientations préférentielles de la N-lauroyl-L-lysine ne respecte l'ensemble des critères admis comme conduisant à une catalyse de la réaction d'hydrolyse, que ce soit à cause d'un éloignement du substrat vis-à-vis du site actif ou en raison d'une mauvaise orientation de ce dernier. Dans le cas de la N-ε-lauroyl-L-lysine, ces résultats sont cohérents avec les observations expérimentales qui mentionnent une absence d'hydrolyse de ce substrat en présence de SamAA. Toutefois, les résultats issus du docking sont contradictoires avec les résultats expérimentaux concernant l'hydrolyse de la N-α-lauroyl-L-lysine. Ils sont également en désaccord avec les résultats obtenus lors de la simulation de l'hydrolyse des dérivés acylés de la lysine à courte chaîne (voir article).

3.4.2.2. Etude de la synthèse de la N-lauroyl-L-lysine

La sélectivité de synthèse de SamAA envers les dérivés N-α-acyl-L-lysine a été démontrée expérimentalement (Bourkaib et al., 2020a). Afin de simuler cette réaction et d'expliquer cette sélectivité, les orientations et les modes d'interaction des substrats acide laurique et L-lysine avec SamAA ont été étudiés en l'absence de la molécule d'eau catalytique. Deux séries de simulations ont été réalisées, la première débutant par le docking du substrat donneur d'acyle suivi du docking de l'accepteur d'acyle, la deuxième étant menée dans l'ordre inverse d'introduction des substrats dans la cavité de SamAA.

3.4.2.2.1. Docking de l'acide laurique puis de la L-lysine

Le docking de l'acide laurique sous sa forme déprotonée a conduit aux résultats suivants. Parmi les poses de docking obtenues, une seule orientation préférentielle de l'acide laurique a été observée. La chaine carbonée est à proximité des résidus Glu154.A et Glu155.A et est orientée vers la région hydrophile de la cavité catalytique. La fonction carboxyle est en interaction avec les deux ions métalliques (entre 2 et 2,5 Å) (Figure III.17A).

Malgré un positionnement surprenant de l'acide laurique dans la région hydrophile de la cavité, les simulations ont été poursuivies par le docking de la L-lysine avec la pose la mieux classée comme cible. Les résultats obtenus montrent quatre orientations préférentielles de la L-lysine, laquelle se positionne toujours dans la partie hydrophobe de la cavité catalytique. Quelle que soit son orientation, les deux fonctions amines de la L-lysine sont distantes de 3 à 7 Å de la fonction carboxyle de l'acide laurique et de 6 à 11 Å du résidu acide-base Glu154.A (Figure III.17 B et C).

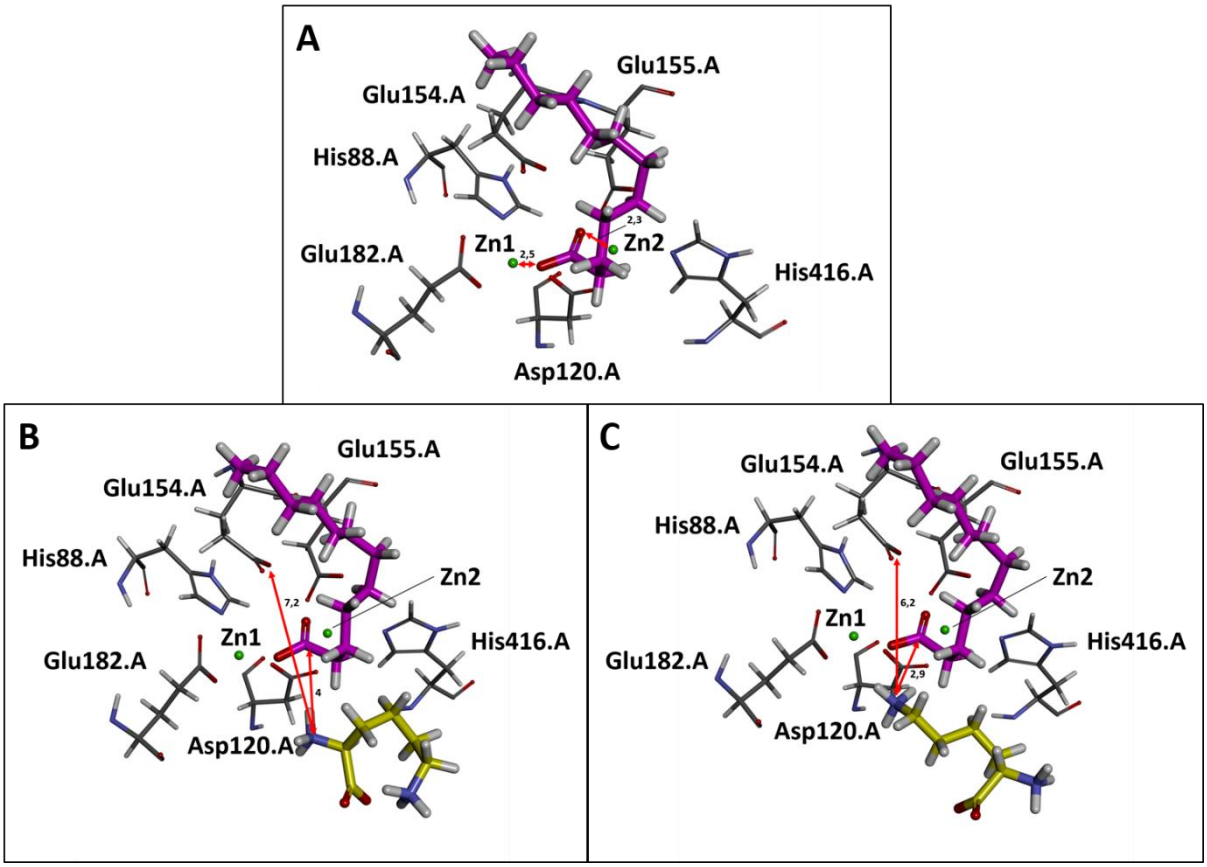


Figure III.17 : A : Orientation préférentielle de l'acide laurique dans le site actif de SamAA. B et C : Exemples de deux orientations préférentielles de la L-lysine dans le site actif de SamAA complexé avec le donneur d'acyle lauroate. Les fragments lysyl et lauroyl sont respectivement en jaune et en rose. Les flèches rouges indiquent les distances entre le substrat et le résidu Glu154.A ou le carbone de l'attaque nucléophile. Les traits noirs indiquent le nom des éléments visibles ou la distance de la flèche rouge.

3.4.2.2.2. Docking de la L-lysine en premier puis de l'acide laurique

Le docking de la L-lysine dans la cavité de SamAA conduit à quatre modes d'interactions principaux. Ces résultats sont détaillés dans l'article présenté précédemment. Pour rappel, la L-lysine se localise préférentiellement dans la partie hydrophile de la cavité catalytique de SamAA. L'extrémité α de la L-lysine est proche des résidus Glu154.A, Glu155.A et Gly384.A, formant des liaisons ioniques et hydrogène. La fonction amine α est à 2 Å de l'extrémité terminale du résidu catalytique Glu154.A, ce qui est favorable à sa déprotonation et par la suite à son acylation. La fonction carboxylate est stabilisée par une liaison ionique avec Arg319.A et par des liaisons hydrogène avec Arg319.A, Gly384.A et Asn306.B. L'orientation inverse dans laquelle la fonction amine ε de la lysine pointerait et serait à une distance favorable des résidus catalytiques n'est jamais observée. Cette dernière forme des interactions ioniques et des liaisons

hydrogène avec Gly157.A, Asp348.A et Asp386.A. La pose présentant le meilleur score a été choisie comme cible pour le docking de l'acide laurique (Figure III.18).

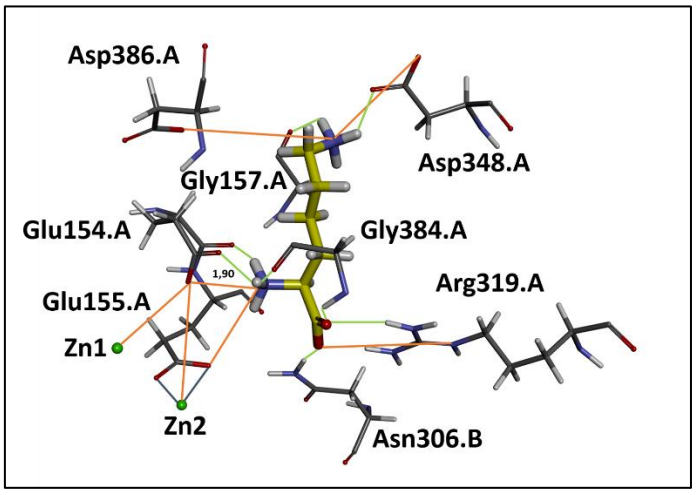


Figure III.18 : Orientation et interactions de la L-lysine dans le site actif de SamAA. Pour davantage de lisibilité, seuls les résidus interagissant avec les substrats sont montrés. Les deux ions métalliques et la L-lysine sont respectivement en vert et jaune. Les liaisons hydrogène, les interactions ioniques et les liaisons datives sont représentées en traits pleins de couleur verte, orange et grise respectivement.

Une seule orientation préférentielle est obtenue lors du docking de l'acide laurique sur le complexe formé entre SamAA et la L-lysine. L'acide laurique interagit avec la partie hydrophobe de la cavité et forme des interactions hydrophobes avec les résidus Val183.A, Leu285.A, Leu381.A, Phe415.A, His219.B et Met222.B. Sa fonction carboxyle forme une liaison ionique avec la fonction amine α de la L-lysine (distance de 3,5 Å). Cette fonction interagit également avec les deux ions métalliques à une distance de 2 à 4 Å. Le RMSD de l'ensemble des poses est inférieur à 2 Å, ce qui appuie l'hypothèse d'une orientation préférentielle unique. Seules quelques variations de position des deux atomes d'oxygène de la fonction carboxyle de l'acide laurique en direction des deux ions métalliques sont relevées. Dans les poses les mieux classées, les deux atomes d'oxygènes pointent vers Zn2, entraînant un angle de Bürgi-Dunitz très fermé. Toutefois, dans certaines poses présentant des scores inférieurs, les deux atomes d'oxygène sont chacun orientés vers un ion métallique, conduisant ainsi à un angle de Bürgi-Dunitz compris entre 80 et 100 ° et suggérant la possibilité d'acyler la lysine en position α (Figure III.19 A, B et C).

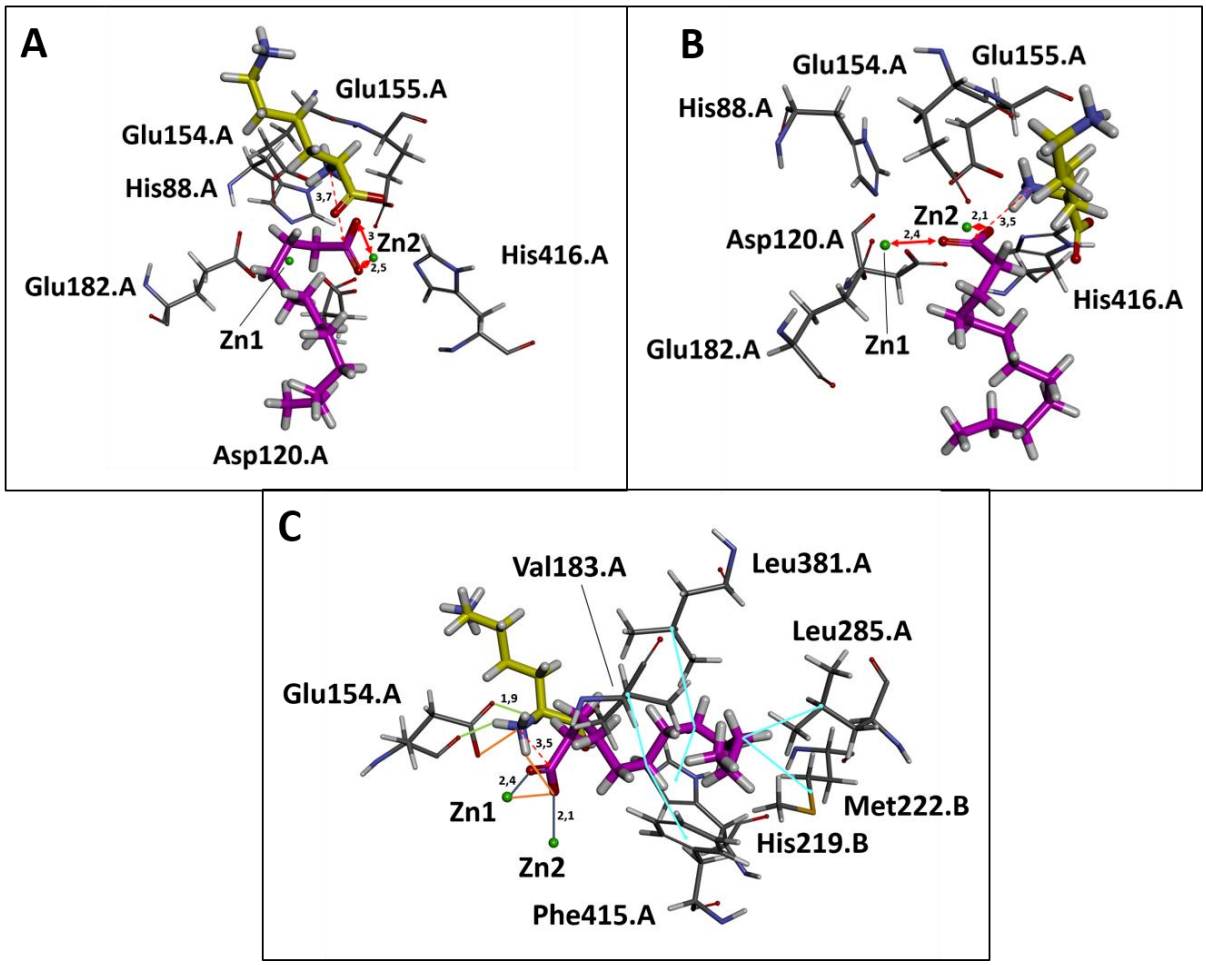


Figure III.19 : Orientations préférentielles (A et B) et interactions (C) de l'acide laurique dans le site actif de SamAA. A et B : Les doubles flèches rouges indiquent les distances entre le substrat et les deux ions métalliques. La flèche rouge en pointillés mime l'attaque nucléophile de l'acide laurique par la fonction amine α de la L-lysine. Les traits noirs indiquent le nom des éléments visibles. C : Pour davantage de clarté, seuls les résidus interagissant avec les substrats sont montrés. Les deux ions métalliques, la L-lysine et l'acide laurique sont respectivement en vert, jaune et rose. Les liaisons hydrogènes, les interactions ioniques et les liaisons datives sont représentées en traits pleins de couleur verts, orange et grise respectivement. Les traits noirs indiquent le nom des éléments visibles.

3.4.2.3. Conclusions

Lors de la thèse de Dettori, l'acylation de la L-lysine sur sa fonction amine α par l'acide laurique a été démontrée grâce à l'analyse par MS^2 du milieu de synthèse en présence des aminoacylases (dont SamAA), issues du surnageant de culture de *Streptomyces ambofaciens* ATCC 23877 (Dettori, 2017). Ce résultat a également été confirmé au cours des travaux menés par Bourkaib (Bourkaib et al., 2020a). Si les preuves expérimentales de la sélectivité de SamAA pour l'amine en position α de la L-lysine existent, les simulations de docking menées pour simuler les

réactions d'hydrolyse ou de synthèse ne sont pas concluantes. Elles suggèrent tout au plus une orientation préférentielle de la L-lysine dans la cavité, sa fonction amine α pointant vers les résidus catalytiques. Seules quelques poses proximales de l'acide laurique suggèrent une acylation possible. Lorsque l'acide laurique est introduit seul dans le site, son positionnement préférentiel dans la région hydrophile de la cavité est surprenant et suggère un manque d'espace pour ce substrat encombrant.

La structure cristallographiée de la succinyl diaminopimelate desuccinylase de *Haemophilus influenzae* en interaction avec les produits issus de l'hydrolyse de son substrat naturel (acide succinique et L,L-2,6-diaminopimelate) (entrée PDB 5VO3) a permis la découverte d'un changement conformationnel significatif de l'enzyme. Lors de la liaison au substrat, le domaine catalytique tourne de presque 50 ° par rapport au domaine de dimérisation et les acides aminés de ce domaine se déplacent de 10 Å au niveau du site actif et jusqu'à 29 Å à l'extérieur de la protéine, provoquant la fermeture du site actif binucléaire au niveau du fragment succinate et réduisant la taille de l'enzyme (Figure III.20). Ce changement conformationnel entraîne une forme d'entrelacement entre les résidus de la sous-unité B et le site actif de la sous-unité A (et inversement) et l'apparition de nouvelles interactions protéine-ligand. Plus précisément, la distance entre le résidu His194 de la chaine B et le Zn2 du site actif de la chaine A passe de 10 Å en absence de ligand (entrée PDB 3IC1) à 2,9 Å en présence de deux produits. L'acide succinique est stabilisé dans la conformation fermée de l'enzyme par le trou d'oxyanion constitué de Zn2 et du résidu His194.B. Le changement conformationnel de l'enzyme et notamment le déplacement du résidu His194.B sont décrits comme contribuant à l'activation de la liaison devant être hydrolysée et à la stabilisation de l'état de transition (Nocek et al., 2018).

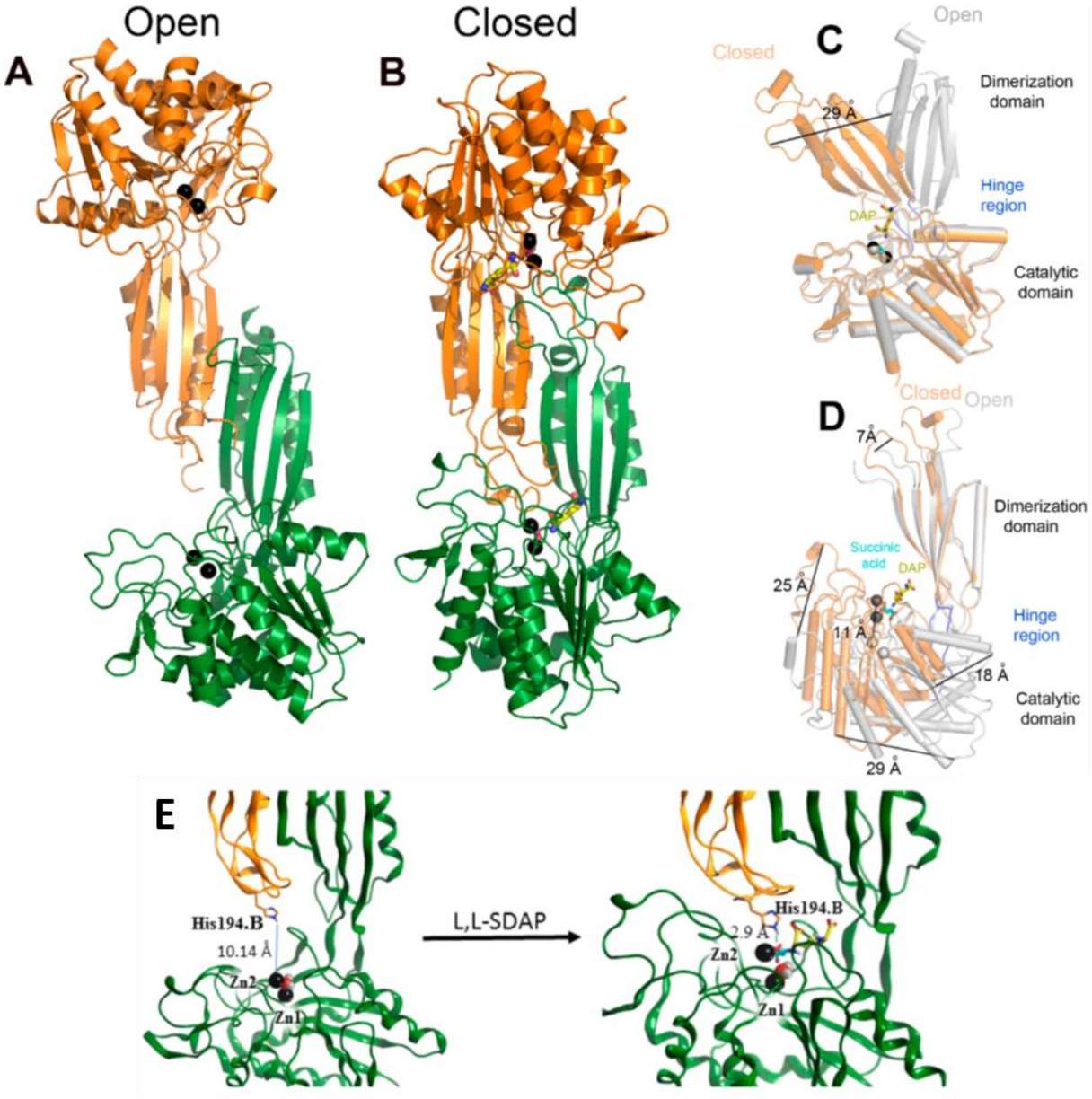


Figure III.20 : Comparaison des structures de la succinyl-diaminopimelate desuccinylase de *H. influenzae* dans sa conformation « ouverte » ou « fermée ». Les sous-unités individuelles de l'homodimère sont colorées en vert et orange. (A) et (B) : Diagramme en ruban de la structure de l'enzyme dans sa conformation « ouverte » (entrée PDB 3IC1) et « fermée » en interaction avec les produits acide succinique (magenta) et acide L-diaminopimélique (jaune) (entrée PDB 5VO3). (C) et (D) Mise en évidence des changements conformationnels par superposition des domaines catalytiques de dimérisation des monomères dans la structure ouverte (gris) ou fermée (orange). (E) : Zoom sur le passage d'une position distale à une position proximale du résidu His194 après l'arrivée du substrat dans le site actif (Nocek et al., 2018).

Le résidu His194 est un acide aminé conservé chez SamAA correspondant au résidu His219. Le modèle utilisé pour les simulations de docking présentées précédemment est caractérisé par une distance entre le His219.B et l'ion Zn2 de la chaine A de 5,3 (+/- 0,1) Å. Cette valeur est intermédiaire entre les distances relevées dans les conformations ouverte et fermée de la succinyl-diaminopimelate desuccinylase de *H. influenzae*. Si l'hypothèse est que SamAA subit

des changements conformationnels similaires au cours de la catalyse, il est probable que la structure de SamAA prédite par ColabFold corresponde à un état de l'enzyme intermédiaire entre la forme ouverte et la forme fermée. Ceci pourrait expliquer les difficultés rencontrées lors du docking de l'acide laurique et du dérivé N-lauroyl-L-lysine dans une cavité catalytique en partie refermée.

3.4.3. Prédiction de la structure de SamAA en conformation ouverte

Pour tenter de confirmer l'hypothèse formulée précédemment selon laquelle le mauvais positionnement de l'acide laurique serait dû à une conformation trop fermée de SamAA, de nouvelles simulations visant l'obtention d'une structure plus ouverte de l'enzyme ont été menées en suivant deux stratégies : la première a consisté à utiliser la structure ouverte de la succinyl-diaminopimelate desuccinylase de *H. influenzae* (entrée PDB 3IC1) comme template pour guider la prédiction de la structure de SamAA ; la seconde repose sur l'utilisation d'AlphaFold3.

3.4.3.1. Description de la structure tridimensionnelle de SamAA modélisée sur la base du template 3IC1

L'acylation de la L-lysine par l'acide laurique ou l'hydrolyse de la N-α-lauroyl-L-lysine catalysée par SamAA a été prouvée expérimentalement. Or, les résultats des simulations mimant ces réactions ne sont pas concluants lorsque la structure de SamAA prédite par ColabFold est utilisée comme cible. La construction d'un modèle de SamAA en conformation ouverte a été entrepris en utilisant comme template la structure de la succinyl-diaminopimelate desuccinylase de *H. influenzae* obtenue en absence des produits de l'hydrolyse de son substrat naturel (entrée PDB 3IC1).

La structure de cette entrée PDB est incomplète : les fragments Gly193-Ala196 et Asn244-245 de la chaine A et Gln192-198 de la chaine B sont absents du cristal publié (Nocek et al., 2010). En utilisant le logiciel Discovery Studio, ces trois fragments manquants ont été identifiés depuis les données SEQRES et ont été ajoutés. La conformation des extrémités des boucles concernées et des résidus Thr73.B, Glu165.A, Arg329.A et Asn21.A a été optimisée. La chaine latérale de Tyr197.A a été complétée. Globalement, la structure tridimensionnelle du template a été améliorée, passant d'une valeur de DOPE Score normalisé de -1,42 à -1,50 après optimisation,

celle-ci sera utilisée comme template lors de la modélisation par homologie de la structure de SamAA.

Les séquences du template 3IC1 et de SamAA ont été alignées sur Discovery Studio avec l'outil d'alignement « Align Sequences » et l'option « DSC ». La structure tridimensionnelle de SamAA a été modélisée en se basant sur la structure dimérique optimisée de 3IC1, sur l'alignement de séquences obtenu précédemment, et en utilisant Modeller avec l'option « RefineLoop ». La distance entre Zn2 du site actif de la chaine A et His219.B est égale à 33,7 Å tandis que la distance entre le Zn2 de la chaine B et His219.A est de 21,4 Å, suggérant une conformation très ouverte de l'enzyme modélisée. Le modèle obtenu a une valeur de DOPE Score normalisé égale à 0,58 indiquant une structure de mauvaise qualité (Shen et al., 2006, Eramian et al., 2008). Plusieurs interruptions de chaine peptidique sans résidu manquant sont observées : deux interruptions entre les résidus 410 et 411 de la chaine A et entre les résidus 230 et 231 de la chaine B. Plusieurs structures secondaires prédites par ColabFold sont manquantes : toutes les hélices η, le regroupement des hélices α 6 à 8, le feuillet β à l'extrémité du domaine de dimérisation, le brin β17, l'hélice α12 pour les deux chaines peptidiques et le brin β8 pour la chaine peptidique B (Figure III.21 A).

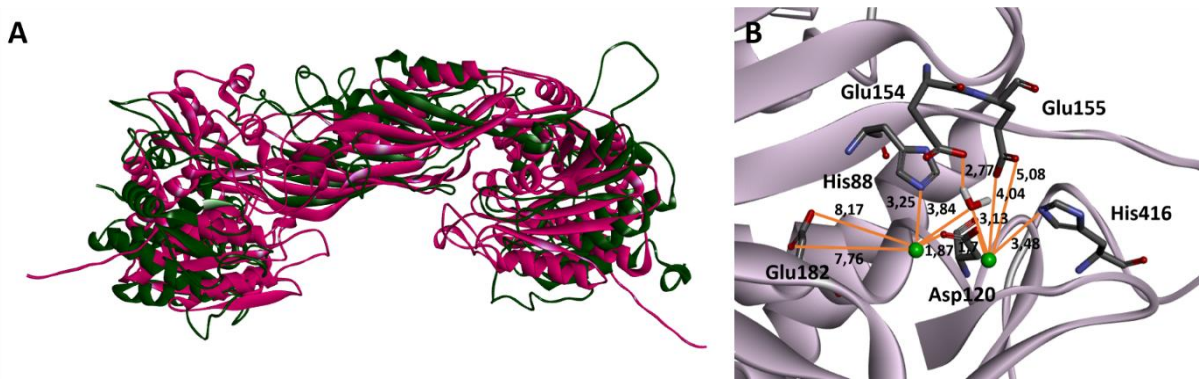


Figure III.21 : A : Superposition des structures 3D de SamAA prédite par ColabFold (en rose) et Modeller en utilisant 3IC1 comme template (en vert foncé). B : Site actif avec la molécule d'eau catalytique, le résidu catalytique et les résidus fixateurs des ions métalliques (en vert clair) dans le modèle de SamAA prédit par Modeller sur la base de 3IC1 comme template.

La géométrie de coordination de type bipyramidale à base trigonale ou carrée au niveau des deux ions métalliques n'est pas respectée. L'architecture du site actif diffère de celle obtenue avec ColabFold. Dans le modèle généré par ColabFold, l'hélice α12 permet la délimitation de la fente catalytique et le bon positionnement du résidu His416, jouant un rôle dans la fixation de Zn2, *via* la présence d'une boucle Phe415-His416. Cette hélice et cette boucle ne sont pas structurées dans le modèle inspiré par 3IC1, provoquant l'éloignement du résidu His416 du

centre métallique (Figure III.21 B). Ce modèle est globalement non-structuré à proximité du site actif, ce qui crée un encombrement stérique dans la fente catalytique et limite l'approche des substrats au plus près du site actif (résultats de docking non montrés ici). Au vu des résultats obtenus, ce modèle a été abandonnée pour la suite de l'étude.

3.4.3.2. Description de la structure de SamAA modélisée par AlphaFold3

La troisième version du logiciel AlphaFold est disponible depuis le printemps 2024 sur le serveur : https://alphafoldserver.com/. Cette nouvelle version permet d'ajouter certains éléments comme des ions métalliques lors de la modélisation d'une structure tridimensionnelle, ce qui évite d'effectuer cette opération manuellement.

Une structure tridimensionnelle supplémentaire de SamAA a été prédite avec AlphaFold3, dans l'espoir d'obtenir un modèle de meilleure qualité pour SamAA, correspondant à une structure suffisamment ouverte pour accueillir l'acide laurique, même s'il est mentionné que le logiciel propose préférentiellement des conformations protéiques fermées (Abramson et al., 2024). Le modèle obtenu a un DOPE score normalisé de -1,66 indiquant une structure tridimensionnelle de meilleure qualité que celle prédite par ColabFold (Figure III.22 A). La distance entre Zn2 d'une chaine et His219 de l'autre chaine est de 7,9 Å, suggérant une meilleure ouverture des sites catalytiques comparativement au modèle issu de ColabFold. Quelques différences sont observées au niveau des structures secondaires : le brin β4 est modélisé alors que l'hélice α12 est absente.

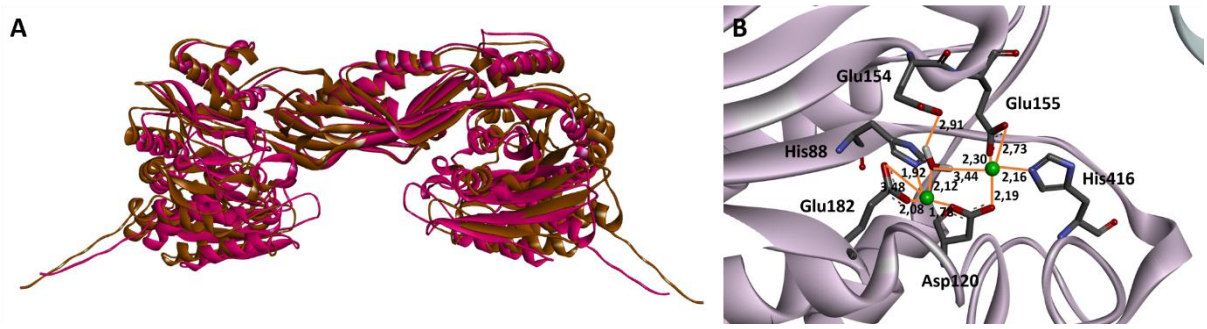


Figure III.22 : A : Superposition des structures 3D des modèles de SamAA issus de ColabFold (en rose) et AlphaFold3 (en marron). B : Site actif avec la molécule d'eau catalytique, le résidu catalytique et les résidus fixateurs des ions métalliques (en vert clair) du modèle de SamAA généré par AlphaFold3.

L'architecture des sites actifs des modèles d'AlphaFold3 et de ColabFold sont similaires. Les cinq résidus fixateurs des deux ions métalliques sont à 2 Å de ces derniers. Le résidu acide base Glu154 est proche de la molécule d'eau catalytique, qui a été ajoutée manuellement. Les deux

ions métalliques sont dans une géométrie de coordination de type bipyramidale trigonale ou carrée déformée (Figure III.22 B). L'ensemble de ces éléments indiquent l'obtention d'une structure tridimensionnelle de SamAA de grande qualité ayant une conformation plus ouverte que celle utilisée jusque-là pour les simulations de docking moléculaire. Malgré une valeur de RMSD faible (0,12 Å sur les chaines peptidiques), la superposition des deux monomères constitutifs de SamAA indique de légères différences structurales entre les deux chaines peptidiques, dont certaines concernent les deux sites actifs. Le site actif de la chaine B présente l'architecture la plus proche de celle attendue, c'est pourquoi celle-ci a été utilisée comme cible de docking.

3.4.4. Résultats de docking avec le modèle de SamAA obtenu avec AlphaFold3

De nouvelle simulations de docking ont été réalisées avec ce dernier modèle de SamAA obtenu avec AlphaFold3 pour tenter d'expliquer la sélectivité de l'enzyme observée expérimentalement envers l'amine α de la lysine et tenter de contourner les difficultés rencontrées précédemment lors des simulations impliquant l'acide laurique. Pour rappel, le modèle comportant la molécule d'eau a été utilisé pour mimer l'hydrolyse de la N-α- et de la N-ε-lauroyl-L-lysine. Le modèle exempt de la molécule d'eau catalytique a servi de cible pour mimer la réaction de N-acylation. Seuls les 20 meilleurs complexes substrats-enzyme ont été analysés. Les mêmes critères que ceux cités précédemment ont été appliqués pour analyser les résultats de docking et conclure sur la faisabilité et la sélectivité des réactions étudiées (Bürgi et al., 1974 ; Lodge et al., 1986 ; Radisky et al., 2002 ; Dutta et al., 2016 ; Nocek et al., 2018 ; Lubkowski et al., 2019).

3.4.4.1. Etude de l'hydrolyse de la N-lauroyl-L-lysine

Les résultats de docking montrent que le substrat N-α-lauroyl-L-lysine a deux orientations préférentielles dans la cavité catalytique de SamAA, correspondant à deux groupes de poses.
Un premier ensemble regroupant les 17 poses les mieux classées, est observé. Dans cette orientation, la liaison amide est proche de la molécule d'eau catalytique (4,2 Å) et de Zn2 (2,2 Å) et l'angle entre le centre électrophile et le nucléophile est de 67 ° (Figure III.23 A).
Dans le second groupe de 3 poses, la fonction carboxyle est orientée vers les deux ions métalliques et la molécule d'eau catalytique. Alors que l'angle pour l'attaque nucléophile est

de 90 °, les distances entre la liaison amide et l'ion Zn2 ou la molécule d'eau catalytique sont comprises entre 6 et 7 Å (Figure III.23 B).

Dans toutes ces poses, la chaine lauroyl et le fragment lysyl interagissent respectivement avec la région hydrophobe et la région hydrophile de la cavité.

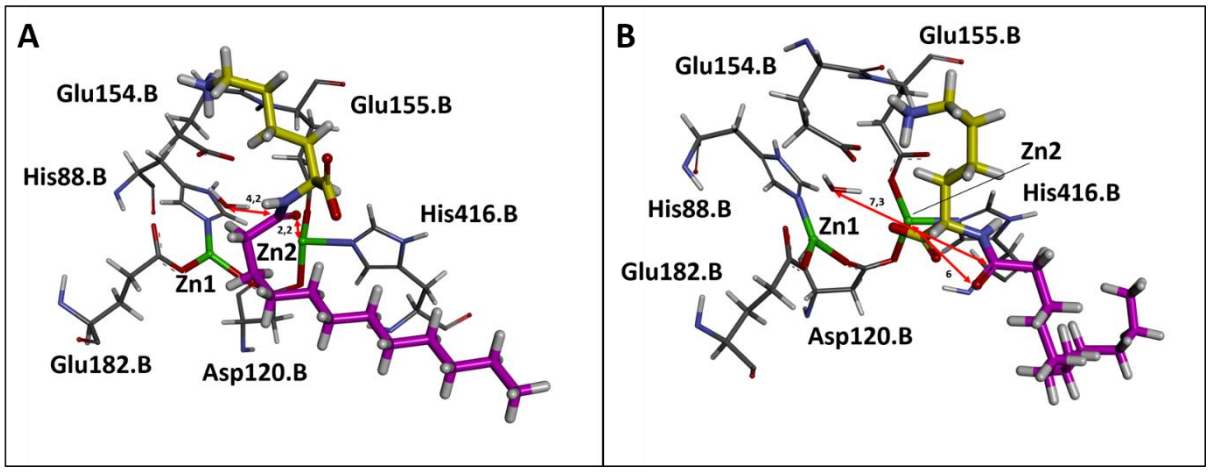


Figure III.23 (A et B) : Orientations préférentielles de la N-α-lauroyl-L-lysine dans le site actif de SamAA. Les deux ions métalliques sont en vert. Les fragments lysyl et lauroyl sont respectivement en jaune et en rose. Les flèches rouges indiquent les distances entre le substrat et Zn2 ou la molécule d'eau catalytique. Les traits noirs indiquent le nom des éléments visibles. Dans le cas de la N-ε-lauroyl-L-lysine, trois principales orientations sont observées parmi les poses issues du docking.

Dans le premier groupe correspondant aux 6 poses les mieux classée, les fragments lauroyl et lysyl s'insèrent respectivement dans les parties hydrophile et hydrophobe de la cavité catalytique. La liaison amide du substrat est à 2,1 Å de Zn2 et à 4,6 Å de la molécule d'eau catalytique avec laquelle elle forme un angle de 37 ° (Figure III.24 A).

Dans le second groupe de 6 poses, le fragment lauroyl est localisé à nouveau dans la partie hydrophile de la cavité. La liaison amide est à environ 8,5 Å du centre catalytique (Figure III.24 B).

Dans le dernier groupe de 8 poses, le fragment lauroyl est dans la partie hydrophobe tandis que le fragment lysyl est dans la partie hydrophile de la cavité. La fonction carbonyle est en interaction avec l'ion Zn2 (distance de 2,1 Å) et la liaison amide est à 4,0 Å du résidu Glu154.B et à 3,9 Å de la molécule d'eau catalytique. L'angle de Bürgi-Dunitz est de 47 ° (Figure III.24 C).

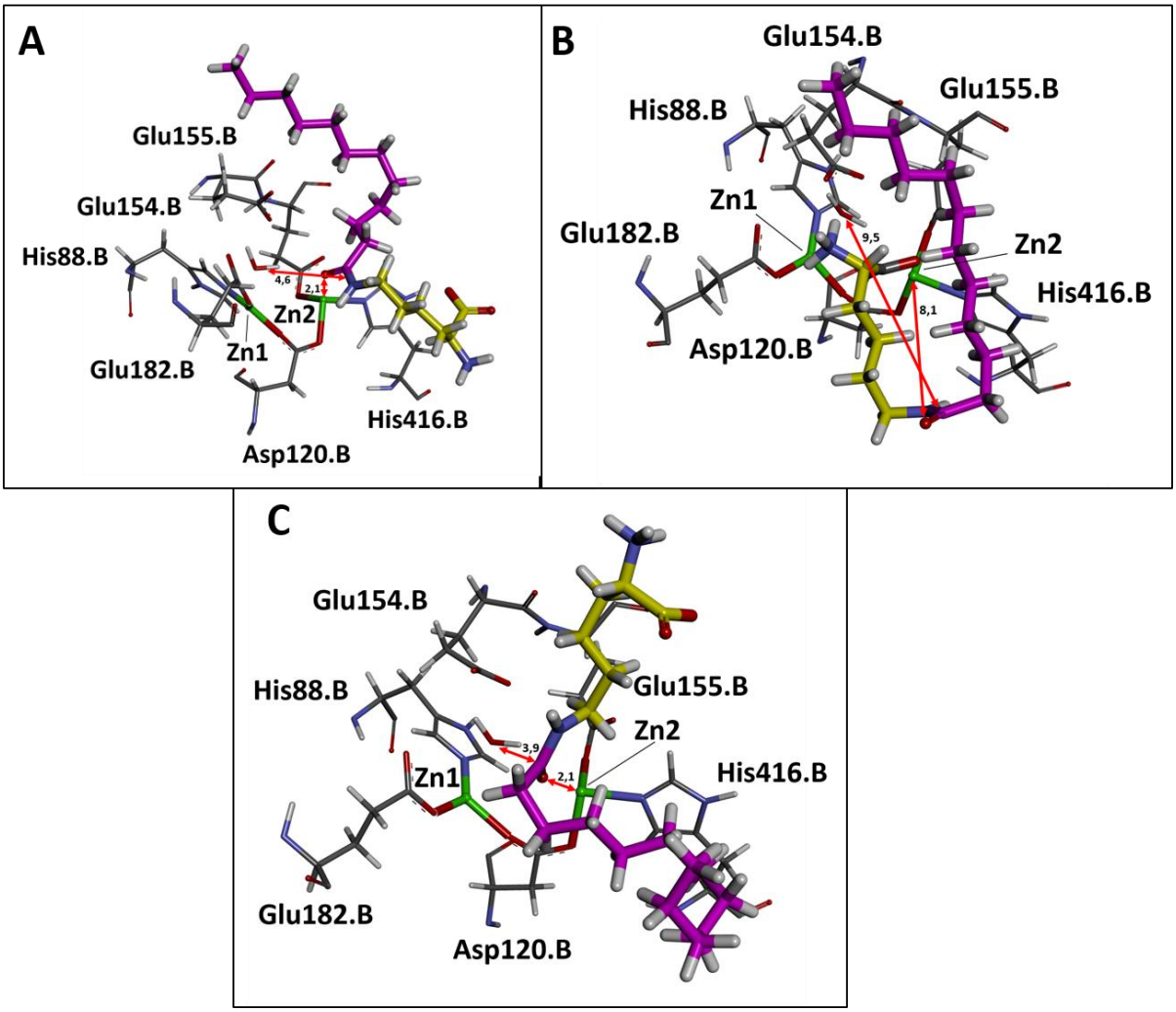


Figure III.24 (A, B et C) : Orientations préférentielles de la N-ε-lauroyl-L-lysine dans le site actif de SamAA. Les deux ions sont en vert. Les fragments lysyl et lauroyl sont respectivement en jaune et en rose. Les flèches rouges indiquent les distances entre le substrat et Zn2 ou la molécule d'eau catalytique. Les traits noirs indiquent le nom des éléments visibles.

En conclusion, trois cas de figure principaux sont relevés parmi les poses mimant l'hydrolyse de la N-lauroyl-L-lysine : i) les substrats adoptent une orientation logique au regard des propriétés de surface de la cavité, ils s'approchent des résidus catalytiques mais l'angle pour l'attaque de la molécule d'eau est plus faible que les valeurs rapportées dans la littérature pour une attaque nucléophile (observé pour la N-α-lauroyl-L-lysine et pour la N-ε-lauroyl-L-lysine), ii) les substrats adoptent une orientation logique au regard des propriétés de surface de la cavité mais ils sont distants du site catalytique (observé pour la N-α-lauroyl-L-lysine), iii) les substrats prennent une orientation illogique au regard des propriétés de surface de la cavité (observé pour la N-ε-lauroyl-L-lysine). Même si aucune conclusion solide ne peut être tirée des modèles quant à la faisabilité de l'hydrolyse, les résultats suggèrent une meilleure orientation du substrat dans le cas de la N-α-lauroyl-L-lysine comparativement à son homologue ε-acylé. Le caractère rigide

de la cible de docking permet très peu d'ajustements de la structure de la protéine en présence du ligand dont la chaîne acyle encombrée peine à se placer dans la cavité. A cela viennent s'ajouter les imprécisions de structure du modèle. Tout ceci pourrait expliquer l'absence de pose « parfaite » sur le plan structural dans le cas de la N-α-lauroyl-lysine. En ce qui concerne la N-ε-lauroyl-lysine, les résultats issus des simulations suggèrent une mauvaise orientation du substrat dans la cavité catalytique de SamAA, pouvant expliquer l'absence d'hydrolyse constatée expérimentalement.

3.4.4.2. Etude de la synthèse de la N-lauroyl-L-lysine

3.4.4.2.1. Docking de l'acide laurique puis de la L-lysine

Une orientation préférentielle de l'acide laurique dans la cavité est observée. La fonction carboxyle de l'acide laurique interagit avec les deux ions métalliques, à une distance moyenne de 2,5 Å. Contrairement à ce qui avait été obtenu avec le modèle issu de ColabFold, la chaine carbonée est positionnée dans la partie hydrophobe de la cavité catalytique et adopte diverses conformations (Figure III.25). La pose ayant obtenu le meilleur score CDOCKER est utilisée pour le docking ultérieur de la lysine.

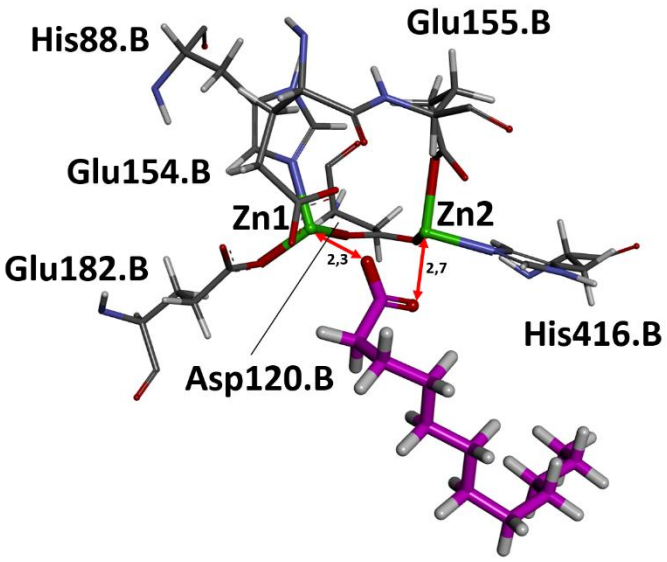


Figure III.25 : Orientation préférentielle de l'acide laurique dans le site actif de SamAA, choisie comme cible de docking de la L-lysine. Les deux ions métalliques sont en vert. L'acide laurique est représenté en rose. Les flèches rouges indiquent les distances entre le substrat et les deux ions métalliques. Les traits noirs indiquent le nom des éléments visibles.

Deux orientations préférentielles de la L-lysine sont obtenues dans le complexe formé par SamAA et l'acide laurique. Pour l'ensemble des poses, la L-lysine est positionnée dans la partie hydrophile de la cavité catalytique.

Dans le premier groupe des 11 poses les mieux classées, la fonction amine α interagit avec la fonction carboxyle de l'acide laurique (4,7 Å) et le résidu Glu154.B (2,3 Å). Ces résultats suggèrent l'acylation de la fonction amine α de la L-lysine même si l'angle d'attaque nucléophile de l'acide laurique est égale à 70 ° (Figure III.26 A).

Dans le second groupe de 9 poses, la fonction amine ε est orientée vers le résidu acide-base Glu154.B avec lequel elle interagit. Toute fois elle est distante de plus de 7 Å du carboxylate de l'acide laurique (Figure III.26 B).

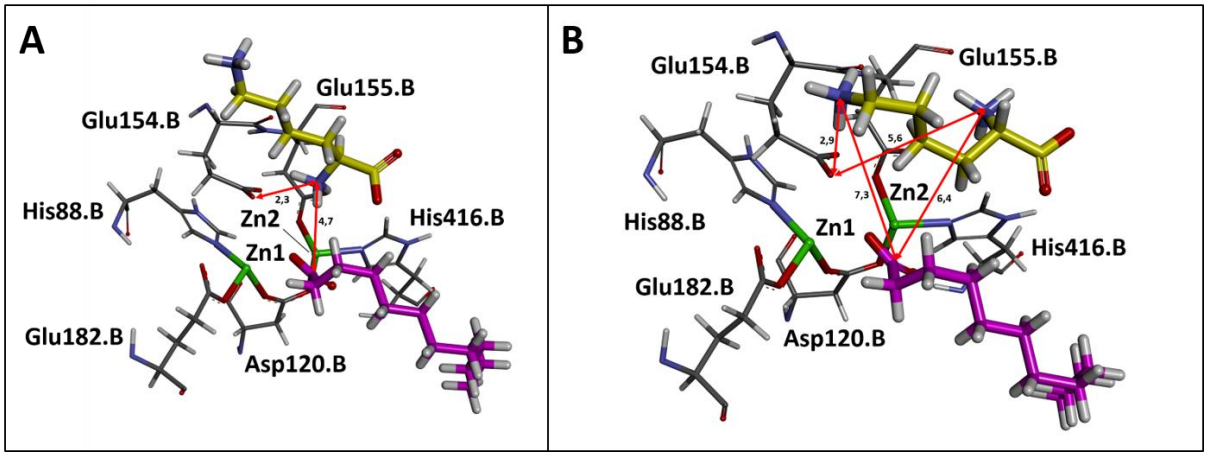


Figure III.26 (A et B) : Orientations préférentielles de la L-lysine dans le complexe SamAA-acide laurique. Les deux ions métalliques sont en vert. La L-lysine et l'acide laurique sont respectivement en jaune et en rose. Les flèches rouges indiquent les distances entre les fonctions amines de la L-lysine et le résidu Glu154 ou la fonction carboxyle de l'acide laurique. Les traits noirs indiquent le nom des éléments visibles.

Ces résultats suggèrent que la fonction amine α de la L-lysine et la fonction carboxylate du donneur d'acyle peuvent s'approcher l'une de l'autre tout en étant à proximité du site catalytique. Seul l'angle correspondant à la réaction entre les deux substrats est plus faible que la valeur attendue de 90 °. L'acylation de la fonction ε de la L-lysine semble inenvisageable du fait de son éloignement par rapport au donneur d'acyle.

3.4.4.2.2. Docking de la L-lysine puis de l'acide laurique

Afin de vérifier si l'ordre d'introduction des substrats dans la cavité catalytique de SamAA impacte les résultats de docking, les modes d'interaction de la L-lysine en tant que premier substrat atteignant la cavité catalytique de SamAA pendant la réaction de synthèse ont été

étudiés selon le même protocole que précédemment. Une seule orientation préférentielle est obtenue dans laquelle la fonction amine ε de la L-lysine pointe vers le résidu Glu154.B (distance de 2,7 Å) tandis que la fonction amine α est orientée vers l'extérieur du site actif, à une distance de 4,2 Å du résidu acide base (Figure III.27 A).

Un complexe SamAA-L-lysine est utilisé pour une simulation de docking de l'acide laurique, où la position de la L-lysine décrite au paragraphe précédent et ayant les meilleurs scores est choisie. De nombreuses et différentes poses de l'acide laurique dans la cavité catalytique sont obtenues, mais aucune d'elles ne montrent le carboxylate de l'acide à proximité de la fonction amine ε de la L-lysine. Certaines poses présentent la fonction carboxylate de l'acide interagissant avec la fonction amine α (distance de 3,1 Å), mais cette fonction carboxylate est à plus de 6 Å des ions métalliques, empêchant son activation et la création du trou oxyanion nécessaire à la réaction de N-acylation (Figure III.27 B).

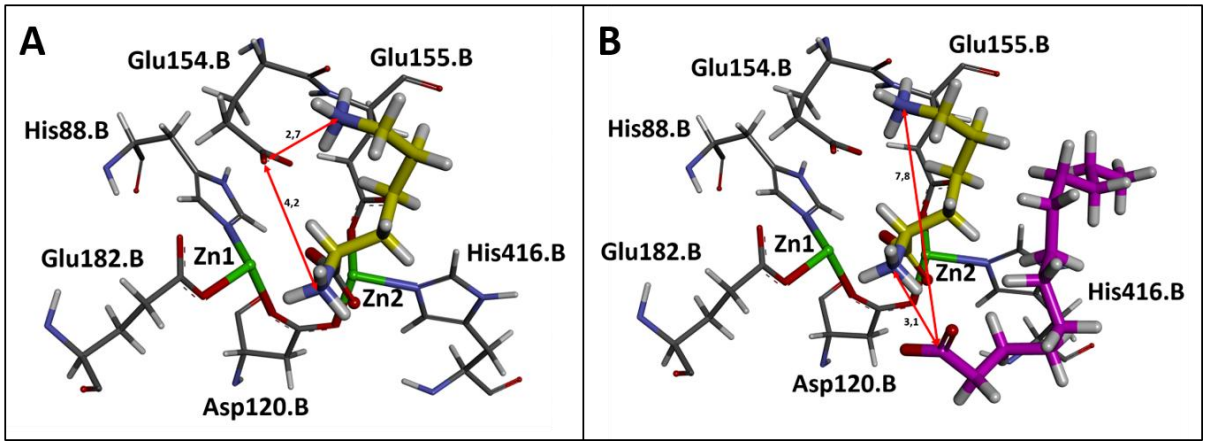


Figure III.27 : A : Orientation préférentielle de la L-lysine dans le site actif de SamAA. B : Exemple d'une orientation de l'acide laurique au sein du complexe SamAA-L-lysine. Les deux ions métalliques sont en vert. La L-lysine et l'acide laurique sont respectivement en jaune et en rose. Les flèches rouges indiquent les distances entre les fonctions amines de la L-lysine et le résidu Glu154 ou le carboxylate de l'acide laurique.

La N-acylation de la L-lysine par l'acide laurique et la réaction inverse d'hydrolyse du dérivé acylé ont été étudiées par des simulations de docking en utilisant deux structures de SamAA prédites respectivement par ColabFold et AlphaFold3. Les résultats issus du modèle de ColabFold suggèrent un positionnement des substrats défavorable à l'hydrolyse de la N-lauroyl-L-lysine avec ce modèle comme cible de docking, que ce soit pour le dérivé α-acylé ou pour son homologue ε-acylé. Les simulations de docking menées sur la structure de SamAA prédite par Alphafold3 suggèrent l'hydrolyse possible de la N-α-lauroyl-L-lysine. Les modes d'interaction entre SamAA et la N-ε-lauroyl-L-lysine semblent quant à eux incompatibles avec son hydrolyse. Des modes d'interaction cohérents ont été obtenus entre l'acide laurique, la L-

lysine et les structures de SamAA générées par Alphafold3 et Colabfold, suggérant un placement des substrats favorable à l'acylation de la fonction amine α de la lysine, ce qui est en accord avec les résultats expérimentaux (Dettori et al., 2018a ; Bourkaib et al., 2020a). L'ordre d'introduction des substrats a également eu un impact au cours des simulations de docking, les meilleurs résultats étant obtenus avec le modèle issu d'Alphafold3 avec docking initial de l'acide laurique suivi du docking de la L-lysine. Il semble donc que la structure de SamAA prédite par Alphafold3 conduit globalement à des résultats cohérents avec les observations expérimentales. Ceci pourrait s'expliquer par une structure plus ouverte que celle générée par Colabfold, et pouvant accueillir l'acide laurique. D'après l'étude de Nocek et al., l'arrivée du substrat dans le site actif de la succinyl diaminopimelate desuccinylase de *H. influenzae* qui est un template de SamAA, provoque le passage d'une conformation ouverte à une conformation fermée de l'enzyme. La distance entre le résidu His194.B et l'ion Zn2.A passe de 10,1 Å dans la conformation ouverte à 2,9 Å dans la conformation fermée (Nocek et al., 2018). Dans le premier modèle de SamAA généré par ColabFold, la distance entre ces deux éléments est de 5,3 Å, suggérant une conformation semi-ouverte avec un site actif trop petit pour accueillir des substrats encombrants tels que l'acide laurique.

Deux stratégies ont été suivies pour l'obtention d'une conformation ouverte de SamAA. La première a consisté à utiliser un template présentant une conformation ouverte (la succinyl diaminopimelate desuccinylase de *H. influenzae* ; entrée PDB 3IC1) pour guider la modélisation d'homologie. Toutefois la structure de référence est de mauvaise qualité, entraînant des imprécisions structurales notamment au niveau du site actif. La deuxième stratégie a été d'utiliser la version la plus récente d'AlphaFold (AlphaFold3). Le modèle obtenu peut être qualifié de robuste et correspond à une structure de SamAA plus ouverte que celles précédemment générées (distance His219.B-Zn2.A de 7,9 Å). Comme dit précédemment, les simulations mimant la réaction de synthèse menées avec cette structure de SamAA sont globalement compatibles avec les observations expérimentales si l'on se base sur les distances de référence impliquées dans le mécanisme catalytique supposé. Toutefois pour tous les modèles, des valeurs d'angle d'attaque nucléophile inférieures ont été rapportées à celles dans la littérature. Deux valeurs sont couramment citées pour l'angle de Bürgi-Dunitz : 105 ° dans le cas d'une attaque nucléophile impliquant des petites molécules et 90 ° dans le cas des enzymes (Burgi et al., 1972 ; Lubkowski et al., 2019). Quel que soit la valeur admise, l'angle mesuré au sein des complexes issus des simulations de docking semble trop fermé pour permettre une attaque nucléophile. Ceci pourrait s'expliquer par une ouverture de site encore insuffisante pour accueillir l'acide laurique correctement.

3.5. Conclusions

Au cours de ce chapitre, un modèle robuste de la structure tridimensionnelle de SamAA prédite par modélisation par homologie a été proposé. Des templates proches de SamAA du point de vue de leur séquence ont été identifiés et utilisés pour orienter la modélisation de sa structure tridimensionnelle et pour identifier des résidus conservés ayant un rôle dans la catalyse ou dans la fixation des ions métalliques. Les premiers éléments de compréhension et d'information sur la sélectivité de SamAA lors de la synthèse/hydrolyse de dérivés N-acylés de la L-lysine ont été apportés par l'étude des modes d'interaction entre l'enzyme et ses substrats grâce à des simulations de docking.

Cinq templates ont été identifiés pour SamAA par recherche d'homologie : l'aminoacylase 1 humaine (entrée PDB 1Q7L), les succinyl-diaminopimelate desuccinylases de *Neisseria meningitidis* MC85 (entrée PDB 4PPZ), de *Haemophilus influenzae* (entrée PDB 5VO3) et de *Shigella flexneri* (entrée PDB 7LGP), et la carnosinase 2 humaine (entrée PDB 4RUH). En faisant l'hypothèse d'une organisation similaire chez SamAA, l'étude de la structure des templates a permis d'obtenir des informations sur l'organisation du site actif. Il serait présent dans chaque chaine peptidique et contiendrait une molécule d'eau catalytique et deux ions Zn^{2+} dans une géométrie de coordination bipyramidale trigonale déformée. Des résidus clés ont été identifiés, jouant un rôle dans la fixation des ions métalliques (His88.A, Asp120.A, Glu155.A, Glu182.A et His416.A), dans la catalyse (Asp90.A, Glu154.A, Asp317.A, Arg319.A, His219.B et Asn306.B) ou encore au niveau de la boucle de spécificité de l'enzyme (Ser382.A, Gly383.A, Gly384.A et Thr385.A).

Quatre outils de modélisation par homologie ont été utilisés pour prédire la structure tridimensionnelle de SamAA : Modeller, Swiss Model, ColabFold (interface web de AlphaFold2) et AlphaFold3. La qualité des modèles générés a été évaluée en se basant sur une fonction de score (DOPE Score normalisé) et la similarité de l'organisation des sites actifs de SamAA et de la succinyl-diaminopimelate desuccinylase de *H. influenzae*. Les deux meilleurs modèles de la structure tridimensionnelle de SamAA sont ceux produits par ColabFold et AlphaFold3, avec des valeurs respectives de DOPE Score normalisé égales à -1,50 et à -1,66. La géométrie des sites actifs dans ces deux modèles est similaire à celle observée dans la structure résolue du template. SamAA serait un homodimère binucléaire dont le monomère en forme de sifflet présente un domaine catalytique et un domaine de dimérisation.

Pour expliquer la sélectivité de SamAA, les modes d'interactions entre l'enzyme et les substrats des réactions de N-acylation et de N-déacylation ont été étudiés. Ainsi, la structure de SamAA

prédite par ColabFold a été utilisée comme cible pour le docking de plusieurs substrats : N-α/ε-acétyl-L-lysine, N-α/ε-caproyl-L-lysine, L-lysine, acide acétique, acide caproïque. L'acylation et la déacylation sélective de la fonction amine α de la L-lysine est fortement suggérée par les modèles, tandis que l'acylation et la déacylation de la fonction amine ε paraît peu probable, ce qui est cohérent avec les résultats expérimentaux obtenus. Sur la base des interactions enzyme-substrat identifiées et des informations disponibles dans la littérature sur la succinyl-diaminopimelate desuccinylase de *H. influenzae*, un mécanisme de l'hydrolyse et de la synthèse de la N-α-caproyl-L-lysine a été proposé.

La simulation de l'hydrolyse et de la synthèse de la N-lauroyl-L-lysine a mis en évidence des incohérences dans le placement des substrats au niveau de la structure de SamAA générée par ColabFold. L'hypothèse d'une conformation de l'enzyme insuffisamment ouverte pour accueillir la chaine lauroyle a été formulée. La construction d'un modèle de l'enzyme sur la base d'un template en conformation ouverte a conduit à une structure tridimensionnelle incomplète et peu fiable. En revanche, le modèle de SamAA prédit par AlphaFold3 est robuste et présente une ouverture de site supérieure à celle observée dans le modèle utilisé jusque-là. Les résultats des simulations indiquent globalement une orientation et une proximité des substrats compatibles avec l'acylation/déacylation de la fonction amine α de la L-lysine, observée expérimentalement.

Ces travaux apportent donc des premiers éléments de réponse concernant la sélectivité particulière de SamAA. Ils ont par ailleurs permis de progresser dans la connaissance de la structure de l'enzyme, de l'organisation de son site actif et de ses interactions avec les substrats de la réaction de N-acylation. Il faut toutefois souligner qu'ils sont issus de modèles, ces derniers présentant par définition des imperfections et des imprécisions qu'il faut garder à l'esprit. La validation du modèle de la structure tridimensionnelle de SamAA et des résultats associés ne pourra être effective qu'après l'élucidation expérimentale de la structure de l'enzyme. SamELA est une autre aminoacylase produite de *S. ambofaciens*, dont la sélectivité envers la fonction amine ε de la L-lysine a été démontrée (Dettori et al., 2018a). Elle est intéressante également, car elle catalyse la N-acylation en milieu aqueux avec une sélectivité similaire à celle des lipases en milieu organique. En attendant la résolution de sa structure, des travaux ont été entrepris pour progresser dans la découverte de sa structure et tenter d'expliquer ses propriétés de sélectivité. La démarche suivie est similaire à celle décrite pour SamAA ; les résultats obtenus font l'objet du chapitre suivant.

4. Chapitre IV : Modélisation de la N-acylation de la L-lysine catalysée par SamELA

4.1. Introduction

Des travaux antérieurs menés sur la N-acylation d'acides aminés et de peptides ont montré le potentiel de quelques enzymes pour catalyser cette réaction. Parmi elles, deux familles ressortent plus particulièrement, les lipases et les aminoacylases. Les lipases sont mises en œuvre dans différents milieux comme les solvants organiques pauvres en eau, les systèmes sans solvant appelés aussi milieux fondus ou encore les liquides ioniques (Montet et al., 1990 ; Soo et al., 2004 ; Husson et al., 2008 ; Husson et al., 2009 ; Kua et al., 2023). Les aminoacylases quant à elles, sont des enzymes décrites pour leur capacité à catalyser la réaction de N-acylation en milieu aqueux, favorisant la solubilisation du substrat aminé et réduisant l'impact environnemental du procédé comparativement aux réactions enzymatiques menées en milieu organique.

Plusieurs aminoacylases (Sm-ELA, Sm-AA et Sm-PVA) catalysant la N-acylation de plusieurs acides aminés ont été découvertes chez *Streptomyces mobaraensis* par l'équipe de Koreishi (Koreishi et al., 2005a ; Koreishi et al., 2007 ; Koreishi et al., 2009a ; Koreishi et al., 2009b). Depuis 2014, d'autres aminoacylases produites par *Streptomyces ambofaciens* ont été identifiées, dont les gènes partagent une forte identité avec les gènes de Sm-AA, Sm-ELA et Sm-PVA. Parmi ces enzymes, deux d'entre elles ont particulièrement retenu l'attention des chercheurs. SamAA catalyse sélectivement l'acylation et la déacylation de la fonction amine α de la L-lysine comme montré dans le chapitre précédent. SamELA montre une sélectivité stricte vis-à-vis de la fonction amine ε de la L-lysine. Bien que ces enzymes aient été utilisées jusqu'ici en mélange, leurs propriétés de sélectivité ont pu être démontrées grâce à des expériences menées après délétion du gène codant pour l'une ou l'autre enzyme. Ainsi, une diminution spectaculaire de l'hydrolyse de la N-α-acétyl-L-lysine a été observée en présence des fractions provenant du mutant dépourvu du gène de SamAA. De manière similaire, la délétion du gène codant pour SamELA conduit à une chute de l'activité catalytique d'hydrolyse de la N-ε-acétyl-L-lysine (Dettori et al., 2018a). Par ailleurs, d'autres travaux ont confirmé la spécificité de SamELA pour l'acylation de la fonction amine ε de la L-lysine par des donneurs d'acyle de moins huit atomes de carbone (Bourkaib et al., 2020a). A ce jour, aucune explication n'a jamais

été donnée pour ces propriétés catalytiques, que ce soit pour les aminoacylases de *S. mobaraensis* ou celles produites par *S. ambofaciens*.

Ce chapitre est consacré à l'étude de l'aminoacylase SamELA. Cette enzyme catalyse en milieu aqueux la N-acylation et la déacylation de la L-lysine, avec la même régiosélectivité que celle observée en présence de lipases en milieu organique. L'objectif principal de cette étude est d'expliquer la sélectivité particulière de SamELA à l'échelle moléculaire, à partir d'éléments structurels. Cependant, aucune structure élucidée de l'enzyme n'est à ce jour disponible. Pour contourner cette absence de données expérimentales, une stratégie de modélisation par homologie a été adoptée pour prédire la structure de SamELA, en s'appuyant sur l'approche méthodologique développée pour l'étude de SamAA (Muhammed et al., 2018 ; Haddad et al., 2020). Une fois le modèle construit, celui-ci a été utilisé comme cible de docking pour étudier les modes d'interaction les plus probables avec les substrats de la réaction de N-acylation de la L-lysine en présence de l'acide caprylique ou l'acide laurique comme donneurs d'acyle, et tenter d'expliquer la sélectivité observée expérimentalement (Selvam et al., 2017). En parallèle de cet objectif principal, les travaux menés visaient à apporter des informations sur cette enzyme très peu connue, notamment sa structure, l'organisation de son site catalytique, les résidus clés pour la catalyse et la fixation des substrats, en se basant sur le modèle construit et sur le peu de données de la littérature portant sur les enzymes homologues.

4.2. Identification des templates de SamELA et étude de leurs principales caractéristiques structurelles

Les templates ont été identifiés après avoir effectué une recherche de similarité à l'aide des outils NCBI et Swiss Model, permettant d'obtenir un alignement des séquences protéiques et indiquant le pourcentage d'identité avec la séquence de SamELA. Les pourcentages d'identité et de similarité de séquences entre les templates identifiés et SamELA sont récapitulés dans le tableau IV.1.

PfH et PhH sont des hydrolases non-caractérisées dépendantes des métaux de *Pyrococcus furiosus* et de *Pyrococcus horikoshii* (Bonanno et al., 2009 ; Patskovsky et al., 2009). NatAM est une amidohydrolase de *Streptomyces* sp. Tu 6176 et catalyse l'amidation de l'ester dimérique 3-HAA (acide 3-hydroxyanthranilique) en benzoxazole qui est une étape clé de la biosynthèse du nataxazole (agent antitumoral) (Song et al., 2020). Afa est un mutant D366A de la D-aminoacylase d'*Alcaligenes faecalis* qui est une enzyme utilisée pour la désacétylation des acides N-acétyl-D-aminés (Lai et al., 2004). hCRMP-2 est un mutant S522D de la protéine

médiatrice de la réponse à la collapsine 2 humaine, ayant un rôle de régulation dans la croissance et la stabilité des microtubules neuronaux (Sumi et al., 2018).

Tableau IV.1 : Pourcentages d'identité et de similarité de séquences entre les templates identifiés et SamELA.

Nom de l'enzyme	Nom abrégé	Entrée PDB	Pourcentage d'identité	Pourcentage de similarité
Hydrolase métallo dépendante non-caractérisée de *Pyrococcus furiosus*	PfH	3ICJ	29 %	34 %
Hydrolase métallo dépendante non-caractérisée de *Pyrococcus horikoshii*	PhH	3IGH	26 %	34 %
Amidohydrolase de *Streptomyces* sp. Tu 6176	NatAM	6SJ0	17 %	27 %
Mutant de la D-aminoacylase d'*Alcaligenes faecalis*	AfA	1RJQ	15 %	26 %
Mutant de la protéine médiatrice de la réponse à la collapsine 2 humaine	hCRMP-2	5YZA	12 %	25 %

Les cinq templates ont des structures tertiaires différentes. PfH et Afa sont des monomères tandis que PhH et NatAM sont des homodimères et hCRMP-2 est un homotétramère. L'analyse visuelle des structures montre que la présence et le nombre d'ions métalliques et la présence d'une molécule d'eau catalytique diffère également selon les templates. PhH et hCRMP-2 n'ont pas d'ion métallique ni de molécule d'eau catalytique au sein de leur structure élucidée. Afa présente un ion métallique et deux ions acétate, mais pas de molécule d'eau catalytique. L'ion métallique est lié à un ion acétate, deux histidines et une cystéine. NatAM a un ion métallique lié à trois histidines, un aspartate et à une molécule d'eau catalytique, elle-même lié à une histidine et à l'aspartate précédemment évoqué. PfH a deux ions métalliques fixés par quatre histidines, un aspartate et une N(6)-carboxy-L-lysine et n'a pas de molécule d'eau catalytique.

4.3. Analyse de la séquence protéique de SamELA

Les templates identifiés pour SamELA sont peu décrits dans la littérature et leur activité catalytique n'a pas été caractérisée. A notre connaissance, aucune étude par mutagénèse dirigée n'a été menée pour confirmer le rôle des résidus cités précédemment dans la fixation des ions métalliques dans le site actif ou pour identifier les résidus essentiels à la catalyse. Toutefois, à partir des alignements de séquences des templates et de SamELA, il est possible d'identifier

des résidus conservés et de supposer qu'ils jouent un rôle analogue chez SamELA. Ainsi, les résidus His99, His101, Lys294, His346, His371 et Asp432 de PfH seraient impliqués dans la fixation des deux ions métalliques, tout comme les résidus conservés His71, His73, Asp265, His316 et His351 de SamELA (Figure IV.1).

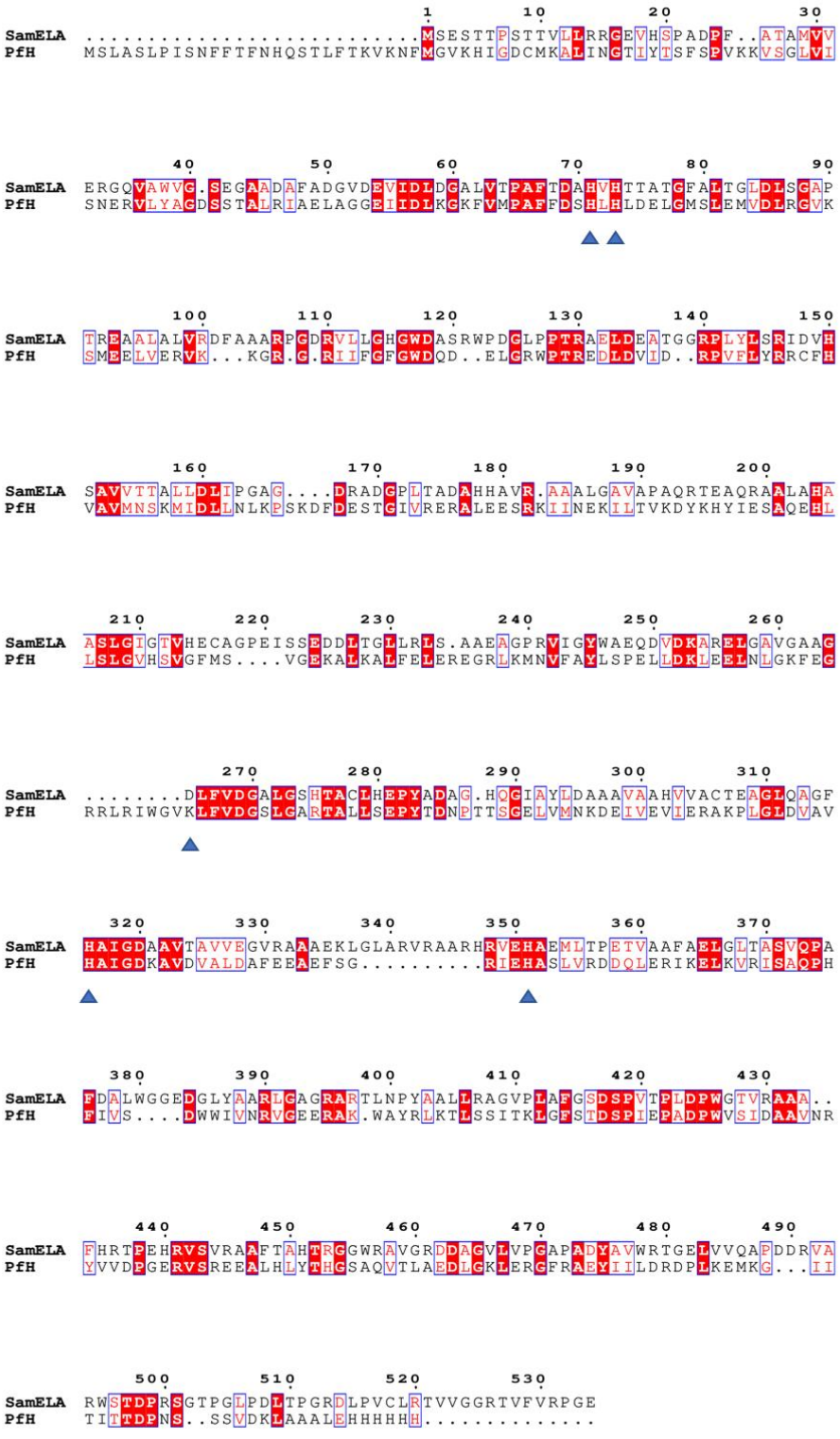


Figure IV.1 : Alignement des séquences de SamELA et du template PfH. Les résidus supposés être impliqués dans la fixation des ions métalliques sont annotés avec un triangle (Robert et al., 2014).

Le motif His71-X-His73 identifié précédemment et impliqué dans la fixation des ions métalliques semble un élément particulier. Seule la famille des métalloenzymes M38 présente un motif similaire, situé approximativement à la même place dans la séquence protéique. La β-aspartyl dipeptidase est une enzyme produite par le gène *iadA* chez *Escherichia coli*. Elle est l'enzyme de référence de cette famille des peptidases (EcIAD ; entrée PDB 1ONW). Les enzymes de cette famille sont des métalloenzymes présentant quatre histidines et une lysine modifiée impliquées dans la fixation de deux ions métalliques dans le site actif et un aspartate jouant le rôle de base catalytique. Le site actif contient par ailleurs une molécule d'eau catalytique liée aux deux ions Zn^{2+}. Les résidus His68, His70, Lys162, His201 et His230 fixent les deux ions métalliques de EcIAD. Le résidu Asp285 est un résidu base essentiel à l'activité de l'enzyme (Thoden et al., 2003 ; Jozic et al., 2003). A l'exception du résidu Lys162 qui n'est pas correctement identifié dans la séquence de EcIAD, les quatre autres résidus (His68, His70, His316 et His351) sont conservés chez SamELA, confirmant le rôle probable de ces résidus comme fixateurs des ions métalliques. Par ailleurs, le résidu conservé Asp417 de SamELA constituerait un résidu catalytique essentiel (Figure IV.2).

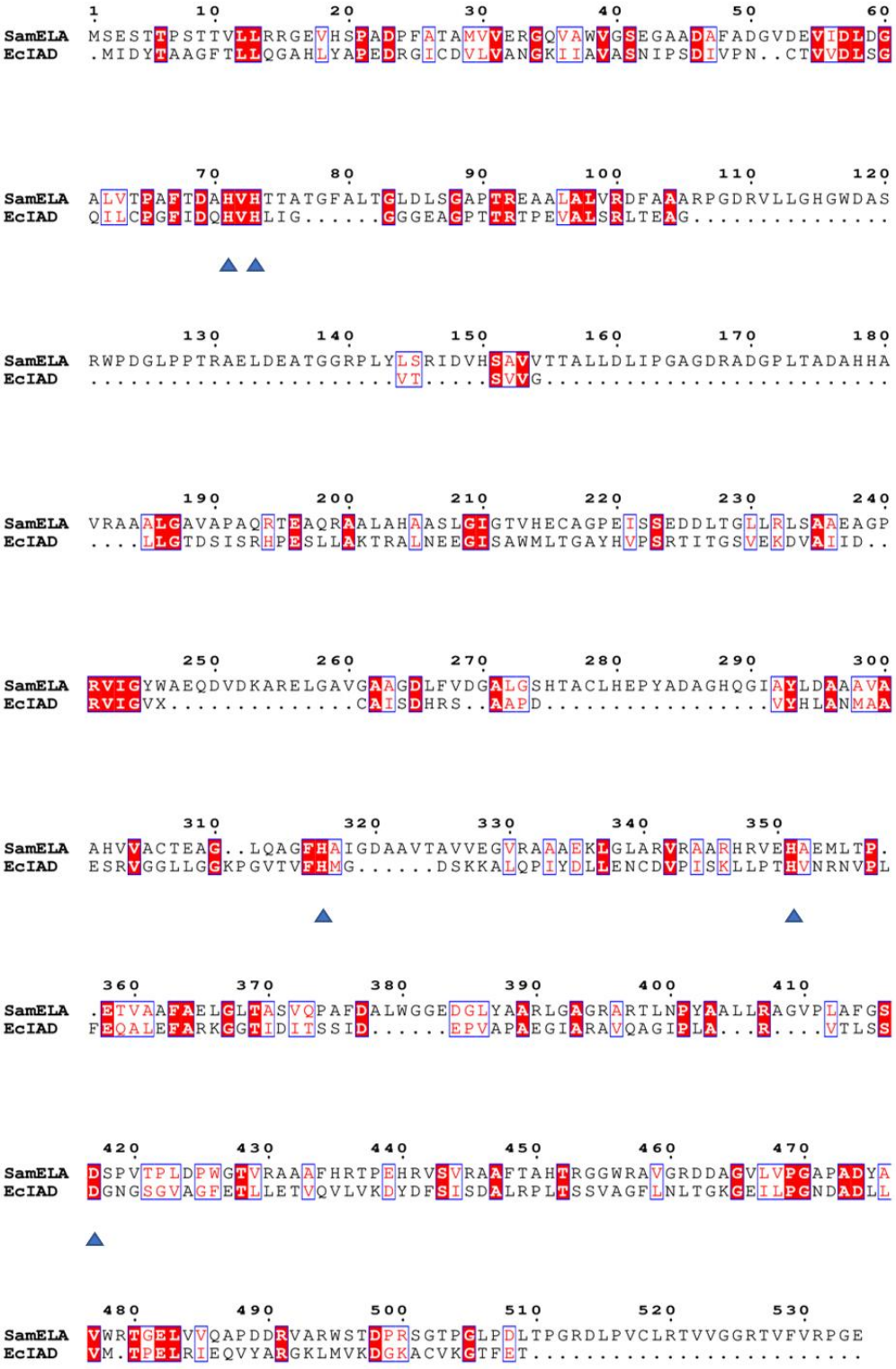


Figure IV.2 : Alignement des séquences de SamELA et de EcIAD. Les résidus supposés être impliqués dans la fixation des ions métalliques ou dans la catalyse sont annotés avec un triangle (Robert et al., 2014).

4.4. Modélisation de la structure tridimensionnelle de SamELA

Les performances de trois outils de modélisation par homologie (Modeller, Swiss Model et ColabFold) ont été étudiées et discutées dans le cadre de la modélisation de la structure de SamAA présentée dans le chapitre précédent. Le choix de ces outils avait été guidé par leur excellent classement lors des concours CASP13 et CASP14 (Kryshtafovych et al., 2019 ; Krystafovysh et al., 2021). Swiss Model avait permis de prédire la structure homodimérique de SamAA tandis ColabFold avait conduit aux modèles de SamAA les plus robustes. En 2024, DeepMind de Google présente la troisième version d'AlphaFold, AlphaFold3, dont la principale caractéristique est de pouvoir prédire la structure des protéines et leurs interactions avec d'autres types de molécules (métaux, ligands, …) en utilisant des méthodes d'intelligence artificielle. Dans ce chapitre, CobalFold, Swiss Model et AlphaFold3 ont été utilisés pour prédire la structure tertiaire et quaternaire de SamELA. Comme précédemment, la qualité des modèles tridimensionnels obtenus a été évaluée par le DOPE Score normalisé et en se basant sur le maintien des éléments structuraux identifiés lors des alignements de séquences présentés juste avant.

Quatre modèles de SamELA ont été obtenus en utilisant Swiss Model avec les templates PfH, NatAM, Afa et hCRMP-2. Alors que ces derniers sont caractérisés par des structures quaternaires différentes, allant du monomère à l'homotétramère, tous les modèles construits présentent une structure monomérique. L'hypothèse selon laquelle SamELA serait un monomère a donc été posée.

4.4.1. Description de la structure de SamELA prédite par ColabFold

Une séquence monomérique a été soumise à ColabFold pour la prédiction de la structure tridimensionnelle de SamELA. 11847 alignements entre la séquence de SamELA et des séquences partielles ou complètes issues des bases de données PDB70 et UniRef ont été générées par MMseq2. Le modèle ayant la valeur de DOPE Score la plus basse (-1,12) est décrit ci-dessous ; il a été choisi comme cible pour les simulations de docking ultérieures. La structure a été optimisée par minimisation de son énergie potentielle. Deux ions métalliques et une molécule d'eau catalytique ont été ajoutés pour respecter l'architecture du site actif des peptidases de la famille des M38.

Le monomère de SamELA comporte 534 résidus. Il est composé de trois domaines distincts. La première partie centrale (résidus 67-82 et 190-465) est constituée d'un tonneau $(\beta/\alpha)_8$, c'est-

à-dire un feuillet β de huit brins (β7, β13, β14, β15, β16, β17, β18 et β19) entouré par huit hélices α (α2, α6, α7, α8, α9, α10, α11 et α12) et quelques hélices 3_{10} (η3, η4, η5, η6, η7, η8 et η9). Cette organisation est caractéristique de la famille des peptidases M38. Le second domaine cylindrique (résidus 1-66 et 466-534) est formé d'une hélice α (α1), d'un feuillet β de neuf brins (β1, β2, β3, β4, β5, β6, β20, β23 et β24) et d'une boucle peu structurée (β21, β22, η10 et η11). Le troisième élément globulaire (résidus 83-189) est composé d'un feuillet β de cinq brins (β8, β9, β10, β11 et β12), de trois hélices α (α3, α4 et α5) et de 2 hélices 3_{10} (η1 et η2) (Figure IV.3A).

Le site actif est localisé à la surface du premier domaine et à l'interface du premier et du troisième domaines. Il est placé dans une cavité très fermée. Les ions métalliques n'adoptent pas une géométrie de coordination de type bipyramidale ou carrée car ils sont distants de plus de 2 Å des résidus conservés (Figure IV.3B).

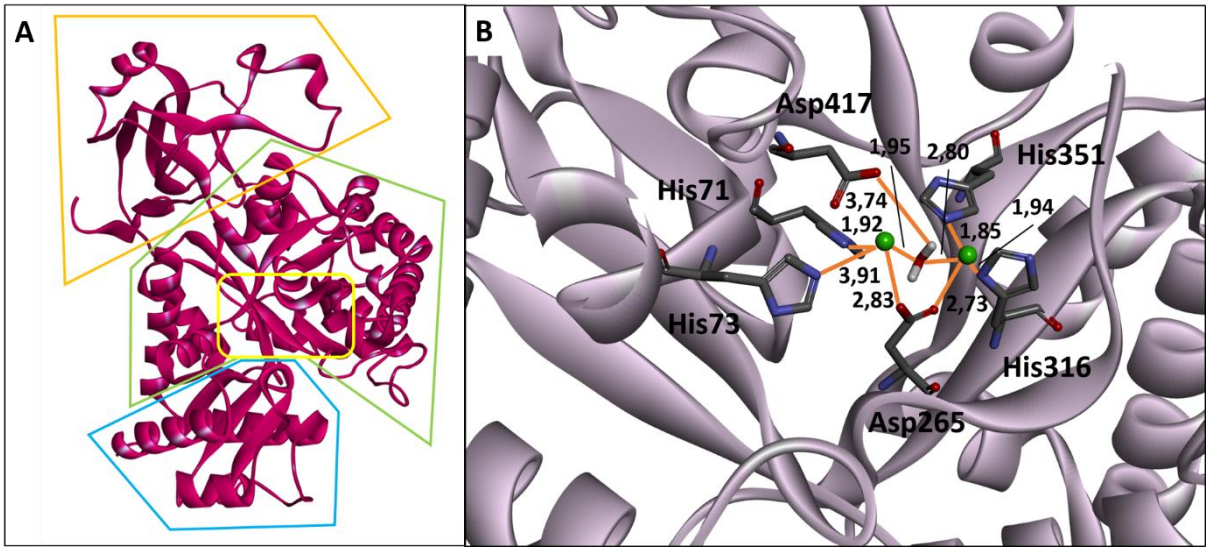


Figure IV.3 : A : Structure tridimensionnelle monomérique de SamELA prédite par ColabFold. Les formes de couleur verte, orange et bleue indiquent respectivement la position du premier, second et troisième domaine de SamELA. Le rectangle jaune indique la position du site actif. B : Site actif de SamELA prédit par ColabFold, avec la molécule d'eau catalytique, le résidu catalytique, les résidus fixateurs des ions métalliques et les deux ions métalliques (en vert). Les traits noirs indiquent la longueur des liaisons.

Malgré l'ajout des deux ions métalliques et de la molécule d'eau catalytique et la mise en œuvre d'une minimisation d'énergie, les différentes simulations de docking réalisées en utilisant ce modèle comme cible ne permettent pas d'expliquer la sélectivité de SamELA observée expérimentalement (résultats non-montrés). La qualité insuffisante du modèle suggérant des imprécisions structurales pourrait expliquer ces résultats. C'est pourquoi de nouveaux modèles de SamELA ont été générés en utilisant AlphaFold3, qui pour rappel, permet d'intégrer la présence des ions métalliques dans la prédiction de la structure.

4.4.2. Description de la structure de SamELA prédite par AlphaFold3

Une séquence monomérique de SamELA contenant deux ions métalliques a été soumise à l'interface web d'AlphaFold3 pour la construction de nouveaux modèles de SamELA. Une valeur de DOPE score normalisé de -1,28 a été obtenue pour le meilleur modèle, indiquant une qualité de prédiction supérieure à celle obtenue précédemment avec ColabFold. Cependant, peu de différences structurales sont observées entre les deux modèles : une boucle 3_{10} (η10) devient un brin β dans le second domaine ; une valeur faible de RMSD égale à 1,32 Å est obtenue lors de la superposition de deux meilleurs modèles issus de ColabFold et d'AlphaFold3 (Figure IV.4A).

La localisation du site actif n'est pas modifiée. Les deux ions métalliques ont été placés automatiquement par le logiciel à proximité des acides aminés précédemment identifiés (distance moyenne de 2,1 Å). Une molécule d'eau a été ajoutée manuellement à proximité près des deux ions Zn^{2+} et du résidu Asp417 de sorte qu'elle se situe à environ 2 Å de ces trois éléments. Zn1 est lié aux résidus His 71, His73, Asp265 et à la molécule d'eau catalytique tandis que Zn2 est lié aux résidus Asp265, His316, His351 et à la molécule d'eau catalytique. Les deux ions présentent une géométrie de coordination de type tétraédrique dans laquelle la valeur moyenne des angles est d'environ 107 ° pour les deux centres métalliques. La distance entre les deux ions métalliques est égale à 3,36 Å (Figure IV.4B). Ce modèle a été utilisé comme cible pour miner la réaction de synthèse de dérivés acylés de la L-lysine.

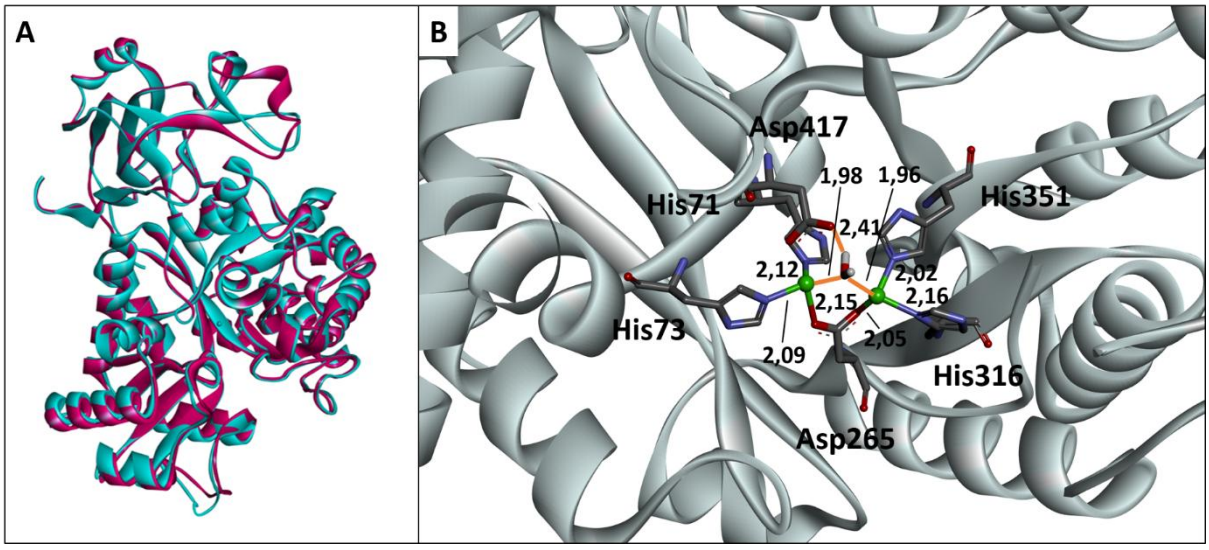


Figure IV.4 : A : Superposition des structures tridimensionnelles de SamELA prédites par ColabFold (en rose) et par AlphaFold3 (en bleu). B : Site actif de SamELA prédit par AlphaFold3, avec la molécule d'eau catalytique, le résidu catalytique, les résidus fixateurs des ions métalliques et les deux ions métalliques (en vert). Les traits noirs indiquent la longueur des liaisons.

4.5. Modélisation de la synthèse de la N-capryl- et de la N-lauroyl-L-lysine

Lors de travaux précédents, la sélectivité de SamELA envers la fonction amine ε de la L-lysine a été démontrée expérimentalement en présence de l'acide caprylique qui est un donneur d'acyle à huit atomes de carbone (Bourkaib et al., 2020a). Dans cette étude, des simulations de docking ont été menées pour miner cette réaction et identifier des éléments déterminants pouvant expliquer la sélectivité de l'enzyme. Le modèle de SamELA prédit par AlphaFold3, sans molécule d'eau catalytique, a été choisi comme cible pour le docking des substrats de la réaction de synthèse de la N-ε-capryl-L-lysine. Parmi les éléments analysés au sein des poses obtenues, une attention particulière a été portée à l'orientation de la L-lysine ainsi qu'au positionnement de la fonction carboxylate du donneur d'acyle. Pour compléter cette analyse, en émettant l'hypothèse de mécanismes similaires pour les métalloenzymes de la famille M20 et celles de la famille M38, la proximité de la fonction carboxylate aux ions Zn^{2+} et à l'une des fonctions amine de la L-lysine, ainsi que l'angle de l'attaque nucléophile entre les deux substrats ont été mesurés pour évaluer la faisabilité et la sélectivité de la réaction de N-acylation. Pour rappel, la fonction carboxylate du donneur d'acyle interagit avec les ions Zn^{2+} (distance de 2 Å) ; les groupes respectivement donneur et accepteur d'acyle des deux substrats sont suffisamment proches l'un de l'autre et forment un angle compatible avec une attaque nucléophile (distance de 2,5 Å et angle de Bürgi-Dunitz de 90 °) (Bürgi et al., 1974 ; Lodge et al., 1987 ; Radisky et al., 2002 ; Xiang et al., 2010 ; Nocek et al., 2018 ; Lubkowski et al., 2019). Les simulations de docking de la L-lysine sur le modèle de SamELA menées avec CDOCKER ont conduit à des poses du substrat très éloignées du site actif (distance supérieure à 6 Å) (résultats non-montrés). C'est pourquoi de nouvelles simulations ont été réalisées en utilisant un autre logiciel de docking, appelé LibDock. Dans ce cas, plusieurs orientations de la L-lysine ont été observées au sein de la cavité catalytique, dont deux majoritaires : la première, minoritaire et bien classée, est caractérisée par l'orientation de la fonction amine ε de la L-lysine vers le résidu catalytique Asp417 (orientation ε ; distance de 2,0 Å) ; la seconde, majoritaire et moins bien classée, montre la fonction amine α à proximité du résidu catalytique (orientation α ; distance de 1,7 Å) (Figures IV.5 A et B). La pose la mieux classée de chacune de ces deux orientations a été choisie pour former deux complexes SamELA-L-lysine qui sont ensuite utilisés comme cible pour le docking ultérieur de l'acide caprylique. Quelle que soit l'orientation α ou ε de la L-lysine dans le site actif de l'enzyme, plusieurs modes d'interaction de l'acide caprylique sont observés. Ce dernier est majoritairement éloigné du site actif. Quelques poses minoritaires montrent la fonction carboxylate du donneur d'acyle interagissant

avec les ions métalliques et la fonction amine de la L-lysine. Les poses correspondant à l'orientation ε de la L-lysine sont mieux classées que celles issues de l'orientation α de la L-lysine. De même, les critères de distance sont davantage respectés lorsque la L-lysine présente sa fonction amine ε au résidu catalytique. Cependant, dans tous les cas, l'attaque nucléophile entre l'accepteur et le donneur d'acyle semble peu envisageable en raison d'un angle trop fermé (entre 50 et 65 °) (Figures IV.6 A et B).

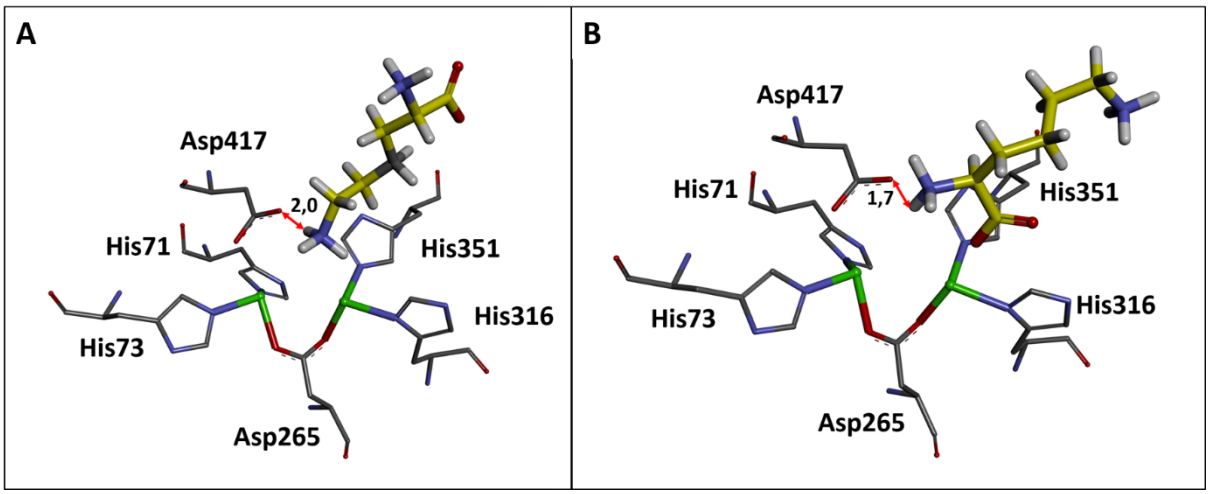


Figure IV.5 : Orientations préférentielles de la L-lysine dans la cavité catalytique de SamELA. A : orientation ε dans laquelle la fonction amine ε de la chaine latérale pointe vers le site actif. B : orientation α où la fonction amine α de la L-lysine est dirigée vers le site actif. Les deux ions métalliques et la chaine carbonée de la L-lysine sont colorés respectivement en vert et en jaune. La double flèche rouge indique la distance entre la fonction amine de la L-lysine dirigée vers le site actif et le résidu catalytique Asp417.

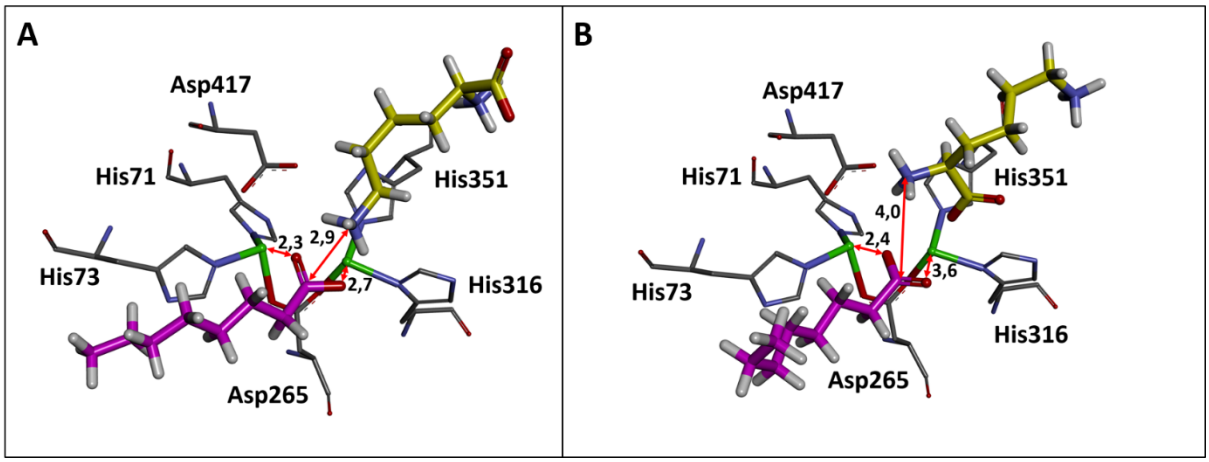


Figure IV.6 : Orientations préférentielles de l'acide caprylique dans la cavité catalytique de SamELA en présence de la L-lysine en orientation ε (A) ou α (B). Les deux ions métalliques, la chaine carbonée de la L-lysine et de l'acide caprylique sont respectivement en vert, jaune et rose. Les doubles flèches rouges indiquent les distances entre la fonction carboxylate du donneur d'acyle et les ions métalliques ou la fonction amine de la L-lysine dirigée vers le site actif.

Pour vérifier l'effet de l'ordre d'arrivée des substrats sur les résultats de docking, des simulations débutant par l'introduction de l'acide caprylique, suivie de l'introduction de la L-lysine ont été réalisées. Les résultats obtenus montrent que l'acide caprylique adopte de nombreuses orientations dans la cavité de SamELA. Dans l'orientation majoritaire qui est aussi la mieux classée, la fonction carboxylate du donneur d'acyle interagit avec les deux ions métalliques à une distance moyenne de 2,8 Å. La chaine acyle se positionne dans la région occupée par la L-lysine lors des précédentes simulations (Figure IV.7A). Le docking de la L-lysine dans la cavité de SamELA contenant l'acide caprylique conduit à des poses éloignées du site actif, incompatibles avec le mécanisme supposé de N-acylation (Figure IV.7B).

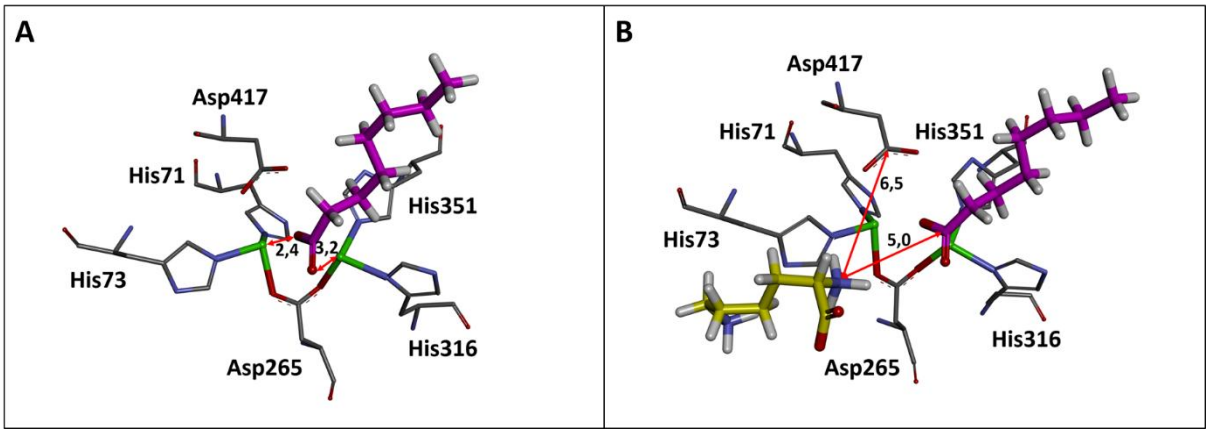


Figure IV.7 : Orientation préférentielle de l'acide caprylique seul (A) et en présence de la L-lysine (B) dans le site actif de SamELA. Les deux ions métalliques, la chaine carbonée de la L-lysine et de l'acide caprylique sont colorés respectivement en vert, jaune et rose. Les doubles flèches rouges indiquent les distances d'intérêt impliquant la fonction carboxylate du donneur d'acyle, les ions métalliques, la fonction amine de la L-lysine dirigée vers le site actif et le résidu catalytique Asp417.

Lors de travaux précédents, la spécificité de SamELA est spécifique pour des donneurs d'acyles dont la chaine carbonée comporte au plus, huit atomes de carbone, avait été montrée expérimentalement (Bourkaib et al., 2020a). Pour tenter de mettre en évidence l'effet de la longueur de chaine du donneur d'acyle sur l'interaction de celui-ci avec la cavité catalytique de SamELA, des simulations de docking de l'acide laurique ont été menées.

Comme précédemment, quelle que soit l'orientation de la L-lysine dans la cavité, l'acide laurique semble pouvoir s'approcher de la machinerie catalytique. Toutefois, l'attaque nucléophile entre les deux substrats semble peu envisageable en raison d'un angle trop fermé (entre 45 et 50 °) (Figure IV.8 A et B). Lorsque l'acide laurique est introduit seul dans la cavité de SamELA, l'orientation la mieux classée et minoritaire positionne le donneur d'acyle dans la même région que celle qu'il occupe en présence de la L-lysine. Sa fonction carboxylate interagit avec les deux ions métalliques (distance de 2,2 Å) (Figure IV.8C). L'ajout de la L-lysine conduit

à trois groupes de poses. Dans les poses les mieux classées et les plus nombreuses, la fonction amine ε de la L-lysine est dirigée vers la fonction carboxylate de l'acide et vers le résidu catalytique Asp417 (distances respectives de 2,8 Å et de 2,6 Å) et forme avec la fonction électrophile un angle de Bürgi-Dunitz égale à 90 ° (Figure IV.8D). Ces éléments suggèrent la possibilité d'une acylation en position ε de la L-lysine par l'acide laurique. Or, ce résultat n'est pas observé expérimentalement.

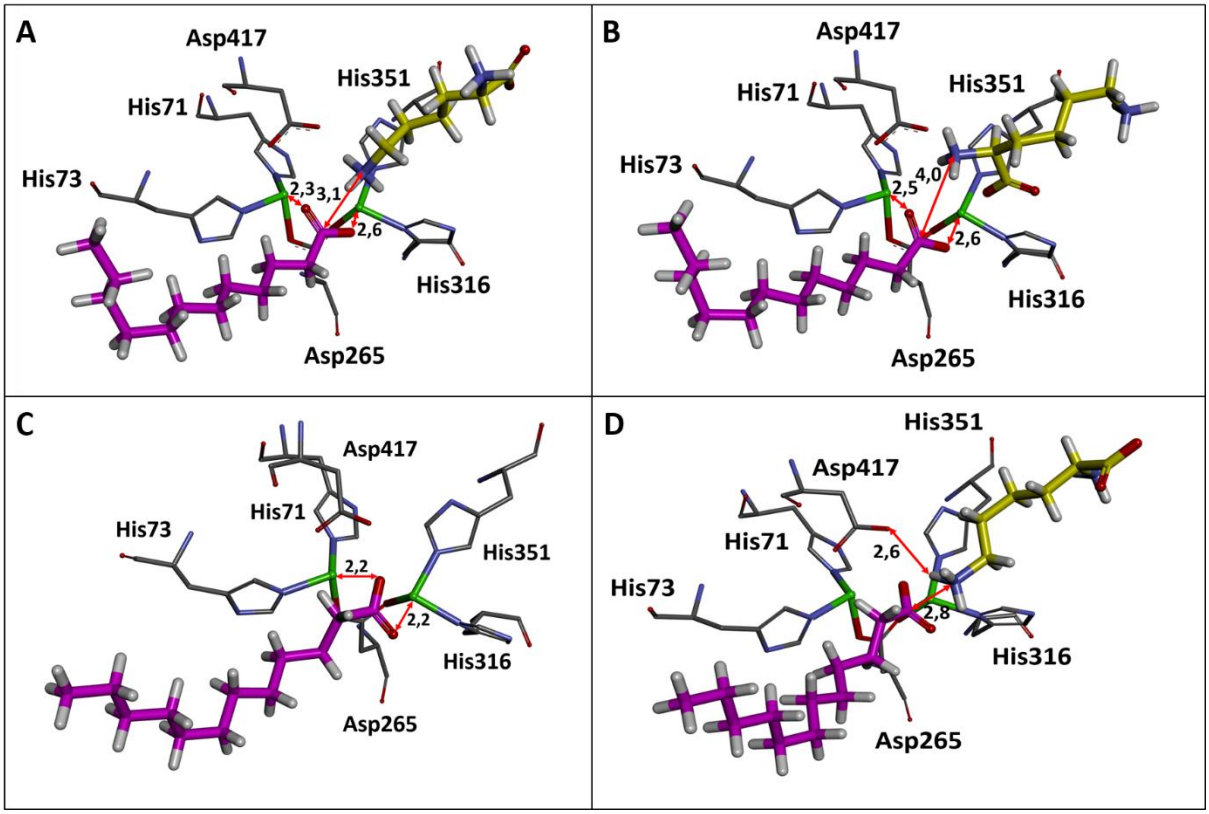


Figure IV.8 : Orientations préférentielles des substrats de la réaction de N-acylation de la L-lysine par l'acide laurique, catalysée par SamELA. A et B : Modes d'interaction de l'acide laurique en présence de la L-lysine en orientation ε (A) ou en orientation α (B). C : Orientation de l'acide laurique seul. D : Mode d'interaction majoritaire de la L-lysine en présence de l'acide laurique. Les deux ions métalliques, la chaine carbonée de la L-lysine et de l'acide laurique sont colorés respectivement en vert, jaune et rose. Les doubles flèches rouges indiquent les distances d'intérêt impliquant la fonction carboxylate du donneur d'acyle, les ions métalliques, la fonction amine de la L-lysine dirigée vers le site actif et le résidu catalytique Asp417.

A la vue de ces résultats, il apparaît que les modèles ne permettent pas d'expliquer les données expérimentales relatives à la sélectivité de SamELA vis-à-vis de la fonction ε de la L-lysine lors de son acylation par l'acide caprylique. Par ailleurs, alors qu'aucune réaction n'a été observée avec l'acide laurique comme donneur d'acyle, les modèles suggèrent la possibilité d'une acylation de la fonction amine ε de la L-lysine. Ces résultats contradictoires semblent indiquer un problème d'origine structural concernant la cible de docking utilisée et remettent en question la qualité du modèle de SamELA. Il faut mentionner notamment la présence d'une

boucle comportant 33 résidus (Ala263-Asp295) en périphérie du centre métallique. Une boucle constitue un élément de structure très complexe à modéliser. Son positionnement peut impacter le positionnement des substrats dans le site actif (Muhammed et al., 2018 ; Haddad et al., 2020).

4.6. Conclusions

Ce chapitre avait pour but d'expliquer la sélectivité de SamELA pour la fonction amine ε de la L-lysine lors de son acylation. La structure de l'enzyme n'étant pas élucidée, celle-ci a été prédite par modélisation par homologie pour être ensuite utilisée comme cible dans des simulations de docking. La recherche d'enzymes homologues a permis d'identifier cinq templates : les hydrolases métallo dépendantes non-caractérisées de *Pyrococcus furiosus* (entrée PDB 3ICJ) et de *Pyrococcus horikoshii* (entrée PDB 3IGH), l'amidohydrolase de *Streptomyces* sp. Tu 6176 (entrée PDB 6SJ0), la D-aminoacylase d'*Alcaligenes faecalis* (entrée PDB 1RJQ) et la protéine médiatrice de la réponse à la collapsine 2 humaine (entrée PDB 5YZA). SamELA serait une enzyme monomèrique, comportant deux ions métalliques et une molécule d'eau catalytique dans son site acitf. Elle appartiendrait à la famille des métalloenzymes M38. Des résidus conservés ont été découverts ; ils sont supposés jouer un rôle clé dans la fixation des ions métalliques (His71, His73, Asp265, His316 et His351) et/ou dans la catalyse (Asp417).

Deux outils de modélisation par homologie ont été utilisés pour prédire la structure tridimensionnelle de SamELA : ColabFold (interface web de AlphaFold2) et AlphaFold3. Très proches d'un point de vue structural, ces deux modèles suggèrent une organisation autour de trois domaines. La cavité catalytique a été localisée grâce à la localisation des résidus fixateurs des ions métalliques et du résidu catalytique. Celle-ci a été définie comme cible pour étudier les modes d'interaction entre l'enzyme et les substrats de la réaction de N-acylation de la L-lysine et tenter d'expliquer la sélectivité et la spécificité observées expérimentalement. Cependant, les résultats issus des simulations de docking ne corroborent pas les résultats expérimentaux et ne permettent pas d'expliquer l'acylation exclusive de l'amine ε de la L-lysine ou la prise en charge préférentielle de donneurs d'acyle à courte chaine par l'enzyme. Cet écart entre les modèles et les observations expérimentales trouverait son origine dans la qualité insuffisante du modèle de SamELA, notamment au niveau d'une région proche du site catalytique. A ce jour, ces enzymes restent très peu décrites et malgré des recherches bibliographiques poussées, aucune donnée n'a pu être trouvée pour orienter l'amélioration du modèle de SamELA.

Malgré cela, l'étude constitue un premier pas vers une meilleure connaissance de cette enzyme. La complémentarité de SamAA et de SamELA du point de vue de leur sélectivité constitue un atout important dans le domaine de la biocatalyse, permettant d'appréhender l'hydrolyse ou la synthèse contrôlée de toute une diversité d'acides aminés et de peptides, en milieu aqueux. La validation des modèles produits au cours de ce travail constitue une étape cruciale dans la démarche. Ceci implique l'élucidation de la structure tridimensionnelle des enzymes et dans l'idéal, la confirmation du rôle des résidus clés identifiés par une approche de mutagénèse dirigée. Pour ce faire, les deux aminoacylases doivent être produites sous forme soluble, dans des quantités permettant leur purification et leur caractérisation. Les travaux portant sur la production de SamAA et de SamELA par voie hétérologue font l'objet du chapitre suivant.

5. Chapitre V : Production hétérologue des deux aminoacylases de *Streptomyces ambofaciens*, SamAA et SamELA

5.1. Introduction

Pour compléter et valider les modèles des aminoacylases SamAA et de SamELA développés par l'approche numérique présentée dans les chapitres précédents, cette partie du travail visait à produire les deux enzymes par expression hétérologue, à les purifier et à élucider leur structure. Par ailleurs, les enzymes de *Streptomyces ambofaciens* ayant été utilisées jusqu'ici en mélange, disposer des enzymes pures dans des quantités suffisantes permettrait d'envisager des campagnes expérimentales visant une caractérisation plus approfondie de leurs propriétés catalytiques. L'expression de protéines recombinantes est une technique de biotechnologie très répandue, permettant la production et la purification d'une protéine d'intérêt en grande quantité dans un but de caractérisation ou de commercialisation industrielle. Les systèmes d'expression bactériens sont majoritairement employés car ils sont simples à utiliser et abordables (Pouresmaeil et al., 2023).

Plusieurs tentatives d'expression de SamAA et de SamELA avaient été effectuées dans des études antérieures :

- La levure *Pichia pastoris* et le vecteur pPICZαA ont d'abord été utilisés comme hôte et vecteur plasmidique pour l'expression hétérologue de SamELA. Une altération de la régio-spécificité de l'enzyme pour la N-ε-acétyl-L-lysine a été observée après une optimisation de son gène en raison de la différence d'usage de codons entre *Streptomyces ambofaciens* et *P. pastoris* (Ferrari, 2014).
- La culture d'*E. coli* BL21 (DE3) transformée par le plasmide pET-15b-SamELA a été testée à 28 °C, pendant 24 h, et induite par un ajout d'IPTG (1 mM) à une DO_{600} de 0,6, mais l'enzyme est produite sous forme insoluble (George, 2017).
- *E. coli* Origami B (DE3) a alors été employée comme hôte d'expression pour SamAA et SamELA. La température de culture après l'induction a été abaissée à 16 °C. La concentration en inducteur a été également réduite à 0,1 mM. La surexpression de SamELA et de SamAA a été observée sous la forme d'une bande de faible intensité à 50 kDa et à 61 kDa sur les gels SDS-PAGE. Les activités hydrolytiques spécifiques des extraits bruts de ces cultures ont été étudiées sur les substrats N-α- et N-ε-acétyl-L-lysine et sont visibles dans le tableau V.1. Le plasmide pET-15b contient une séquence codant pour un tag 6-His

en position N-terminale des sites de restriction utilisés pour l'insertion du gène de SamAA. Ceci était censé faciliter la purification de l'enzyme, toutefois, la tentative de purification de SamAA par chromatographie d'affinité au nickel n'a pas été concluante (Nili, 2018).

Tableau V.1 : Activités spécifiques apparentes d'hydrolyse de la N-α-acétyl-L-lysine et de la N-ε-acétyl-L-lysine en présence des fractions solubles des cultures d'*E. coli* Origami B (DE3) transformées (Nili, 2018).

Activité spécifique apparente (nmole Lys/min/mg)	En présence de N-α-acétyl-L-lysine	En présence de N-ε-acétyl-L-lysine
E. coli Origami B (DE3)/SamELA	0,26	3,71
E. coli Origami B (DE3)/SamAA	33,24	0

Jusqu'ici, l'expression hétérologue de SamAA et de SamELA a donc très majoritairement conduit à des corps d'inclusion et n'a pas permis d'obtenir une quantité suffisante d'enzyme purifiée pour poursuivre les travaux de caractérisation structurale et mécanistique. Dans le cadre de cette thèse, les travaux ont été poursuivis en faisant varier différents paramètres comme l'hôte d'expression et les conditions de culture, l'objectif étant d'identifier un système de production favorisant la forme soluble et active des deux enzymes. Le système d'expression *E. coli* a été choisi en première intention pour les deux enzymes, en raison de sa simplicité d'utilisation (Koehn et al., 2009). Les travaux sur SamAA ont par ailleurs porté sur un hôte du même genre que la bactérie productrice de l'enzyme dans l'espoir de favoriser sa surexpression sous une forme soluble (Haeger et al., 2024b). Enfin, dans le but de valoriser les corps d'inclusion obtenus jusqu'ici, une étude de refolding, suivie d'une purification de SamAA a été menée.

Une approche méthodologique commune a été suivie dans toutes les campagnes expérimentales réalisées dans le cadre de ce travail. Pour rappel, celle-ci s'articule autour d'étapes clés de l'expression hétérologue : suivi de la croissance bactérienne par mesure de la densité optique à 600 nm (uniquement avec *E. coli*), lyse cellulaire, dosage des protéines, vérification de l'expression par SDS-PAGE dans les fractions solubles et insolubles (poids moléculaire de 50 kDa pour SamAA et de 61 kDa pour SamELA), et test de l'activité enzymatique dans les fractions solubles. La réaction de référence pour SamAA est la synthèse de la N-α-lauroyl-L-lysine, tandis que l'hydrolyse de la N-ε-acétyl-L-lysine a permis d'évaluer l'activité de

SamELA. Les données issues de ces tests enzymatiques ont permis d'évaluer la présence des enzymes sous une forme soluble et fonctionnelle dans les fractions étudiées.

5.2. Production hétérologue de SamAA

Les travaux d'expression se sont concentrés sur la production hétérologue de SamAA en raison de sa spécificité enzymatique rare envers les dérivés N-α-acyl-L-lysine ; ils ont principalement porté sur la mise en œuvre de cultures impliquant divers systèmes *E. coli* puis, l'hôte *S. lividans*.

5.2.1. Production chez *Escherichia coli*

De nombreux paramètres et conditions opératoires ont été étudiés pour surproduire SamAA sous une forme soluble et active ; ils sont présentés dans la figure V.1. Les premières tentatives d'expression hétérologue de SamAA ont été menées chez *E. coli*, en utilisant plusieurs souches commerciales disponibles : *E. coli* Origami B (DE3), *E. coli* Arctic Express (DE3), *E. coli* Rosetta 2 (DE3) et *E. coli* BL21 Gold (DE3) (Merck, USA ; Agilent Technologies, USA). Certaines souches possèdent également le plasmide pLysS (Studier, 1991). Deux types de TAG ont été étudiés : le tag poly-His (6-His) et le tag Streptavidin-binding tag (ou StrepTag) (Kosobakova et al., 2016). Ce dernier est placé uniquement en position N-terminal alors que le tag 6-His a été positionné soit en position N-terminal, soit en position C-terminal. Dans certaines cultures, le gène de SamAA optimisé pour une production chez *E. coli* a été utilisé. Les plasmides pET-15b et pET-28a contiennent respectivement une version non-optimisée et une version optimisée du gène de SamAA. Quelle que soit la version du gène, ce dernier est sous le contrôle d'un promoteur inductif. Les milieux LB, Overnight Express™ Instant TB (Overnight) et Terrific Broth AIM (TB-AIM) ont été utilisés lors des cultures bactériennes. Les milieux Overnight et TB-AIM sont des milieux autoinductifs basés sur une utilisation différentielle du lactose (Studier, 2005 ; Grabski et al., 2005). Le milieu LB nécessite une induction préalablement optimisée de la manière suivante : ajout de l'IPTG (0,1 mM) quand la DO_{600} est à 0,6 en absence de cobalt, et à 0,4 en présence de cobalt (Incir et al., 2024). SamAA étant probablement une métalloenzyme, du cobalt a été ajouté à certains milieux en début de culture ou au moment de l'induction, en faisant l'hypothèse que les ions métalliques joueraient un rôle structural en évitant la formation des corps d'inclusion. Des études antérieures ont permis de déterminer qu'une concentration de 0,25 mM permet de limiter les effets délétères des métaux pour la croissance bactérienne, tout en favorisant la production d'enzymes sous

forme soluble et active (Sari et al., 2007 ; Matjan et al., 2011 ; Takakura et al., 2019). L'arrêt des cultures est intervenu 4 h ou 20 h après l'induction, ou après 16 h d'incubation en absence d'induction. Les cultures bactériennes ont été menées à 37 °C, sauf dans certains cas où la température a été abaissée à 20 °C après l'induction ou en cas de choc osmotique et thermique juste avant l'induction (ajout de NaCl 4 M et de bétaïne 0,2 M et une heure à 47 °C avant l'induction puis poursuite de la culture à 20 °C) (Rosano et al., 2014 ; Papp et al., 2006 ; Bhatwa et al., 2021).

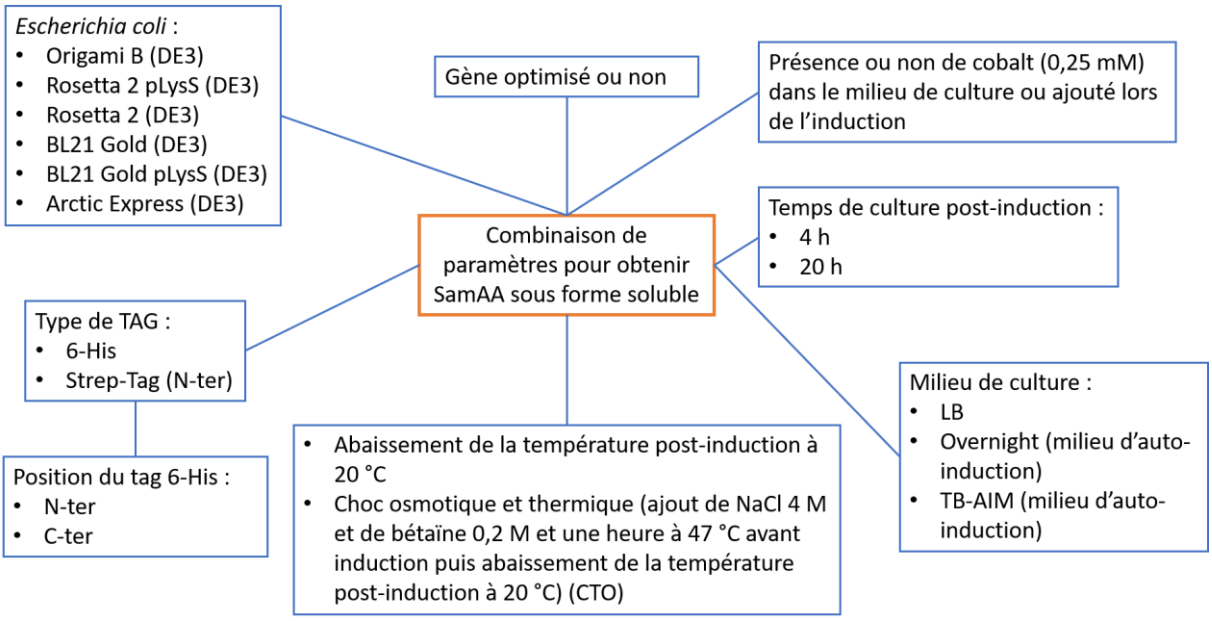


Figure V.1 : Récapitulatif des conditions étudiées pour la surexpression de SamAA chez *E. coli*.

Par rappel, une fois les cultures bactériennes réalisées, les cellules sont séparées du milieu de culture par centrifugation et sont resuspendues dans du tampon Tris-HCl 25 mM, NaCl 50 mM, pH 8. Si la croissance bactérienne s'est effectuée en présence de cobalt, ce dernier est ajouté au tampon à la concentration de 0,25 mM. Les bactéries sont lysées avec la presse de French. En vue d'une analyse SDS-PAGE, des prélèvements sont effectués en fin de culture, correspondant aux extraits bruts non-lysés (EBNL), et après la lyse conduisant aux fractions insolubles (FI) et solubles (FS). La concentration protéique a été déterminée par un dosage de Bradford. Si la concentration protéique obtenue est faible (inférieure à 0,5 g/L), la fraction soluble est concentrée dix fois par ultrafiltration (FSc). Toutes les fractions recueillies dans le cadre de l'étude ont été analysées par SDS-PAGE. Par ailleurs, la réaction de synthèse de la N-α-lauroyl-L-lysine a été utilisée comme référence pour évaluer l'activité enzymatique spécifique de SamAA dans les fractions solubles. Du cobalt a été ajouté lors de cette synthèse, car il a été démontré que sa présence augmente l'activité catalytique de l'enzyme (Bourkaib et al., 2020a).

5.2.1.1. Production avec *E. coli* Origami B (DE3)

La souche *E. coli* Origami B (DE3) contenant le plasmide pET-15b-SamAA était déjà préparée avant le début des travaux de thèse. Cette souche bactérienne possède plusieurs mutations génétiques entrainant l'absence de la thiorédoxine réductase et de la glutathion réductase et une modification de la protéine AhpC d'une peroxydase en une disulfure réductase. Ces changements améliorent la formation des ponts disulfures dans le cytoplasme bactérien devenu oxydant et limitent les mauvais repliements protéiques, et par suite, la formation des corps d'inclusion (Hatahet et al., 2014 ; Binder et al., 2016 ; Pouresmaeil et al., 2023). Ce système avait permis d'obtenir des résultats plutôt encourageants dans une étude précédente de l'expression hétérologue de SamAA : la présence d'une bande pouvant indiquer la présence de SamAA à la taille attendue de l'enzyme sur un gel SDS-PAGE et la mise en évidence d'une activité spécifique apparente de l'hydrolyse de la N-α-acétyl-L-lysine. Les travaux ont débuté avec ce système disponible et ayant conduit à des résultats prometteurs.

L'expression de SamAA a été menée chez *E. coli* Origami B (DE3) suite à une induction à l'ITPG (0,1 mM) à une DO_{600} de 0,6, selon les différentes conditions présentées sur la figure V.2. Toutes les cultures ont été réalisées une seule fois, sauf la culture d'*E. coli* Origami B (DE3) avec le plasmide pET-15b-SamAA sur milieu LB qui a été répétée trois fois. La température des cultures bactériennes avec le milieu LB a été abaissée à 16 °C après l'induction. La croissance de ces cultures a été suivie par mesure de la DO_{600} dont les résultats sont présentés dans le graphique suivant (Figure V.3). Les cultures avec le milieu Overnight ont été conduites à 37 °C pendant 16 h, selon les recommandations du fournisseur du milieu. La DO_{600} à 16h de croissance était égale à 2,4. Il apparaît que SamAA ne présente pas de toxicité vis-à-vis d'*E. coli*, car la croissance bactérienne n'est pas modifiée suite à l'ajout de l'IPTG. La croissance bactérienne est peu élevée ce qui s'explique par les mutations de cette souche qui diminuent sa vitesse de croissance (Wardenga et al., 2008). Pour augmenter la production de biomasse bactérienne, le temps de culture sur milieu LB est prolongé de 24 h et 48 h. La DO_{600} à 48 h et à 72 h est alors égale à 1,1 et à 0,9 respectivement. A partir de 24 h, une diminution de la DO_{600} est observée indiquant une phase de déclin de la culture pendant laquelle le risque de dégradation de SamAA recombinante est élevé.

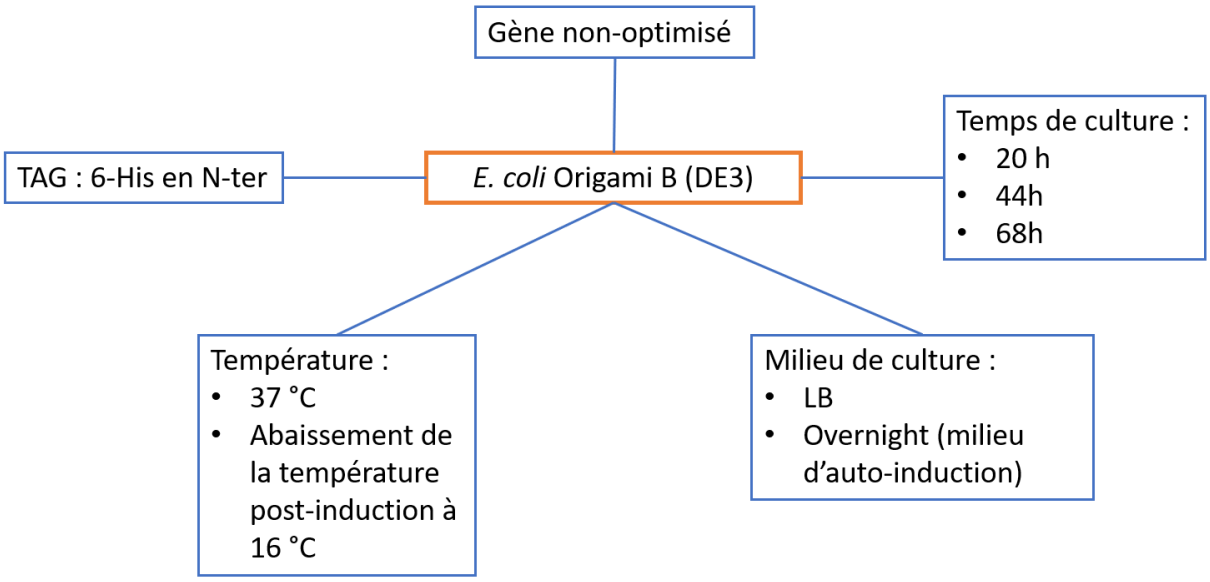


Figure V.2 : Récapitulatif des conditions étudiées pour la surexpression de SamAA chez *E. coli* Origami B (DE3).

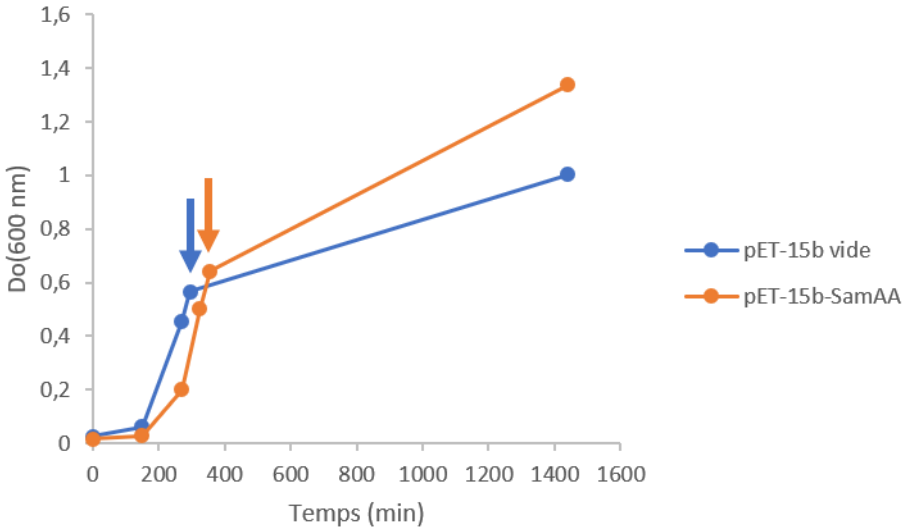


Figure V.3 : Suivi de la densité optique du milieu de culture LB, à 600 nm, lors de l'expression protéique chez les souches *E. coli* Origami B (DE3) transformées par les plasmides pET-15b vide ou pET-15b-SamAA. Le moment d'induction est indiqué par la flèche bleue (plasmide pET-15b vide) et par la flèche orange (plasmide pET-15b-SamAA).

A la fin des cultures bactériennes, les cellules ont été extraites puis lysées. Un dosage des protéines a montré une faible concentration dans les fractions solubles (inférieure à 0,5 g/L). Celles-ci ont été concentrées. Les résultats de l'analyse SDS-PAGE des différents prélèvements des cultures d'*E. coli* Origami B (DE3) en milieu LB est visible en figure V.4. Une bande différentielle à la taille attendue est visible dans les fractions insolubles indiquant une production de l'enzyme sous forme de corps d'inclusion. Cette bande n'est pas retrouvée dans les fractions solubles concentrées ou non, montrant l'absence de la forme soluble de SamAA.

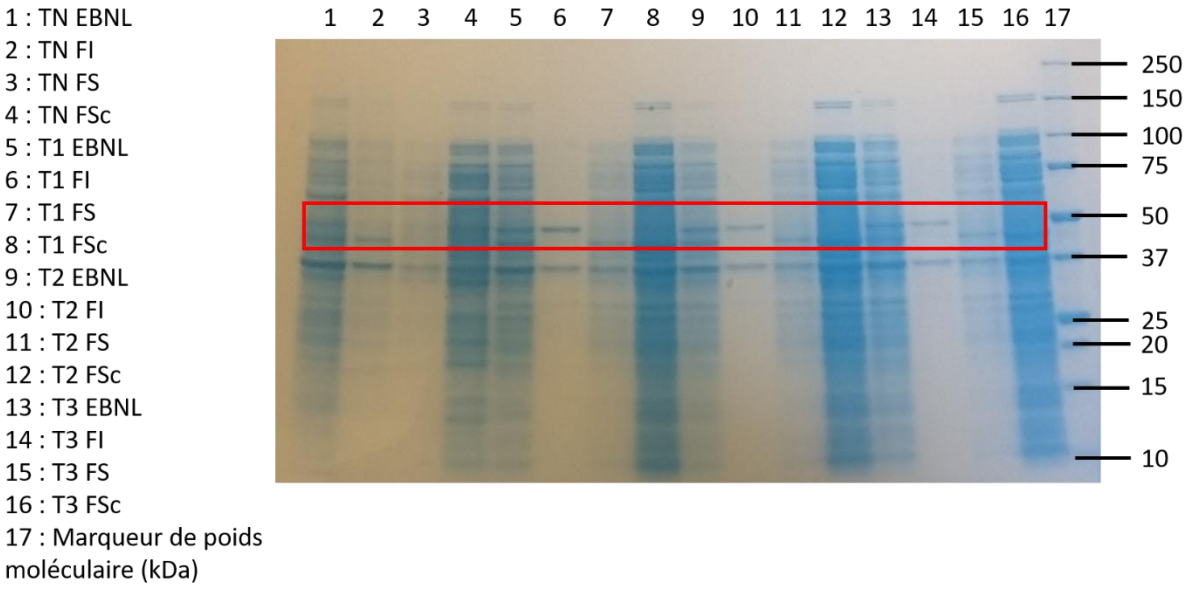


Figure V.4 : Analyse par SDS-PAGE des protéines issues des cultures de 24 h d'*E. coli* Origami B (DE3) avec le plasmide pET-15b vide et pET-15b-SamAA en milieu LB. TN : témoin négatif (plasmide vide) ; T1, T2, T3 : erlen 1, 2, 3 avec le plasmide pET-15b-SamAA ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée. Le cadre rouge indique la position de SamAA recombinante (50 kDa).

L'analyse des fractions issues des cultures menées en milieu Overnight est visible en annexe 1 ; elle montre l'absence de production de SamAA, que ce soit sous forme soluble, ou sous la forme de corps d'inclusion. Overnight est un milieu d'auto-induction dans lequel l'induction se fait par une utilisation différentielle du lactose. Il est nécessaire pour la souche de produire une lactose perméase (codée par le gène *lacY*) pour le transport efficace du lactose et une β-galactosidase (codée par le gène *lacZ*) pour la transformation de ce lactose en allolactose qui joue le rôle d'inducteur (Studier, 2005 ; Grabski et al., 2005). Or ces deux gènes (*lacY* et *lacZ*) sont mutés chez *E. coli* Origami B (DE3), créant une incompatibilité entre le milieu de culture Overnight et la souche *E. coli* Origami B (DE3) (Pouresmaeil et al., 2023). Ces informations ont été mises en évidence au cours de l'essai, c'est pourquoi l'analyse de la culture n'a pas été poursuivie.

L'activité des différentes fractions solubles concentrées obtenues issues des cultures en milieu LB pour la synthèse de la N-α-lauroyl-L-lysine a été vérifiée par une analyse quantitative (HPLC). Quelles que soient les conditions de culture et quel que soit le plasmide utilisé, aucune activité de synthèse n'a été observée en présence des fractions provenant de la culture de la souche *E. coli* Origami B (DE3). Ce système n'a donc pas permis l'expression de SamAA sous forme soluble et active. Il est caractérisé par une croissance bactérienne moyenne. L'analyse SDS-PAGE et le test d'activité enzymatique spécifique confirment la production de SamAA

sous la forme de corps d'inclusion uniquement. A la vue de ces résultats, les travaux ont été poursuivis avec un autre système d'expression.

5.2.1.2. Production avec *E. coli* Rosetta 2 pLysS (DE3)

Le système *E. coli* Rosetta 2 pLysS (DE3) avec le plasmide pET-15b-SamAA était également construit avant le début des travaux de thèse. La souche *E. coli* Rosetta 2 (DE3) contient le plasmide pRARE2 permettant la surexpression des ARNt rares d'*E. coli* dans le but de contourner le biais d'utilisation des codons, la pénurie d'ARNt rares et les problèmes associés au niveau de la traduction protéique (Nakamura et al., 2000 ; Kim et al., 2020). Le plasmide pLysS est un plasmide supplémentaire permettant l'augmentation de la stabilité, du maintien et de l'expression des plasmides porteurs de gènes de protéines recombinantes toxiques (Dunn et al., 1983 ; Huang et al., 1999).

SamAA a été exprimée chez *E. coli* Rosetta 2 pLysS (DE3) à 37 °C, suite à une induction à l'IPTG (0,1 mM) à une valeur de DO_{600} de 0,6 dans les conditions présentées en figure V.5. Toutes les cultures ont été répétées trois fois, sauf les cultures longues de 48 h et de 72 h et la culture du témoin négatif (TN ; avec le plasmide vide). Le suivi de la croissance est présenté sur la figure V.6. Lors de l'utilisation du milieu Overnight, au bout de 16 h, la DO_{600} moyenne est égale à 2,8. Une DO_{600} de 1,8 est relevée après 24h de culture d'un témoin négatif sans induction. Ces résultats confirment l'absence de toxicité de SamAA vis-à-vis d'*E. coli* et indiquent un impact de l'IPTG sur la croissance des bactéries ; en effet la vitesse de croissance est ralentie en présence de l'inducteur (Dvorak et al., 2015). Dans une tentative d'augmenter la production de biomasse bactérienne, le temps de culture sur LB est prolongé de 24 h et de 48 h au bout desquelles la DO_{600} est environ égale à 0,5. Ces cultures sont alors en phase de déclin, et comme dit précédemment, le risque de dégradation de SamAA recombinante est important.

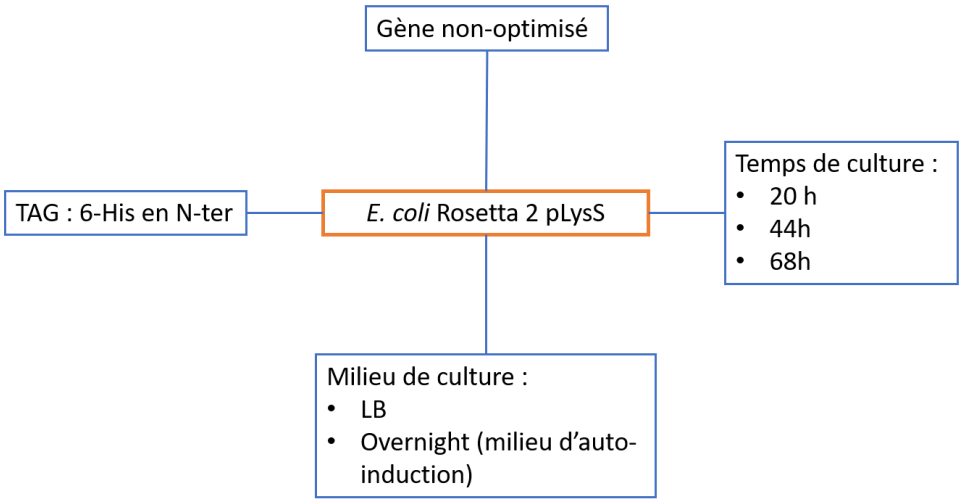


Figure V.5 : Récapitulatif des conditions étudiées pour la surexpression de SamAA chez *E. coli* Rosetta 2 pLysS (DE3).

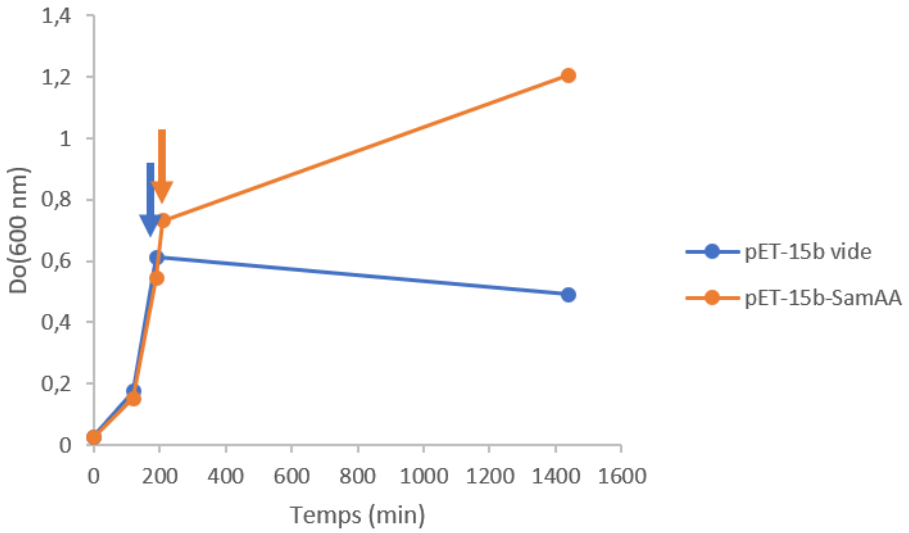


Figure V.6 : Suivi de la densité optique du milieu de culture LB, à 600 nm, lors de l'expression protéique chez les souches *E. coli* Rosetta 2 pLysS (DE3) transformées par les plasmides pET-15b vide ou pET-15b-SamAA. Le moment d'induction est indiqué par la flèche bleue (plasmide pET-15b vide) et par la flèche orange (plasmide pET-15b-SamAA).

L'analyse des fractions solubles indiquent une faible concentration protéique (inférieure à 0,5 g/L) ; celles-ci ont été concentrées. Les différents prélèvements (EBNL, FI, FS et FSc) sont analysés par SDS-PAGE. Le gel correspondant aux cultures de 24 h, en milieu LB, est visible en figure V.7. SamAA (50 kDa) est uniquement visible dans les fractions insolubles indiquant une production exclusivement sous la forme de corps d'inclusion.

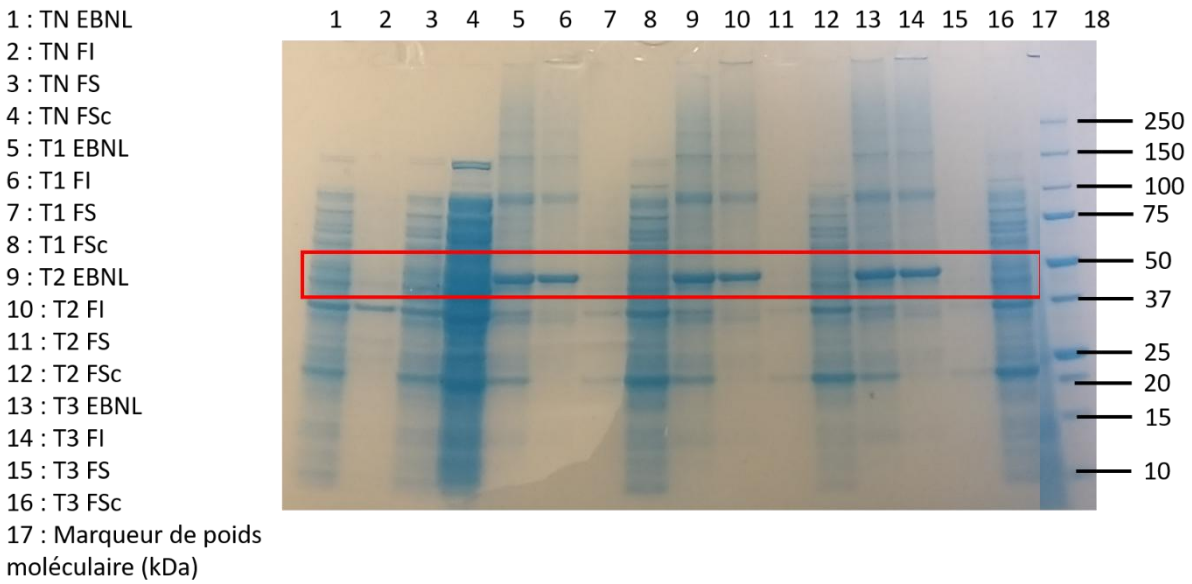


Figure V.7 : Analyse par SDS-PAGE des protéines issues des cultures de 24 h d'*E. coli* Rosetta 2 pLysS (DE3) avec le plasmide pET-15b vide et pET-15b-SamAA, en milieu LB. TN : témoin négatif (plasmide vide) ; T1, T2, T3 : erlen 1, 2, 3 avec le plasmide pET-15b-SamAA ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée. Le cadre rouge indique la position de SamAA exprimée.

Le gel SDS-PAGE correspondant aux cultures avec le milieu Overnight est visible en annexe 1 et présente un défaut lors de la préparation des dépôts qui débordent dans les puits adjacents. Au vu des résultats de la synthèse enzymatique présenté ci-dessous, le gel n'a pas été refait.

Comme précédemment, même si la présence de la forme soluble de l'enzyme n'a pu être mise en évidence, l'activité enzymatique des fractions conduisant à la synthèse de la N-α-lauroyl-L-lysine a été recherchée. Toutefois, aucune synthèse n'a été observée, indiquant l'absence de production de SamAA soluble et active, quelles que soient les conditions appliquées.

Ces travaux montrent que le système *E. coli* Rosetta 2 pLysS (DE3) avec le plasmide pET-15b-SamAA permet l'expression de SamAA mais uniquement sous la forme de corps d'inclusion uniquement. Il a par ailleurs été observé que l'ajout de l'IPTG a un impact négatif sur la croissance bactérienne.

5.2.1.3. Production avec *E. coli* Arctic Express (DE3)

Abaisser la température d'incubation a pour conséquence une diminution de la vitesse de croissance bactérienne mais cela permet par ailleurs la baisse de la formation des corps d'inclusion au profit de l'expression protéique soluble et recombinante. L'idée sous-jacente est de ralentir le métabolisme cellulaire pour limiter une surexpression trop intense d'une protéine

recombinante qui n'a pas le temps d'acquérir sa conformation tridimensionnelle active. Cette stratégie s'accompagne d'une augmentation du temps de culture ou de l'utilisation de la souche *E. coli* Arctic Express (DE3) (San-Miguel et al., 2013). Celle-ci coexprime les protéines chaperonnes Cpn10 et Cpn60 de la bactérie *Oleispira antarctica*, qui sont homologues des protéines GroEL et GroES d'*E. coli* et qui sont actives à des températures comprises entre 4 et 12 °C. Cette souche est alors capable de croître à basse température en utilisant Cpn10 et Cpn60 pour le repliement des protéines endogènes et recombinantes (Ferrer et al., 2003 ; Ferrer et al., 2004a ; Ferrer et al., 2004b).

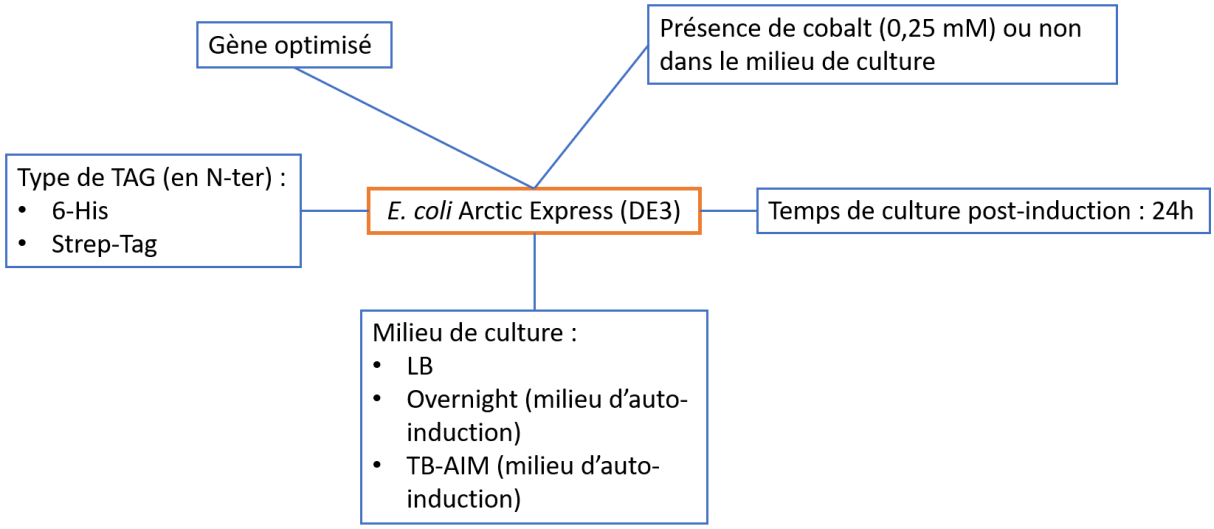


Figure V.8 : Récapitulatif des conditions étudiées pour la surexpression de SamAA chez *E. coli* Arctic Express (DE3).

SamAA a été exprimée chez *E. coli* Arctic Express (DE3) en utilisant le plasmide pET-28a dans les différentes conditions présentées à la figure V.8. Ce système a été construit à partir de cellules compétentes commandées auprès d'Agilent Technologies et de deux plasmides préparés par Genscript. Ces plasmides contiennent une version optimisée du gène de SamAA pour son expression chez *E. coli*. Un second milieu autoinductif par utilisation de lactose (TB-AIM) a été également étudié avec ce système (Haeger et al., 2024a). Le milieu de culture est supplémenté en cobalt dans le but d'augmenter la quantité d'enzyme soluble par la présence d'ions métalliques (Takakura et al., 2019). Un Streptavidin-binding tag en position N-terminale a été testé en remplacement du tag 6-His car ce dernier a une affinité importante pour les ions métalliques divalents, ce qui pourrait se révéler problématique pour le repliement correct d'une métalloenzyme. La culture en milieu LB a été induite avec l'ajout d'IPTG (0,1 mM) et sa température est abaissée à 11 °C jusqu'au lendemain. La température des cultures avec les milieux d'auto-induction est de 30 °C pendant 5 h, puis de 11 °C pendant 16 h avec le milieu Overnight, et pendant 24 h avec le milieu TB-AIM.

Ces cultures ont été répétées deux fois, en présence et en absence d'antibiotiques pour maintenir les plasmides, car le fournisseur de la souche (Agilent Technologies) recommande une culture sans antibiotique. Les résultats suivent la même tendance, mais indiquent assez logiquement une meilleure croissance bactérienne en absence d'antibiotique. Les résultats présentés ci-dessous concernent uniquement les cultures réalisées en absence d'antibiotique.

Le suivi de la croissance des cultures en milieu LB supplémenté en cobalt (LB-Co) est visible en figure V.9. Avec le milieu Overnight supplémenté en cobalt (OV-Co), après 16h de croissance, la DO_{600} est égale à 0,5 et 1,7, en présence du tag 6-His et du tag Strep-Tag respectivement. Les résultats indiquent que le tag 6-His a une influence positive sur la culture en milieu LB-Co tandis que le tag Strep-Tag a une influence positive sur la culture en milieu OV-Co. Les cultures avec le tag Strep-Tag et le milieu TB-AIM après 24 h à 11 °C présentent une DO_{600} de 5,4 en absence de cobalt et de 3,6 en présence du cobalt, montrant l'influence négative du cobalt sur la croissance bactérienne. Cette dernière est de façon générale faible à cause de l'abaissement de la température à 11 °C, ce qui peut être compensé par une augmentation du temps de culture.

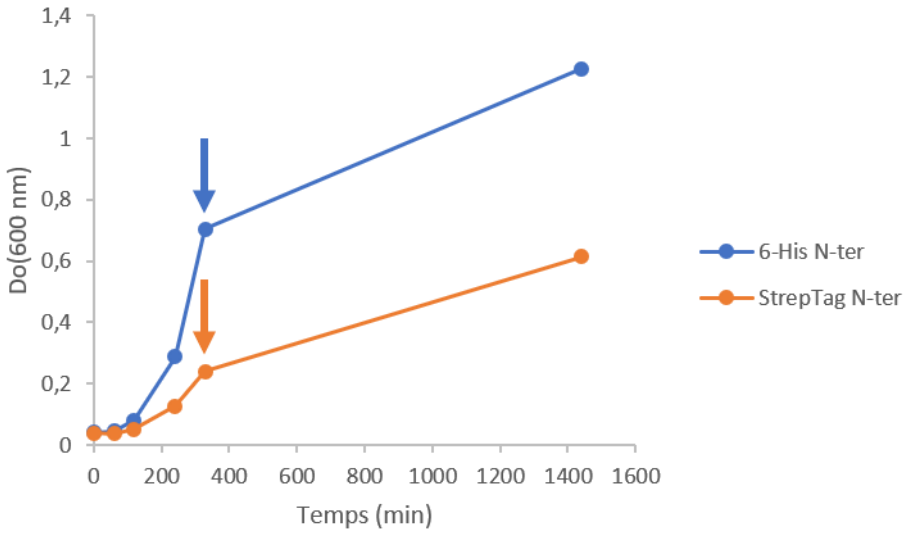


Figure V.9 : Suivi de la densité optique du milieu de culture LB-Co, à 600 nm, lors de l'expression protéique chez les souches *E. coli* Arctic Express (DE3) transformées par les plasmides pET-28a-SamAA 6-His N-ter ou Strep-Tag N-ter, à 11 °C après l'induction. Le moment d'induction est indiqué par la flèche bleue (6-His N-ter) et par la flèche orange (Strep-Tag N-ter).

La concentration protéique des fractions solubles étant faible, celles-ci ont été concentrées. Chaque prélèvement issu des cultures d'*E. coli* Arctic Express (DE3) en milieu LB avec le plasmide pET-28a-SamAA contenant le tag 6-His ou le tag Strep-Tag en position N-terminale a été analysé par SDS-PAGE dont le gel est visible en figure V.10. Les deux autres gels SDS-

PAGE réalisés à partir des prélèvements issus des cultures menées dans les deux milieux d'auto-induction ont le même profil et sont visibles en annexe 1. Une bande comprise entre les bandes 50 kDa et 75 kDa du marqueur de poids moléculaire est visible dans tous les puits ; elle correspond à la protéine chaperonne Cpn60 dont le poids moléculaire est de 56 kDa. La seconde protéine chaperonne Cpn10 (environ 10 kDa) n'est pas visible sur les gels SDS-PAGE. Aucune bande correspondant à SamAA (50 kDa) n'est visible sur les gels, que ce soit dans les fractions insolubles ou dans les fractions solubles concentrées.

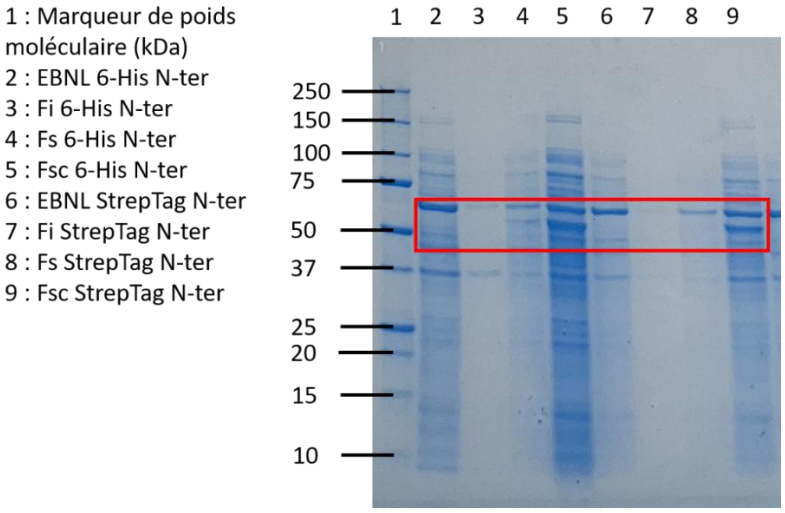


Figure V.10 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Arctic Express (DE3) avec le plasmide pET-28a-SamAA 6-His ou Strep-Tag N-ter en milieu LB supplémenté en cobalt (0,25 mM). EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée. Le cadre rouge indique la position de SamAA exprimée.

Malgré l'absence de bande indiquant la présence de SamAA sous une forme soluble dans les prélèvements, la synthèse de la N-α-lauroyl-L-lysine en présence des fractions solubles concentrées a été étudiée. L'analyse HPLC des milieux montre une activité spécifique apparente très faible voire nulle, dans l'ensemble des fractions (Tableau V.2). Le milieu LB et le tag 6-His semblent avoir une légère influence positive sur l'activité apparente enzymatique spécifique, mais dans tous les cas l'activité reste faible.

Tableau V.2 : Activités spécifiques apparentes pour la synthèse de la N-α-lauroyl-L-lysine en présence des fractions solubles concentrées des cultures d'*E. coli* Arctic Express (DE3) transformées par le plasmide pET-28a-SamAA 6-His ou StrepTag N-ter en milieu LB, OV ou TB-AIM, supplémenté ou non en cobalt.

Fractions solubles concentrées				Activité spécifique apparente (mg/g*h)
Milieu utilisé	Type de TAG	Température de culture	Présence de cobalt	
LB	6-His	11°C	Oui	3,04
	Strep-Tag			4,20
OV	6-His	30°C pendant 5h puis 16h à 11°C	Oui	5,13
	Strep-Tag			0,55
TB-AIM	Strep-Tag	30°C pendant 6h puis 24h à 11°C	Non	0,54
			Oui	0,88

L'utilisation de la souche *E. coli* Arctic Express (DE3) surexprimant deux protéines chaperonnes actives à basse température, du milieu TB-AIM et du tag Strep-Tag n'ont pas permis l'expression de SamAA. Par ailleurs, l'abaissement de la température de culture à 11 °C a dû être compensé par une augmentation du temps de culture pour obtenir une quantité de biomasse bactérienne similaire à celle obtenue lors d'une culture à 37 °C, sans que cela conduise toutefois à la production de la protéine recherchée.

5.2.1.4. Production avec *E. coli* Rosetta 2 (DE3)

Plusieurs études présentées dans le chapitre I ont montré la possibilité d'améliorer l'expression recombinante de certaines aminoacylases sous une forme soluble en utilisant *E. coli* Rosetta 2 (DE3) (Liu et al., 2006 ; Hou et al., 2016). Les travaux d'expression de SamAA se sont donc poursuivis en étudiant cette souche en absence du plasmide pLysS (Figure V.11). Plusieurs constructions plasmidiques ont été testées : le plasmide pET-15b avec le gène non-optimisé de SamAA et un tag 6-His en position N-terminale du gène et les plasmides pET-28a avec le gène de SamAA optimisé pour son expression chez *E. coli* et un tag 6-His en position N-terminale ou C-terminale du gène (commandés chez Genscript). Les milieux de culture LB et Overnight ont été utilisés pour les cultures bactériennes ; ceux-ci peuvent être supplémentés ou non en cobalt (0,25 mM). Les cultures bactériennes ont été menées à 37 °C puis sont arrêtées 4 h ou 20 h après l'induction. La température de certaines cultures avec le milieu Overnight a été abaissée à 20 °C après l'induction pour ralentir la croissance bactérienne et favoriser le bon repliement de l'enzyme recombinante. Des témoins négatifs ont été réalisés avec les systèmes contenant le gène de SamAA sans faire l'induction dans le milieu LB sans cobalt pendant 24 h et à 37 °C. Les cultures n'ont pas été répétées.

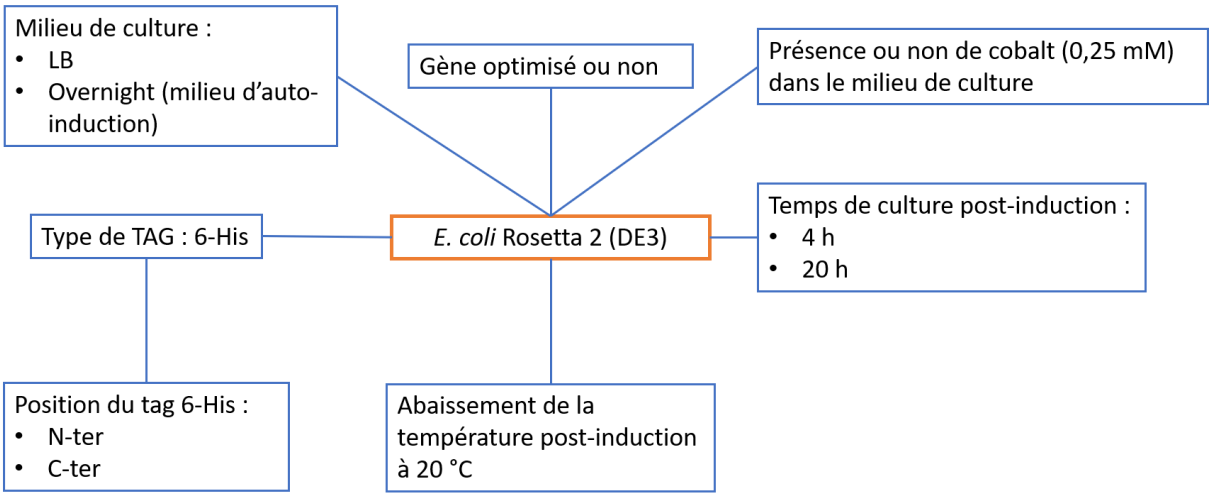


Figure V.11 : Récapitulatif des conditions étudiées pour la surexpression de SamAA chez *E. coli* Rosetta 2 (DE3).

La croissance bactérienne a été suivie par mesure de la DO à 600 nm (Figure V.12). Les cultures en milieu LB ont été induites par ajout d'IPTG (0,1 mM) à 0,4 de DO_{600} en présence de cobalt, et à 0,6 de DO_{600} en absence de cobalt. La présence ou l'absence de cobalt dans le milieu de culture a un effet significatif sur la croissance. Au bout de 24 h, les cultures sans cobalt sont déjà en phase stationnaire ou en phase de déclin alors que les cultures avec cobalt sont encore en phase exponentielle. Ce résultat confirme l'influence négative du cobalt présent dans le milieu de culture sur la vitesse de la croissance bactérienne. Les DO_{600} mesurées après 16 h de culture avec le milieu Overnight sont récapitulées dans le tableau V.3. L'impact du cobalt semble moins marqué que précédemment. Le tag en position C-terminale du gène est favorable à la croissance bactérienne. La croissance bactérienne est très importante en présence du gène de SamAA et de l'enzyme recombinante chez *E. coli* Rosetta 2 (DE3) quand la température est de 37 °C. Elle devient très faible quand la température est abaissée à 20 °C, malgré un milieu de culture enrichi (Overnight). La croissance bactérienne n'est pas influencée par la nature optimisée ou non du gène, ou par le changement de plasmide.

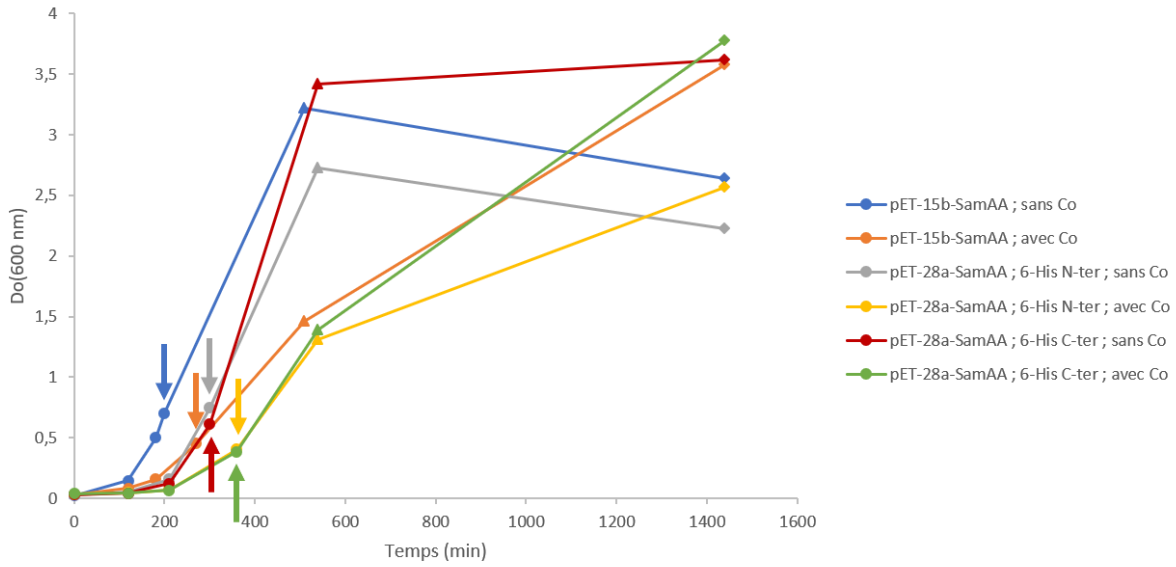


Figure V.12 : Suivi de la densité optique du milieu de culture LB, avec et sans cobalt, à 600 nm, lors de l'expression protéique chez les souches *E. coli* Rosetta 2 (DE3) transformées par les plasmides pET-15b-SamAA, pET-28a-SamAA 6-His N-ter ou pET-28A-SamAA 6-His C-ter. Chaque couleur renvoie à une condition particulière comme mentionné dans la légende. Le moment d'induction est indiqué par les flèches. La densité optique relevée à 4 h post induction est indiquée par un symbole en triangle. La densité optique relevée à 24 h de culture est indiquée par un symbole en losange.

Tableau V.3 : Mesure de la densité optique du milieu de culture Overnight (OV) avec et sans cobalt, en fin de culture (16 h), à 600 nm, lors de l'expression protéique chez les souches *E. coli* Rosetta 2 (DE3) transformées par les plasmides pET-15b-SamAA, pET-28a-SamAA 6-His N-ter ou pET-28a-SamAA 6-His C-ter.

Souche	Gène optimisé	Tag	Ajout de cobalt	Température post-induction	Do_{600}
E. coli Rosetta 2 (DE3) avec OV	Oui	6-His N-ter	Non	37 °C	19,1
			Oui		21,2
		6-His C-ter	Non		22,2
			Oui		22,4
	Non	6-His N-ter	Non		21,3
			Oui		18
			Oui	20 °C	1

Les cultures ont été arrêtées 4 h ou 20 h après l'induction dans le cas du milieu LB et au bout de 16 h dans le cas du milieu Overnight. Les résultats de l'analyse SDS-PAGE des prélèvements issus des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter et le milieu LB sont visibles en figure V.13. Une bande différentielle indiquant la présence de SamAA (50 kDa) est visible dans les puits des fractions EBNL et FI issues des cultures induites. L'absence de l'enzyme dans les fractions solubles indique à nouveau une production sous la forme de corps d'inclusion. Il n'y a pas d'expression de SamAA sans faire une induction avec le plasmide pET-28a-SamAA car cette bande n'est pas visible dans les puits EBNL et FI de la

culture non-induite du témoin négatif contenant ce plasmide. Ceci est confirmé lors de l'analyse des prélèvements issus des autres cultures menées avec ce plasmide (voir en annexe 1). SamAA avec le plasmide pET-15b-SamAA est également exprimée sous la forme de corps d'inclusion car les bandes de surexpression sont seulement visibles dans les puits EBNL et FI induits. Elles sont également visibles dans les puits EBNL et FI du témoin négatif non-induit, ce qui est le signe d'une fuite au niveau du promoteur inductif (Figure V.14). Les autres analyses SDS-PAGE des cultures réalisées avec ce plasmide confirment ce résultat ; elles sont visibles en annexe 1.

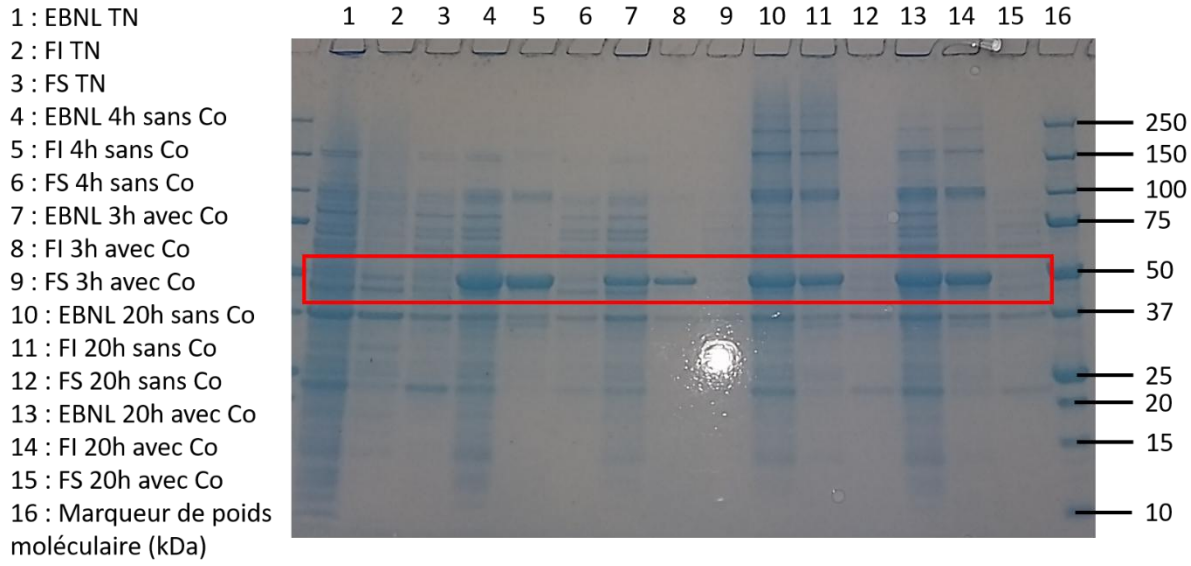


Figure V.13 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu LB supplémenté ou non en cobalt (0,25 mM). Les cultures ont été stoppées 3-4 h ou 20 h post induction. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

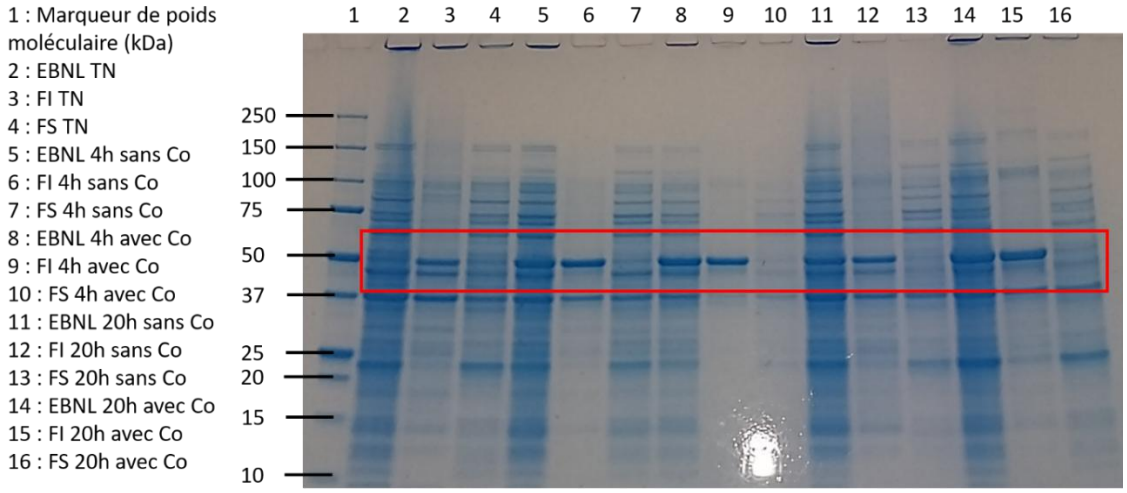


Figure V.14 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-15b-SamAA 6-His N-ter en milieu LB supplémenté ou non en

cobalt (0,25 mM). Les cultures ont été stoppées 3-4 h ou 20 h post induction. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

L'activité de synthèse de la N-α-lauroyl-L-lysine a été vérifiée dans l'ensemble des fractions solubles concentrées. L'analyse des milieux réactionnels montre que certaines fractions solubles concentrées conduisent à la synthèse du produit attendu (Tableau V.4). Une activité spécifique apparente de 80 mg de N-α-lauroyl-L-lysine /g*h est obtenue en utilisant la souche *E. coli* Rosetta 2 (DE3) avec le plasmide pET-15b-SamAA 6-His N-ter et le milieu Overnight supplémenté en cobalt. En présence du tag en position N-terminale du gène et du milieu LB supplémenté en cobalt, l'activité est plus élevée à 4 h qu'à 20 h après l'induction. L'ajout de cobalt conduit à une augmentation de l'activité spécifique apparente tandis que le milieu LB et le tag en position C-terminale du gène semblent engendrer une activité moindre. L'abaissement de la température à 20 °C n'a pas eu d'impact positif sur la production de SamAA sous forme soluble.

Tableau V.4 : Activités spécifiques apparentes pour la synthèse de la N-α-lauroyl-L-lysine en présence des fractions solubles concentrées des cultures d'*E. coli* Rosetta 2 (DE3) transformées par le plasmide pET-15b-SamAA 6-His ou pET-28a-SamAA 6-His N-ter ou C-ter en milieu LB ou OV, supplémenté ou non en cobalt.

Fractions solubles concentrées					Activité spécifique
Plasmide	Place du TAG	Milieu utilisé	Conditions particulières de culture	Présence de cobalt	apparente (mg/g*h)
pET-28a	N-ter	LB	Témoin négatif	Non	0,00
			Durée de 4h post-induction	Non	0,00
				Oui	41,37
			Durée de 20h post-induction	Non	0,00
				Oui	18,75
		OV	***	Non	18,20
				Oui	61,75
	C-ter	LB	Témoin négatif	Non	0,00
			Durée de 4h post-induction	Non	0,00
				Oui	0,00
			Durée de 20h post-induction	Non	0,00
				Oui	2,24
		OV	***	Non	2,80
				Oui	15,01
pET-15b	N-ter	LB	Témoin négatif	Non	0,00
			Durée de 4h post-induction	Non	1,85
				Oui	18,53
			Durée de 20h post-induction	Non	2,57
				Oui	10,37
		OV	***	Non	29,66
				Oui	**80,12**
			Température de culture à 20°C	Oui	0,00

Si les analyses SDS-PAGE montrent la présence majoritaire de l'enzyme sous forme de corps d'inclusion, SamAA est produite sous forme soluble en faible quantité car une activité enzymatique spécifique apparente est détectée. L'utilisation de la souche *E. coli* Rosetta 2

(DE3) avec le plasmide pET-15b et le milieu de culture Overnight supplémenté en cobalt a conduit à l'activité spécifique apparente la plus élevée (80 mg de N-α-lauroyl-L-lysine / g*h). Le cobalt a une influence négative sur la croissance bactérienne mais sa présence dans le milieu de culture augmente l'activité spécifique apparente, ce qui laisse supposer qu'il aide au bon repliement de l'enzyme. Le positionnement du tag en position N-terminale semble favorable à l'expression de SamAA. L'utilisation d'un milieu enrichi autoinductif a un coût financier supérieur à celui du milieu LB. Son utilisation et la présence d'un gène de SamAA non-optimisé pour son expression chez *E. coli* favorise l'expression et l'activité de l'enzyme recombinante avec la souche *E. coli* Rosetta 2 (DE3).

5.2.1.5. Production avec *E. coli* BL21 Gold pLysS (DE3)

E. coli BL21 Gold est la dernière souche issue d'*E. coli* étudiée dans le cadre de l'expression recombinante de SamAA. Elle présente plusieurs mutations augmentant l'efficacité de la transformation des cellules et du maintien de l'intégrité de l'ADN plasmidique (Greener et al., 2004). A notre connaissance, aucune étude n'a été menée sur l'expression d'une aminoacylase recombinante avec cette souche. Elle nous a été conseillée par une équipe de la TU Delft où elle est utilisée de façon régulière. Cette partie rapporte les résultats obtenus avec la souche *E. coli* BL21 Gold (DE3) contenant le plasmide pLysS et le plasmide pET-28a-SamAA portant la version optimisée du gène de SamAA pour son expression chez *E. coli*, et un tag 6-His en position N-terminale ou C-terminale du gène. Les cultures bactériennes ont été menées à 37 °C uniquement en milieu LB supplémenté ou non en cobalt (0,25 mM) ; elles ont été stoppées 4 h ou 20 h après l'induction (Figure V.15). L'expression de SamAA a été induite par ajout d'IPTG (0,1 mM) à une DO_{600} de 0,6 en absence de cobalt, et de 0,4 en présence de cobalt. Un témoin négatif a été réalisé sans induction à l'IPTG. Les cultures n'ont pas été répétées.

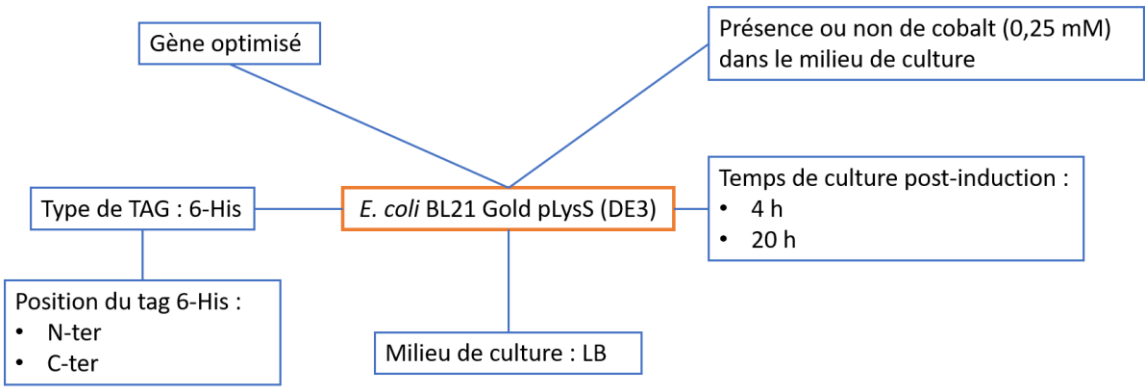


Figure V.15 : Récapitulatif des conditions étudiées la surexpression de SamAA chez *E. coli* BL21 Gold pLysS (DE3).

Le suivi de croissance des quatre cultures est présenté sur la figure V.16. Deux profils sont observés. La position du tag ne semble pas avoir d'influence sur la croissance, contrairement au cobalt. En absence de cobalt dans le milieu de culture, la croissance bactérienne suit un profil classique. En revanche, la présence de cobalt ralentit fortement la croissance bactérienne, affaiblissant la phase exponentielle.

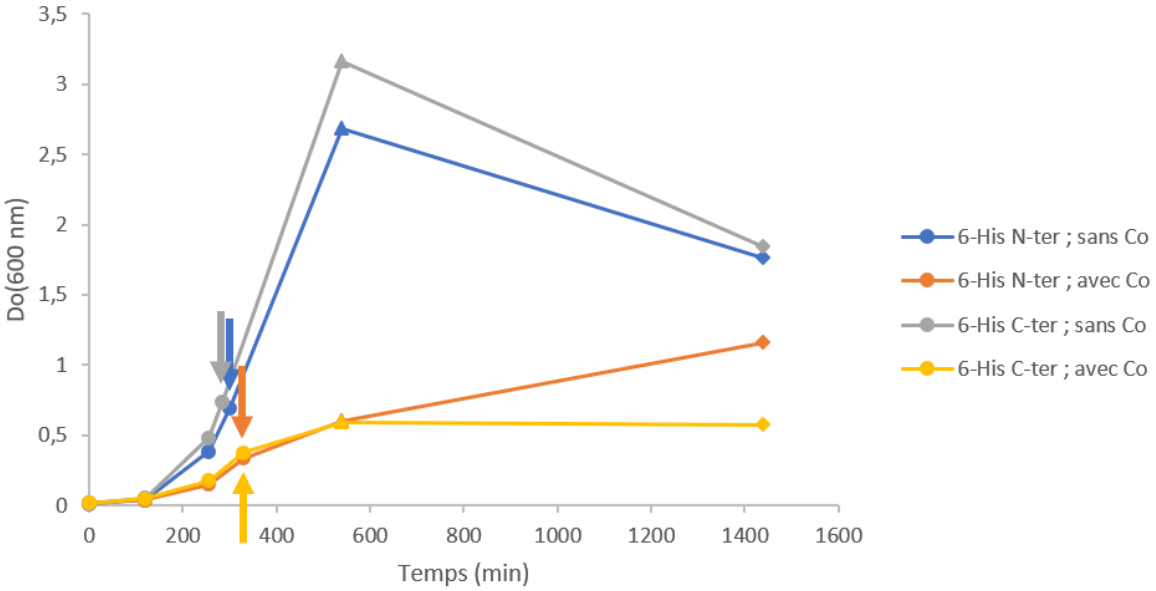


Figure V.16 : Suivi de la densité optique à 600 nm du milieu de culture LB, supplémenté ou non en cobalt, lors de l'expression protéique chez les souches *E. coli* BL21 Gold pLysS (DE3) transformées par les plasmides pET-28a-SamAA 6-His N-ter ou pET-28A-SamAA 6-His C-ter. Chacune des couleurs renvoie à une condition particulière précisée en légende. Le moment d'induction est indiqué par les flèches. La densité optique relevée à 4 h post induction est indiquée par un symbole en triangle. La densité optique relevée à 24 h de culture est indiquée par un symbole en losange.

Les cultures ont été arrêtées 4 h ou 20 h après l'induction. Le résultat de l'analyse SDS-PAGE des prélèvements issus des cultures avec le plasmide pET-28a-SamAA 6-His N-ter et le milieu LB avec ou sans cobalt est visible en figure V.17. L'analyse des fractions solubles concentrées obtenues avec le plasmide pET-SamAA 6-His C-ter est visible en annexe 1. SamAA (50 kDa) est produite sous la forme de corps d'inclusion visibles dans les fractions insolubles. Aucune surexpression significative de SamAA n'est visible dans les fractions solubles, qu'elles soient concentrées ou non. Il n'y a pas de fuite du promoteur en absence d'induction. La présence de cobalt dans le milieu de culture conduit à des fractions moins riches en protéines que celles issues de milieux sans cobalt.

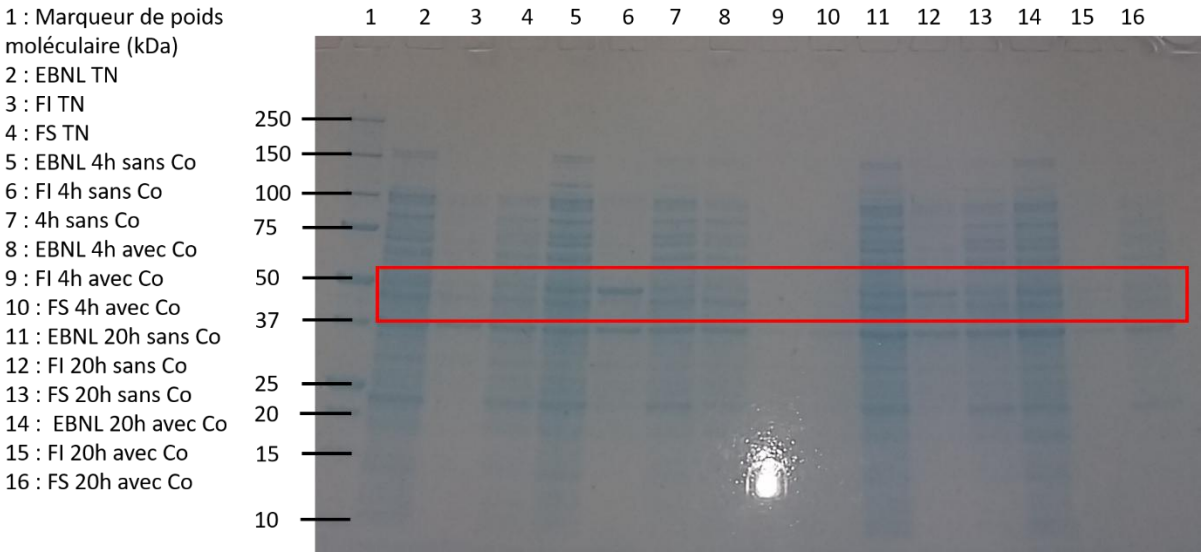


Figure V.17 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold pLysS (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu LB supplémenté ou non en cobalt (0,25 mM). Les cultures ont été stoppées 3-4 h ou 20 h post induction. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

Les résultats des tests d'activité enzymatique en présence des fractions solubles concentrées sont présentés au tableau V.5. Seules les fractions issues des cultures supplémentées en cobalt présentent une activité spécifique apparente de synthèse de la N-α-lauroyl-L-lysine. Cette activité est augmentée pour un temps de culture de 24 h ou lors de l'utilisation du tag 6-His en position N-terminale du gène. Ainsi, une activité spécifique apparente de 55 mg de N-α-lauroyl-L-lysine / g*h est obtenue en utilisant la fraction soluble concentrée issue de la culture de 24 h d'*E. coli* BL21 Gold pLysS avec le plasmide pET-28a-SamAA 6-His N-ter, en milieu de culture LB supplémenté en cobalt.

Tableau V.5 : Activités spécifiques apparentes pour la synthèse de la N-α-lauroyl-L-lysine en présence des fractions solubles concentrées des cultures d'*E. coli* BL21 Gold pLysS (DE3) transformées par le plasmide pET-28a-SamAA 6-His N-ter ou C-ter avec le milieu LB supplémenté ou non en cobalt.

Extraits de culture					Activité spécifique apparente (mg/g*h)
Plasmide	Place du TAG	Milieu utilisé	Temps de culture post induction	Présence de cobalt	
pET-28a	N-ter	LB	Témoin négatif	Non	0,00
			4h	Non	0,00
				Oui	34,56
			20h	Non	0,00
				Oui	**54,82**
	C-ter		Témoin négatif	Non	0,00
			4h	Non	0,00
				Oui	6,52
			20h	Non	0,00
				Oui	15,00

Ces résultats montrent que le système *E. coli* BL21 Gold pLysS (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter et le milieu LB supplémenté en cobalt permet d'obtenir la meilleure activité spécifique apparente, malgré une croissance bactérienne très faible et l'absence d'enzyme soluble visible en SDS-PAGE. L'ajout de cobalt dans le milieu a un impact très positif sur l'activité spécifique apparente des fractions solubles, renforçant l'hypothèse d'un rôle bénéfique au bon repliement de SamAA. Cependant à l'inverse, la présence de cobalt a un impact négatif très important sur la vitesse de croissance bactérienne. Les travaux ont été poursuivis avec la souche dépourvue du plasmide pLysS afin de vérifier son effet éventuel sur la production de l'enzyme.

5.2.1.6. Production avec *E. coli* BL21 Gold (DE3)

Des cellules compétentes d'*E. coli* BL21 Gold (DE3) ont été transformées par le plasmide pET-28a-SamAA 6-His N-ter ou 6-His C-ter, qui est porteur de la version optimisée du gène de SamAA pour son expression chez *E. coli*. Les caractéristiques de cette souche ont été présentées dans la sous-partie précédente (Greener et al., 2004). L'expression de SamAA a été induite par l'ajout d'IPTG (0,1 mM ; respectivement à 0,6 et 0,4 de DO_{600} en absence et en présence de cobalt) ; les cultures ont été menées selon les différentes conditions présentées en figure V.18. Les études numériques de SamAA suggèrent que l'enzyme serait une métalloenzyme, c'est pourquoi l'ajout de cobalt dans le milieu de culture dès le début de culture ou au moment de l'induction (COI) est étudié. Pour tenter de favoriser le bon repliement de SamAA et limiter la formation de corps d'inclusion, deux conditions supplémentaires particulières ont été étudiées : un abaissement de la température à 20 °C après l'induction et la réalisation d'un choc thermique et osmotique (CTO) (Papp et al., 2006 ; Bhatwa et al., 2021). Une culture faisant office de témoin négatif avec un plasmide pET-28a-SamAA 6-His N-ter vide est conduite dans les conditions permettant d'obtenir la meilleure activité spécifique apparente. Une seconde culture témoin négatif avec le plasmide pET-28a-SamAA a été réalisée sans induction. Toutes les cultures ont été réalisées une fois, sauf celles permettant l'obtention de la meilleure activité apparente avec le plasmide pET-28a-SamAA et avec le plasmide pET-28a vide qui ont été répétées trois fois.

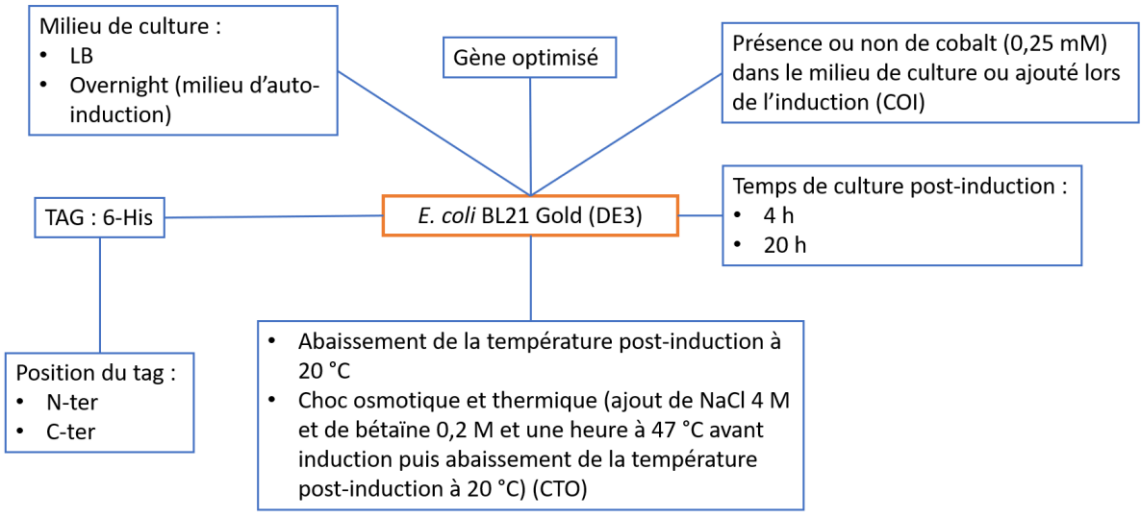


Figure V.18 : Récapitulatif des conditions étudiées pour la surexpression de SamAA chez *E. coli* BL21 Gold (DE3).

Les courbes de croissance obtenues sont visibles en figure V.19. Les DO_{600} mesurées après 16 h de culture avec le milieu Overnight sont récapitulées dans le tableau V.6. Une bonne croissance bactérienne est observée dans les cultures sans cobalt, car les cellules bactériennes avec le milieu LB sont en phase stationnaire dès 24 h. La présence du cobalt a un impact négatif important sur la croissance bactérienne : la phase de latence est allongée quand il est ajouté au début de la culture ; la croissance est fortement perturbée quand il est ajouté au moment de l'induction. De même, l'abaissement de la température au moment de l'induction et le choc thermique et osmotique ont un impact négatif sur la croissance bactérienne, comme le montre la cassure dans la courbe de la croissance.

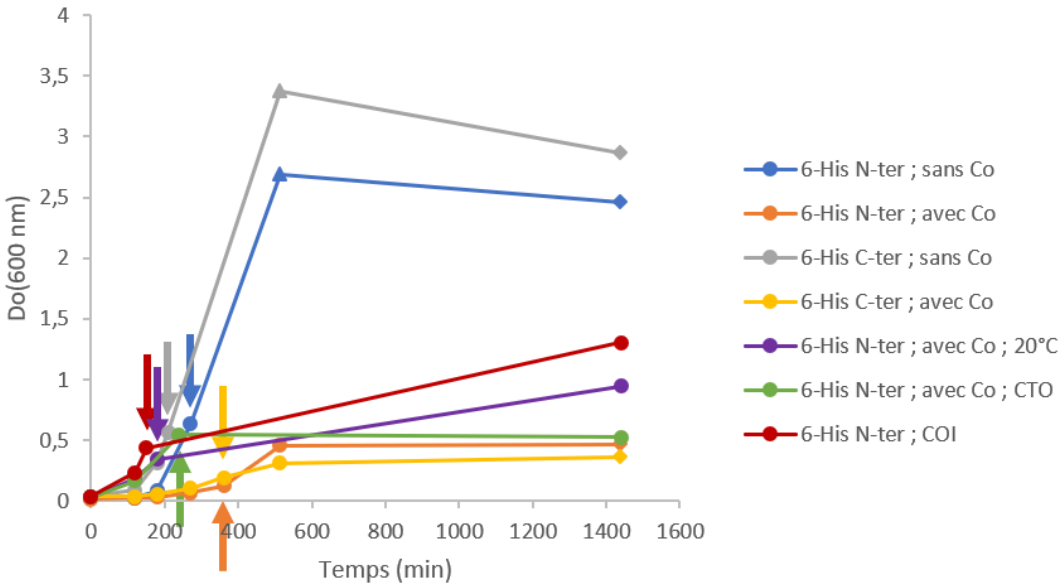


Figure V.19 : Suivi de la densité optique du milieu de culture LB, avec et sans cobalt, à 600 nm, lors de l'expression protéique chez les souches *E. coli* BL21 Gold (DE3) transformées par

les plasmides pET-28a-SamAA 6-His N-ter ou C-ter. Chaque couleur renvoie à une condition particulière précisée en légende. Le moment d'induction est indiqué par les flèches. La densité optique relevée à 4 h post induction est indiquée par un symbole en triangle. La densité optique relevée à 24 h de culture est indiquée par un symbole en losange.

Tableau V.6 : Mesure de la densité optique du milieu de culture Overnight (OV) en fin de culture (16 h) avec et sans cobalt, à 600 nm, lors de l'expression protéique chez les souches *E. coli* BL21 Gold (DE3) transformées par les plasmides pET-28a-SamAA 6-His N-ter ou pET-28a-SamAA 6-His C-ter.

Souche	Tag	Ajout du cobalt	Do_{600}
E. coli BL21 Gold (DE3)	6-His N-ter	Non	19,6
		Oui	12,6
	6-His C-ter	Non	14,9
		Oui	15,7

Après plusieurs répétitions des cultures ayant une croissance normale, un arrêt provisoire de la croissance a été observé en pleine phase exponentielle (DO_{600} d'environ 0,2) avec la souche *E. coli* BL21 Gold (DE3), pendant plusieurs heures (à minima 4 h), en présence du cobalt, et en présence ou non du gène de SamAA dans le plasmide pET-28a. La croissance reprend ensuite au cours de la nuit ; la DO_{600} mesurée après 24 h de culture atteint une valeur de 2. Le phénomène reste à ce jour inexpliqué. Il évoque un arrêt de croissance provoqué par l'épuisement d'un composé organique dans le milieu et une adaptation du métabolisme bactérien, ce qui est incompatible avec l'utilisation du milieu LB. Plusieurs éléments du protocole ont été modifiés comme la préparation d'un nouveau milieu LB, de nouvelles solutions d'antibiotiques et d'une nouvelle solution de cobalt, le changement d'incubateur (et la vérification du premier incubateur utilisé) et la réalisation de nouveaux clones. Cependant, malgré tous ces changements, le problème persiste. Ce phénomène d'arrêt de croissance n'est jamais observé dans les cultures conduites en l'absence de cobalt. L'ajout de cobalt au moment de l'induction quand la DO_{600} est à 0,4 a été la seule solution pour retrouver des profils de croissance normaux.

Les résultats de l'analyse SDS-PAGE des prélèvements est visible en figure V.20. Seul le gel correspondant à la culture d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter et le milieu de culture LB supplémenté en cobalt (0,25 mM) est visible. Les autres gels présentent un profil similaire ; ils sont visibles en annexe 1. Il n'y a pas de production de SamAA dans le témoin négatif non-induit. La production de SamAA (50 kDa) a lieu sous la forme de corps d'inclusion avec les milieux de culture LB et Overnight ; la bande différentielle relative à l'enzyme est en effet visible uniquement dans les extraits bruts non-lysés et les fractions insolubles. Il n'y a pas de surexpression visible de SamAA dans les fractions solubles

concentrées ou non, quelles que soient les conditions testées. L'abaissement de la température après l'induction et l'application d'un choc thermique et osmotique n'ont aucun effet positif sur la production de SamAA, que ce soit sous forme soluble ou insoluble.

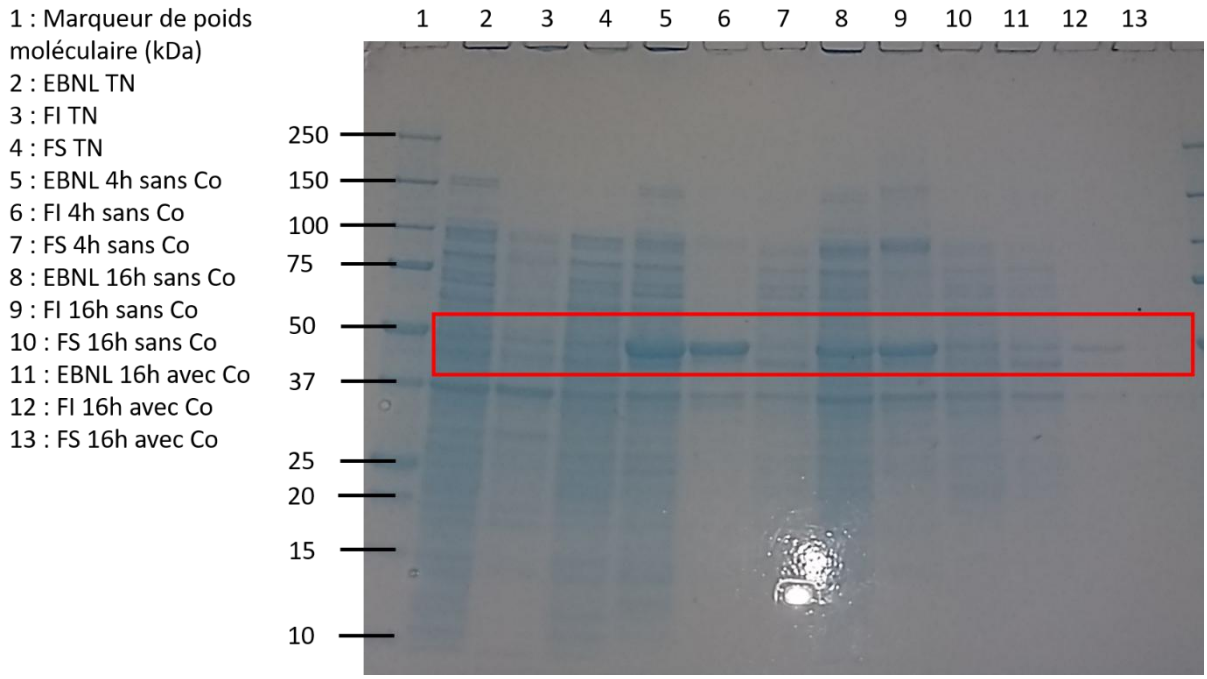


Figure V.20 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu LB supplémenté ou non en cobalt (0,25 mM). Les cultures ont été stoppées 4h ou 20h post induction. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

Les valeurs d'activité spécifique apparente liée à la présence de SamAA soluble sont présentées dans le Tableau V.7. Une activité spécifique apparente de 339 (± 50) mg de N-α-lauroyl-L-lysine / g*h est obtenue en utilisant la fraction soluble concentrée issue de la culture d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter et le milieu de culture LB supplémenté en cobalt dès le début de la culture. L'analyse du milieu réactionnel par HPLC confirme la présence du produit attendu (Figure V.21). Ces résultats constituent les meilleurs jamais obtenus avec la souche *E. coli*. Il est possible de contourner l'arrêt de croissance de la souche en présence de cobalt en ajoutant ce dernier au moment de l'induction. Dans ces conditions, une perte de 40 % de l'activité spécifique apparente est observée comparativement à celle obtenue quand le cobalt est ajouté au début de la culture bactérienne. Les résultats indiquent une influence positive du cobalt et du tag en position N-terminale du gène sur la production de SamAA soluble, de même que l'absence du plasmide pLysS. Aucune activité enzymatique n'est détectée suite à l'abaissement de la température après l'induction ou suite au choc thermique et osmotique.

Tableau V.7 : Activités spécifiques apparentes pour la synthèse de la N-α-lauroyl-L-lysine en présence des fractions solubles concentrées des cultures d'*E. coli* BL21 Gold (DE3) transformées par le plasmide pET-28a-SamAA 6-His N-ter ou C-ter, en milieu LB ou OV supplémenté ou non en cobalt.

Extraits de culture					Activité spécifique apparente (mg/g*h)
Plasmide	Place du TAG	Milieu utilisé	Conditions de culture	Présence de cobalt	
pET-28a	N-ter	LB	Témoin négatif	Non	0,00
			4h	Non	0,00
			20h	Non	0,00
				Oui	**339,00**
			20°C post induction	Oui	0,00
			CTO	Oui	0,00
			COI	Oui	210,23
		OV	***	Non	20,64
				Oui	148,68
	C-ter	LB	Témoin négatif	Non	0,00
			4h	Non	0,00
			20h	Non	0,00
				Oui	105,27
		OV	***	Non	0,00
				Oui	25,37

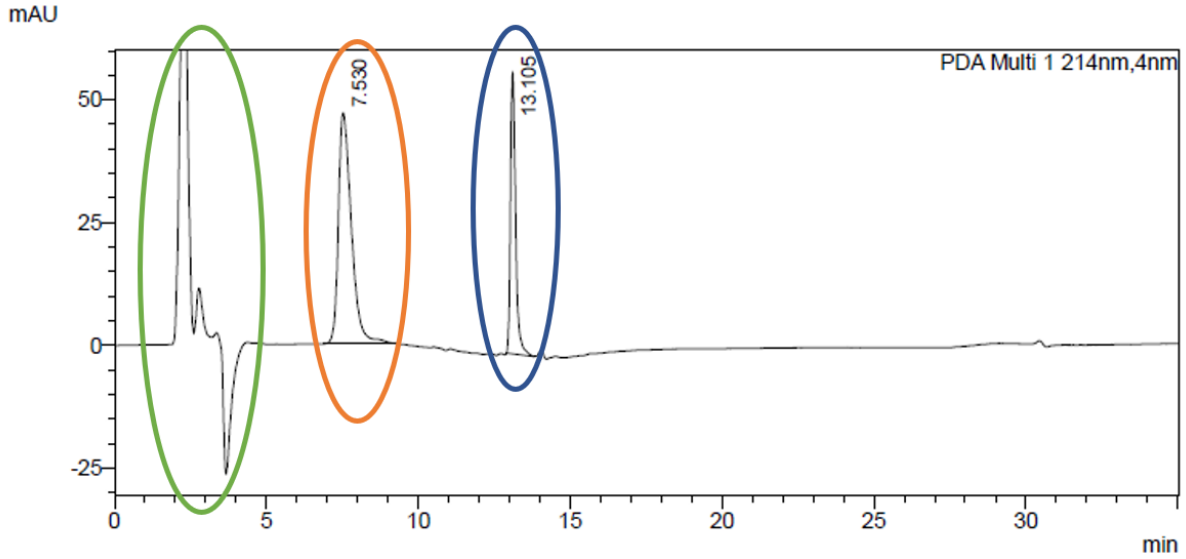


Figure V.21 : Analyse par chromatographie du milieu de synthèse de la N-α-lauroyl-L-lysine en présence de la fraction soluble concentrée issue de la culture d'E. coli BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter, en milieu LB supplémenté en cobalt dès le début de la culture. Le temps et la température réactionnels sont respectivement de 48 h et 45 °C. Les pics correspondant à la N-α-lauroyl-L-lysine, l'acide laurique et au solvant sont encerclés respectivement en orange, bleu et vert. La L-lysine n'est pas détectable dans ces conditions.

A la vue de ces résultats, nous pouvons dire que le cobalt a un impact très négatif sur la vitesse de la croissance bactérienne d'*E. coli* BL21 Gold (DE3). La production de SamAA avec cette souche a lieu sous forme de corps d'inclusion d'après l'analyse SDS-PAGE. Toutefois, une activité spécifique apparente jamais obtenue jusqu'ici est mesurée pour la fraction soluble concentrée d'*E. coli* BL21 Gold (DE3), avec le plasmide pET-28a-SamAA 6-His N-ter, et en

milieu de culture LB supplémenté en cobalt dès le début de la culture. Le problème de croissance rencontré avec ce système en présence de cobalt peut être contourné en ajoutant le cobalt au moment de l'induction ; ceci conduit à une baisse de l'activité spécifique apparente, toutefois cette perte reste limitée (40 %).

5.2.1.7. Conclusion de la production de SamAA chez *E. coli*

Plusieurs paramètres et conditions ont été étudiés pour surexprimer SamAA sous une forme soluble chez *E. coli* : différentes souches (Origami B (DE3), Rosetta 2 pLysS (DE3), Arctic Express (DE3), Rosetta 2 (DE3), BL21 Gold pLysS (DE3), BL21 Gold (DE3)), différents milieux de culture (LB, Overnight ou TB-AIM), deux versions du gène de SamAA (optimisé ou non pour son expression chez *E. coli*), deux tags différents (6-His ou Strep-Tag) pouvant être placés à l'extrémité N- ou C-terminale du gène et différentes conditions de culture (temps, température, ajouts d'éléments supplémentaires au milieu de culture). Une activité spécifique apparente de synthèse de la N-α-lauroyl-L-lysine en présence des fractions solubles concentrées issues des cultures a été mise en évidence, indiquant la présence de l'enzyme. Cette activité constitue un bon indicateur de la production de SamAA recombinante sous une forme soluble et active.

La meilleure activité a été mesurée dans la fraction soluble concentrée d'*E. coli* BL21 Gold (DE3) dans les conditions suivantes : gène optimisé de SamAA ayant le tag 6-His en position N-ter, cloné dans le plasmide pET-28a ; milieu LB supplémenté en cobalt (0,25 mM) ; à 37 °C pendant 24 h ; induction à l'IPTG (0,1 mM) à une DO_{600} de 0,4. La répétition des cultures a mis en évidence un problème de croissance de la souche *E. coli* BL21 Gold (DE3) en présence du cobalt, qu'il est possible de contourner en ajoutant le cobalt au moment de l'induction, ce qui entraîne une baisse de l'activité d'environ 40 %.

Malgré certains résultats encourageants, force est de constater qu'il n'y a pas de surexpression de SamAA sous forme soluble chez *E. coli*. L'enzyme reste produite majoritairement sous forme de corps d'inclusion, ce qui a orienté la suite du travail vers la production de l'enzyme dans un nouvel hôte d'expression.

5.2.2. Production chez *Streptomyces lividans*

Les travaux d'expression de SamAA se sont poursuivis en changeant d'hôte d'expression. *Streptomyces lividans* a été choisi pour plusieurs raisons. SamAA est une protéine provenant

du micro-organisme *Streptomyces ambofaciens*. L'hypothèse formulée est qu'un hôte proche sur le plan génétique pourrait limiter la formation de corps d'inclusion et favoriser le bon repliement de l'enzyme, en se basant sur le principe que les machineries métaboliques des deux bactéries devraient être proches également. Les bactéries du genre de *Streptomyces* ont une faible activité protéolytique endogène et une capacité de sécrétion élevée. Elles sont adaptées à l'expression des gènes riches en GC et forment peu de corps d'inclusion. Des modifications génétiques et des conditions de culture peuvent diminuer les caractères hétéromorphes et visqueux de ces cultures bactériennes filamenteuses, qui créent des conditions défavorables à une utilisation industrielle (van Wezel et al., 2006 ; Sevillano et al., 2016 ; Berini et al., 2020 ; Haeger et al., 2024b). Quatre systèmes de *S. lividans* TK24 ont été construits à partir de différents plasmides présentés dans le tableau V.8, chaque plasmide existant avec et sans le gène de SamAA. Un vecteur intégratif permet l'insertion du gène de SamAA au sein du génome de *S. lividans*, tandis qu'un vecteur réplicatif maintenu par une pression sélective (ici, l'apramycine) contient le gène d'intérêt. Un promoteur constitutif permet l'expression stable du gène dans des conditions variables tandis qu'un promoteur inductible permet l'expression du gène d'intérêt uniquement après l'ajout de l'inducteur.

Tableau V.8 : Principales caractéristiques des quatre plasmides étudiés pour l'expression de SamAA chez *S. lividans*.

Nom du plasmide	Caractéristiques
pIB139	Vecteur intégratif avec un promoteur fort et constitutif (*ermE*p*)
pIJ8600	Vecteur intégratif avec un promoteur fort et inductible au thiostrepton
pIJ86	Vecteur réplicatif avec un promoteur fort et constitutif (*ermE*p*)
pIJ8660	Vecteur intégratif avec un promoteur constitutif et une séquence *vsi* codant pour un peptide signal pour la sécrétion

Les quatre systèmes de *S. lividans* ont été étudiés en erlenmeyers bafflés de 500 mL. Après l'obtention de spores fraiches sur milieu solide MSF (Mannitol Soja Flour), celles-ci ont été inoculées en pré-culture pendant deux jours, puis en culture pendant trois jours dans le milieu de fermentation. Aucun suivi de croissance n'a été réalisé pendant et à la fin des cultures. Les cultures avec les plasmides pIB139 pIJ8600 vides et pIB139-SamAA ont été réalisées en triplicats, contrairement aux autres systèmes qui n'ont fait l'objet d'aucune répétition. La production de SamAA a été vérifiée dans les milieux extra- et intra-cellulaires lors de l'étude de la souche contenant le plasmide pIJ8660 capable de sécréter l'enzyme recombinante.

Après 72 h de culture, extraction et lyse des cellules, la concentration protéique mesurée est supérieure à 2,5 g/L dans certaines fractions solubles qui n'ont pas été concentrées. Les autres

fractions ont été concentrées. Pour étudier la surexpression de SamAA, tous les prélèvements ont été analysés par SDS-PAGE. Les résultats obtenus avec le plasmide pIB139-SamAA sont visibles en figure V.22. Les gels correspondant aux autres cultures bactériennes sont visibles en annexe 1. Tous les gels présentent le même profil ; ils montrent l'absence de surexpression de SamAA, que ce soit sous forme soluble ou insoluble.

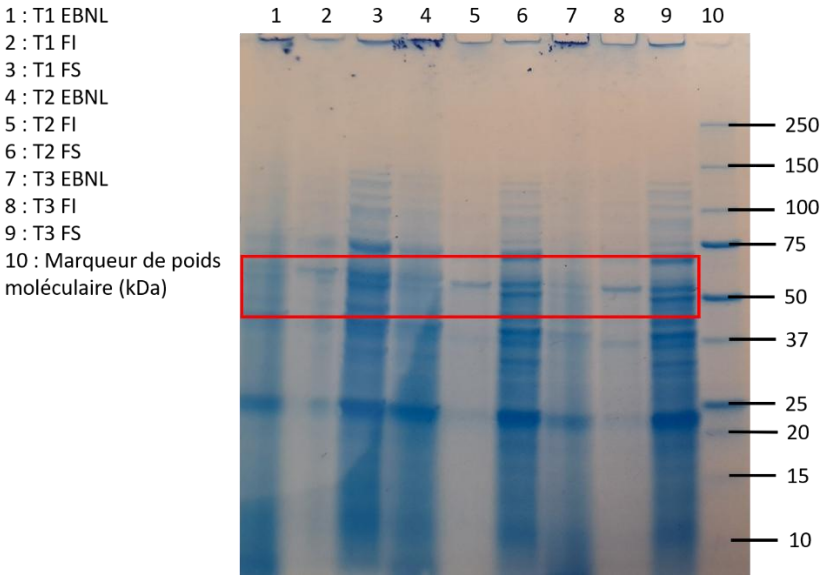


Figure V.22 : Analyse par SDS-PAGE des protéines issues des cultures de *S. lividans* avec le plasmide pIB139-SamAA StrepTag N-ter. T1, T2, T3 : erlenmeyer 1, 2, 3 ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble. Le cadre rouge indique la position de SamAA exprimée.

L'activité de synthèse de la N-α-lauroyl-L-lysine a été étudiée dans les fractions solubles concentrées ou non issues des cultures de *S. lividans* menées avec un plasmide vide ou contenant le gène de SamAA. Il est nécessaire de réaliser un témoin négatif sur les fractions obtenues avec un plasmide vide, car *S. lividans* présente une activité endogène de N-acylation de la fonction amine α de la L-lysine. La synthèse a été menée en présence de cobalt, car il a été démontré précédemment que ce dernier augmente l'activité catalytique spécifique de l'enzyme. Les résultats de l'activité spécifique apparente de chaque fraction sont récapitulés dans le tableau V.9. La meilleure activité spécifique apparente égale à 635 mg de N-α-lauroyl-L-lysine / g*h est obtenue avec la fraction concentrée (facteur dix) issue de la culture avec le plasmide pIJ8600-SamAA. Parmi les fractions non concentrées, la fraction soluble obtenue avec le plasmide pIB139-SamAA présente l'activité spécifique apparente la plus élevée (214 mg de N-α-lauroyl-L-lysine / g*h).

Tableau V.9 : Activités spécifiques apparentes pour la synthèse de la N-α-lauroyl-L-lysine en présence des fractions solubles issues des cultures de *S. lividans* transformées par les plasmides pIB139, pIJ8600, pIJ86, pIJ8660 vides ou avec SamAA.

Extraits de cultures			Activité spécifique apparente (mg/g*h)	Ecart-type
Plasmide	Présence du gène de SamAA	Type de fractions		
pIB139	Non	**Fractions solubles**	8,01	4,7
	Oui		**214,14**	16,51
pIJ8600	Non	**Fractions solubles concentrées**	209,74	13,47
	Oui		**635,62**	Pas de réplicat de culture bactérienne
pIJ86	Non	Fractions solubles	11,56	
	Oui		31,38	
pIJ8660	Non	Fractions solubles concentrées	16,15	
	Oui		38,05	
	Non	Surnageants concentrés	Pic non-visible	
	Oui		Pic non-visible	

5.2.3. Valorisation des corps d'inclusion de SamAA

A la vue des résultats obtenus lors de la surexpression de SamAA, il apparaît clairement que l'enzyme est produite très majoritairement sous la forme de corps d'inclusion. Une partie du travail réalisé a porté sur leur valorisation. Le refolding est une technique permettant la valorisation des corps d'inclusion produits lors de l'expression d'une protéine recombinante. Les corps d'inclusion sont produits en grande quantité et faciles à isoler ; toutefois, ces agrégats contiennent la protéine mal repliée et, donc généralement, inactive. Ils peuvent être transformés en protéines solubles et actives en suivant deux étapes (solubilisation des agrégats puis repliement sous une forme active) en utilisant différentes techniques (Kaur et al., 2018).

Le système *E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamAA 6-His C-ter permet une importante production de SamAA sous la forme de corps d'inclusion comme indiqué par les gels SDS-PAGE. Les fractions insolubles issues de ce système avec les milieux de culture LB et Overnight sans cobalt ont été choisies pour mener les essais de refolding. Celles-ci ont été resuspendues dans un tampon alcalin supplémenté en cobalt favorisant leur resolubilisation. Du cobalt est ajouté dans le tampon de refolding car sa présence dans le milieu de culture a permis précédemment d'augmenter la quantité de SamAA sous une forme soluble (Khan et al., 1998 ; Patra et al., 2001).

Après séparation de la fraction insoluble refoldée (FIR) et de la fraction soluble refoldée (FSR) par centrifugation, puis concentration de cette dernière par ultrafiltration (FSRc), une analyse SDS-PAGE a été réalisée, dont les résultats sont visibles en figure V.23. La présence d'une

bande de surexpression nette pouvant correspondre à SamAA (50 kDa) est observée dans les fractions solubles refoldées, concentrées ou non. Les corps d'inclusion se sont donc bien solubilisés dans le tampon alcalin. Une bande équivalente est également visible dans les puits des fractions insolubles refoldées, indiquant qu'une partie des corps d'inclusion n'a pas été solubilisée et suggérant la nécessité d'augmenter la durée de l'étape de refolding.

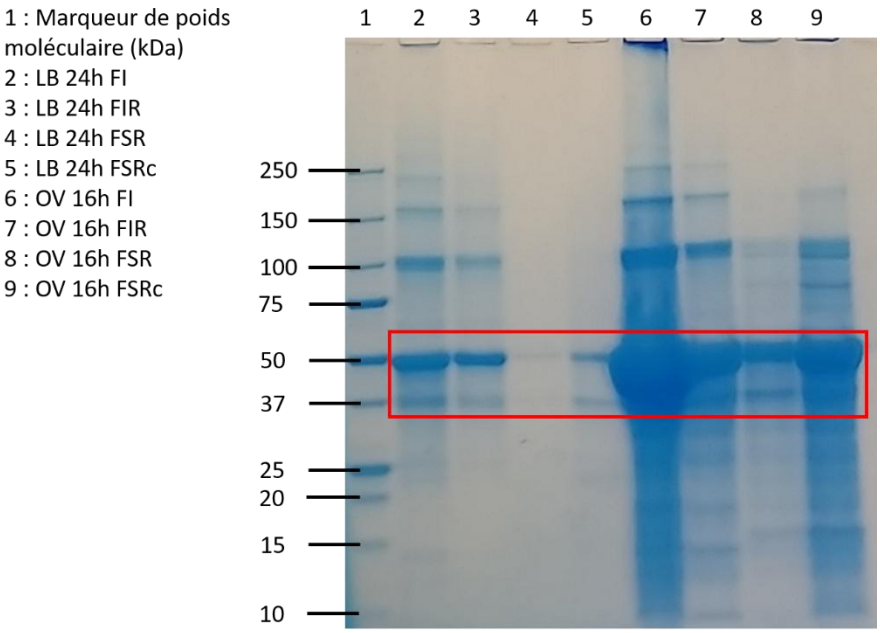


Figure V.23 : Analyse par SDS-PAGE des protéines issues du refolding des fractions insolubles issues de la culture avec *E. coli* Rosetta 2 (DE3), le plasmide pET-28a-SamAA 6-His C-ter, et le milieu LB ou OV sans cobalt. FI : fraction insoluble ; FIR : fraction insoluble refoldée ; FSR : fraction soluble refoldée ; FSRc : fraction soluble refoldée concentrée. Le cadre rouge indique la position de SamAA exprimée.

Le potentiel catalytique des fractions solubles refoldées et concentrées pour la synthèse de la N-α-lauroyl-L-lysine a été étudiée et quantifiée par HPLC. Aucune activité n'a été détectée. Plusieurs hypothèses ont été formulées pour expliquer ce résultat. Selon la première, les corps d'inclusion se sont solubilisés, mais SamAA ne s'est pas repliée correctement. Il serait alors nécessaire d'ajouter une étape et/ou des additifs dans le tampon de refolding pour favoriser le repliement correct de l'enzyme. Les résultats de l'étude numérique de SamAA ont montré que cette enzyme serait un homodimère dans lequel les deux chaines peptidiques sont impliquées dans la catalyse. Un gel d'électrophorèse en condition native aurait permis de vérifier la structure quaternaire de l'enzyme refoldée. La deuxième hypothèse est liée au caractère alcalin de la fraction protéique contenant l'enzyme. Son pH de 11,5 pourrait perturber l'activité de l'enzyme pour laquelle un pH optimal de 8 a été déterminé. Bien qu'un ajout d'acide ait été effectué pour compenser l'alcalinité apportée par la fraction refoldée, le pH du milieu de synthèse n'a pas été vérifié.

La procédure de refolding appliquée aux corps d'inclusion de SamAA a donc conduit des résultats encourageants dans la mesure où les corps d'inclusion ont pu être partiellement resolubilisés. Toutefois, aucune activité enzymatique n'a été détectée, suggérant un mauvais repliement de la protéine et/ou de mauvaises conditions de mise en œuvre du test d'activité enzymatique.

5.3. Purification de SamAA recombinante

Une tentative de purification de SamAA recombinante avait déjà été effectuée avant le démarrage de cette thèse. L'enzyme recombinante possédant un tag 6-His en position N-terminale, celle-ci avait été purifiée sur colonne IMAC par chromatographie d'affinité au nickel. Après élution en présence d'imidazole (de 80 mM à 250 mM), l'analyse SDS-PAGE des fractions recueillies n'avait révélé aucune bande correspondant à SamAA (50 kDa). L'analyse de l'activité d'hydrolyse de la N-α-acétyl-L-lysine par CCM avait montré que seule la fraction éluée avec 80 mM d'imidazole contenait de l'activité, suggérant une absence de purification (Nili, 2018). Ces résultats n'ayant jamais été confirmés, une nouvelle tentative de purification par chromatographie d'affinité au nickel de SamAA recombinante a été réalisée.

D'après les résultats précédents relatifs à la production de SamAA, la plus forte activité de synthèse de la N-α-lauroyl-L-lysine a été observée à l'issue d'une culture avec *E. coli* BL21 Gold (DE3), le plasmide pET-28a-SamAA 6-His N-ter et le milieu LB supplémenté en cobalt. Plusieurs fractions solubles issues de ce système ont été rassemblées et purifiées par chromatographie d'affinité au nickel, d'une part de façon manuelle (à la seringue), d'autre part sur un équipement automatisé (ÄKTA pure).

Les fractions éluées manuellement sont analysées par SDS-PAGE (Figure V.24). Aucune bande n'est visible dans les fractions éluées avec l'imidazole, indiquant l'absence de SamAA.

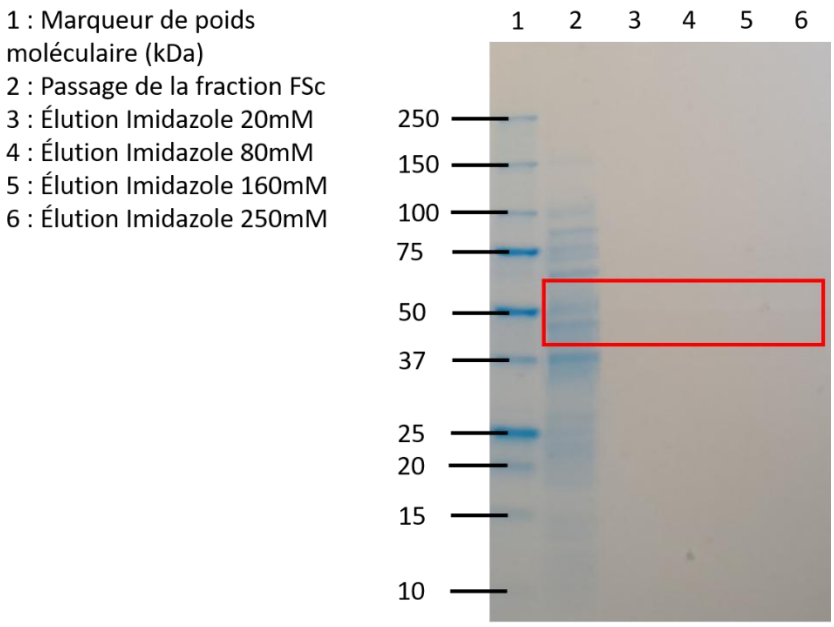


Figure V.24 : Analyse par électrophorèse SDS-PAGE des protéines issues des élutions ultrafiltrées sur colonne IMAC (à la seringue) des fractions solubles issues des cultures avec *E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu LB avec cobalt. FSc : fraction soluble concentrée. Le cadre rouge indique la position de SamAA exprimée.

Les concentrations protéiques dans les fractions éluées sont inférieures au seuil de détection du réactif de Bradford, sauf dans la fraction correspondant au passage de la fraction soluble dont la concentration protéique est de 5 g/L. Un dosage BCA adapté aux petites concentrations protéiques a été réalisé et confirme ce résultat.

L'activité catalytique de synthèse de la N-α-lauroyl-L-lysine est vérifiée dans toutes les fractions éluées. La fraction de passage de la fraction soluble concentrée présente une activité de 12 mg/g*h, tandis que la première fraction éluée à l'imidazole (20 mM) présente une très faible activité (0,19 mg/g*h). Aucune autre fraction ne présente l'activité enzymatique recherchée. Il apparaît donc que SamAA recombinante n'est pas purifiée. Pour confirmer ce résultat, une seconde purification par chromatographie d'affinité au nickel est réalisée en utilisant un système automatique avec détecteur (ÄKTA pure) ; le chromatogramme obtenu est visible en figure V.25. L'élution protéique est suivie à 214 et 280 nm. Lorsque la fraction soluble concentrée est injectée, un pic est visible, signe du passage de toutes les protéines non retenues sur la colonne. Ensuite plus aucun pic n'est observé, ce qui montre qu'il n'y a plus d'élution protéique lorsque la concentration en imidazole augmente. SamAA recombinante ne s'accroche donc pas sur la colonne et demeure en mélange avec les autres protéines.

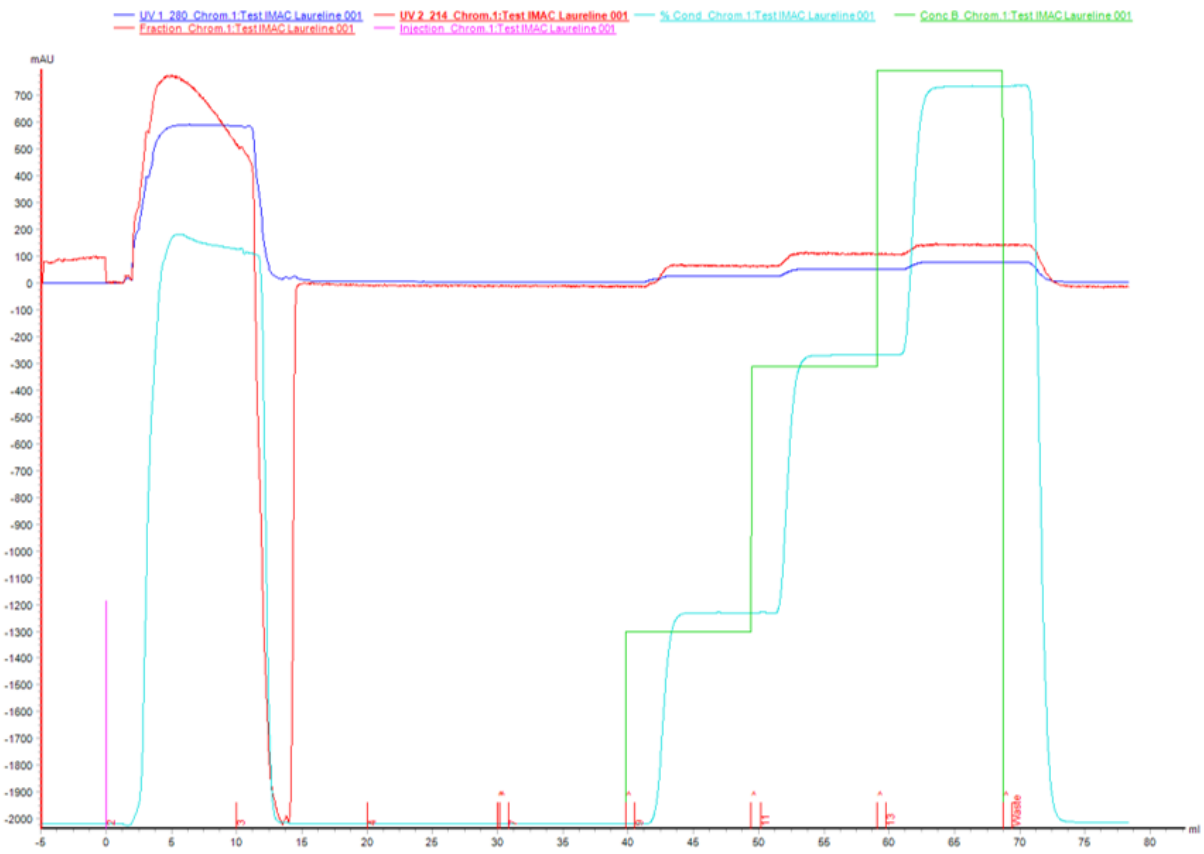


Figure V.25 : Chromatogramme obtenu lors de la purification de SamAA sur ÄKTA Pure. Les courbes rouge et bleue donnent l'absorbance de la phase éluée à 214 et 280 nm, respectivement. La courbe verte suit l'évolution de la concentration en imidazole. Le pic rose correspond au moment de l'injection. Les pics rouges indiquent la fréquence de collecte des fractions.

Comme précédemment, l'activité enzymatique de synthèse de la N-α-lauroyl-L-lysine est détectée uniquement dans la fraction de passage de la fraction soluble (41,35 mg/g*h) et dans une moindre mesure, dans la première fraction d'élution à l'imidazole (10,23 mg/g*h).

Les deux tentatives de purification de SamAA par chromatographie d'affinité au nickel en présence d'un tag 6-His en position N-terminale de l'enzyme ne sont donc pas concluantes.

L'accessibilité et la persistance du tag 6-His en position N-terminale à la surface de la protéine recombinante ont été vérifiées en réalisant un dot blot sur la fraction soluble concentrée et les fractions éluées. Les anticorps utilisés sont dirigés contre le tag 6-His. Aucune tâche n'est visible sur la membrane après l'étape de révélation, indiquant l'absence de protéine taguée dans les fractions déposées. L'absence de témoin positif (protéine avec tag 6-His) rend impossible la conclusion sur la présence ou l'absence du tag 6-His à la surface de SamAA qui pourrait justifier que les tentatives de purification de l'enzyme n'ont pas abouti.

5.4. Production hétérologue de SamELA chez *Escherichia coli*

La production de SamELA a également été abordée dans ces travaux. Pour rappel, l'intérêt de cette enzyme est qu'elle catalyse l'hydrolyse ou la synthèse de dérivés N-ε-acylés de la lysine, selon la même sélectivité que les lipases, mais en milieu aqueux. Les différentes conditions de culture étudiées sont présentées en figure V.26. La stratégie adoptée s'inspire des travaux relatifs à l'expression hétérologue de SamAA. Plusieurs souches commerciales ont été utilisées : *E. coli* Rosetta 2 (DE3), *E. coli* BL21 Gold (DE3) et *E. coli* Arctic Express (DE3) (Merck, USA ; Agilent Technologies, USA). Le plasmide pLysS a été ajouté dans la souche *E. coli* BL21 Gold (DE3). L'effet de la position N-terminale ou C-terminale du tag 6-His a été étudiée. Le gène de SamELA provient du génome de *S. ambofaciens* ; son optimisation est prévue pour une expression chez *E. coli*. La version non-optimisée du gène est portée par le plasmide pET-15b pendant que la version optimisée du gène est contenue dans le plasmide pET-28a. Les deux plasmides contiennent un promoteur inductif à l'IPTG. Les milieux de culture LB et Overnight Express^{TM} Instant TB (Overnight) ont été utilisés, en présence ou non de cobalt (0,25 mM) (Sari et al., 2007 ; Matjan et al., 2011 ; Takakura et al., 2019). Le milieu Overnight est un milieu autoinductif basé sur une utilisation différentielle du lactose. L'utilisation du milieu LB nécessite une induction provoquée par l'ajout d'IPTG (0,1 mM) quand la DO_{600} est à 0,6 en absence de cobalt et à 0,4 en présence de cobalt (Incir et al., 2024). Les cultures ont été arrêtées 4 h ou 20 h après l'induction en milieu LB ou ont duré 16 h en milieu Overnight. Les cultures ont été conduites à 37 °C, sauf en cas d'abaissement de la température à 20 °C après induction ou en cas de choc osmotique et thermique (Papp et al., 2006 ; Rosano et al., 2014 ; Bhatwa et al., 2021). Les cultures bactériennes n'ont pas été répétées. Des témoins négatifs ont été réalisés avec un plasmide contenant le gène de SamELA mais sans réaliser l'induction.

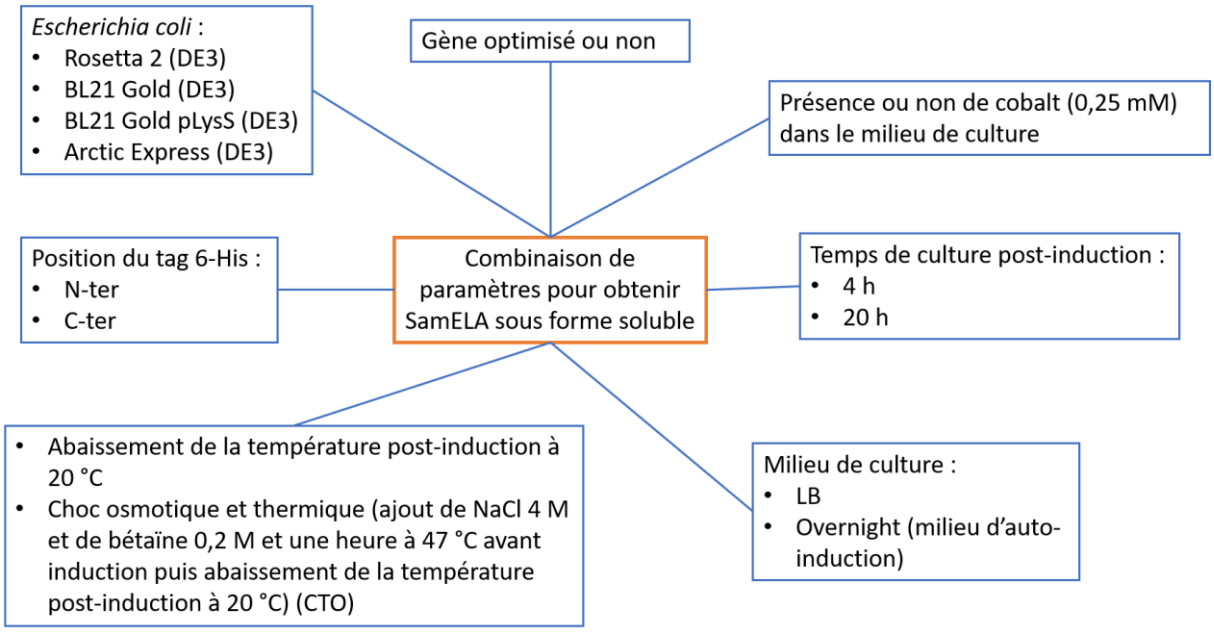


Figure V.26 : Récapitulatif des conditions étudiées pour la surexpression de SamELA chez *E. coli*.

Les résultats de la croissance bactérienne sont similaires à ceux obtenus lors de la production de SamAA recombinante et ne sont pas montrés.

SamELA présente un poids moléculaire de 61 kDa. L'analyse par SDS-PAGE des fractions issues des cultures chez *E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter, et les milieux LB ou OV supplémentés ou non en cobalt sont visibles dans les figures V.27 et V.28. Tous les autres gels SDS-PAGE sont disponibles en annexe 2. SamELA recombinante est produite sous la forme de corps d'inclusion visibles dans les fractions insolubles. Une bande peu intense pouvant correspondre à SamELA recombinante est visible dans certaines fractions solubles concentrées ou non. Une fuite du promoteur inductif est observée en particulier avec les souches *E. coli* Rosetta 2 (DE3) et *E. coli* BL21 Gold (DE3). La position C-terminale du tag 6-His semble favoriser la production de l'enzyme sous sa forme insoluble.

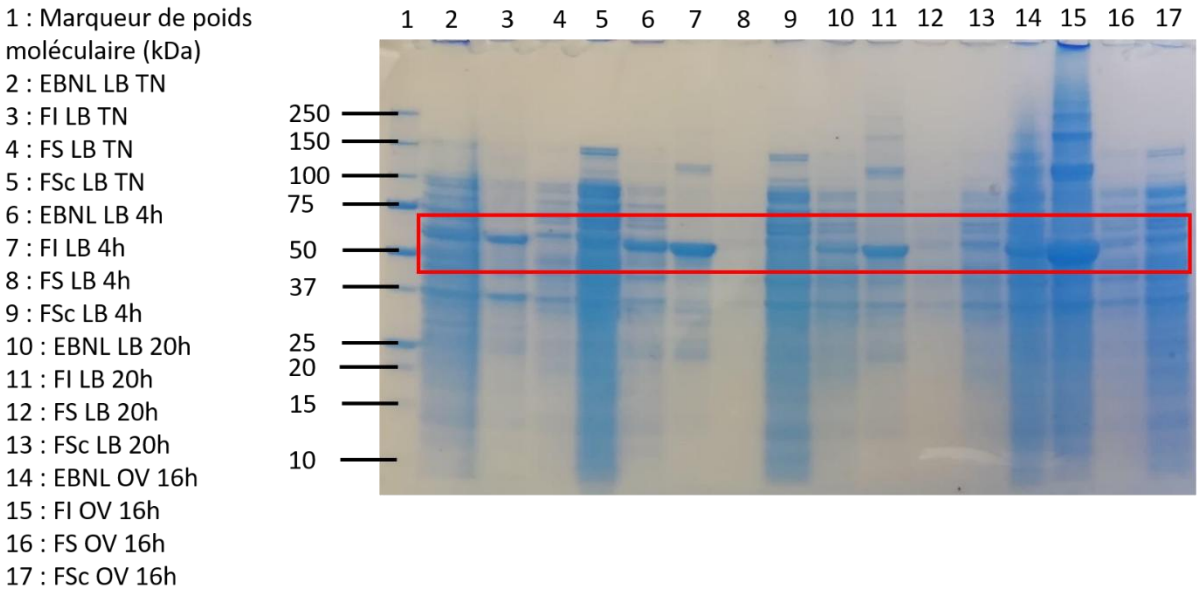


Figure V.27 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter, en milieu LB ou OV sans cobalt. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight. Le cadre rouge indique la position de SamELA exprimée.

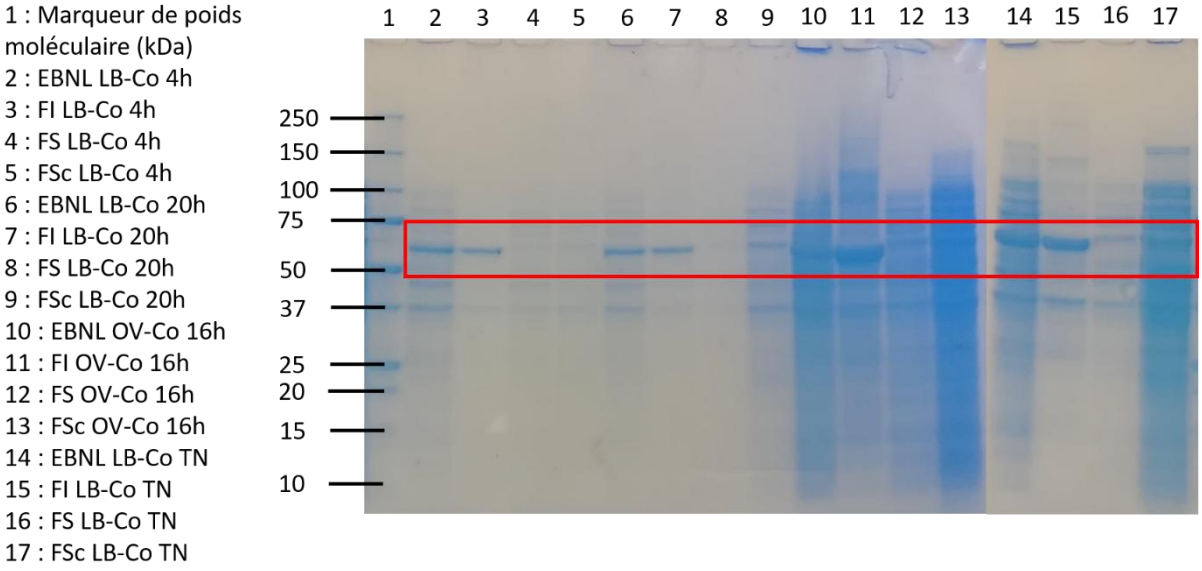


Figure V.28 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter, en milieu LB ou OV supplémenté en cobalt. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt. Le cadre rouge indique la position de SamELA exprimée.

La présence de SamELA soluble et fonctionnelle a été vérifiée dans les différentes fractions en suivant la synthèse de la N-ε-lauroyl-L-lysine et l'hydrolyse de la N-ε-acétyl-L-lysine. Les résultats de la réaction de synthèse ont été analysés par chromatographie sur couche mince

(CCM) et par HPLC. La CCM obtenue en présence des fractions solubles concentrées issues d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter est visible en figure V.29. Les résultats des CCM obtenues à partir des autres fractions solubles concentrées sont visibles en annexe 2. Une très faible production de N-ε-lauroyl-L-lysine est observée, qui n'a pas été confirmée par les analyses HPLC. Indépendamment de la quantité d'enzyme présente dans les fractions, ce résultat n'est pas surprenant compte tenu de la spécificité démontrée de SamELA pour les donneurs d'acyle à courte chaine (inférieure à huit atomes de carbone) (Bourkaib et al., 2020a). Ce résultat ne permet donc pas de tirer des conclusions solides quant à la production de l'enzyme.

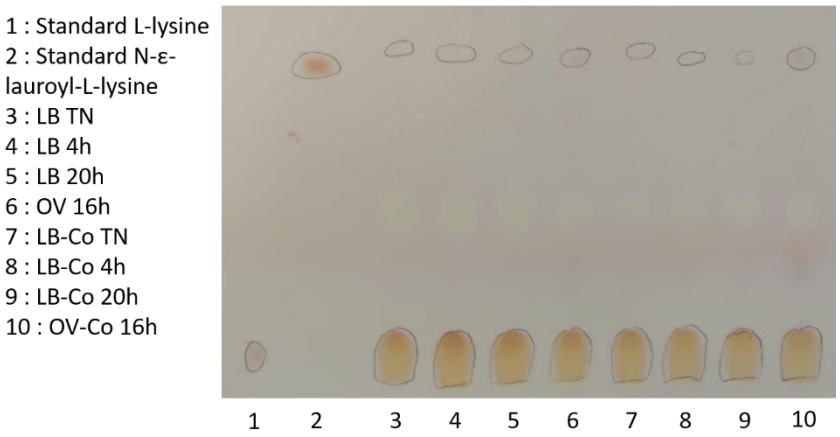


Figure V.29 : Analyse par CCM du milieu de synthèse de la N-ε-lauroyl-L-lysine en présence des fractions solubles concentrées issues d'E. *coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter, en milieu LB ou Overnight, avec ou sans cobalt. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

Les résultats relatifs à l'hydrolyse de la N-ε-acétyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter sont visibles en figure V.30. Les résultats des autres CCM sont très similaires ; ils sont montrés en annexe 2. L'hydrolyse de la N-ε-acétyl-L-lysine paraît particulièrement importante dans les fractions issues du milieu Overnight. La supplémentation en cobalt dans le milieu de culture conduit à l'hydrolyse quasi-complète de la N-ε-acétyl-L-lysine en 24 h. Cette activité est révélatrice de la présence de SamELA sous une forme fonctionnelle au sein des fractions solubles.

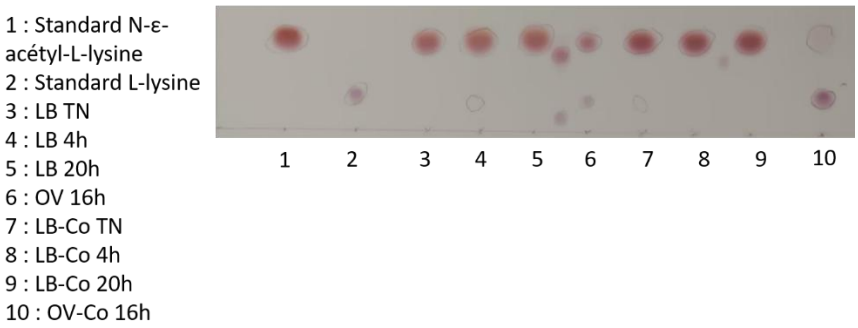


Figure V.30 : Analyse par CCM du milieu d'hydrolyse de la N-ε-acétyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter, en milieu LB ou OV, avec ou sans cobalt. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

5.5. Conclusions sur la production de SamAA et de SamELA

Lors de travaux antérieurs, il a été montré que l'expression de SamAA et de SamELA a lieu très majoritairement sous la forme de corps d'inclusion. Dans ce contexte, l'obtention d'enzymes pures caractérisables sur le plan structural et mécanistique représente un enjeu scientifique majeur. Pour tendre vers cet objectif, de nombreuses campagnes expérimentales de production des deux enzymes ont été menées en utilisant *E. coli* et *S. lividans* comme hôtes d'expression.

Quelles que soient les systèmes d'expression et les conditions utilisés, tous les essais de production de SamAA ont abouti à une production de l'enzyme majoritairement sous la forme de corps d'inclusion. Aucune surexpression de SamAA sous une forme soluble n'est observée. Néanmoins, une activité spécifique apparente de synthèse de la N-α-lauroyl-L-lysine est détectée dans certaines fractions solubles obtenues, indiquant la présence d'une petite quantité de SamAA recombinante sous une forme soluble et fonctionnelle. Deux hôtes d'expression ont été étudiés, l'un du genre *Escherichia*, l'autre du genre *Streptomyces*. La plus forte activité spécifique apparente (339 mg de N-α-lauroyl-L-lysine / g*h) a été obtenue dans la fraction soluble concentrée issue d'une culture avec la souche *E. coli* BL21 Gold (DE3), le plasmide pET-28a 6-His N-ter, en milieu LB supplémenté en cobalt (0,25 mM), à 37 °C pendant 24 h, et avec une induction à l'IPTG (0,1 mM) à 0,4 de DO_{600}. Il a été montré que la présence de cobalt dans le milieu de culture favorise l'activité enzymatique de la fraction soluble, même si la présence de SamAA n'a pas pu être mise en évidence. Toutefois, cet ajout de cobalt peut avoir une incidence négative sur la croissance bactérienne. De plus, un arrêt provisoire de la

croissance au début de la phase exponentielle est apparu lors des essais en présence de cobalt. L'ajout de cobalt au moment de l'induction constitue un compromis intéressant pour obtenir une croissance bactérienne normale, tout en conservant une activité enzymatique élevée dans la fraction soluble. L'expression de SamAA au sein d'un organisme génétiquement proche de la souche produisant l'enzyme naturellement a été étudiée, avec *S. lividans* comme hôte d'expression. Cette bactérie transformée par le vecteur intégratif pIB139-SamAA StrepTag N-ter dans le milieu de fermentation, à 28 °C et pendant trois jours, a permis d'obtenir l'activité spécifique apparente de synthèse de la N-α-lauroyl-L-lysine la plus élevée dans la fraction soluble non diluée (206 mg/g*h).

Un test de refolding a été effectué dans le but de valoriser les corps d'inclusion produits lors de l'expression de SamAA recombinante chez *E. coli*. L'analyse SDS-PAGE indique une importante resolubilisation des corps d'inclusion mais une absence d'activité catalytique dans la fraction soluble refoldée est observée. Ce résultat est malgré tout prometteur et mériterait d'être approfondi pour améliorer le protocole de refolding.

La purification de SamAA tagguée avec un tag 6-His en position N-terminale par chromatographie d'affinité au nickel n'a pas permis d'isoler l'enzyme. Il est nécessaire de poursuivre les travaux pour vérifier l'accessibilité et la présence du tag à la surface de SamAA et d'adapter la technique de purification en conséquence.

Des résultats similaires ont été obtenus au cours de l'expression de SamELA chez *E. coli*. L'enzyme est majoritairement produite sous la forme de corps d'inclusion, quels que soient les systèmes et les conditions utilisés. L'analyse SDS-PAGE des fractions solubles montre la présence d'une bande pouvant correspondre à l'enzyme recombinante. Une activité d'hydrolyse de la N-ε-acétyl-L-lysine a été mise en évidence, démontrant la présence de SamELA sous une forme soluble et fonctionnelle. Les travaux doivent être poursuivis pour identifier précisément le système d'expression et les conditions de culture favorisant la production de l'enzyme dans des quantités permettant d'envisager sa purification et sa caractérisation.

D'une manière générale, malgré tous les systèmes (souche + construction plasmidique avec gène optimisé ou non, et présence et position de deux tags) et toutes les conditions de culture testés, SamAA et SamELA n'ont pas été surexprimés sous forme soluble et active, même si quelques-unes des combinaisons « système-condition de culture » ont permis d'obtenir des activités de synthèse de la N-α-lauroyl-L-lysine ou d'hydrolyse de la N-ε-acétyl-L-lysine dans les fractions solubles. Ces résultats décevants constituent un obstacle à l'atteinte des objectifs qui étaient d'élucider la structure des deux enzymes par cristallographie afin d'établir un parallèle avec les études *in silico* menées par ailleurs durant la thèse.

Conclusion générale et perspectives

Ce travail de thèse s'inscrit dans une thématique de recherche visant le développement d'un procédé enzymatique de N-acylation d'acides aminés, pour la production de dérivés tensioactifs très recherchés dans différents secteurs industriels. Ce procédé à faible impact environnemental constitue une alternative au procédé chimique actuel. Dans ce contexte, l'identification et la caractérisation d'enzymes capables de catalyser ce type de réaction représente un enjeu majeur. Parmi elles figurent les aminoacylases de *Streptomyces ambofaciens* qui ont été découvertes au LRGP il y a une dizaine d'années. Les travaux se sont focalisés sur deux des quatre aminoacylases produites par cette bactérie, dont la sélectivité et la spécificité ont été déterminées expérimentalement. SamAA présente la particularité de catalyser l'acylation sélective de la fonction amine α de la lysine, tandis que SamELA présente une sélectivité comparable à celle des lipases favorisant l'acylation de la fonction amine ε de cet acide aminé. Cette thèse visait d'une part à identifier les éléments responsables de cette sélectivité à l'échelle moléculaire (chapitres III et IV), d'autre part à déterminer les conditions optimales de production et de purification des deux enzymes recombinantes (chapitre V).

En l'absence de données structurales sur ces enzymes, un premier axe d'investigation a consisté à prédire leur structure tridimensionnelle par une approche de modélisation par homologie. Les modèles ainsi construits ont ensuite été utilisés comme cible dans des simulations de docking pour étudier leurs modes d'interaction avec les substrats de la réaction de synthèse ou d'hydrolyse de dérivés N-acylés de la L-lysine, et tenter d'expliquer la sélectivité observée expérimentalement.

Une recherche par homologie de séquence a permis d'identifier plusieurs templates de SamAA et de SamELA. L'étude de la séquence des templates, de leur structure et des données bibliographiques les concernant a permis d'obtenir des informations sur la présence de résidus conservés et sur l'organisation du site actif des deux enzymes (Tableau CCL.1). Chaque chaine peptidique contiendrait un site actif organisé autour de deux ions Zn^{2+} et d'une molécule d'eau catalytique. SamAA appartiendrait à la famille des métalloenzymes M20 tandis que SamELA ferait partie de la famille des métalloenzymes M38. Des hypothèses sur la structure et la fonction des enzymes, en particulier en se basant sur l'identification de résidus conservés, ont été formulées. Toutefois, celles-ci n'ont pas pu être vérifiées expérimentalement.

Tableau CCL.1: Récapitulatif des templates identifiés et des résidus conservés chez SamAA et SamELA.

		SamAA	SamELA
Templates identifiés		aminoacylase 1 humaine (entrée PDB 1Q7L), succinyl-diaminopimelate desuccinylases de *Neisseria meningitidis* MC85 (entrée PDB 4PPZ), de *Haemophilus influenzae* (entrée PDB 5VO3) et de *Shigella flexneri* (entrée PDB 7LGP) et carnosinase 2 humaine (entrée PDB 4RUH)	hydrolases métallodépendantes non-caractérisées de *Pyrococcus furiosus* (entrée PDB 3ICJ) et de Pyrococcus horikoshii (entrée PDB 3IGH), amidohydrolase de *Streptomyces* sp. Tu 6176 (entrée PDB 6SJ0), D-aminoacylase d'*Alcaligenes faecalis* (entrée PDB 1RJQ) et protéine médiatrice de la réponse à la collapsine 2 humaine (entrée PDB 5YZA)
Résidus conservés	Fixateurs des ions métalliques	His88.A, Asp120.A, Glu155.A, Glu182.A et His416.A	His71, His73, Asp265, His316 et His351
	Rôle dans la catalyse	Asp90.A, Glu154.A, Asp317.A, Arg319.A, His219.B et Asn306.B	Asp417
	Boucle de spécificité	Ser382.A, Gly383.A, Gly384.A et Thr385.A	///

Quatre outils de modélisation par homologie (Modeller, Swiss Model, ColabFold et AlphaFold3) ont été étudiés pour prédire la structure tridimensionnelle de SamAA. La qualité des modèles générés a été évaluée en se basant sur une fonction de score (DOPE Score normalisé) et la similarité entre l'architecture du site actif de SamAA et celle du site de la succinyl-diaminopimelate desuccinylase de *H. influenzae*. Les deux meilleurs modèles ont été produits par ColabFold et AlphaFold3 ; ces derniers se distinguent par une conformation plus ou moins ouverte de l'enzyme. SamAA serait un homodimère binucléaire dont le monomère en forme de sifflet présente un domaine catalytique et un domaine de dimérisation. Une approche méthodologique similaire a été appliquée à SamELA, conduisant également à deux modèles très similaires l'un à l'autre. SamELA serait un monomère dont la structure s'organise autour de trois domaines.

Pour expliquer la sélectivité des deux aminoacylases à l'échelle moléculaire, leurs modes d'interaction avec les substrats de la réaction de N-acylation et de N-déacylation de la L-lysine ont été étudiés grâce à des simulations de docking moléculaire.

Ainsi, la réaction d'acylation de la L-lysine par l'acide caprylique ou l'acide laurique a été modélisée en utilisant la structure de SamELA prédite par AlphaFold3. Malheureusement, les résultats issus des modèles ne corroborent pas les données expérimentales précédemment obtenues ; ils ne permettent donc pas d'expliquer la sélectivité de l'enzyme envers la fonction amine ε de la L-lysine ou sa spécificité vis-à-vis des donneurs d'acyle présentant moins de huit atomes de carbone. Ce manque de corrélation entre les modèles et les observations expérimentales a été attribué à des imprécisions dans la structure prédite de l'enzyme, notamment au niveau d'une boucle de 33 résidus située à proximité du site de fixation des ions métalliques.

Dans le cas de SamAA, la synthèse et l'hydrolyse des dérivés N-α/ε-acétyl-L-lysine, N-α/ε-caproyl-L-lysine et N-α/ε-lauroyl-L-lysine a été modélisée. Les résultats des simulations menées sur la structure fermée de SamAA prédite par ColabFold sont cohérents avec les résultats expérimentaux, dans le cas de l'acide acétique comme donneur d'acyle. Ils suggèrent fortement l'acylation et la déacylation sélective de la fonction amine α de la L-lysine tandis que la prise en charge de la fonction amine ε de l'acide aminé semble improbable. Un mécanisme de synthèse et d'hydrolyse de la N-α-caproyl-L-lysine a été proposé sur la base des interactions enzyme-substrats identifiées et des informations disponibles dans la littérature sur la succinyl-diaminopimelate desuccinylase de *H. influenzae*. En présence d'un donneur d'acyle plus encombrant comme l'acide laurique, le placement des substrats de la réaction dans la cavité catalytique de l'enzyme est incorrect, suggérant un manque d'espace dans la structure modélisée. De nouvelles simulations menées sur une conformation plus ouverte de l'enzyme générée par Alphafold3 ont conduit à un placement des substrats favorable à la synthèse ou l'hydrolyse de la N-α-lauroyl-lysine, ce qui corroborent les résultats expérimentaux. Ces observations renforcent l'hypothèse de l'existence d'une conformation ouverte et d'une conformation fermée de SamAA. Comme suggéré dans la littérature, le passage d'une conformation à l'autre interviendrait lors de l'arrivée des substrats dans le site actif de l'enzyme.

Le second axe de recherche portait sur la production et la purification des deux aminoacylases recombinantes de *S. ambofaciens*, SamAA et SamELA. L'objectif était d'identifier un système d'expression approprié limitant la formation des corps d'inclusion et permettant l'obtention d'une quantité suffisante d'enzymes solubles à purifier et à caractériser sur le plan structural et

catalytique. Toutes ces tentatives de production hétérologue des deux aminoacylases se sont conclues par une expression majoritaire sous forme de corps d'inclusion.

Dans le cas de SamAA, une activité spécifique apparente de synthèse de la N-α-lauroyl-L-lysine a été détectée dans certaines fractions solubles. Cette activité indique la présence d'une faible quantité d'enzyme, indétectable sur gel SDS-PAGE. L'activité la plus forte mesurée (339 mg/g*h) a été obtenue dans une fraction soluble concentrée dix fois, issue d'une culture avec *E. coli*. Les conditions de cette culture étaient les suivantes : souche *E. coli* BL21 Gold (DE3) avec le plasmide pET-28a 6-His N-ter, en milieu LB supplémenté en cobalt (0,25 mM), à 37 °C, pendant 24 h, et avec une induction par l'IPTG (0,1 mM) à 0,4 de DO_{600}. Par ailleurs, des résultats prometteurs ont été obtenu en utilisant *S. lividans* transformée par le vecteur intégratif pIB139-SamAA StrepTag N-ter et cultivée dans le milieu de fermentation à 28 °C, pendant trois jours. Dans ces conditions, l'activité spécifique apparente est de 206 mg/g*h, dans la fraction soluble non concentrée.

Une stratégie de refolding des corps d'inclusion de SamAA a été amorcée. Toutefois, si l'analyse SDS-PAGE des fractions refoldées montre une resolubilisation importante des corps d'inclusion, ceux-ci sont dépourvus d'activité enzymatique.

Enfin, les travaux de purification de SamAA par chromatographie d'affinité au nickel n'ont pas permis jusqu'à présent d'isoler l'enzyme.

Concernant l'expression de SamELA chez *E. coli*, les résultats obtenus montrent qu'elle est également produite majoritairement sous la forme de corps d'inclusion. Toutefois, l'analyse de certaines fractions solubles et les tests enzymatiques associés ont mis en évidence la présence de l'enzyme sous une forme soluble et fonctionnelle.

Sur la base des résultats obtenus au cours de cette étude, différents axes de travail qu'il serait pertinent d'approfondir ou d'explorer ont été identifiés. Ils concernent principalement la vérification et la validation des hypothèses formulées lors de l'étude structurale numérique des deux aminoacylases, et la poursuite des travaux visant l'identification du système d'expression et des conditions de culture favorisant la production des deux enzymes dans des quantités permettant d'envisager leur purification et leur caractérisation structurale.

Des hypothèses ont été émises au sujet de plusieurs résidus conservés présents chez les deux enzymes et jouant potentiellement un rôle dans la catalyse ou dans la fixation des ions métalliques. Pour vérifier le rôle de chacun des résidus, des expériences de mutagénèse dirigée pourraient être entreprises, ainsi que cela a été fait pour certains templates. En effet, la

substitution d'un résidu par un autre est susceptible d'induire des changements dans l'activité enzymatique, ce qui permet de conclure sur le rôle précis du résidu muté.

Le travail numérique présenté ici est basé exclusivement sur des modèles. Même si les outils de prédiction de structure sont devenus performants, les modèles ne corroborent pas toujours les résultats expérimentaux, ce qui indiquent des imprécisions, voire des erreurs, au sein des modèles. Dans tous les cas, ces modèles doivent être confrontés à des données expérimentales pour validation. Notamment, l'élucidation de la structure tridimensionnelle de SamAA et de SamELA par cristallographie, est nécessaire pour pouvoir les comparer avec les structures prédites *in silico*. De plus, quand il sera possible de surproduire les deux enzymes, leur structure quaternaire pourra être confirmée en réalisant des analyses PAGE non dénaturantes. Combinées aux données cristallographiques, celles-ci permettraient de confirmer la nature monomérique de SamELA et la nature homodimérique de SamAA ainsi que la localisation précise de son domaine de dimérisation.

L'étude des modes d'interaction entre SamAA et ses substrats a été menée avec des donneurs d'acyle présentant différentes longueurs de chaine. Si la N-acylation de la L-lysine avec un donneur d'acyle à deux et à douze atomes de carbone catalysée par SamAA a déjà été démontrée expérimentalement ; la réaction avec l'acide caproïque n'a jamais été réalisée. Cela mériterait d'être fait pour valider les résultats obtenus lors des simulations de docking.

Concernant ces simulations, plusieurs pistes d'amélioration ont été identifiées pour affiner les résultats. Ainsi un protocole de docking flexible dans lequel la flexibilité de la protéine réceptrice est prise en compte partiellement, au niveau de résidus choisis, est susceptible d'améliorer la robustesse des modèles d'interaction enzyme/substrats. Une autre méthode pourrait être la dynamique moléculaire. Cette dernière peut être mise en œuvre en amont du docking, pour étudier le comportement dynamique de l'enzyme dans son environnement, et/ou après le docking, afin d'examiner le comportement dynamique des complexes enzyme-substrats. La dynamique moléculaire, bien que coûteuse en temps de calcul, contribuerait à l'amélioration des modèles en prenant en compte la flexibilité des enzymes et leur environnement. De plus, dans le cas particulier de SamAA, l'hypothèse du passage d'une conformation ouverte à une conformation fermée pourrait être vérifiée en réalisant une dynamique moléculaire de longue durée, comme cela a été possible pour l'amidohydrolase BrLL2 de *Brassica rapa* (Simunovic et al., 2010).

Enfin, des calculs d'énergie d'interaction électrostatiques pourraient être réalisés sur les complexes enzyme-substrats obtenus à l'issue du docking. Ceux-ci conduiraient à

l'identification des interactions clés expliquant la sélectivité de SamAA et de SamELA et permettraient de confirmer les hypothèses formulées dans le cadre de ce travail.

Concernant le volet expérimental de cette thèse, tous les essais de production de SamAA et de SamELA ont conduit à une production des enzymes majoritairement sous la forme des corps d'inclusion.

Les tentatives de production de SamAA ont débuté avec la souche *E. coli*. Des conditions conduisant à une faible production de l'enzyme soluble et active ont été identifiées. Cependant, ce résultat n'est plus reproductible suite à une interruption provisoire dans la croissance des bactéries en phase exponentielle. L'origine de ce phénomène demeure inconnue et mériterait d'être expliquée et résolue.

Les essais d'expression de SamAA chez *S. lividans* menés jusqu'ici en erlenmeyers semblent prometteurs, mais les conditions opératoires nécessitent d'être optimisées. La mise en place d'un suivi de la croissance bactérienne et l'ajout de cobalt à une concentration consensus entre croissance et production sont deux éléments nécessaires à l'optimisation de la production de SamAA avec ce système. La mise en œuvre d'un bioréacteur en conditions contrôlées pourrait être l'étape suivante de la mise en place d'un procédé de production à plus grande échelle. Dans les travaux portant sur la surexpression de l'aminoacylase SgAA, il a été montré que la forte agitation améliorant l'oxygénation du milieu, couplée à une alimentation partiellement effectuée en mode fed-batch, s'est révélée beaucoup plus efficace que la culture en erlemeyer (Haeger et al., 2023a). Un premier essai de culture en bioréacteur avec le système *S. lividans* et le plasmide pIJ86-SamAA a été réalisé au cours de cette thèse. Cependant différents problèmes techniques ont été rencontrés si bien que les résultats obtenus n'ont pas pu être exploités. Toutefois, une culture en bioréacteur, en conditions contrôlées, constitue une piste intéressante pour augmenter la production de SamAA sous une forme soluble et fonctionnelle.

De la même manière que des mutants avaient été créés pour caractériser les propriétés de la sélectivité de SamAA et de SamELA, un triple mutant de *Streptomyces ambofaciens* ne contenant que le gène de SamAA a été constitué dans le cadre de ce projet. Celui-ci pourrait être utilisé comme hôte d'expression dans les conditions classiques et connues pour cette souche. Dans ce cas, il serait également possible d'ajouter un promoteur fort en amont du gène de SamAA afin de stimuler son expression, et ainsi augmenter la production de l'enzyme.

Certaines fractions obtenues lors de la production de SamELA chez *E. coli* ont présenté une activité hydrolytique de la N-ε-acétyl-L-lysine, visible par analyse CCM. L'activité enzymatique pour la réaction inverse de synthèse de dérivés acylés de la L-lysine n'a pas été vérifiée. Ceci pourrait être entrepris notamment avec des donneurs d'acyle comportant moins

de huit atomes de carbone afin de respecter la spécificité de l'enzyme démontrée antérieurement. Ce résultat quantitatif permettrait d'identifier le système d'expression conduisant à la forte activité enzymatique. Si malgré cela, la production de SamELA recombinante reste faible, il serait alors envisageable de poursuivre les essais chez *S. lividans*.

Le test de refolding des corps d'inclusion de SamAA a montré une solubilisation importante de l'enzyme ; toutefois celle-ci ne présente aucune activité. Le processus de repliement doit être retravaillé de façon à valoriser ces corps d'inclusion et à produire l'aminoacylase sous une forme partiellement purifiée. Cette technique pourrait être également appliquée aux corps d'inclusion de SamELA qui sont eux aussi produits de façon majoritaire.

Pour faire face aux difficultés rencontrées pour surexprimer les deux aminoacylases, la solution consistant à faire appel à un partenaire public ou privé spécialisé dans la production des protéines récalcitrantes reste envisageable. Un screening de conditions de surexpression mené sur une plateforme adaptée permettrait certainement d'accélérer le travail et l'accès aux enzymes pures.

Enfin, dans le prolongement de ce qui a été entrepris dans le cadre de ce travail, les essais de purification des enzymes recombinantes doivent être poursuivis pour SamAA et initiés pour SamELA. L'accès au tag utilisé pour faciliter la purification devra notamment être vérifié. Il n'est pas possible d'effectuer cette vérification à l'aide d'outils numériques, car la modélisation des extrémités N- et C-terminales des protéines manque de précision. Une approche expérimentale possible pourrait être de réaliser un Dot Blot ou un Western Blot, incluant un témoin positif.

Références bibliographiques

Abe A, Asakura K and Osanai S (2004) « Synthesis and characterization of novel amphiphiles containing amino acid and carbohydrate. » Journal of surfactants and detergents 7 (3) : 297-303. https://doi.org/10.1007/s11743-004-0315-9.

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, Bodenstein SW, Evans DA, Hung C-C, O'Neill M, Reiman D, Tunyasuvunakool K, Wu Z, Zemgulyte A, Arvaniti E, Beattie C, Bertolli O, Bridgland A, Cherepanov A, Congreve M, Cowen-Rivers AI, Cowie A, Figurnov M, Fuchs FB, Gladman H, Jain R, Khan YA, Low CMR, Perlin K, Potapenko A, Savy P, Singh S, Stecula A, Thillaisundaram A, Tong C, Yakneen S, Zhong ED, Zielinski M, Zidek A, Bapst V, Kohli P, Jaderberg M, Hassabis D and Jumper JM (2024) « Accurate structure prediction of biomolecular interactions with AlphaFold3. » Nature 630 : 493-500. https://doi.org/10.1038/s41586-024-07487-w.

Agarwal S and Mehrotra R (2016) « An overview of molecular docking. » JSM Chemistry 4 (2) : 1024. https://doi.org/10.47739/2334-1831/1024

Al-Hawash AB, Zhang X and Ma F (2017) « Strategies of codon optimization for high-level heterologous protein expression in microbial expression systems. » Gene Reports 9 : 46-53. https://doi.org/10.1016/j.genrep.2017.08.006.

Alcantara AR, de Maria PD, Littlechild JA, Schürmann M, Sheldon RA and Wohlgemuth R (2022) « Biocatalysis as key to sustainable industrial chemistry. » ChemSusChem 15 (9) : e202102709. https://doi.org/10.1002/cssc.202102709.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) « Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. » Nucleic Acids Research 25 (17) : 3389-3402. https://doi.org/10.1093/nar/25.17.3389.

Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schäffer AA and Yu Y-K (2005) « Protein database searches using compositionally adjusted substitution matrices. » The FEBS Journal 272 (20) : 5101-5109. https://doi.org/10.1111/j.1742-4658.2005.04945.x.

Alves JS, Garcia-Galan C, Schein MF, Silva AM, Barbosa O, Ayub MAZ, Fernandez-Lafuente R and Rodrigues RC (2014) « Combined effects of ultrasound and immobilization protocol on butyl acetate synthesis catalyzed by CALB. » Molecules 19 : 9562-9576. https://doi.org/10.3390/molecules19079562.

Andreev YA, Kozlov SA Vassilevski AA and Grishin EV (2010) « Cyanogen bromide cleavage of proteins in salt and buffer solutions. » Analytical Biochemistry 407 (1) : 144-146. https://doi.org/10.1016/j.ab.2010.07.023.

Arima J, Isoda Y, Hatanaka T and Mori N (2013) « Recombinant production and characterization of an N-acyl-D-amino acid amidohydrolase from *Streptomyces* sp. 64E6. » World Journal of Microbiology Biotechnology 29 : 899-906. https://doi.org/10.1007/s11274-012-1245-5.

Avinash VS, Pundle AV, Ramasamy S and Suresh CG (2014) « Penicillin acylases revisited: importance beyond their industrial utility. » Critical Reviews in Biotechnology 36 (2) : 303-316. https://doi.org/10.3109/07388551.2014.960359.

Badger J, Sauder JM, Adams JM, Antonysamy S, Bain K, Bergseid MG, Buchnan SG, Buchanan MD, Batiyenko Y, Christopher JA, Emtage S, Eroshkina A, Feil I, Furlong EB, Gajiwala KS, Gao X, He D, Hendle J, Huber A, Hoda K, Kearins P, Kissinger C, Laubert B, Lewis HA, Lin J, Loomis K, Lorimer D, Louie G, Maletic M, Marsh CD, Miller I, Molinari J, Muller-Dieckmann HJ, Newman JM, Noland BW, Pagarigan B, Park F, Peat TS, Post KW, Radojiciv S, Ramos A, Romero R, Rutter ME, Sanderson WE, Schwinn KD, Tresser J, Winhoven J, Wright TA, Xu J and Harris TJR (2005) « Structural analysis of a set of proteins resulting from a bacterial genomics project. » Proteins: Structure, Functions, and Bioinformatics 60 (4) : 787-796. https://doi.org/10.1002/prot.20541.

Benkert P, Biasini M and Schwede T (2011) « Toward the estimation of the absolute quality of individual protein structure models. » Bioinformatics 27 (3) : 343-350. https://doi.org/10.1093/bioinformatics/btq662.

Berini F, Marinelli F and Binda E (2020) « Streptomycetes: attractive hosts for recombinant protein production. » Frontiers in Microbiology 11 : 1958. https://doi.org/10.3389/fmicb.2020.01958.

Berkmen M (2012) « Production of disulfide-bonded proteins in *Escherichia coli*. » Protein Expression and Purification 82 (1) : 240-251. https://doi.org/10.1016/j.pep.2011.10.009.

Bertoni M, Kiefer F, Biasini M, Bordoli L and Schwede T (2017) « Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. » Scientific Reports 7 : 10480. https://doi.org/10.1038/s41598-017-09654-8.

Bessette PH, Aslund F, Beckwith J and Georgiou G (1999) « Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. » Proceedings of the National Academy of Sciences 96 (24) : 13703-13708. https://doi.org/10.1073/pnas.96.24.13703.

Bhatwa A, Wang W, Hassan YI, Abraham N, Li X-Z and Zhou T (2021) « Challenges associated with the formation of recombinant protein inclusion bodies in Escherichia coli and strategies to address them for industrial applications. » Frontiers in Bioengineering and Biotechnology 9 : 630551. https://doi.org/10.3389/fbioe.2021.630551.

Bibb MJ, Janssen GR and Ward JM (1985) « Cloning and analysis of the promoter region of the erythromycin resistance gene (*ermE*) of *Streptomyces erythraeus*. » Gene 38 (1-3) : 215-226. https://doi.org/10.1016/0378-1119(85)90220-3.

Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L and Schwede T (2017) « The SWISS-MODEL Repository – new features and functionality. » Nucleic Acids Research 45 (D1) : D313-D319. https://doi.org/10.1093/nar/gkw1132.

Bienvenue DL, Gilner DM, Davis RS, Bennett B and Holz RC (2003) « Substrate specificity, metal binding properties, and spectroscopic characterization of the DapE-encoded N-succinyl-L,L-diaminopimelic acid desuccinylase from *Haemophilus influenzae*. » Biochemistry 42 (36) : 10756-10763. https://doi.org/10.1021/bi034845+.

Binder D, Grünberger A, Loeschcke A, Probst C, Bier C, Pietruszka J, Wiechert W, Kohlheyer D, Jaeger K-E and Drepper T (2014) « Light-responsive control of bacterial gene expression: precise triggering of the *lac* promoter activity using photocaged IPTG. » Integrative Biology 6 (8) : 755-765. https://doi.org/10.1039/c4ib00027g.

Binder D, Probst C, Grünberger A, Hilgers F, Loeschcke A, Jaeger K-E, Kohlheyer D and Drepper T (2016) « Comparative single-cell analysis of different *E. coli* expression systems during microfluidic cultivation. » Plos One 11 (8) : e0160711. https://doi.org/10.1371/journal.pone.0160711.

Bitto E, Bingman CA, Wesenberg GE and Phillips Jr GN (2007) « Structure of aspartoacylase, the brain enzyme impaired in Canavan disease. » Proceedings of the National Academy of Sciences 104 (2) : 456-461. https://doi.org/10.1073/pnas.0607817104.

Bitto E, Bingman CA, Bittova L, Houston HL, Boston RS, Fox BG and Phillips Jr GN (2008) « X-ray structure of ILL2, an auxin-conjugate amidohydrolase from *Arabidopsis thaliana*. » Proteins: Structure, Functions, and Bioinformatics 74 (1) : 61-71. https://doi.org/10.1002/prot.22124.

Blackwell JR and Horgan R (1991) « A novel strategy for production of a highly expressed recombinant protein in an active form. » FEBS 295 (1-3) : 10-12. https://doi.org/10.1016/0014-5793(91)81372-F.

Blommel PG and Fox BG (2007) « A combined approach to improving large-scale production of tobacco etch virus protease. » Protein Expression and Purification 55 (1) : 53-68. https://doi.org/10.1016/j.pep.2007.04.013.

Bonanno JB, Patskovsky Y, Freeman J, Bain KT, Hu S, Ozyurt S, Wasserman S, Sauder JM, Raushel FM, Burley SK and Almo SC (2009) « Crystal structure of an uncharacterized metal-dependent hydrolase from *Pyrococcus furiosus*. » Full wwPDB X-ray Structure Validation Report

Bordes R, Tropsch J and Holmberg K (2009) « Counterion specificity of surfactants based on dicarboxylic amino acids. » Journal of Colloid and Interface Science 338 : 529-536. https://doi:10.1016/j.jcis.2009.06.032.

Bordes R and Holmberg K (2015) « Amino acid-based surfactants – do they deserve more attention ? » Advances in Colloid and Interface Science 222 : 79-91. http://dx.doi.org/10.1016/j.cis.2014.10.013.

Born TL, Zheng R and Blanchard JS (1998) « Hydrolysis of N-succinyl-L,L-diaminopimeic acid by the *Haemophilus influenzae* dapE-encoded desuccinylase: metal activation, solvent isotope effects, and kinetic mechanism. » Biochemistry 37 (29) : 10478-10487. https://doi.org/10.1021/bi9806807.

Botelho TO, Guevara T, Marrero A, Arêde P, Fluxa VS, Reymond J-L, Oliveira DC and Gomis-Rüth FX (2011) « Structural and functional analyses reveal that *Staphylococcus aureus* antibiotic resistance factor HmrA is a zinc-dependant endopeptidase. » Journal of Biological Chemistry 286 (29) : 25697-25709. https://doi.org/10.1074/jbc.M111.247437.

Bourgeois G (2011) « Etude des propriétés enzymatiques et du rôle dans la biogénèse des ribosomes de la protéine humaine p120, marqueur de l'évolution tumorale, et de Nop2p son

homologue chez *Saccharomyces cerevisiae*. » Université de Lorraine. https://hal.univ-lorraine.fr/tel-01746261.

Bourkaib MC (2020) « Prodédé enzymatique intensifié et durable de N-acylation d'acides aminés en milieux écocompatibles : caractérisation, cinétiques et immobilisation de nouveaux biocatalyseurs. » Université de Lorraine. https://theses.fr/2020LORR0114.

Bourkaib MC, Delaunay S, Framboisier X, Hotel L, Aigle B, Humeau C, Guiavarc'h Y and Chevalot I (2020a) « N-acylation of L-amino acids in aqueous media: evaluation of the cataltyic performances of *Streptomyces ambofaciens* aminoacylases. » Enzyme and Microbial Technology 137 : 109536. https://doi.org/10.1016/j.enzmictec.2020.109536.

Bourkaib MC, Delaunay S, Framboisier X, Humeau C, Guilbot J, Bize C, Illous E, Chevalot I and Guiavarc'h Y (2020b) « Enzymatic synthesis of N-10-undecenoyl-phenylalanine catalysed by aminoacylases from *Streptomyces ambofaciens*. » Process Biochemistry 99 : 307-315. https://doi.org/10.1016/j.procbio.2020.09.009.

Bourkaib MC, Gaudin P, Vibert F, Guiavarc'h Y, Delaunay S, Framboisier X, Humeau C, Chevalot I and Blin J-C (2021) « APTES modified SBA15 and meso-macro silica materials for the immobilization of aminoacylases from *Streptomyces ambofaciens*. » Microporous and Mesoporous Materials 323 : 111226. https://doi.org/10.1016/j.micromeso.2021.111226.

Bowie JU, Lüthy R and Eisenberg D (1991) « A method to identify protein sequence that fold into a known three-dimensional structure. » Science 253 (5016) : 164-170. https://doi.org/10.1126/science.1853201.

Bradford MM (1976) « A rapd and sensitive method for the quantitification of microgram quantities of protein utilizing the principle of protein-dye binding. » Analytical Biochemistry 72 (1-2) : 248-254. https://doi.org/10.1016/0003-2697(76)90527-3.

Branneby C, Carlqvist P, Magnusson A, Hult K, Brinck T and Berglund P (2002) « Carbon-carbon bonds by hydrolytic enzymes. » Journal of the American Chemical Society 125 (4) : 874-875. https://doi.org/10.1021/ja028056b.

Brinkmann U, Mattes RE and Buckel P (1989) « High-level expression of recombinant genes in *Escherichia coli* is dependant on the availability of the *dnaY* gene product. » Gene 85 (1) : 109-114. https://doi.org/10.1016/0378-1119(89)90470-8.

Brito RO, Marques EF, Silva SG, do Vale ML, Gomes P, Araújo MJ, Rodriguez-Borges JE, Infante MR, Garcia MT, Ribosa I, Vinardell MP and Mitjans M (2009) « Physicochemical and toxicological properties of the novel amino acid-based amphiphiles and their spontaneously formed catanionic vesicles. » Colloids and SurfacesB: Biointerfaces 72 : 80-87. https://doi:10.1016/j.colsurfb.2009.03.017.

Brito RO, Silva SG, Fernandes RMF, Marques EF, Enrique-Borges J and do Vale MLC (2011) « Enhanced interfacial properties of novel amino acid-derived surfactants: Effects of headgroup chemistry and of alkyl chain length and unsaturation. » Colloids and Surfaces B: Biointerfaces 86 : 65-70. https://doi:10.1016/j.colsurfb.2011.03.017.

Brooks BR, Brooks CL, Mackerell Jr AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post

CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM and Karplus M (2009) « CHARMM: The biomolecular simulation program. » Journal of Computational Chemistry 30 (10) : 1545-1614. https://doi.org/10.1002/jcc.21287.

Bürgi HB, Dunitz JD, Lehn JM and Wipff G (1974) « Stereochemistry of reaction paths at carbonyl centres. » Tetrahedron 30 (12) : 1563-1572. https://doi.org/10.1016/S0040-4020(01)90678-7.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K and Madden TL (2009) « BLAST+: architecture and applications. » BMC Bioinformatics 10 : 421. https://doi.org/10.1186/1471-2105-10-421.

Cao C-H, Cheng F, Xue Y-P and Zheng Y-Guo (2020) « Efficient synthesis of L-phosphinothricin using a novel aminoacylase mined from *Stenotrophomonas maltophilia*. » Enzyme and Microbial Technology 135 : 109493. https://doi.org/10.1016/j.enzmictec.2019.109493.

Case DA, Cheatham III TE, Darden T, Gohlke H, Luo R, Merz Jr KM, Onufriev A, Simmerling C, Wang B and Woods RJ (2005) « The Amber biomolecular simulation programs. » Journal of Computational Chemistry 26 (16) : 1668-1688. https://doi.org/10.1002/jcc.20290.

Castillo JA, Infante MR, Manresa À, Vinardell MP, Mitjans M and Clapés P (2006) « Chemoenzymatic synthesis and antimicrobial and haemolytic activities of amphiphilic bis(phenylacetylarginine) derivatives. » ChemMedChem 1 : 1091-1098. https://doi:10.1002/cmdc.200600148.

Cavasotto CN and Phatak SS (2009) « Homology modeling in drug discovery: current trends and applications. » Drug Discovery Today 14 (13-14) : 676-683. https://doi.org/10.1016/j.drudis.2009.04.006.

Chandra N and Tyagi VK (2013) « Synthesis, properties, and applications of amino acids based surfactants: a review. » Journal of Dispersion Science and Technology 34 : 800-808. htpps://doi:10.1080/01932691.2012.695967.

Chang ACY and Cohen SN (1978) « Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. » Journal of Biotechnology 134 (3) : 1141-1156. https://doi.org/10.1128/jb.134.3.1141-1156.1978.

Chang C-Y, Hsieh Y-C, Wang T-Y, Chen Y-C, Wang Y-K, Chiang T-W, Chen Y-J, Chang C-H, Chen C-J and Wu T-K (2010) Crystal structure and mutational analysis of aminoacylhistidine dipeptidase from *Vibrio alginolyticus* reveal a new architecture of M20 metallopeptidases. » Journal of Biological Chemistry 285 (50) : 39500-39510. https://doi.org/10.1074/jbc.M110.139683.

Chatzileontiadou DSM, Szeto C, Jayasinghe D and Gras S (2021) « Protein purification and crystallization of HLA-A*02:01 in complex with SARS-CoV-2 peptides. » STAR Protocols 2 (3) : 100635. https://doi.org/10.1016/j.xpro.2021.100635.

Chen Z, Gao J, Sun J, Wu Z and Wang B (2021) « Aminoacylase 1 (ACY-1) mediates the proliferation and migration of neuroblastoma cells in humans through the ERK/transforming growth factor β (TGF-β) signaling pathways. » Medical Science Monitor 27 : e928813. https://doi.org/10.12659/MSM.928813.

Chen Z, Sun T-Y and Wu Y-D (2023) « New insights into the cooperativity and dynamics of dimeric enzymes. » Chemical Reviews 123 (16) : 9940-9981. https://doi.org/10.1021/acs.chemrev.3c00042.

Chen Z, Zhang P, Liu Y, Mekkaoui A, Wang C, Zhao L and Xu B (2020) « Interfacial rheological behavior of N-acyl amino acid surfactants derived from vegetable oils. » Colloids and Surface A 592 : 124600. https://doi.org/10.1016/j.colsurfa.2020.124600.

Cheng T, Li Q, Zhou Z, Wang Y and Bryant SH (2012) « Structure-based virtual screening for drug discovery: a problem-centric review. » The American Association of Pharmaceutical Scientists Journal 14 : 133-141. https://doi.org/10.1208/s12248-012-9322-0.

Chiu YC and Hwang HJ (1994) « Rapid dissolution of vitamin E by using sodium N-lauroylsarcosinate ionic surfactant. » Colloids and Surfaces A: Physicochemical and Engineering Aspects 90 : 155-165. https://doi.org/10.1016/0927-7757(94)02921-0.

Christodoulou E and Vorgias CE (2002) « Understanding heterologous protein overproduction under the T7 promoteur: A practical exercice. » Biochemistry and Molecular Biology Education 30 (3) : 189-191. https://doi.org/10.1002/bmb.2002.494030030066.

Clapés P, Morán C and Infante MR (1998) « Enzymatic synthesis of arginine-based cationic surfactants. » Biotechnology and Bioengineering 63 (3) : 333-343. https://doi.org/10.1002/(SICI)1097-0290(19990505)63:3<333::AID-BIT10>3.0.CO;2-G.

Clark DS (2004) « Characteristics of nearly dry enzymes in organic solvents: implications for biocatalysis in the absence of water. » Philosophical Transactions of the Royal Society B 359 : 1299-1307. https://doi.org/10.1098/rstb.2004.1506.

Cole JC, Murray CW, Nissink JWM, Taylor RD and Taylor R (2005) « Comparing protein-ligand docking programs is difficult. » Proteins : Structure, Function, and Bioinformatics 60 (3) : 325-332. https://doi.org/10.1002/prot.20497.

Contente ML, Pinto A, Molinari F and Paradisi F (2018) « Biocatalytic N-acylation of amines in water using an acyltransferase from *Mycobacterium smegmatis*. » Advanced Synthesis & Catalysis 360 (24) : 4814-4819. https://doi.org/10.1002/adsc.201801061.

Costa S, Almeida A, Castro A and Domingues L (2014) « Fusion tags for protein solubility, purification, and immunogenicity in *Escherichia coli*: the novel Fh8 system. » Frontiers in Microbiology 5 : 63. https://doi.org/10.3389/fmicb.2014.00063.

Cummings JA, Nguyen TT, Fedorov AA, Kolb P, Xu C, Fedorov EV, Shoichet BK, Barondeau DP, Almo SC and Raushei FM (2009) « Structure, mechanism, and substrate profile for Sco3085: the closest bacterial homologue to human renal dipeptidase. » Biochemistry 49 (3) : 611-622. https://doi.org/10.1021/bi901935y.

D'Ambrosio C, Talamo F, Vitale RM, Amodeo P, Tell G, Ferrara L and Scaloni A (2003) « Probing the dimeric structure of porcine aminoacylase 1 by mass spectrometric and modeling precedures. » Biochemistry 42 (15) : 4430-4443. https://doi.org/10.1021/bi0206715.

Da Silva GFZ and Ming L-J (2005) « Catechol oxidase activity of Di-Cu^{2+}-substituted aminopeptidase from *Streptomyces griseus*. » Journal of the American Chemical Society 127 (47) : 16380-16381. https://doi.org/10.1021/ja056034u.

Dalton JAR and Jackson RM (2007) « An evaluation of automated homology modelling methods at low target-template sequence similarity. » Bioinformatics 23 (15) : 1901-1908. https://doi.org/10.1093/bioinformatics/btm262.

Damasceno LM, Huang C-J and Batt CA (2012) « Protein secretion in *Pichia pastoris* and advances in protein production. » Applied Microbiology and Biotechnology 93 : 31-39. https://doi.org/10.1007/s00253-011-3654-z.

Damen M, Aarbiou J, van Dongen SFM, Buijs-Offerman RM, Spijkers PP, van den Heuvel M, Kvashnina K, Nolte RJM, Scholte BJ and Feiters M (2010) « Delivery of DNA and siRNA by novel gemini-like amphiphilic peptides. » Journal of Controlled Release 145 : 33-39. htpps://doi:10.1016/j.jconrel.2010.03.028.

Dan A, Ghosh S and Moulik SP (2010) « Interaction of cationic hydroxyethylcellulose (JR400) and cationic hydrophobically modified hydroxyethylcellulose (LM200) with the amino-acid based anionic amphiphile Sodium N-Dodecanoyl Sarcosinate (SDDS) in aqueous medium. » Carbohydrate Polymers 80 : 44-52. https://doi:10.1016/j.carbpol.2009.10.061.

Davis R, Bienvenue D, Swierczek SI, Gilner DM, Rajagopal L, Bennett B and Holz RC (2006) « Kinetic and spectroscopic characterization of the E134A- and E134D-altered dapE-encoded N-succinyl-L,L-diaminopimelic acid desuccinylase from *Haemophilus influenzae*. » Journal of Biological Inorganic Chemistry 11 : 206-216. https://doi.org/10.1007/s00775-005-0071-8.

De Léon, Jimérez-Islas H, Gonzalez-Cuevas M and de la Rosa APB (2004) « Analysis of the expression of the *Trichoderma harzianum ech42* gene in two isogenic clones of *Echerichia coli* by surface response methodology. » Process Biochemistry 39 (12) : 2173-2178. https://doi.org/10.1016/j.procbio.2003.11.013.

De Lima JY, dos Santos MDM, Andreassa EC, Kurt LU, Carvalho PC and De Souza TAC (2022) « Refolding of metacaspase 5 from *Trypanosoma cruzi*, structural characterization and the influence of c-terminal in protein recombinant production. » Protein Expression and Purification 191 : 106007. https://doi.org/10.1016/j.pep.2021.106007.

De los Ríos AP, van Rantwijk F and Sheldon RA (2012) Effective resolution of 1-phenyl ethanol by *Candida antarctica* lipase B catalysed acylation with vinyl acetate in protic ionic liquids (PILs). » Green Chemistry 14 : 1584-1588. https://doi.org/10.1039/C2GC35196J.

Del Tito Jr BJ, Ward JM, Hodgson J, Gershater CJ, Edwards H, Wysocki LA, Watson FA, Sathe G and Kane JF (1995) « Effects of a minor isoleucyl tRNA on heterologous protein translation in *Escherichia coli*. » Journal of Biotechnology 177 (24) : 7086-7091. https://doi.org/10.1128/jb.177.24.7086-7091.1995.

Deljooei M, Zargar G, Nooripoor V, Takassi MA and Esfandiarian A (2021) « Novel green surfactant made from L-aspartic acid as enhancer of oil production from sandstone reversoirs: wettability, IFT, microfluidic, and core flooding assessments. » Journal of Molecular Liquids 323 : 115037. https://doi.org/10.1016/j.molliq.2020.115037.

Desmarais W, Bienvenue DL, Bzymek KP, Petsko GA, Ringe D and Holz RC (2006) « The high-resolution structures of the neutral and the low pH crystals of aminopeptidase from *Aeromonas proteolytica*. » Journal of Biological Inorganic Chemistry 11 : 398-408. https://doi.org/10.1007/s00775-006-0093-x.

Dettori L, Ferrari F, Framboisier X, Paris C, Guiavarc'h Y, Hôtel L, Aymes A, Leblond P, Humeau C, Kapel R, Chevalot I, Aigle B and Delaunay S (2018a) « An aminoacylase activity from *Streptomyces ambofaciens* catalyzes the acylation of lysine on α-position and peptides on N-terminal position. » Engineering in Life Sciences 18 (8) : 589-599. https://doi.org/10.1002/elsc.201700173.

Dettori L, Vibert F, Guiavarc'h Y, Delaunay S, Humeau C, Blin J-L and Chevalot I (2018b) « N-α-acylation of lysine catalyzed by immobilized aminoacylases from Streptomyces ambofaciens in aqueous medium. » Microporous and Mesoporous Materials 267 : 24-34. https://doi.org/10.1016/j.micromeso.2018.03.018.

Diller DJ and Merz Jr KM (2001) « High throughput docking for library design and library prioritization. » Proteins: Structure, Function, and Bioinformatics 43 (2) : 113-124. https://doi.org/10.1002/1097-0134(20010501)43:2%3C113::AID-PROT1023%3E3.0.CO;2-T.

Dong A, Dobrovetsky E, Seitova A, He H, Tempel W, Kozieradzki I, Arrowsmith CH, Weigelt J, Bountra C, Edwards AM, Bochkarev A and Cossar D (2008) « Crystal structure of human carnosine dipeptidase 1. » Full wwPDB X-ray Structure Validation Report.

Dorr BM and Fuerst DE (2018) « Enzymatic amidation for industrial applications. » Biocatalysis and biotransformation 43 : 127-133. https://doi.org/10.1016/j.cbpa.2018.01.008.

Dunn JJ, Studier FW and Gottesman M (1983) « Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. » Journal of Molecular Biology 166 (4) : 477-535. https://doi.org/10.1016/S0022-2836(83)80282-4.

Dupuy J, Dutoit R, Durisotti V, Demarez M, Borel F, Van Elder D, Legrain C and Bauvois C (2013) « The crystal structure of a M20 family metallo-carboxypeptidase Sso-CP2 from *Sulfolobus solfataricus*. » Full wwPDB X-ray Structure Validation Report.

Durmus B, Temurtas H, Yumusak N, Temurtas F and Kazan R (2008) « The cost function minimization for predictive control by Newton-Raphson method. » Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS) Vol II.

Dutta D and Mishra S (2014) « The structural and energetic aspects of substrate binding and the mechanism of action of the DapE-encoded N-succinyl-L,L-diaminopimelic acid desuccinylase (DapE) investigated using a hybrid QM/MM method. » Physical Chemistry Chemical Physics 16 : 26348-26358. https://doi.org/10.1039/C4CP03986F.

Dutta D and Mishra S (2016) « Loss of catalytic activity in the E134D, H67A, and H349A mutants of DapE: mechanistic analysis with QM/MM investigation. » The Journal of Physical Chemistry B 120 (45) : 11654-11664. https://doi.org/10.1021/acs.jpcb.6b07446.

Dvorak P, Chrast L, Nikel PI, Fedr R, Soucek K, Sedlackova M, Chaloupkova R, de Lorenzo V, Prokop Z and Damborsky J (2015) « Exacerbation of substrate toxicity by IPTG in *Escherichia coli* BL21 (DE3) carrying a synthetic metabolic pathway. » Microbial Cell Factories 14 : 201. https://doi.org/10.1186/s12934-015-0393-3.

Eagon RG (1962) « Pseudomas natriegens, a marine bacterium with a generation time of less than 10 minutes. » Journal of Bacteriology 83 (4) : 736-737. https://doi.org/10.1128/jb.83.4.736-737.1962.

Ercan A, Park HI and Ming L-J (2006) « A « Mmoonlighting » dizinc aminopeptidase from *Streptomyces griseus*: mechanisms for peptide hydrolysis and 4 x 1010-fold acceleration of the alternative phosphodiester hydrolysis. » Biochemistry 45 : 13779-13793. https://doi.org/10.1021/bi061086x.

Eramian D, Eswar N, Shen M-Y and Sali A (2009) « How well can the accuracy of comparative protein structure models be predicted? » Protein Science 17 (11) 1881-1893. https://doi.org/10.1110/ps.036061.108.

Fan J, Fu A and Zhang L (2019) « Progress in molecular docking. » Quantitative Biology 7 : 83-89. https://doi.org/10.1007/s40484-019-0172-y.

Fatima S, Faryad A, Ataa A, Joyia FA and Parvaiz A (2020) « Microbial lipase production: a deep insight into the recent advances of lipase production and purification techniques. » Biotechnology and Applied Biochemistry 68 (3) : 445-458. https://doi.org/10.1002/bab.2019.

Fernandez-Castané A, Vine CE, Camical G and Lopez-Santin J (2012) « Evidencing the role of lactose permease in IPTG uptake by *Escherichia coli* in fed-batch high cell density cultures. » Journal of Biotechnology 157 (3) : 391-398. https://doi.org/10.1016/j.jbiotec.2011.12.007.

Fernandez-Fuentes N, Zhai J and Fiser A (2006) « ArchPRED: a template based loop structure prediction server. » Nucleic Acids Research 34 (2) : W173-W176. https://doi.org/10.1093/nar/gkl113.

Ferrari F (2014) « Etude de la sélectivité d'acylation enzymatique de peptides : prédiction de la sélectivité de la lipase B de *Candida antarctica* par modélisation moléculaire et recherche de nouvelles enzymes spécifiques de type aminoacylases. » Université de Lorraine. https://theses.hal.science/tel-01614465v1.

Ferrari F, Paris C, Maigret B, Bidouil C, Delaunay S, Humeau C and Chevalot I (2014) « Molecular rules for chemo- and regio-selectivity of *Candida antarctica* lipase B in peptide acylation reactions. » Journal of Molecular Catalysis B : Enzymatic 101 : 122-132. http://dx.doi.org/10.1016/j.molcatb.2013.12.007.

Ferrer M, Chernikova TN, Yakimov MM, Golyshin PN and Timmis KN (2003) « Chaperonins govern growth of *Escherichia coli* at low temperatures. » Nature Biotechnology 21 : 1266-1267. https://doi.org/10.1038/nbt1103-1266.

Ferrer M, Chernikova TN, Timmis KN and Golyshin PN (2004a) « Expression of a temperature-sensitive esterase in a novel chaperone-based *Escherichia coli* strain. » Applied and Environmental Microbiology 70 (8) : 4499-4504. https://doi.org/10.1128/AEM.70.8.4499-4504.2004.

Ferrer M, Lünsdorf H, Chernikova TN, Yakimov M, Timmis KN and Golyshin PN (2004b) « Functional consequences of single: double ring transitions in chaperonins: life in the cold. » Molecular Microbiology 53 (1) : 167-182. https://doi.org/10.1111/j.1365-2958.2004.04077.x.

Fiser A, Do RKG and Sali A (2000) « Modeling of loops in protein structures. » Protein Science 9 (9) : 1753-1773. https://doi.org/10.1110/ps.9.9.1753.

Franceschi S, de Viguerie N, Riviere M and Lattes A (1999) « Synthesis and aggregation of two-headed surfactants bearing amino acid moieties. » New Journal of Chemistry 23 : 447-452. https://doi.org/10.1039/A809079C.

Fukatsu H, Herai S, Hashimoto Y, Maseda H, Higashibata H and Kobayashi M (2005) « High-level expression of a novel amine-synthesizing enzyme, N-substituted foramide deformylase, in Streptomyces with a strong protein expression system. » Protein Expression and Purification 40 (1) : 212-219. https://doi.org/10.1016/j.pep.2004.11.013.

Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DR, Fogel LJ and Freer ST (1995) « Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. » Chemistry & Biology 2 (5) : 317-324. https://doi.org/10.1016/1074-5521(95)90050-0

Gehlhaar DK, Verkhivker GM and Rejto PA (1999) « Rational drug design: novel methodology and practical applications. » American Chemical Society symposium series 719 : 292-311.

George S (2017) « Production hétérologue d'aminoacylase issues de *Streptomyces ambofaciens*. » Rapport de stage M2 Mention BioMANE. Université de Lorraine.

Gerova M, Rodrigues F, Lamère J-F, Dobrev A and Fery-Forgues S (2008) « Self-assembly properties of some chiral N-palmitoyl amino acid surfactants in aqueous solution. » Journal of Colloid and Interface Science 319 : 526-533. https://doi.org/10.1016/j.jcis.2007.12.004.

Girish TS, Colaco BM, Misquith S and Gopal B (2013) « Structure of an amidohydrolase, SACOL0085, from methicillin-resistant *Staphylococcus aureus* COL. » Acta Crystallographica Section F, Structural Biology Communications 69 : 103-108. https://doi.org/10.1107/S1744309112049822.

Godehard SP, Badenhorst CPS, Müller H and Bornscheuer UT (2020) « Protein engineering for enhanced acyltransferase activity, substrate scope, and selectivity of the *Mycobacterium smegmatis* acyltransferase MsAcT. » ACS Catalysis 10 (14) : 7552-7562. https://doi.org/10.1021/acscatal.0c01767.

Gottesman S (1996) « Protease and their targets in *Escherichia coli*. » Annual Review of Genetics 30 : 465-506. https://doi.org/10.1146/annurev.genet.30.1.465.

Goujard L, Figueroa MC, and Villeneuve P (2004) « Chemo-enzymatic synthesis of N-arachidonoyl glycine. » Biotechnology Letters 26 : 1211-1216. https://doi.org/10.1023/B:BILE.0000036597.25673.33.

Grabski A, Mehler M and Drott D (2005) « The Overnight Express Autoinduction System: high density cell growth and protein expression while you sleep. » Advertising Feature : Application Note 2 : 233-235. https://doi.org/10.1038/nmeth0305-233.

Graebin NG, Martins AB, Lorenzoni ASG, Garcia-Galan C, Fernandez-Lafuente R, Ayub MAZ and Rodrigues RC (2011) « Immobilization of lipase B from *Candida antarctica* on porous styrene-divinylbenzene beads improves butyl acetate synthesis. » Biotechnology Progress 28 (2) : 406-412. https://doi.org/10.1002/btpr.1508.

Greener A (2004) « BL21-Gold competent cells, BL21-Gold (DE3) competent cells, and BL21-Gold (DE3) pLysS competent cells. » Instruction Manual – Agilent Technologies

Grodberg J and Dunn JJ (1988) « *ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. » Journal of Bacteriology 170 : 1245-1253. https://doi.org/10.1128/jb.170.3.1245-1253.1988.

Grossman TH, Kawasaki ES, Punreddy SR, Osburne MS (1998) « Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. » Gene 209 (1-2) : 95-103. https://doi.org/10.1016/S0378-1119(98)00020-1.

Guex N, Peitsch MC and Schwede T (2009) « Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. » Electrophoresis 30 (S1) : S162-S173. https://doi.org/10.1002/elps.200900140.

Guo J, Sun L, Zhang F, Sun B, Xu B and Zhou Y (2022) « Review: Progress in synthesis, properties and application of amino acid surfactants. » Chemical Physics Letters 794 : 139499. https://doi.org/10.1016/j.cplett.2022.139499.

Gustafsson C, Govindarajan S and Minshull J (2004) « Codon bias and heterologous protein expression. » Trends in Biotechnology 22 (7) : 346-353. https://doi.org/10.1016/j.tibtech.2004.04.006.

Guzenko D, Lafita A, Monastyrskyy B, Kryshtafovych A and Duarte JM (2019) « Assessment of protein assembly prediction in CASP13. » Proteins: Structure, Function, and Bioinformatics 87 (12) : 1190-1199. https://doi.org/10.1002/prot.25795.

Haas J, Barbato A, Behringer D, Studer G, Roth S, Bertoni M, Mostaguir K, Gumienny R and Schwede T (2018) « Continuous Automated Model EvaluatiOn (CAMEO) complementing the critical assessment of structure prediction in CASP12. » Proteins: Structure, Function, and Bioinformatics 86 (S1) : 387-398. https://doi.org/10.1002/prot.25431.

Haas J, Gumienny R, Barbato A, Ackermann F, Tauriello G, Bertoni M, Studer G, Smolinski A and Schwede T (2019) « Introducing « best single template » models as reference baseline for the Continuous Automated Model Evaluation (CAMEO). » Proteins: Structure, Function, and Bioinformatics 87 (12) : 1378-1387. https://doi.org/10.1002/prot.25815.

Haddad Y, Heger Z and Adam V (2016) « Guidelines for homology modeling of dopamine, norepinephrine, and serotonin transporters. » ACS Chemical Neuroscience 7 (11) : 1607-1613. https://doi.org/10.1021/acschemneuro.6b00242.

Haddad Y, Adam V and Heger Z (2020) « Ten quick tips for homology modeling of high-resolution protein 3D structures. » Plos Computational Biology 16 (4) : e1007449. https://doi.org/10.1371/journal.pcbi.1007449.

Haeger G, Probst J, Jaeger K-E, Bongaerts J and Siegert P (2023a) « Novel aminoacylases from *Streptomyces griseus* DSM 40236 and their recombinant production in *Streptomyces lividans*. » FEBS Open Bio 13 (12) : 2224-2238. https://doi.org/10.1002/2211-5463.13723.

Haeger G, Wirges J, Tanzmann N, Oyen S, Jolmes T, Jaeger K-E, Schörken U, Bongaerts J and Siegert P (2023b) « Chaperonne assisted recombinant expression of a mycobacterial aminoacylase in *Vibrio natriegens* and *Escherichia coli* capable of N-lauroyl-L-amino acid synthesis. » Microbial Cell Factories 22 : 77. https://doi.org/10.1186/s12934-023-02079-1.

Haeger G, Jolmes T, Oyen S, Jaeger K-E, Bongaerts J, Schörken U and Siegert P (2024a) « Novel recombinant aminoacylase from *Paraburkholderia monticola* capable of N-acyl-amino acid synthesis. » Applied Microbiology and Biotechnology 108 : 93. https://doi.org/10.1007/s00253-023-12868-8.

Haeger G, Wirges J, Bongaerts J, Schörken U and Siegert P (2024b) « Perspectives of aminoacylases in biocataytic synthesis of N-acyl-amino acids surfactants. » Applied Microbiology and Biotechnology 108 : 495. https://doi.org/10.1007/s00253-024-13328-7.

Hassan ME (2016) « Production, immibilization and industrial uses of penicillin g acylase. » International Journal of Current Research and Review 8 (15) : 11-16.

Hatahet F, Boyd D and Beckwith J (2014) « Disulfide bond formation in prokaryotes: history, diversity and design. » Biochimica et Biophysica Acta (BBA) – Proteins and Proteomics 1844 (8) : 1402-1414. https://doi.org/10.1016/j.bbapap.2014.02.014.

Henseling J and Röhm K-H (1988) « Aminoacylase I from hog kidney: anion effects and the pH dependance of kinetic parameters. » Biochimica et Biophysica Acta (BBA) – Lipids and Lipid Metabolism 959 (3) : 370-377. https://doi.org/10.1016/0005-2760(88)90211-1.

Herai S, Hashimoto Y, Higashibata H, Maseda H, Ikeda H, Omura S and Kobayashi M (2004) « Hyper-inducible expression system for streptomycetes. » Proceedings of the National Academy of Sciences 101 (39) : 14031-14035. https://doi.org/10.1073/pnas.0406058101.

Herga S, Berrin J-G, Perrier A and Giardina PT (2006) « Identification of the zinc binding ligands and the catalytic residue in human aspartoacylase, an enzyme involved in Canavan disease. » FEBS Letters 580 (25) : 5899-5904. https://doi.org/10.1016/j.febslet.2006.09.056.

Hilderbrandt RL (1977) « Application of Newton-Raphson optimization techniques in molecular mechanics calculations. » Computers & Chemistry 1 (3) : 179-186. https://doi.org/10.1016/0097-8485(77)85008-0.

Hoff J, Daniel B, Stukenberg D, Thuronyi BW, Waldminghaus T and Fritz G (2020) « *Vibrio natriegens*: an ultrafast-growing marine bacterium as emerging synthetic biology chassis. » Environmental Microbiology 22 (10) : 4394-4408. https://doi.org/10.1111/1462-2920.15128.

Hoffart E, Grenz S, Lange J, Nitschel R, Müller F, Schwentner A, Feith A, Lenfers-Lücker, Takors R and Blombach B (2017) « High substrate uptake rates empower *Vibrio natriegens* as production host for industrial biotechnology. » Applied and Environmental Microbiology 83 : e01614-17. https://doi.org/10.1128/AEM.01614-17.

Holmberg K, Bilén F and Bordes R (2025) « Development of amino acid-based surfactants: from synthesis to applications. » Current Opinion in Colloid & Interface Science 75 : 101884. https://doi.org/10.1016/j.cocis.2024.101884.

Hou J, Wu T, Cao R and Cheng J (2019) « Protein tertiary structure modeling driven by deep learning and contact distance prediction in CASP13. » Proteins: Structure, Function, and Bioinformatics 87 (12) : 1165-1178. https://doi.org/10.1002/prot.25697.

Hou X, Li J, Nuo Z, Ren Y, Gao C (2016) « The study on D-aminoacylase gene synthesis via chemical and enzimatic combined method. » Bulgarian Chemical Communications 48 (2) : 219-224.

Hsieh JM, Tsirulnikov K, Sawaya MR and Pushkin A (2010) « Structures of aminoacylase 3 in complex with acetylated substrates. » Proceedings of the National Academy of Sciences 107 (42) : 17962-17967. https://doi.org/10.1073/pnas.1006687107.

Huang J, Villemain J, Padilla R and Sousa R (1999) « Mechanisms by which T7 lysozyme specifically regulates T7 RNA polymerase during different phases of transcription. » Journal of Molecular Biology 293 (3) : 457-475. https://doi.org/10.1006/jmbi.1999.3135.

Hult K and Berglund P (2007) « Enzyme promiscuity: mechanism and applications. » Trends in Biotechnology 25 (5) : 231-238. https://doi.org/10.1016/j.tibtech.2007.03.002.

Husson E, Humeau C, Blanchard F, Framboisier X, Marc I and Chevalot I (2008) « Chemo-selectivity of the N,O-enzymatic acylation in organic media and in ionic liquids. » Journal of Molecular Catalysis B: Enzymatic 55 (3-4) : 110-117. https://doi.org/10.1016/j.molcatb.2008.02.004.

Husson E, Humeau C, Paris C, Vanderesse R, Framboisier X, Marc I and Chevalot I (2009) « Enzymatic acylation of polar dipeptides: influence of reaction media and molecular environment of functionel groups. » Process Biochemistry 44 (4) : 428-434. https://doi.org/10.1016/j.procbio.2008.12.011.

Husson E, Humeau C, Harscoat C, Framboisier X, Paris C, Dubreucq E, Marc I and Chevalot I (2011) « Enzymatic acylation of the polar dipeptide, carnosine: reaction performances in organic and aqueous media. » Process Biochemistry 46 (4) : 945-952. https://doi.org/10.1016/j.procbio.2011.01.007.

Hwang PM, Pan JS and Sykes BD (2014) « Targeted expression, purification, and cleavage of fusion proteins from inclusion bodies in *Escherichia coli*. » FEBS Letters 588 (2) : 247-252. https://doi.org/10.1016/j.febslet.2013.09.028.

Incir I and Kaplan O (2024) « *Escherichia coli* as a versatible cell factory: advances and challenges in recombinant protein production. » Protein Expression and Purification 219 : 106463. https://doi.org/10.1016/j.pep.2024.106463.

Infante R, Dominguez JG, Erra P and Prats JM (1984) « Surface active molecules: preparation and properties of long chain nα-acyl-L-α-amino-ω-guanidine alkyl acid derivatives. » International Journal of Cosmetic Science 6 (6) : 275-282. https://doi.org/10.1111/j.1467-2494.1984.tb00385.x.

Inouye M, Arnheim N and Sternglanz R (1973) « Bacteriophage T7 lysozyme is an N-acetylmuramyl-L-alanine amidase. » Journal of Biological Chemistry 248 (20) : 7247-7252. https://doi.org/10.1016/S0021-9258(19)43385-1.

Ismail H, Lau RM, van Rantwijk F and Sheldon RA (2008) « Fully enzymatic resolution of chiral amines: acylation and deacylation in the presence of *Candida antarctica* lipase B. » Advanced Synthesis & Catalysis 350 (10) : 1511-1516. https://doi.org/10.1002/adsc.200800091.

Jackson DA, Symons RH and Berg P (1972) « Biochemical method for inserting new genetic information into DNA of Simian virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. » Proceedings of the National Academy of Sciences 69 (10) : 2904-2909. https://doi.org/10.1073/pnas.69.10.2904.

Jain AN (1996) « Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. » Journal of Computer-Aided Molecular Design 10 : 427-440. https://doi.org/10.1007/BF00124474.

Jansen RS, Addie R, Merkx R and van de Wetering K (2015) « N-lactoyl-amino acids are ubiquitous metabolites that originates from CNDP2-mediated reverse proteolysis of lactate and amino acids. » Proceedings of the National Academy of Sciences 112 (21) : 6601-6606. https://doi.org/10.1073/pnas.1424638112.

Jekabsons MB, Echtay KS and Brand MD (2002) « Nucleotide binding to human uncoupling protein-2 refolded from bacterial inclusion bodies. » Biochemincal Journal 366 (2) : 565-571. https://doi.org/10.1042/bj20020469.

Jenny RJ, Mann KG and Lundblad RL (2003) « A critical review fo the methods for cleavage of fusion proteins with thrombin and factor Xa. » Protein Expression and Purification 31 (1) : 1-11. https://doi.org/10.1016/S1046-5928(03)00168-2.

Jeschek M, Reuter R, Heinisch T, Trindler C, Klehr J, Panke S and Ward TR (2016) « Directed evolution of artificial metalloenzymes for *in vivo* metathesis. » Letter 537 : 661-665. https://doi.org/10.1038/nature19114.

Jia B and Jeon CO (2016) « High-throughput recombinant protein expression in *Escherichia coli* : current status and future perspectives. » Open Biology 6 : 160196. https://doi.org/10.1098/rsob.160196.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S ans Madden TL (2008) « NCBI BLAST: a better web interface. » Nucleic Acids Research 36 (S2) : W5-W9. https://doi.org/10.1093/nar/gkn201.

Joint Center for Structural Genomics (2010) « Crystal structure of a putative acetylornithine deacetylase (RPA2325) from *Rhodopseudomonas palustris* CGA009 at 1.90 A resolution. » Full wwPDB X-ray Structure Validation Report.

Joondan N, Jhaumeer-Laulloo S, Caumul P and Akerman M (2016) « Synthesis, physicolchemical, and biological activities of novel N-acyl tyrosine monomeric and gemini surfactants in single and SDS/CTAB-mixed micellar system. » Journal of Physical Organic Chemistry 30 (10) : e3675. https://doi.org/10.1002/poc.3675.

Joondan N, Laulloo SJ and Caumul P (2018) « Amino acids: building blocks for the synthesis of greener amphiphiles. » Journal of Dispersion Science and Technology 39 (11) : 1550-1564. https://doi.org/10.1080/01932691.2017.1421085.

Jost E, Kazemi M, Mrkonjic V, Himo F, Winkler CK and Kroutil W (2020) « Variants of the acyltransferase from *Mycobacterium smegmatis* enable enantioselective acyl transfer in water. » ACS Catalysis 10 (18) : 10500-10507. https://doi.org/10.1021/acscatal.0c02981.

Jozic D, Bourenkow G, Bartunik H, Schloze H, Dive V, Henrich B, Huber R, Bode W and Maskos K (2002) « Crystal structure of the dinuclear zinc aminopeptidase PepV from *Lactobacillus delbrueckii* unravels its preference for dipeptides. » Structure 10 (8) : 1097-1106. https://doi.org/10.1016/S0969-2126(02)00805-5.

Jozic D, Kaiser JT, Huber R, Bode W and Maskos K (2003) « X-ray structure of isoaspartyl dipeptidase from *E. coli* : a dinuclear zinc peptidase evolved from amidohydrolases. » Journal of Molecular Biology 332 (1-5) : 243-256. https://doi.org/10.1016/S0022-2836(03)00845-3.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P and Hassabis D (2021) « Highly accurate protein structure prediction with AlphaFold. » Nature 596 : 583-589. https://doi.org/10.1038/s41586-021-03819-2.

Juturu V and Wu JC (2017) « Heterologous protein expression in *Pichia pastoris*: latest research progress and applications. » ChemBioChem 19 (1) : 7-21. https://doi.org/10.1002/cbic.201700460.

Kaar JL, Jesionowski AM, Berberich JA, Moulton R and Russell AJ (2003) « Impact of ionic liquid physical properties on lipase activity and stability. » Journal of the American Chemical 125 (14) : 4125-4131. https://doi.org/10.1021/ja028557x.

Kamat S, Barrera J, Beckman J and Ruseel AJ (1992) « Biocatalytic synthesis of acrylates in organic solvents and supercritical fluids: I. Optimization of enzyme environment. » Biotechnology and Bioengineering 40 (1) : 158-166. https://doi.org/10.1002/bit.260400122.

Kandathil SM, Greener JG and Jones DT (2019) « Recent developments in deep learning applied to protein structure prediction. » Proteins: Structure, Function, and Bioinformatics 87 (12) : 1179-1189. https://doi.org/10.1002/prot.25824.

Kane JF (1995) « Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. » Current Opinion in Biotechnology 6 (5) : 494-500. https://doi.org/10.1016/0958-1669(95)80082-4.

Karayannis D, Papanikolaou S, Vatistas C, Paris C and Chevalot I (2023) « Yeast lipid produced through glycerol conversions and its use for enzymatic synthesis of amino acid-based biosurfactants. » International Journal of Molecular Sciences 24 (1) : 714. https://doi.org/10.3390/ijms24010714.

Karbalaei M, Rezaee SA and Farsiani H (2020) « *Pichia pastoris*: a highly successul expression system for optimal synthesis of heterologous proteins. » Journal of Cellular Physiology 235 (9) : 5867-5881. https://doi.org/10.1002/jcp.29583.

Katiyar SS, Kushwah V, Dora CP, Patil RY and Jain S (2019) « Design and toxity evaluation of novel fatty acid-amino acid-based biocompatible surfactants. » AAPS PharmSciTech 20 : 186. https://doi.org/10.1208/s12249-019-1396-x.

Katz E, Thompson CJ and Hopwood DA (1983) « Cloning and expression of the tyrosinase gene from *Streptomyces antibioticus* in *Streptomyces lividans*. » Journal of General Microbiology 129 (9) : 2703-2714. https://doi.org/10.1099/00221287-129-9-2703.

Kaur H, Kumar C, Junot C, Toledano MB and Bachhawat A (2009) « Dug1p is a Cys-Gly peptidase of the γ-glutamyl cycle of *Saccharomyces cerevisiae* and represents a novel family of Cys-Gly peptidases. » Journal of Biological Chemistry 284 (21) : 14493-14502. https://doi.org/10.1074/jbc.M808952200.

Kaur H, Datt M, Ekka MK, Mittal M, Singh AK and Kumaran S (2011) « Cys-Gly specific dipeptidase Dug1p from *S. cerevisiae* binds promiscuously to di-, tri- and tetra-peptides : peptide-protein interaction, homology modeling, and activity studies reveal a latent promiscuity in substrate recognition. » Biochimie 93 (2) : 175-186. https://doi.org/10.1016/j.biochi.2010.09.008.

Kaur J, Kumar A and Kaur J (2018) « Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. » International Journal of Biological Macromolecules 106 : 803-822. https://doi.org/10.1016/j.ijbiomac.2017.08.080.

Khan RH, Rao KBCA, Eshwari ANS, Totey SM and Panda AK (1998) « Solubilization of recombinant ovine growth hormone with retention of native-like secondary structure and its refolding from the inclusion bodies of *Escherichia coli*. » Biotechnology Progress 14 (5) : 722-728. https://doi.org/10.1021/bp980071q.

Khmelnitsky YL, Welch SH, Clark DS and Dordick JS (1994) « Salts dramatically enhance activity of enzymes suspended in organic solvents. » Journal of the American Chemical 116 (6) : 2647-2648. https://doi.org/10.1021/ja00085a066.

Kim G, Lee S, Karin EL, Kim H, Moriwaki Y, Ovchinnikov S, Steinegger M and Mirdita M (2024) « Easy and accurate protein structure prediction using ColabFold. » Nature Protocols 20 : 620-642. https://doi.org/10.1038/s41596-024-01060-5.

Kim J, Copeland CE, Seki K, Vogeli B and Kwon Y-C (2020) « Tuning the cell-free protein synthesis system for biomanufacturing of monomeric human filaggrin. » Frontiers in Bioengineering and Biotechnology 8 : 590341. https://doi.org/10.3389/fbioe.2020.590341.

Kim PI, Bai H, Bai D, Chae H, Chung S, Kim Y, Park R and Chi Y-T (2004) « Purification and characterization of a lipopeptide produced by *Bacillus thuringiensis* CMB26. » Journal of Applied Microbiology 97 (5) : 942-949. https://doi.org/10.1111/j.1365-2672.2004.02356.x.

Kim PI, Ryu J, Kim YK and Chi Y-T (2010) « Production of biosurfactant lipopeptides iturin A, fengycin, and surfactin A from *Bacillus subtilis* CMB32 for control of *Colletotrichum gloeosporioides*. » Journal of Microbiology and Biotechnology 20 (1) : 138-145. https://doi.org/10.4014/jmb.0905.05007.

Kimura Y, Kobayashi Y, Adachi S and Matsuno R (2001) « Aminoacylase-catalyzed synthesis of N-acyl amino acid from fatty acid or its ethyl ester and amino acid. » Biochemical Engineering for 2001: Proccedings of Asia-Pacific Biochemical Engineering Conference 1992 p.109

Kiran GS, Thomas TA, Selvin J, Sabarothnam B and Lipton AP (2010) « Optimization and characterization of a new lipopeptide biosurfactant produced by marine *Brevibacterium aureum* MSA13 in solid state culture. » Bioresource Technology 101 (7) : 2389-2396. https://doi.org/10.1016/j.biortech.2009.11.023.

Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE and Kolinski A (2016) « Coarse-grained protein models and their applications. » Chemical Reviews 116 (14) : 7898-7936. https://doi.org/10.1021/acs.chemrev.6b00163.

Koehn J and Hunt I (2009) « High-throughput protein production (HTPP): a review of enabling technologies to expedite protein production. » Methods in Molecular Biology: High

throughtput protein expression and purification volume 496 Chapter 1 p1-18. https://doi.org/10.1007/978-1-59745-196-3.

Koreishi M, Kawasaki R, Imanaka H, Imamura K and Nakanishi K (2005a) « A novel ε-lysine acylase from *Streptomyces mobaraensis* for synthesis of N-ε-acyl-L-lysines. » Journal of the American Oil Chemists' Society 82 : 631-637. https://doi.org/10.1007/s11746-005-1121-2.

Koreishi M, Asayama F, Imanaka H, Imamura K, Kadota M, Tsuno T and Nakanishi K (2005b) « Purification and characterization of a novel aminoacylase from *Streptomyces mobaraensis*. » Bioscience, Biotechnology, and Biochemistry 69 (10) 1914-1922. https://doi.org/10.1271/bbb.69.1914.

Koreishi M, Zhang D, Imanaka H, Imamura K, Adachi S, Matsuno R and Nakanishi K (2006) « A novel acylase from *Streptomyces mobaraensis* that efficiently catalyzes hydrolysis/synthesis of capsaicins as well as N-acyl-L-amino acids and N-acyl-peptides. » Journal of Agricultural and Food Chemistry 54 (1) : 72-78. https://doi.org/10.1021/jf052102k.

Koreishi M, Tani K, Ise Y, Imanaka H, Imamura K and Nakanishi K (2007) « Enzymatic synthesis of β-lactam antibiotics and N-fatty-acylated amino compounds by the acyl-transfer reaction catalyzed by penicillin V acylase from *Streptomyces mobaraensis*. » Bioscience, Biotechnology, and Biochemistry 71 (6) : 1582-1586. https://doi.org/10.1271/bbb.70052.

Koreishi M, Kawasaki R, Imanaka H, Imamura K, Takakura Y and Nakanishi K (2009a) « Efficient N-ε-lauroyl-L-lysine production by recombinant ε-lysine acylase from *Streptomyces mobaraensis*. » Journal of Biotechnology 141 (3-4) : 160-165. https://doi.org/10.1016/j.jbiotec.2009.03.008.

Koreishi M, Nakatani Y, Ooi M, Imanaka H, Imamura K and Nakanishi K (2009b) « Purification, characterization, molecular cloning, and expression of a new aminoacylase from *Streptomyces mobaraensis* that can hysrolyze N-(middle/long)-chain-fatty-acyl-L-amino acids as well as N-short-chain-acyl-L-amino acids. » Bioscience, Biotechnology and Biochemistry 73 (9) : 1940-1947. https://doi.org/10.1271/bbb.90081.

Kosobokova EN, Skrypnik KA and Kosorukov VS (2016) « Overview of fusion tags for recombinant proteins. » Biochemistry (Moscow) 81 : 187-200. https://doi.org/10.1134/S0006297916030019.

Kots ED, Khrenova MG, Nemukhin AV and Varfolomeev SD (2019) « Aspartoacylase: a central system enzyme. Structure, catalytic activity and regulation mechanisms. » Russian Chemical Reviews 88 (1) : 1-26. https://doi.org/10.1070/RCR4842.

Krammer A, Kirchhoff PD, Jiang X, Venkatachalam CM and Waldman M (2005) « LigScore: a novel scoring function for predicting binding affinities. » Journal of Molecular Graphics and Modelling 23 (5) : 395-407. https://doi.org/10.1016/j.jmgm.2004.11.007.

Kryshtafovych A, Schwede T, Topf M, Fidelis K and Moult J (2019) « Critical assessment of methods of protein structure prediction (CASP) – Round XIII. » Proteins: Structure, Function, and Bioinformatics 87 (12) : 1011-1020. https://doi.org/10.1002/prot.25823.

Kryshtafovych A, Schwede T, Topf M, Fidelis K and Moult J (2021) « Critical assessment of methods of protein structure prediction (CASP) – Round XIV. » Proteins: Structure, Function, and Bioinformatics 89 (12) : 1607-1617. https://doi.org/10.1002/prot.26237.

Kryshtafovych A, Schwede T, Topf M, Fidelis K and Moult J (2023) « Critical assessment of methods of protein structure prediction (CASP) – Round XV. » Proteins: Structure, Function, and Bioinformatics 91 (12) : 1539-1549. https://doi.org/10.1002/prot.26617.

Kua GKB, Nguyen GKT and Li Z (2023) « Enzyme engineering for high-yielding amide formation: lipase-catalyzed synthesis of N-acyl glycines in aqueous media. » Angewandte Chemie 135 (14) : e202217878. https://doi.org/10.1002/ange.202217878.

Kuhlman B and Bradley P (2019) « Advances in protein structure prediction and design. » Nature Reviews Molecular Cell Biology 20 : 681-697. https://doi.org/10.1038/s41580-019-0163-x.

Lai W-L, Chou L-Y, Ting C-Y, Kirby R, Tsai Y-C, Wang A H-J and Liaw S-H (2004) « The functional role of the binuclear metal center in D-aminoacylase. » Enzyme Catalysis and Regulation 279 (14) : 13962-13967. https://doi.org/10.1074/jbc.M308849200.

Le Coq J, Pavlovsky A, Malik R, Sanishvili R, Xu C and Viola RE (2008) « Examination of the mechanism of human brain aspartoacylase through the binding of an intermadiate analogue. » Biochemistry 47 (11) : 3484-3492. https://doi.org/10.1021/bi702400x.

Le Guenic S, Chaveriat L, Lequart V, Joly N and Martin P (2018) « Renewable surfactants for biochemical applications and nanotechnology. » Journal of Surfactants and Detergents 22 (1) : 5-21. https://doi.org/10.1002/jsde.12216.

Lebendiker M and Danieli T (2014) « Production of prone-to-aggregate proteins. » FEBS Letters 588 (2) : 236-246. https://doi.org/10.1016/j.febslet.2013.10.044.

Lee D-H, Kim M-D, Lee W-H, Kweon D-H and Seo J-H (2004) « Consortium of fold-catalyzing proteins increases soluble expression of cyclohexanone monooxygenase in recombinant *Escherichia coli*. » Applied Microbiology and Biotechnology 63 : 549-552. https://doi.org/10.1007/s00253-003-1370-z.

Leemann M, Sagasta A, Eberhardt J, Schwede T, Robin X and Durairaj J (2023) « Automated benchmarking of combined protein structure and ligand conformation prediction. » Proteins: Structure, Function, and Bioinformatics 91 (12) : 1912-1924. https://doi.org/10.1002/prot.26605.

Leibly DJ, Nguyen TN, Kao LT, Hewitt SN, Barrett LK and van Voorhis WC (2012) « Stabilizing additives added during cell lysis aid in the solubilization of recombinant proteins. » Plos One 7 (12) : e52482. https://doi.org/10.1371/journal.pone.0052482.

Li C, Feng X-W, Wang N, Zhou Y-J and Yu X-Q (2008) « Biocatalytic promiscuity: the first lipase-catalysed asymmetric aldol reaction. » Green Chemistry 10 : 616-618. https://doi.org/10.1039/b803406k.

Lindner HA, Lunin VV, Alary A, Hecker R, Cygler M and Ménard R (2003) « Essential roles of zinc ligation and enzyme dimerization for catalysis in the aminoacylase-1/M20 family. » Journal of Biological Chemistry 278 (45) : 44496-44504. https://doi.org/10.1074/jbc.M304233200.

Lindner HA, Alary A, Boju LI, Sulea T and Ménard R (2005) « Roles of dimerization domain residues in binding and catalysis by aminoacylase-1. » Biochemistry 44 (48) : 15645-15651. https://doi.org/10.1021/bi051180y.

Lindner HA, Alary A, Wilke M and Sulea T (2008) « Probing the acyl-binding pocket of aminoacylase-1. » Biochemistry 47 (14) : 4266-4275. https://doi.org/10.1021/bi702156h.

Liu K-J, Liu K-M and Chang H-M (2007) « Biocatalytic synthesis of palmitoyl vanillylamide in supercritical carbon dioxide through amidation of vanillylamine hydrochloride and palmitic anhydride by lipase. » Food Chemistry 102 (4) : 1020-1026. https://doi.org/10.1016/j.foodchem.2006.06.045.

Liu C, Wang Y, Chai C, Ullah S, Zhang G, Xu B, Liu H and Zhao L (2018) « Interfacial activities and aggregation behaviors of N-acyl amino acid surfactants derived from vegetable oils. » Colloids and Surfaces A : Physicochemical and Engineering Aspects 559 : 54-59. https://doi.org/10.1016/j.colsurfa.2018.09.042.

Liu YC and Roujeinikova A (2015) « Expression, refolding, purification and crystallization of the sensory domain of the TlpC chemoreceptor from *Helicobacter pylori* for structural studies. » Protein Expression and Purification 107 : 29-34. https://doi.org/10.1016/j.pep.2014.11.003.

Liu Z, Zhen Z, Zuo Z, Wu Y, Liu A, Yi Q and Li W (2006) « Probing the catalytic center of porcine aminoacylase 1 by site-directed mutagenesis, homology modeling and substrate docking. » The Journal of Biochemistry 139 (3) : 421-430. https://doi.org/10.1093/jb/mvj047.

Lodge EP and Heatcock CH (1987) « Steric effects, as well as σ*-orbital energies, are important in diastereoface differenciiation in additions to chiral aldehydes. » Journal of the American Chemical Society 109 (11) : 3353-3361. https://doi.org/10.1021/ja00245a027.

Lone MS, Bhat PA, Shah RA, Chat OA and Dar AA (2017) « A green pH-switchable amino acid based smart wormlike micellar system for efficient and controlled drug delivery. » Chemistry Select 2 (3) : 1144-1148. https://doi.org/10.1002/slct.201601861.

Lodola A, Branduardi D, de Vivo M, Capoferri L, Mor M, Piomelli D and Cavalli A (2012) « A catalytic mechnism for cysteine N-terminal nucleophile hydrolases, as revealed by free energy simulations. » Plos One 7 (2) : e32397. https://doi.org/10.1371/journal.pone.0032397.

Lozano Terol G, Gallego-Jara J, Martinez RAS, Vivancos AM, Diaz MC and Puente TD (2021) « Impact of the expression system on recombinant protein production in *Escherichia coli* BL21. » Frontiers in Microbiology 12 : 682001. https://doi.org/10.3389/fmicb.2021.682001.

Lubkowski J and Wlodawer A (2019) « Geometric considerations support the doucle-displacement catalytic mechanism of L-asparaginase. » Protein Science 28 (10) : 1850-1864. https://doi.org/10.1002/pro.3709

Luo C, Yang B, Zhou Y, Yang J, Han F and Xu B (2020) « Gelation properties and application based on amino acids gelators with four kinds of edible oils. » Colloids and Surface A: Physicochemical and Enginerring Aspects 585 : 124184. https://doi.org/10.1016/j.colsurfa.2019.124184.

Lv J, Qiao W and Li Z (2016) « Vesicles from pH-regulated reversible gemini amino-acid surfactants as nanocapsules for delivery. » Colloids and Surfaces B: Biointerfaces 146 : 523-531. https://doi.org/10.1016/j.colsurfb.2016.06.054.

Majtan T, Frerman FE and Kraus JP (2011) « Effect of cobalt on *Escherichia coli* metabolism and metalloporphyrin formation. » Biometals 24 : 355-347. https://doi.org/10.1007/s10534-010-9400-7.

Makrides SC (1996) « Strategies for achieving high-level expression of genes in *Escherichia coli*. » Microbiology and Molecular Biology Reviews 60 : 512-538. https://doi.org/10.1128/mr.60.3.512-538.1996.

Malik NA and Ali A (2016) « Krafft temperature and thermodynamic study of interaction of glycine, diglycine, and triglycine with hexadecylpyridinium chloride and hexadecylpyridinium bromide: a conductometric approach. » Journal of Molecular Liquids 213 : 213-220. https://doi.org/10.1016/j.molliq.2015.11.017.

Mao X, Jiang R, Xiao W and Yu J (2015) « Use of surfactants for the remediation of contaminated soils: a review. » Journal of Hazardous Materials 285 : 419-435. https://doi.org/10.1016/j.jhazmat.2014.12.009.

Marbach and Bettenbrock K (2012) « *lac* operon induction in *Escherichia coli*: systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. » Journal of Biotechnology 157 (1) : 82-88. https://doi.org/10.1016/j.jbiotec.2011.10.009.

Mariani V, Biasini M, Barbato A and Schwede T (2013) « lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. » Bioinformatics 29 (21) : 2722-2728. https://doi.org/10.1093/bioinformatics/btt473.

Marinus MG and Morris NR (1973) « Isolation of deoxyribonucleic acid methylase mutants of *Escherichia coli* K-12. » Journal of Bacteriology 114 : 1143-1150. https://doi.org/10.1128/jb.114.3.1143-1150.1973.

Marschall L, Sagmeister P and Herwig C (2016) « Tunable recombinant protein expression in *E. coli*: enable for continuous processing? » Applied Microbiology and Biotechnology 100 : 5719-5728. https://doi.org/10.1007/s00253-016-7550-4.

Martin JR, Nus M, Gago JVS and Sanchez-Montero JM (2008) « Selective esterification of phtalic acids in two ionic liquids at high temperature using a thermostable lipase of *Bacillus thermocatenulatus*: a comparative study. » Journal of Molecular Catalysis B: Enzymatic 52-53 : 162-167. https://doi.org/10.1016/j.molcatb.2007.11.018.

Matsumoto J and Nagai S (1972) « Amidohydrolases for N-short and long chain fatty acyl-L-amino acids from *Mycobacteria*. » The Journal of Biochemistry 72 (2) : 269-279. https://doi.org/10.1093/oxfordjournals.jbchem.a129906.

Maugeri Z, Leitner W and de Maria PD (2013) « Chymotrypsin-catalyzed peptide synthesis in deep eutectic solvents. » European Journal of Organic Chemistry 2013 (20) : 4223-4228. https://doi.org/10.1002/ejoc.201300448.

May MS and Hattman S (1975) « Analysis of bacteriophage deoxyribonucleic acid sequences methylated by host- and R-factor-controlled enzymes. » Journal of Bacteriology 123 : 768-770. https://doi.org/10.1128/jb.123.2.768-770.1975.

Mendoza JA, Dulin P and Warren T (2000) « The lower hydrolysis of ATP by the stress protein GroEL is a major factor responsible for the diminished chaperonin activity at low temperature. » Cryobiology 41 (4) : 319-323. https://doi.org/10.1006/cryo.2000.2287.

Mirdita M, von den Driesch L, Galiez C, Martin MJ, Söding J and Steinegger M (2017) « Uniclust database of clustered and deeply annotated protein sequences and alignments. » Nucleic Acids Research 45 (D1) : D170-D176. https://doi.org/10.1093/nar/gkw1081.

Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S and Steinegger M (2022) « ColabFold: making protein folding accessible to all. » Nature Methods 19 : 679-682. https://doi.org/10.1038/s41592-022-01488-1.

Minasov G, Shuvalova L, Brunzelle JS, Collart FR and Anderson WF (2005) « Crystal structure of *Bacillus substilis* YXEP protein (APC1829), a dinuclear metal binding peptidase from M20 family. » Full wwPDB X-ray Structure Validation Report.

Moffatt BA and Studier FW (1987) « T7 lysozyme inhibits transcription by T7 RNA polymerase. » 49 (2) : 221-227. https://doi.org/10.1016/0092-8674(87)90563-0

Momany FA and Rone R (1992) « Validation of the general purpose QUANTA®3.2/CHARMm® force field. » Journal of Computational Chemistry 13 (7) : 888-900. https://doi.org/10.1002/jcc.540130714.

Montet D, Servat F, Pina M, Graille J, Galzi P, Arnaud A, Ledon L and Marcou L (1990) Enzymatic synthesis of N-ε-acyllysines. » Journal of the American Oil Chemists' Society 67 (11) : 771-774.

Mora-Pale JM, Perez-Munguia S, Gonzalez-Mejia JC, Dordick JS and Barzana E (2007) « The lipase-catalyzed hydrolysis of lutein diesters in non-aqueous media is favored at extremely low water activities. » Biotechnology and Bioengineering 98 (3) : 535-542. https://doi.org/10.1002/bit.21417.

Moult J, Fidelis K, Kryshtafovych A, Schwede T and Tramontano A (2017) « Critical assessment of methods of protein structure prediction (CASP) – Round XII. » Proteins: Structure, Function, and Bioinformatics 86 (S1) : 7-15. https://doi.org/10.1002/prot.25415.

Muegge I (2006) « PMF Scoring revisited. » Journal of Medicinal Chemistry 49 (20) : 5895-5902. https://doi.org/10.1021/jm050038s.

Muegge I and Martin YC (1999) « A general and fast scoring function for protein-ligand interactions: a simplified potential approach. » Journal of Medicinal Chemistry 42 (5) : 791-804. https://doi.org/10.1021/jm980536j.

Muhammed MT and Aki-Yalcin E (2018) « Homology modeling in drug discovery: overview, current applications, and future perspectives. » Chemical Biology & Drug Design 93 (1) : 12-20. https://doi.org/10.1111/cbdd.13388.

Murmu S, Pandey B, Das R, Sharma S and Kumari S (2022) « In-silico approaches of protein tertiary structure prediction. » Krishi Science 3 (12) : 2563.

Nagao A and Kito M (1989) « Synthesis of O-acyl-L-homoserine by lipase. » Journal of the American Oil Chemists' Society 66 : 710-713. https://doi.org/10.1007/BF02669958.

Nagtode VS, Cardoza C, Yasin HKA, Mali SN, Tambe SM, Roy P, Singh K, Goel A, Amin PD, Thorat BR, Cruz JN and Pratap AP (2023) « Green surfactants (biosurfactants): a petroleum-free substitute for sustainability – Comparaison, applications, market, and future prospects. » ACS Omega 8 (13) : 11674-11699. https://doi.org/10.1021/acsomega.3c00591.

Nakamura Y, Gojobori T and Ikemura T (2000) « Codon usage tabulated from international DNA sequences databases: status for the year 2000. » 28 (1) : 292-293. https://doi.org/10.1093/nar/28.1.292.

Nakanishi K, Imamura K, Imanaka H, Koreishi M and Nio N (2011) « N-ε-acyl-L-lysine-specific aminoacylase » US7.888.079B2. https://patents.google.com/patent/US7888079B2/en. US11/842,457.

Natsch A, Gfeller H, Gygax P, Schmid J and Acuna G (2003) « A specific bacterial aminoacylase cleaves odorant precursors secreted in the human axilla. » Enzyme Catalysis and Regulation 278 (8) : 5718-5727. https://doi.org/10.1074/jbc.M210142200.

Natsch A and Emter R (2020) « The specific biochemistry of human axilla odour formation viewed in an evolutionary context. » Philosophical Transactions of the Royal Society B 375 : 20190269. https://doi.org/10.1098/rstb.2019.0269.

Nguyen HC, Huong DTM, Juan H-Y, Su C-H and Chien C-C (2018) « Liquid lipase catalyzef esterification of oleic acid with methanol for biodiesel production in the presence of superabsorbent polymer: optimization by using response surface methodology. » Energies 11 (5) 1085. https://doi.org/10.3390/en11051085.

Nili M (2018) « Production hétérologue des amino-acylases de *Streptomyces ambofaciens*. » Rapport de stage M2 Mention BioMANE. Université de Lorraine.

Nocek BP, Gillner DM, Fan Y, Holz RC and Joachimiak A (2010) « Structural basis for catalysis by the mono- and di-metalated forms of the dapE-encoded N-succinyl-L,L-diaminopimelic acid desuccinylase. » Journal of Molecular Biology 397 (3) : 617-626. https://doi.org/10.1016/j.jmb.2010.01.062.

Nocek B, Starus A, Makowska-Grzyska M, Gutierrez B, Sanchez S, Jedrzejczak R, Mack JC, Olsen KW, Joachimiak A and Holz RC (2014) « The dimerization domain in dapE enzymes is required for catalysis. » Plos One 9 (5) : e93593. https://doi.org/10.1371/journal.pone.0093593.

Nocek B, Reidl C, Starus A, Heath T, Bienvenue D, Osipiuk J, Jedrzejczak A, Becker DP and Holz RC (2018) « Structural evidence of a major conformational change triggered by substrate binding in dapE enzymes: impact on the catalytic mechanism. » Biochemistry 57 (5) : 574-584. https://doi.org/10.1021/acs.biochem.7b01151.

Nogueira DR, Mitjans M, Infante MR and Vinardell MP (2011) « The role of counterions in the membrane-disruptive properties of pH-sensitive lysine-based surfactants. » Acta Biomaterialia 7 (7) : 2846-2856. https://doi.org/10.1016/j.actbio.2011.03.017.

O'Mahony DJ, Wang HY, McConnel DJ, Jia F, Zhou SW, Yin D and Qi S (1990) « The effect of phage T7 lysozyme on the production of biologically active porcine somatototropin in

Escherichia coli from a gene transcribed by T7 RNA polymerase. » Gene 91 (2) : 275-279. https://doi.org/10.1016/0378-1119(90)90100-6.

Obata S, Furukubo S, Kumagi I, Takahashi H and Miura K-I (1989a) « High-level expression in Streptomyces lividans 66 of a gene encoding *Streptomyces* subtilisin inhibitor from *Streptomyces albogriseolus* S-3253. » The Journal of Biochemistry 105 (3) : 372-376. https://doi.org/10.1093/oxfordjournals.jbchem.a122671.

Obata S, Taguchi S, Kumagai I and Miura K-I (1989b) « Molecular cloning and nucleotide sequence determination of gene encoding *Streptomyces* subtilisin inhibitor (SSI). » The Journal of Biochemistry 105 (3) : 367-371. https://doi.org/10.1093/oxfordjournals.jbchem.a122670.

Obata Y, Suzuki D and Takeoka S (2008) « Evaluation of cationic assemblies constructed with amino acid based lipids for plasmid DNA delivery. » Bioconjugate Chemistry 19 (5) : 1055-1063. https://doi.org/10.1021/bc700416u.

Oganesyan N, Ankoudinova I, Kim S-H and Kim R (2007) « Effect of osmotic stress and heat shock in recombinant protein overexpression and crystallization. » Protein Expression and Purification 52 (2) : 280-285. https://doi.org/10.1016/j.pep.2006.09.015.

Ohta A, Ozawa N, Nakashima S, Asakawa T and Miyagishi S (2003) « Krafft temperature and enthalpy of solution of N-acyl amino acid surfactants and their racemic modifications: effect of the amino acid residue. » Colloid and Polymer Science 281 : 363-369. https://doi.org/10.1007/s00396-002-0784-y.

Okumura N and Takao T (2017) « The zinc form of carnosine dipeptidase 2 (CN2) has dipeptidase activity but ist substrate is different from that of the manganese form. » Biochemical and Biophysical Resaerch Communications 494 (3-4) : 484-490. https://doi.org/10.1016/j.bbrc.2017.10.100.

Osipiuk J, Endres M, Becker DP and Joachimiak A (2021) « DapE enzyme from *Shigella flexneri*. » Full wwPDB X-ray Structure Validation Report.

Paggi JM, Pandit A and Dror RO (2024) « The art and science of molecular docking. » Annual Review of Biochemistry 93 : 389-410. https://doi.org/10.1146/annurev-biochem-030222-120000.

Palmer Ia and Wingfield PT (2012) « Preparation an extraction of insoluble (inclusion-body) proteins from *Escherichia coli*. » Current protocols in Protein Science 70 (1) : 6.3.1-6.3.20. https://doi.org/10.1002/0471140864.ps0603s70.

Paludo N, Alves JS, Altmann C, Ayub MAZ, Fernandez-Lafuente R and Rodrigues RC (2015) « The combined use of ultrasound and molecular sieves improves the synthesis of ethyl butyrate catalyzed by immobilized *Thermomyces lanuginosus* lipase. » Ultrasonics Sonochemistry 22 : 89-94. https://doi.org/10.1016/j.ultsonch.2014.05.004.

Pandya V, Ekka MK, Dutta RK and Kumaran S (2011) « Mass spectrometry assay for studying kinetic properties of dipeptidases: charactterization of human and yeast dipeptidases. » Analytical Biochemistry 418 (1) : 134-142. https://doi.org/10.1016/j.ab.2011.06.029.

Pandya V, Kaushik, Singh AK, Singh RP and Kumaran S (2014) « Crystal structure of human carnosinase-2 (CN2) in complex with inhibitor, bestatin at 2.25 A. » Full wwPDB X-ray Structure Validation Report.

Papp E and Csermely P (2006) « Chemicel chaperones: mechanisms of action and potential use. » Molecular Chaperones in Health and Disease 172 : 405-416. https://doi.org/10.1007/3-540-29717-0_16.

Park S and Kazlauskas RJ (2001) « Improved preparation and use of room-temperature ionic liquids in lipase-catalyzed enantio- and regio-selective acylations. » The Journal of Organic Chemistry 66 (25) : 8395-8401. https://doi.org/10.1021/jo015761e.

Patra AK, Gahlay GK, Reddy BVV, Gupta SK and Panda AK (2001) « Refolding, structural transition and spermatozoa-binding of recombinant bonnet monkey (*Macaca radiata*) zona pellucida glycoprotein-C expressed in *Escherichia coli*. » European Journal of Biochemistry 267 (24) : 7075-7081. https://doi.org/10.1046/j.1432-1327.2000.01808.x.

Patskovsky Y, Toro R, Freeman J, Miller S, Sauder JM, Raushel FM, Burley SK and Almo SC (2009) « Crystal structure of an uncharacterized metal-dependant hydrolase from *Pyrococcus horikoshii* ot3. » Full wwPDB X-ray Structure Validation Report.

Pena LC, Argarana MF, de Zan MM, Giorello A, Antuna S, Prieto C, Veaute CMI and Muller D (2017) « New amphiphilic amino acid derivatives for efficient DNA transfection *in vitro*. » Advances in Chemical Engineering and Science 7 (2) : 191-205. http://doi.org/10.4236/aces.2017.72014.

Perinelli DR, Casettari L, Cespi M, Fini F, Man DKW, Giorgioni G, Canala S, Lam JKW, Bonacucina G and Palmieri GF (2016) « Chemical-physical properties and cytotoxicity of N-decanoyl amino acid-based surfactants: effect of polar heads. » Colloids and Surfaces A: Physicochemical and Engineering Aspects 492 : 38-46. https://doi.org/10.1016/j.colsurfa.2015.12.009.

Phillips TA, van Bogelen RA and Neidhardt FC (1984) « *lon* Gene product of *Escherichia coli* is a heat-shock protein. » Journal of Bacteriology 159 : 283-287. https://doi.org/10.1128/jb.159.1.283-287.1984.

Piera E, Infante MR and Clapès P (2000) « Chemo-enzymatic synthesis of arginine-based gemini surfactants. » Biotechnology and Bioengineering 70 (3) : 323-331. https://doi.org/10.1002/1097-0290(20001105)70:3<323::AID-BIT9>3.0.CO;2-N.

Pinazo A, Manresa MA, Marques AM, Bustelo M, Espuny MJ and Pérez L (2016) « Amino acid-based surfactants : New antimicrobial agents. » Advances in Colloid and Interface Science 228 : 17-39. https://doi.org/10.1016/j.cis.2015.11.007.

Pittelkow S, Lindner H and Röhm K-H (1998) « Human and porcine aminoacylase 1 overproduced in a baculovirus expression vector system: evidence for structural and functional identity with enzymes isolated from kidney. » Protein Expression and Purification 12 (2) : 269-276. https://doi.org/10.1006/prep.1997.0816.

Ponder JW and Case DA (2003) « Force fields for protein simulations. » Advances in Protein Chemistry 66 : 27-85. https://doi.org/10.1016/S0065-3233(03)66002-X.

Pouresmaeil M and Azizi-Dargahlou S (2023) « Factors involved in heterologous expression of proteins in *E. coli* host. » Archives of Microbiology 205 : 212. https://doi.org/10.1007/s00203-023-03541-9.

Powell MJD (1977) « Restart procedures for the conjugate gradient method. » Mathematical Programming 12 : 241-254. https://doi.org/10.1007/BF01593790.

Prinz WA, Aslund F, Holmgren A and Beckwith J (1997) « The role of the thioredoxin and glutaredoxin pathways inreducing protein disulfide bonds in the *Escherichia coli* cytoplasm. » Cell Biology and Metabolism 272 (25) : 15661-15667. https://doi.org/10.1074/jbc.272.25.15661.

Preiss LC, Wagner M, Mastai Y, Landfester K and Munoz-Espi R (2016) « Amino-acid-based polymerizable surfactants for the synthesis of chiral nanoparticles. » Macromolecular Rapid Communications 37 (17) : 1421-1426. https://doi.org/10.1002/marc.201600210.

Puri NK, Cardamore M, Crivelli E and Traegert JC (1993) « Characterization of a truncated form of recombinant porcine growth hormone generated *in vitro* during solubilization of inclusion bodies. » Protein Expression and Purification 4 (2) : 164-175. https://doi.org/10.1006/prep.1993.1023.

Quax TEF, Claassens NJ, Söll D and van der Oost J (2015) « Codon bias as a means to fine-tune gene expression. » Molecular Cell 59 (2) : 149-161. https://doi.org/10.1016/j.molcel.2015.05.035.

Quiros LM, Aguirrezabalaga I, Olano C, Méndez C and Salas JA (1998) « Two glycosyltransferases and a glycosidase are involved in oleandomycin modification during its biosynthesis by *Streptomyces antibioticus*. » Molecular Microbiology 28 (6) : 1177-1185. https://doi.org/10.1046/j.1365-2958.1998.00880.x.

Radisky ES and Koshland Jr DE (2002) « A clogged gutter mechanism for protease inhibitors. » Proceedings of The National Academy of Sciences 99 (16) : 10316-10321. https://doi.org/10.1073/pnas.112332899

Ren H, Shi C, Song S, Zeng Q and Zhang Y (2016) « Synthesis of diacyl amino acid surfactant and evaluation of its potential for surfactant-polymer flooding. » Applied Petrochemical Research 6 : 59-63. https://doi.org/10.1007/s13203-014-0094-6.

Ritz D, Lim J, Reynolds M, Poole LB and Beckwith J (2001) « Conversion of a peroxiredoxin into a disulfide reductase by a triplet repeat expansion. » Science 294 (5540) : 158-160. https://doi.org/10.1126/science.1063143.

Robert X and Gouet P (2014) « Deciphering key features in protein structures with the new ENDscript server. » Nucleic Acids Research 42 (W1) : W320-W324. https://doi.org/10.1093/nar/gku316.

Robin X, Haas J Gumienny R, Smolinski A, Tauriello G and Schwede T (2021) « Continuous Automated Model EvaluatiOn (CAMEO) – Perspectives on the future of fully automated evaluation of structure prediction methods. » Proteins: Structure, Function and Bioinformatics 89 (12) : 1977-1986. https://doi.org/10.1002/prot.26213.

Rosano G and Ceccarelli EA (2014) « Recombinant protein expression in *Escherichia coli*: advances and challenges. » Frontiers in Microbiology 5 : 172. https://doi.org/10.3389/fmicb.2014.00172.

Rosenberg AH, Goldman E, Dunn JJ, Studier FW and Zubay G (1993) « Effects of consécutive AGG codons on translation in *Escherichia coli*, demonstrated with a versatile codon test system. » Journal of Bacteriology 175 : 716-722. https://doi.org/10.1128/jb.175.3.716-722.1993.

Rost B and Sander C (1996) « Bridging the protein sequence-structure gap by structure predictions. » Annual Review of Biophysics and Biomolecular Structure 25 (1) : 113-136.

Rouches MV, Xu Y, Corstes LBG and Lambert G (2022) « A plasmid system with tunable copy number. » Nature Communications 13 : 3908. https://doi.org/10.1038/s41467-022-31422-0.

Rowsell S, Pauptit RA, Tucker AD, Melton RG, Blow DM and Brick P (1997) « Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. » Research Article 5 (3) : 377-347. https://doi.org/10.1016/S0969-2126(97)00191-3.

Rudroff F, Mihovilovic MD, Gröger H, Snajdrova RS, Iding H and Bornscheuer UT (2018) « Opportunities and challenges for combining chemo- and bio-catalysis. » Nature Catalysis 1 : 12-22. https://doi.org/10.1038/s41929-017-0010-4.

Sachdev DP and Cameotra SS (2013) « Biosurfactants in agriculture. » Applied Microbiology and Biotechnology 97 : 1005-1016. https://doi.org/10.1007/s00253-012-4641-8.

Sahdev S, Khattar SK and Saini KS (2008) « Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. » Molecular ad Cellular Biochemistry 307 : 249-264. https://doi.org/10.1007/s11010-007-9603-6.

Sakanyan V, Desmarez L, Legrain C, Charlier D, Mett I, Kochikyan A, Savchenko A, Boyen A, Falmagne P and Pierard A (1993) « Gene cloning, sequence analysis, purification, and characterization of a thermostable aminoacylase from *Bacillus stearothermophilus*. » Applied and Environmental Microbiology 59 : 3878-3888. https://doi.org/10.1128/aem.59.11.3878-3888.1993.

Sali A and Blundell TL (1993) « Comparative protein modelling by satisfaction of spatial restraints. » Journal of Molecular Biology 234 (3,5) : 779-815. https://doi.org/10.1006/jmbi.1993.1626.

Samudrala R and Moult J (1998) « Determinants of side chain conformational preferences in protein structures. » Protein Engineering 11 (11) : 991-997. https://doi.org/10.1093/protein/11.11.991.

San-Miguel T, Pérez-Bermudez P and Gavidia I (2013) « Production of soluble eukaryotic recombinant proteins in *E. coli* is favoured in early log-pahse cultures induced at low temperature. » SpringerPlus 2 : 89. https://doi.org/10.1186/2193-1801-2-89.

Sari M-A, Jaouen M, Saroja NR and Artaud I (2007) « Influence of cobalt substitution on the activity of iron-type nitrile hydratase: are cobalt type nitrile hydratase regulated by arbon monoxide? » Journal of Inorganic Biochemistry 101 (4) : 614-622. https://doi.org/10.1016/j.jinorgbio.2006.12.005.

Saxena A, Sangwan RS and Mishra S (2013) « Fundamentals of homology modeling steps and comparaison among important bioinformatics tools: an overview. » Science International 1 (7) : 237-252. https://doi.org/10.17311/sciintl.2013.237.252.

Seidel HM, Pompliano DL and Knowles JR (1992) « Phosphonate biosynthesis : molecular cloning of the gene for phosphoenolpyruvate mutase from *Tetrahymena pyriformis* and overexpression of the gene product in *Escherichia coli*. » Biochemistry 31 (9) : 2598-2608. https://doi.org/10.1021/bi00124a021.

Selvam K, Senbagam D, Selvankumar T, Sudhakar C, Kamala-Kannan S, Senthilkumar B and Govarthanan M (2017) « Cellulase enzyme: Homology modeling, binding site identification and molecular docking. » Journal of Molecular Structure 1150 : 61-67. https://doi.org/10.1016/j.molstruc.2017.08.067.

Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, Qin C, Zidek A, Nelson AWR, Penedones H, Petersen S, Simonyan K, Crossan S, Kohli P, Jones DT, Silver D, Kavukcuoglu K and Hassabis D (2019) « Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of protein Structure Prediction (CASP13). » Proteins: Structure, Function, and Bioinformatics 87 (12) : 1141-1148. https://doi.org/10.1002/prot.25834.

Sevillano L, Vijgenboom E, van Wezel GP, Diaz M and Santamaria RI (2016) « New approaches to achieve high level enzyme production in *Streptomyces lividans*. » Microbioal Cell Factories 15 : 28. https://doi.org/10.1186/s12934-016-0425-7.

Shapir N, Pedersen C, Gil O, Strong L, Seffernick J, Sadowsky MJ and Wackett LP (2006) « TrzN from *Arthrobacter aurescens* TC1 is a zinc amidohydrolase. » Journal of Bacteriology 188 : 5859-5864. https://doi.org/10.1128/jb.00517-06.

Sheldon RA and Brady D (2018) « The limits to biocatalysis: pushing the envelope. » Chemical Communications 54 : 6088-6104. https://doi.org/10.1039/C8CC02463D.

Sheldon RA and Woodley JM (2018) « Role of biocatalysis in sustainable chemistry. » Chemical Reviews 118 (2) : 801-838. https://doi.org/10.1021/acs.chemrev.7b00203.

Sheldon RA and Brady D (2021) « Streamlining design, engineering, and applications of enzymes for sustainable biocatalysis. » ACS Sustainable Chemistry and Engineering 9 (24) : 8032-8052. https://doi.org/10.1021/acssuschemeng.1c01742.

Sheldon RA and Brady D (2022) « Green chemitry, biocatalysis, and the chemical industry of the future. » ChemSusChem 15 (9) : e202102628. https://doi.org/10.1002/cssc.202102628.

Shen M-Y and Sali A (2006) « Statistical potential for assessment and prediction of protein structure. » Protein Science 15 (1) : 2507-2524. https://doi.org/10.1110/ps.062416606.

Shi H, Hayes MT, Kirana C, Miller RJ, Keating JP and Stubbs RS (2013) « Overexpression of aminoacylase 1 is associated with colorectal cancer progression. » Human Pathology 44 (6) : 1089-1097. https://doi.org/10.1016/j.humpath.2012.09.015.

Shi T-T, Fang Z, Zeng W-B, Yang Z, He W and Guo K (2019) « Design, synthesis and properties investigation of N-α-acylation lysine based derivatives. » RSC Advances 9 : 7578-7593. https://doi.org/10.1039/C9RA00213H.

Shida T, Homma Y, Kamimura A and Misato T (1975) « Degradation of N-lauroyl-L-valine in soil and the effect of sunlight and ultraviolet ray on N-lauroyl-L-lysine. » Agricultural and Biological Chemistry 39 (4) : 879-883. https://doi.org/10.1080/00021369.1975.10861685.

Shilling PJ, Mirzadeh K, Cumming AJ, Widesheim M, Köck Z and Daley DO (2020) « Improved designs for pET expression plasmids increase protein production yield in *Escherichia coli*. » Communications Biology 3 : 214. https://doi.org/10.1038/s42003-020-0939-8

Simunovic M, Zagrovic B and Tomic S (2011) « Mechanism and thermodynamics of ligand binding to auxin amidohydrolase. » Journal of Molecular Recognition 24 (5) : 854-861. https://doi.org/10.1002/jmr.1128.

Singare PU and Mhatre JD (2012) « Cationic surfactants from arginine: synthesis and physicochemical properties. » American Journal of Chemistry 2 (4) : 186-190. https://doi.org/10.5923/j.chemistry.20120204.02.

Singh AK, Singh M, Pandya VK, Singh V, Mittal M and Kumaran S (2014) « Molecular basis of peptide recognition in metallopeptidase dug1p from *Saccharomyces cerevisiae*. » Biochemistry 53 (50) : 7870-7883. https://doi.org/10.1021/bi501263u.

Singh SM, Sharma A, Upadhyay AK, Singh A, Garg LC and Panda AK (2012) « Solubilization of inclusion body proteins using n-propanol and its refolding into bioactive form. » Protein Expression and Purification 81 (1) : 75-82. https://doi.org/10.1016/j.pep.2011.09.004.

Singha TK, Gulati P, Mohanty A, Khasa YP, Kappor RK and Kumar S (2017) « Efficient genetic approaches for improvement of plasmid based expression of recombinant protein in *Escherichia coli*: a review. » Process Biochemistry 55 : 17-31. https://doi.org/10.1016/j.procbio.2017.01.026.

Soo EL, Salleh AB, Basri M, Rahman RNZA and Kamaruddin K (2004) « Response surface methodological study on lipase-catalyzed synthesis of amino acid surfactants. » Process Biochemistry 39 (11) : 1511-1518. https://doi.org/10.1016/S0032-9592(03)00279-6.

Song H, Rao C, Deng Z, Yu Y and Naismith JH (2020) « The biosynthesis of the benzoxazole in nataxazole proceeds via an unstable ester and has synthetic utility. » Angewandte Chemie International Edition 59 (15) : 6054-6061. https://doi.org/10.1002/anie.201915685.

Sreenu M, Rao BVSK, Prasad RBN, Sujitha P and Chityala GK (2014a) « Synthesis, surface and biological properties of sodium N-acyl isoleucines. » European Journal of Lipid Science and Technology 116 (2) : 193-206. https://doi.org/10.1002/ejlt.201300189.

Sreenu M, Nayak RR, Prasad RBN and Sreedhar B (2014b) « Synthesis, surface and micellar properties of sodium N-oleoyl amoni acids. » Colloids and Surfaces A : Physicochemical and Engineering Aspects 449 : 74-81. https://doi.org/10.1016/j.colsurfa.2014.02.037.

Sreenu M, Nayak RR, Prasad RBN, Poornachandra Y and Kumar CG (2015) « Surface and antimicrobial properties of N-palmitoyl amino acid-based surfactants. » Journal of Dispersion Science and Technology 36 (6) : 765-771. https://doi.org/10.1080/01932691.2014.921626.

Srirangan K, Orr V, Akawi L, Westbrook A, Moo-Young M and Chou CP (2013) « Biotechnological advances on penicillin G acylase: pharmaceutical implications, unique

expression mechanism and production strategies. » Biotechnology Advances 31 (8) : 1319-1332. https://doi.org/10.1016/j.biotechadv.2013.05.006.

Stahl E, Schilz W, Schütz E and Willing E (1978) « A quick method for the microanalytical evaluation of the dissolving power of supercritical gases. » Angewandte Cheme 17 (10) : 731-738. https://doi.org/10.1002/anie.197807311.

Starus A, Nocek B, Bennett B, Larrabee JA, Shaw DL, Sae-LEE W, Russo MT, Gillner DM, Makowska-Grzyska M, Joachimiak A and Holz RC (2015) « Inhibition of the dapE-encoded N-succinyl-L,L-diaminopimelic acid desuccinylase from *Neisseria meningitidis* by L-captopril. » Biochemistry 54 (31) 4834-4844. https://doi.org/10.1021/acs.biochem.5b00475.

Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ and Söding J (2019) « HH-suite3 for fast remote homology detection and deep protein annotation. » BMC Bioinformatics 20 : 473. https://doi.org/10.1186/s12859-019-3019-7.

Stewart EJ, Aslund F and Beckwith J (1998) « Disulfide bond formation in the *Escherichia coli* cytoplasm: an *in vivo* reversal for the thioredoxins. » The EMBO Journal 17 : 5543-5550. https://doi.org/10.1093/emboj/17.19.5543.

Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J and Schwede T (2020) « QMEANDisco-distance constraints applied on model quality estimation. » Bioinformatics 36 (6) : 1765-1771. https://doi.org/10.1093/bioinformatics/btz828.

Studer G, Tauriello G, Bienert S, Biasini M, Johner N and Schwede T (2021) « ProMod3 – A versatile homology modelling toolbox. » Plos Computational Biology 17 (1) : e1008667. https://doi.org/10.1371/journal.pcbi.1008667.

Studier FW (1991) « Use of bacteriophage T7 lysozyme to improve an inductible T7 expression system. » Journal of Molecular Biology 219 (1) : 37-44. https://doi.org/10.1016/0022-2836(91)90855-Z.

Studier FW (2005) « Protein production by auto-induction in high-density shaking cultures. » Protein Expression and Purification 41 (1) : 207-234. https://doi.org/10.1016/j.pep.2005.01.016.

Studier FW and Moffatt BA (1986) « Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. » Journal of Molecular Biology 189 (1) : 113-130. https://doi.org/10.1016/0022-2836(86)90385-2.

Sumi T, Imasaki T, Aoki M, Sakai N, Nitta E, Shirouzu M and Nitta R (2018) « Structural insights into the altering function of CRMP2 by phosphorylation. » Cell Structure and Function 43 (1) : 15-23. https://doi.org/10.1247/csf.17025.

Tackie-Otoo BN and Mohammed MAA (2020) « Experimental investigation of the behaviour of a novel amino acid-based surfactant relevant to EOR application. » Journal of molecular liquids 316 : 113848. https://doi.org/10.1016/j.molliq.2020.113848.

Takakura J and Nozaki N (2019) « Method for producing N-ε-acyl-L-lysine » US10.266.857B2. https://patents.google.com/patent/US20180230502A1/en. US15/955,267.

Takakura Y and Asano Y (2019) « Purification, characterization, and gene cloning of a novel aminoacylase from *Burkholderia* sp. strain LP5_18B that efficiently catalyzes the synthesis of

N-lauroyl-L-amino acids. » Bioscienc, Biotechnology, and Biochemistry 83 (10) : 1964-1973. https://doi.org/10.1080/09168451.2019.1630255.

Takehara M, Yoshimura I, Takizawa K and Yoshida (1972) « Surface active N-acylglutamate : I. Preparation of long chain N-acylglutamic acid. » Journal of the American Oil Chemists Society 49 : 157-161. https://doi.org/10.1007/BF02633785.

Tang Z-Y, Yu J-Y, Zhou Q, He B, Wang Z-F and Zhou H-M (1995) « Secondary structure of holo- and apo-aminoacylase from prediction, circular dichroism, and FT-Raman spectroscopy. » The journal of Biochemistry 118 (4) : 706-709. https://doi.org/10.1093/oxfordjournals.jbchem.a124969.

Tanimoto K, Higashi N, Nishioka M, Ishikawa K and Taya M (2008) « Characterization of thermostable aminoacylase from hyperthermophilic archaeon *Pyrococcus horikoshii*. » The FEBS Journal 275 (6) : 1140-1149. https://doi.org/10.1111/j.1742-4658.2008.06274.x.

Tavano L, Infante MR, Riya MA, Pinazo A, Vinardell MP, Mitjans M, Manresa MA and Perez L (2013) « Role of aggregate size in the hemolytic and antimicrobial activity of colloidal solutions based on single and gemini surfactants from arginine. » Soft Matter 9 : 306-319. https://doi.org/10.1039/C2SM26670A.

Teufel M, Saudek V, Ledig J-P, Bernhardt A, Boularand S, Carreau A, Cairns NJ, Carter C, Cowley DJ, Duverger D, Ganzhorn AJ, Guenet C, Heintzelmann B, Laucher V, Sauvage C and Smirnova T (2003) « Sequence identification and characterization of human carnosinase and closely related non-specific dipeptidase. » Enzyme Catalysis and Regulation 278 (8) : 6521-6531. https://doi.org/10.1074/jbc.M209764200.

Thoden JB, Marti-Arbona R, Raushel FM and Holden HM (2003) « High-resolution X-ray structure of isoaspartyl dipeptidase from *Escherichia coli*. » Biochemistry 42 (17) : 4878-4882. https://doi.org/10.1021/bi034233p

Thompson JD, Higgins DG and Gibson TJ (1994) « CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. » Nucleic Acids Research 22 (12) : 4673-4680. https://doi.org/10.1093/nar/22.22.4673.

Torres-Bacete J, Hormigo D, Torres-Guzman R, Arroyo M, Castillon MP, Garcia JL, Acebal C and de la Mata I (2015) « Overexpression of penicillin V acylase from *Streptomyces lavendulae* and elucidation of its catalytics residues. » Applied and Environmental Microbiology 81 : 1225-1233. https://doi.org/10.1128/AEM.02352-14.

Tripathy DB, Mishra A, Clark J and Farmer T (2018) « Synthesis, chemistry, physicochemical properties and industrial applications of amino acid surfactants: a review. » Comptes Rendus Chimie 21 : 112-130. https://doi.org/10.1016/j.crci.2017.11.005.

Tronrud DE (1992) « Conjugate-direction minimization: an improved method for the refinement of macromolcules. » Acta Crystallographica Section A, Foundations and Advances 48 : 912-916. https://doi.org/10.1107/S0108767392005415.

Tufvesson P, Annerling A, Hatti-Kaul and Adlercreutz D (2007) « Solvent-free enzymatic synthesis of fatty alkanolamides. » Biotechnology and Bioengineering 97 (3) : 447-453. https://doi.org/10.1002/bit.21258.

Unver Y and Dagci I (2024) « *Komagataella phaffii* (*Pichia pastoris*) as a powerful yeast expression system for biologics production. » Frontiers in Bioscience (Elite Ed) 16 (2) : 19. https://doi.org/10.31083/j.fbe1602019.

Valivety R, Jauregi P, Gill I and Vulson E (1997) « Chemo-enzymatic synthesis of amino acid-based surfactants. » Journal of the American Oil Chemists' Society 74 (7) : 879-886. https://doi.org/10.1007/s11746-997-0232-8.

Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, Zidek A, Green T, Tunyasuvunakool K, Petersen S, Jumper J, Clancy E, Green R, Vora A, Lutfi M, Figurnov M, Cowie A, Hobbs N, Kohli P, Kleywegt G, Birney E, Hassabis D and Velanlar S (2021) « AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. » Nucleic Acids Research 50 (D1) : D439-D444. https://doi.org/10.1093/nar/gkab1061.

van Coster RN, Gerlo EA, Giardina TG, Engelke UF, Smet JE, de Praeter CM, Meersschaut VA, de Meirleir LJ, Seneca SH, Devreese B, Leroy JG, Herga S, Perrier JP, Wevers RA and Lissens W (2005) « Aminoacylase I deficiency: a novel inborn error of metabolism. » Biochemical and Biophysical Research Communications 338 (3) : 1322-1326. https://doi.org/10.1016/j.bbrc.2005.10.126.

van Rantwijk and Sheldon RA (2004) « Enantioselective acylation of chiral amines catalysed by serine hydrolases. » Tetrahedron 60 (3) : 501-519. https://doi.org/10.1016/j.tet.2003.10.018.

van Roosmalen MJE, Woerlee GF and Witkamp GJ (2004) « Amino acid based surfactants for dry-cleaning with high-pressure carbon dioxyde. » The Journal of Supercritical Fluids 32 (1-3) : 243-254. https://doi.org/10.1016/j.supflu.2004.01.005.

van Wezel GP, Krabben P, Traag BA, Keijser BJF, Kerste R, Vigenboom E, Heijnen JJ and Kraal B (2006) « Unlocking *Streptomyces* spp. for use as sustainable industrial production platforms by morphological enginnering. » Applied and Environmental Microbiology 72 : 5283-5288. https://doi.org/10.1128/AEM.00808-06.

Veiga-da-Cunha M, Chevalier N, Stoobant V, Vertommen D and van Schaftingen E (2014) « Metabolite proofreading in carnosine and homocarnosine synthesis – Molecular identification of PM20D2 as β-alanyl-lysine dipeptidase. » Enzymology 289 (28) : 19726-19736. https://doi.org/10.1074/jbc.M114.576579.

Velho RV, Medina LFC, Segalin J and Brandelli A (2011) « Production of lipopetides among *Bacillus* strains showing growth inhibition of phytopathogenis fungi. » Folia Microbiologica 56 : 297-303. https://doi.org/10.1007/s12223-011-0056-7.

Vera A, Gonzalez-Montalban N, Aris A and Villaverde A (2006) « The conformational quality of insoluble recombinant proteins is enhanced at low groth temperatures. » Biotechnology and Bioengineering 96 (6) : 1101-1106. https://doi.org/10.1002/bit.21218.

Vongerichten KF, Klein JR, Matern H and Plapp R (1994) « Cloning and nucleotide sequence analysis of pepV, a carnosinase gene from *Lactobacillus delbrueckii* subsp. *lactis* DSM 7270, and partial characterization of the enzyme. » Microbiology 140 (10) : 2591-2600. https://doi.org/10.1099/00221287-140-10-2591.

Wang C, Zhang P, Chen Z, Liu Y, Zhao L, Wang N and Xu B (2021) « Effects of fatty acyl chains on the interfacial rheogical behaviors of amino acid surfactants. » Journal of Molecular Liquids 325 : 114823. https://doi.org/10.1016/j.molliq.2020.114823.

Wang G, Li H and Hou Y (2022) « LncRNA MAGI2-AS3 inhibits tumor progression and angiogenesis by regulating ACY1 via interacting with transcription factor HEY1 in clear cell renal cell carcinoma. » Cancer Gene Therapy 29 : 585-596. https://doi.org/10.1038/s41417-021-00339-z.

Wang N, Yao K, Wang Y, Ti J, Tan J, Liu C, Zhang G, Wang C and Xu B (2019) « Green synthesis, characterization, and properties of acyl lysine, serine, threonine, and methionine derived from three types of natural oils. » Journal of surfactants and detergents 23 (2) : 239-250. https://doi.org/10.1002/jsde.12365.

Wang T-Y, Chen Y-C, Kao L-W, Chang C-Y, Wang Y-K, Liu Y-S, Feng J-M and Wu T-K (2008) « Expression and characterization of the biofilm-related and carnosine-hydrolyzing aminoacylhistidine dipeptidase from *Vibrio alginolyticus*. » The FEBS Journal 275 (20) : 5007-5020. https://doi.org/10.1111/j.1742-4658.2008.06635.x.

Wang W, Xi H, Bi Q, Hu Y, Zhang Y and Ni M (2013) « Cloning, expression and characterization of D-aminoacylase from *Achromobacter xylosoxidans* subsp. *denitrificans* ATCC 15173. » Microbiological Research 168 (6) : 360-366. https://doi.org/10.1016/j.micres.2013.01.002.

Wang Z, Zhang M, Lv X, Fan J, Zhang J, Sun J and Shen Y (2018) « GroEL/ES mediated the *in vivo* recovery of TRAIL inclusion bodies in *Escherichia coli*. » Scientific Reports 8 : 15766. https://doi.org/10.1038/s41598-018-34090-7.

Wardenga R, Hollmann F, Thum O and Bornscheuer U (2008) « Functional expression of porcine aminoacylase 1 in *E. coli* using a codon optimized synthetic gene and molecular chaperones. » Applied Microbiology and Biotechnology 81 : 721-729. https://doi.org/10.1007/s00253-008-1716-7.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R and Schwede T (2018) « SWISS-MODEL: homology modelling of protein structures and complexes. » Nucleic Acids Research 46 (W1) : W296-W303. https://doi.org/10.1093/nar/gky427.

Wei X, Li J, Xie H, Ling Q, Wang J, Lu D, Zhou L, Xu X and Zheng S (2014) « Proteomics-based identification of the tumor suppressor role of aminoacylase 1 in hepatocellular carcinoma. » Cancer Letters 351 (1) : 117-125. https://doi.org/10.1016/j.canlet.2014.05.001.

Weinstock MT, Hesek ED, Wilson CM and Gibson DG (2016) « *Vibrio natriegens* as a fast-growing host for molecular biology. » Brief Communication 13 : 849-851. https://doi.org/10.1038/nmeth.3970.

Weinstock MT, Wilson CM and Hesek ED (2019) « Genetically engineered *Vibrio* sp. and uses thereof. » US10377997B1. https://patents.google.com/patent/US10377997B1/en.

Weinstock MT, Wilson CM and Hesek ED (2022) « Genetically engineered *Vibrio* sp. and uses thereof. » US11447755B2. https://patents.google.com/patent/US11447755B2/en.

Werner T, Morris MB, Dastmalchi S and Church WB (2012) « Structural modelling and dynamics of proteins for insights into drug interactions. » Advanced Drug Delivery Reviews 64 (4) : 323-343. https://doi.org/10.1016/j.addr.2011.11.011.

Weuster-Botz D, Altenbach-Rehm J and Arnold M (2001) « Parallel substrate feeding and pH-control in shaking-flasks. » Biochemical Engineering Journal 7 (2) : 163-170. https://doi.org/10.1016/S1369-703X(00)00117-0.

Wu G, Robertson DH, Brooks III CL and Vieth M (2003) « Detailed analysis of grid-based molecular docking: a case study of CDOCKER – a CHARMm-based MD docking algorithm. » Journal of Computational Chemistry 24 (13) : 1549-1562. https://doi.org/10.1002/jcc.10306.

Xiang DF, Xu C, Kumaran D, Brown AC, Sauder JM, Burley SK, Swaminathan S and Raushel FM (2009) « Functionnal annotation of two new carboxypeptidase from the amidohydrolase superfamily of enzymes. » Biochemistry 48 (21) : 4567-4576. https://doi.org/10.1021/bi900453u.

Xiang DF, Patskovsky Y, Xu C, Fedorov AA, Fedorov EV, Sisco AA, Sauder JM, Burley SK, Almo SC and Raushel FM (2010) « Functionnal identification and structure determination of two novel prolidases from cog1228 in the amidohydrolase superfamily. » Biochemistry 49 (31) : 6791-6803. https://doi.org/10.1021/bi100897u.

Xu J and Zhang Y (2010) « How significant is a protein structure similarity with TM-score = 0.5? » Bioinformatics 26 (7) : 889-895. https://doi.org/10.1093/bioinformatics/btq066.

Xu J and Wang S (2019) « Analysis of distance-based protein structure prediction by deep learning in CASP13. » Proteins: Structure, Function, and Bioinformatics 87 (12) : 1069-1081. https://doi.org/10.1002/prot.25810.

Xu J, Dong F, Wu M, Tao R, Yang J, Wu M, Jiang Y, Yang S and Yang L (2021) « *Vibrio natriegens* as a pET-compatible expression host complementary to *Escherichia coli*. » Frontiers in Microbiology 12 : 627181. https://doi.org/10.3389/fmicb.2021.627181.

Yang J, Zhou Y, Yang B, Huang L, Liu C, Han F and Xu B (2020) « Comparative study on the syntheses and properties of three N-lauroyl aromatic amino acid surfactants. » Journal of Surfactants and Detergents 23 (4) : 687-695. https://doi.org/10.1002/jsde.12405.

Yea D, Lee S, Jo S, Yu H and Lim J (2018) « Preparation of environmentally friendly amino acid-based anionic surfactants and characterization of their interfacial properties for detergent products formulation. » Journal of Surfactants and Detergents 21 (4) : 541-552. https://doi.org/10.1002/jsde.12051.

Yen Y-C, Kammeyer AM, Jensen KC, Tirlangi J, Ghosh AK and Mesecar AD (2019) « Development of an efficient enzyme production and structure-based discovery platform for BACE1 in inhibitors. » Biochemistry 58 (44) : 4424-4435. https://doi.org/10.1021/acs.biochem.9b00714.

Yu H, Park K-M and Chang P-S (2020) « Lipase-catalyzed synthesis of lauroyl tripeptide-KHA with multi-functionalities: its surface-active, antibacterial, and antioxydant properties. » Food Chemistry 319 : 126533. https://doi.org/10.1016/j.foodchem.2020.126533.

Yuriev E, Holien J and Ramsland PA (2015) « Improvements, trends, and new ideas in molecular docking: 2012-2013 in review. » Journal of Molecular Recognition 28 (10) : 581-604. https://doi.org/10.1002/jmr.2471.

Zhao H, He C, Zhou Y, Yang J, Luo C and Xu B (2019) « Study on foaming properties of N-acyl amino acid surfactants: sodium N-acyl glycinate and sodium N-acyl phenylalaninate. » Colloids and Surfaces A: Physicochemical and Engineering Aspects 567 : 240-248. https://doi.org/10.1016/j.colsurfa.2019.01.073.

Zhang D, Koreishi M, Imanaka H, Imamura K and Nakanishi Z (2007) « Cloning and characterization of penicillin V acylase from *Streptomyces mobaraensis*. » Journal of Biotechnology 128 (4) : 788-800. https://doi.org/10.1016/j.jbiotec.2006.12.017.

Zhang D, Sun Y, Deng Q, Qi X, Sun H and Li Y (2016) « Study of the environmental responsiveness of amino acid-based surfactant sodium lauroylglutamate and its foam characteristics. » (2016) Colloids and Surfaces A : Physicochemical and Engineering Aspects 504 : 384-392. https://doi.org/10.1016/j.colsurfa.2016.05.097.

Zhang J, Sugisawa N, Felton KC, Fuse S and Lapkin AA (2024) « Multi-objective Bayesian optimisation using q-noisy expected hypervolume improvement (qNEHVI) for the Schotten-Baumann reaction. » Reaction Chemistry & Engineering 9 : 706-712. https://doi.org/10.1039/D3RE00502J.

Zhang T, Ding M, Tao N, Wang X and Zhong J (2020) « Effects of surfactants type and preparation pH on the droplets and emulsion forms of fish oil-loaded gelatin/surfactant-stabilized emulsions. » LWT – Food Science and Technology 117 : 108654. https://doi.org/10.1016/j.lwt.2019.108654.

Zhang X and Studier FW (1997) « Mechanism of inhibition of bacteriophage T7 RNA polymerase by T7 lysozyme. » Journal of Molecular Biology 269 (1) : 10-27. https://doi.org/10.1006/jmbi.1997.1016.

Zhang Y and Skolnick J (2004) « Scoring function for automated assessment of protein structure template quality. » Proteins: Structure, Function, and Bioinformatics 57 (4) : 702-710. https://doi.org/10.1002/prot.20264.

Zhang Z, Miao L, Xin X, Zhang J, Yang S, Miao M, Kong X and Jiao B (2014) « Underexpressed CNDP2 particpates in gastric cancer growth inhibition through activating the MAPK signaling pathway. » Molecular Medicine 20 : 17-28. https://doi.org/10.2119/molmed.2013.00102.

Zhang Z-X, Nong F-T, Wang Y-Z, Yan C-X, Gu Y, Song P and Sun X-M (2022) « Strategies for efficient production of recombinant proteins in *Escherichia coli*: alleviating the host burden and enhancing protein activity. » Microbial Cell Factories 21 : 191. https://doi.org/10.1186/s12934-022-01917-y.

Zhao H, Baker GA, Song Z, Olubajo O, Crittle T and Peters D (2008) « Designing enzyme-compatible ionic liquids that can dissolve carbohydrates. » Green Chemistry 10 : 696-705. https://doi.org/10.1039/B801489B.

Zhou X, Han Y, Lv Z, Tian X, Li H, Xie P and Zheng L (2017) « Simultaneously achieve soluble expression and biomimetic immobilization of *Candida antarctica* lipase B by

introducing polyamine tags. » Journal of Biotechnology 249 : 1-9. https://doi.org/10.1016/j.jbiotec.2017.03.015.

Annexes

Annexe 1 : Résultats SDS-PAGE complémentaires de SamAA

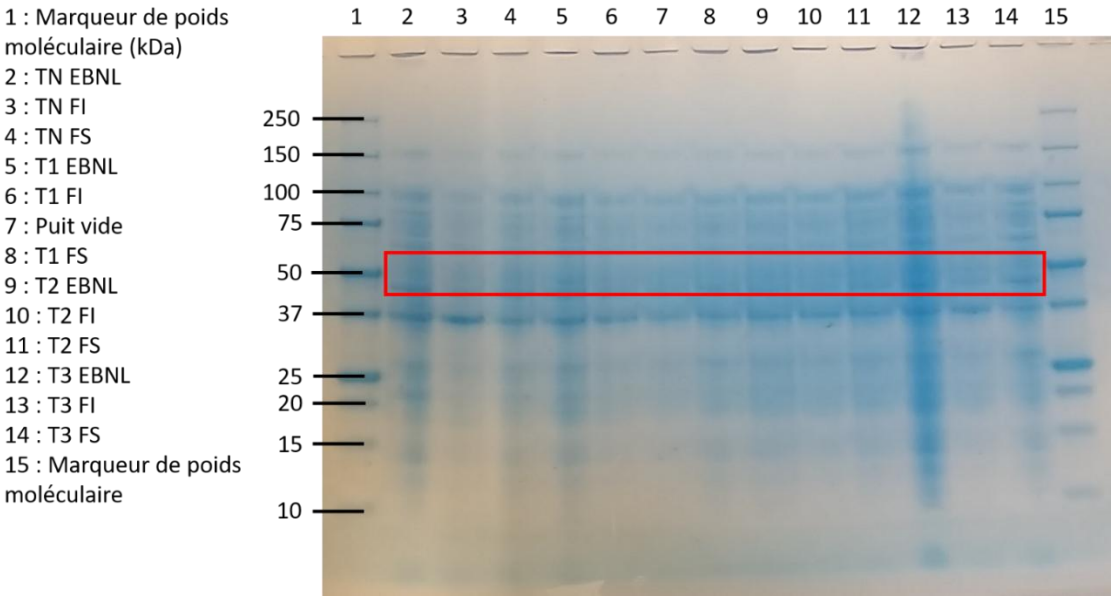


Figure 1 : Analyse par SDS-PAGE des protéines issues des cultures de 16 h d'*E. coli* Origami B (DE3) avec le plasmide pET-15b vide et pET-15b-SamAA en milieu Overnight. TN : témoin négatif (plasmide vide) ; T1, T2, T3 : erlen 1, 2, 3 avec le plasmide pET-15b-SamAA ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée. Le cadre rouge indique la position de SamAA exprimée.

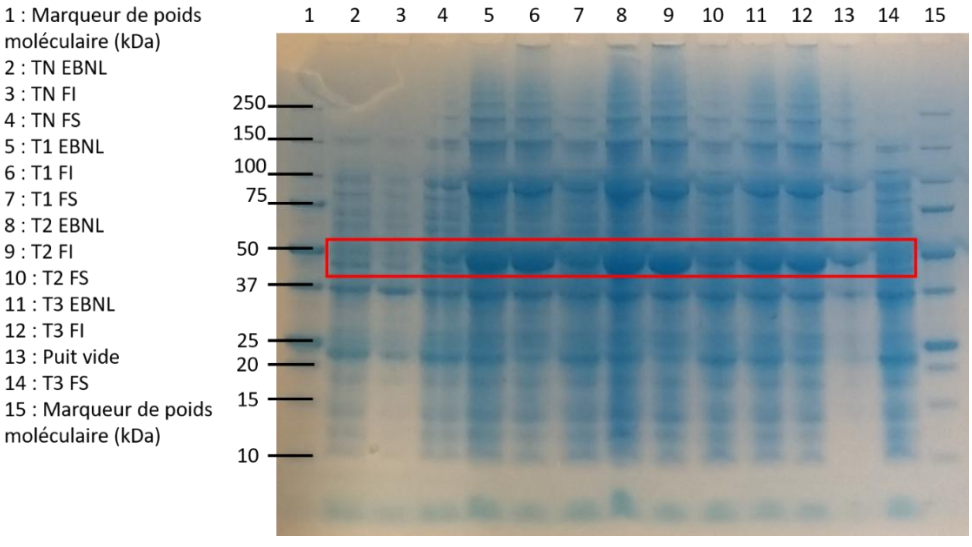


Figure 2 : Analyse par SDS-PAGE des protéines issues des cultures de 24 h d'*E. coli* Rosetta 2 pLysS (DE3) avec le plasmide pET-15b vide et pET-15b-SamAA en milieu OV. TN : témoin négatif (plasmide vide) ; T1, T2, T3 : erlen 1, 2, 3 avec le plasmide pET-15b-SamAA ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble. Le cadre rouge indique la position de SamAA exprimée.

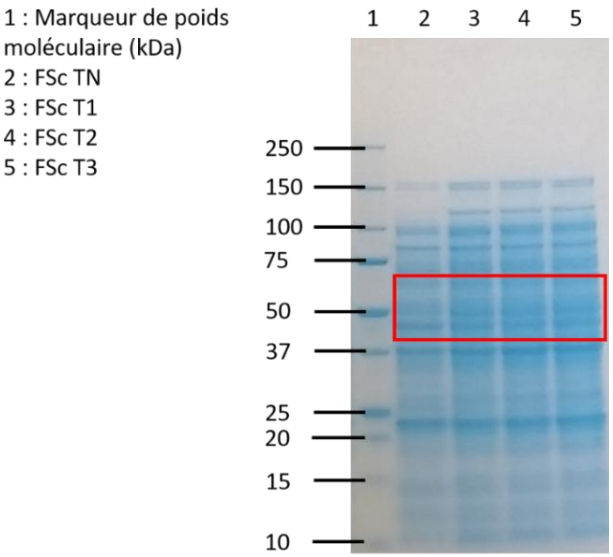


Figure 3 : Analyse par SDS-PAGE des protéines issues des fractions solubles concentrées des cultures de 24 h d'*E. coli* Rosetta 2 pLysS (DE3) avec le plasmide pET-15b vide et pET-15b-SamAA en milieu OV. TN : témoin négatif (plasmide vide) ; T1, T2, T3 : erlen 1, 2, 3 avec le plasmide pET-15b-SamAA ; FSc : fraction soluble concentrée. Le cadre rouge indique la position de SamAA exprimée.

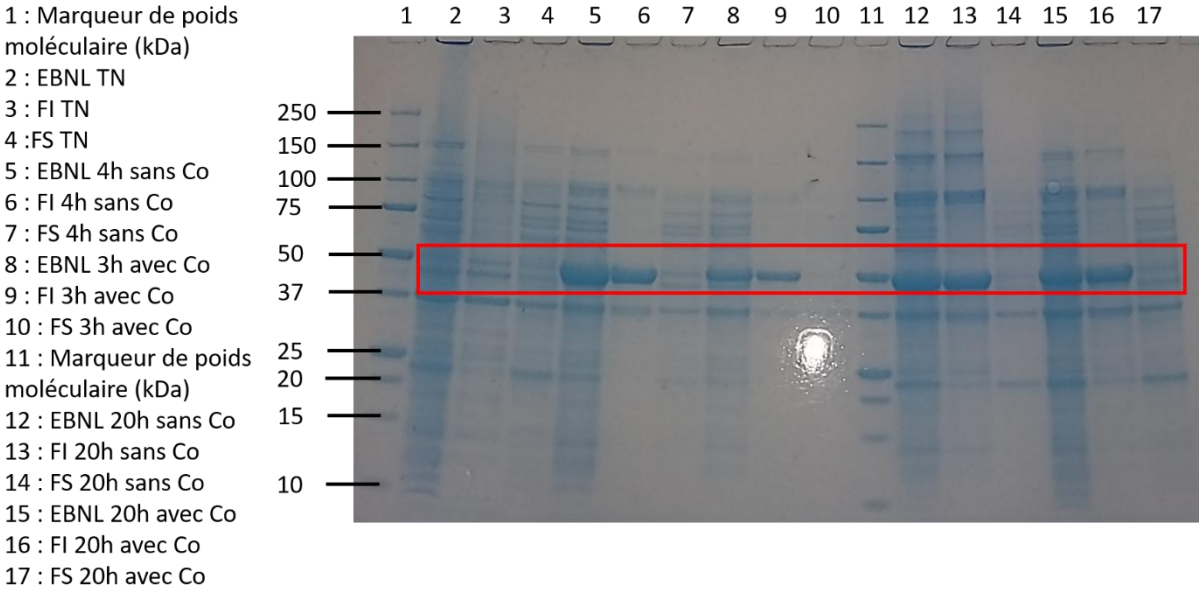


Figure 4 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamAA 6-His C-ter en milieu LB supplémenté ou non en cobalt (0,25 mM). TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

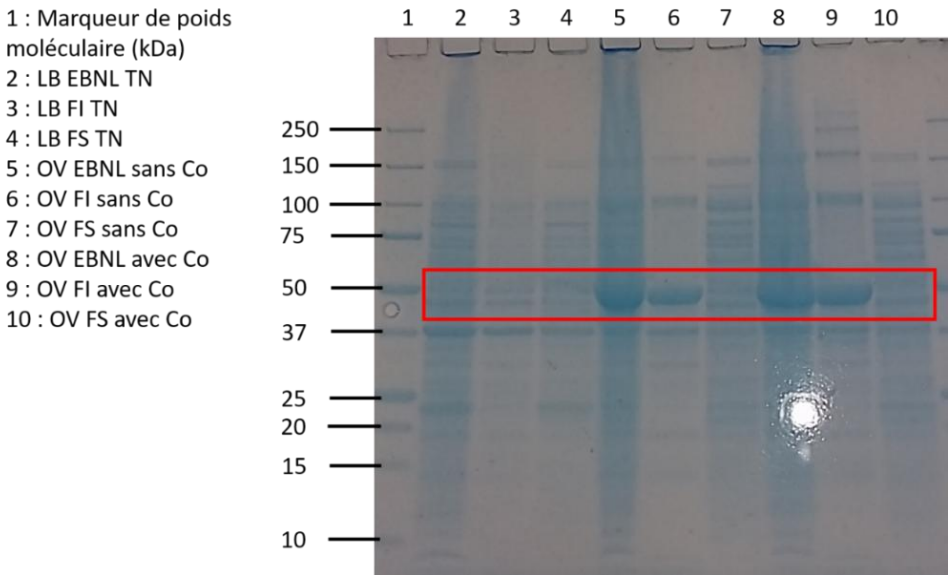


Figure 5 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu LB ou OV supplémenté ou non en cobalt (0,25 mM). OV : Overnight ; TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

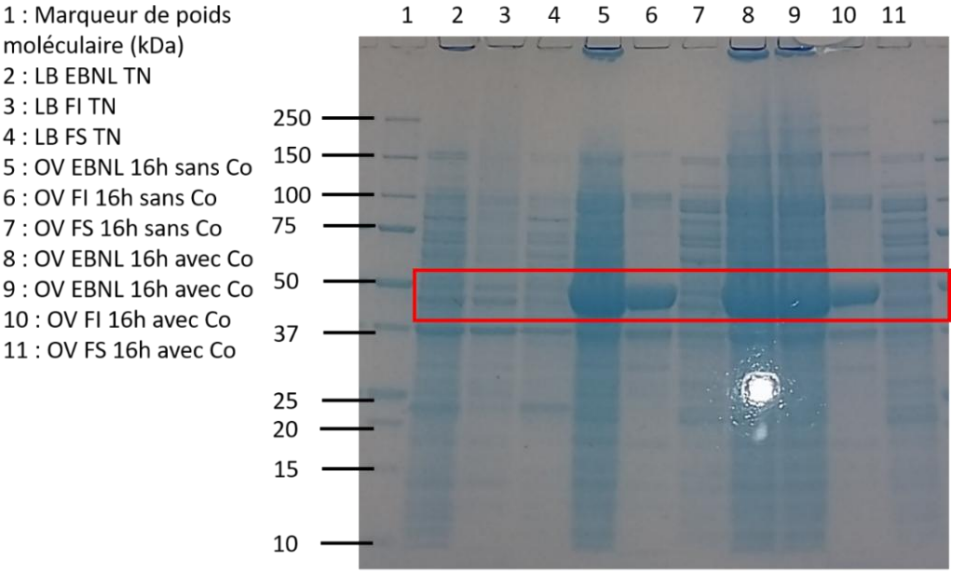


Figure 6 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamAA 6-His C-ter en milieu LB ou OV supplémenté ou non en cobalt (0,25 mM). OV : Overnight ; TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

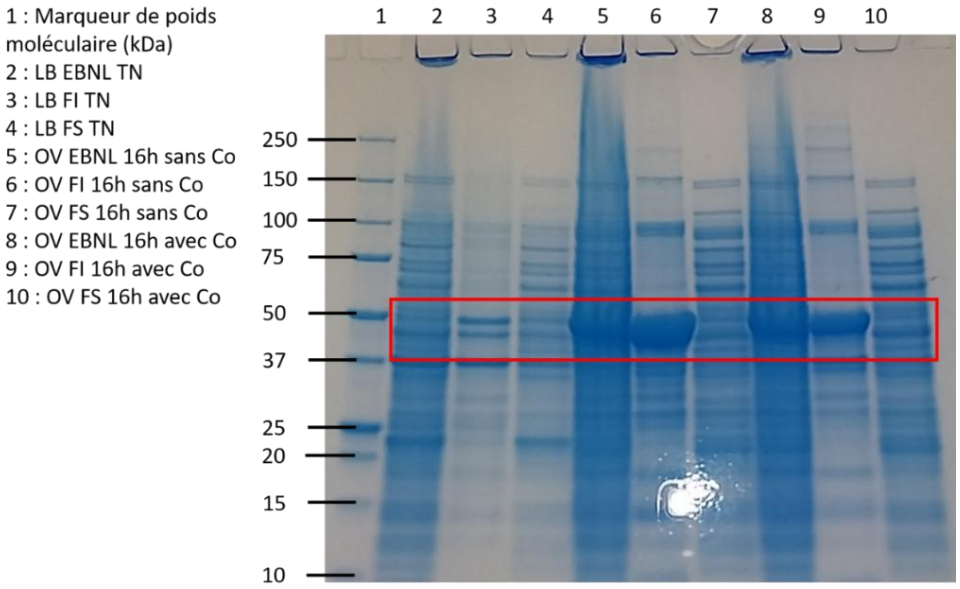


Figure 7 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu LB ou OV supplémenté ou non en cobalt (0,25 mM). OV : Overnight ; TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

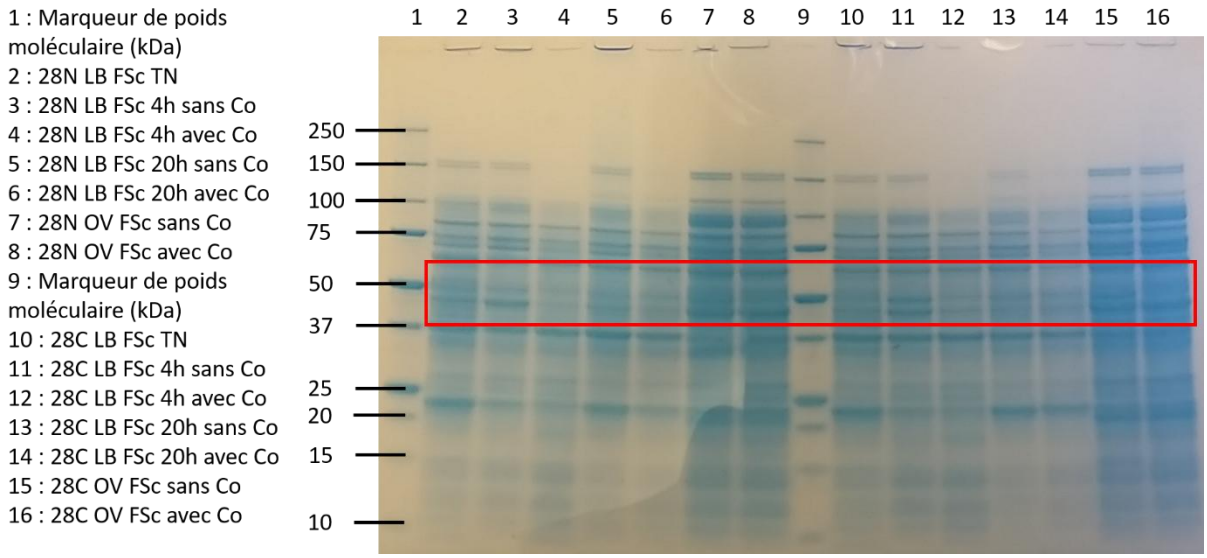


Figure 8 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter ou C-ter en milieu LB ou OV supplémenté ou non en cobalt (0,25 mM). FSc : fraction soluble concentrée ; TN : témoin négatif (sans induction) ; Co : cobalt ; OV : Overnight. Le cadre rouge indique la position de SamAA exprimée.

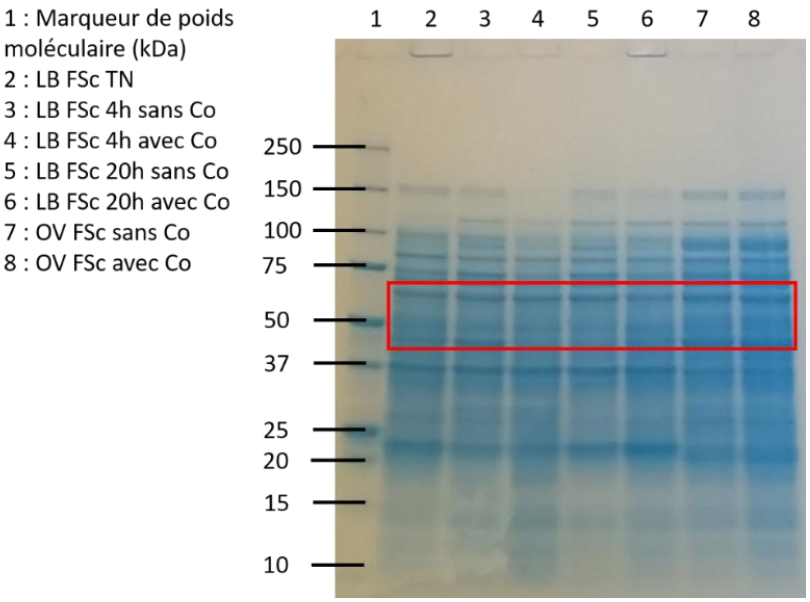


Figure 9 : Analyse par SDS-PAGE des protéines issues des fractions solubles concentrées des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-15b-SamAA 6-His N-ter en milieu LB ou OV supplémenté ou non en cobalt (0,25 mM). FSc : fraction soluble concentrée ; TN : témoin négatif (sans induction) ; Co : cobalt ; OV : Overnight. Le cadre rouge indique la position de SamAA exprimée.

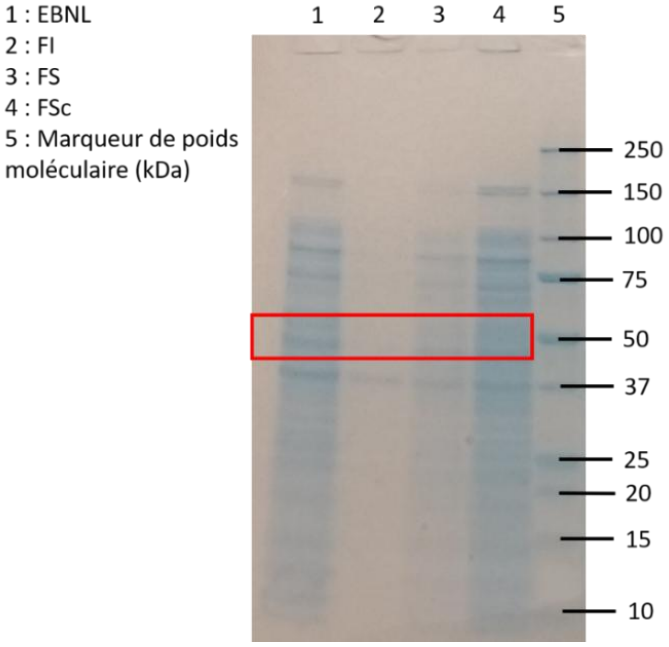


Figure 10 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-15b-SamAA 6-His N-ter en milieu OV supplémenté en cobalt (0,25 mM) à 20°C pendant 16h. EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble. Le cadre rouge indique la position de SamAA exprimée.

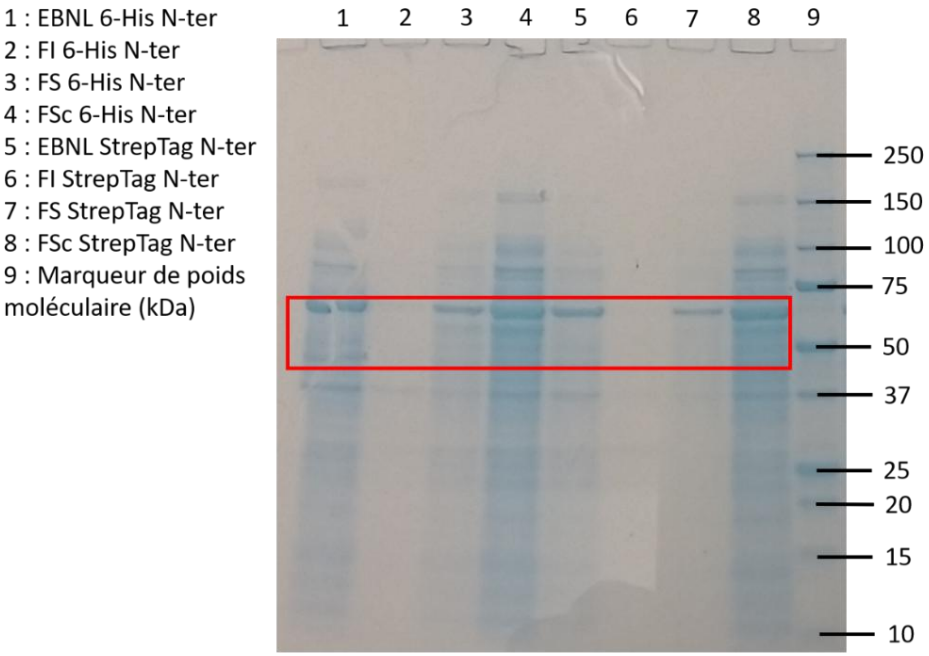


Figure 11 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Arctic Express (DE3) avec le plasmide pET-28a-SamAA 6-His ou StrepTag N-ter en milieu OV supplémenté en cobalt (0,25 mM). La culture a duré 5h à 30°C puis 16h à 11°C. EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée. Le cadre rouge indique la position de SamAA exprimée.

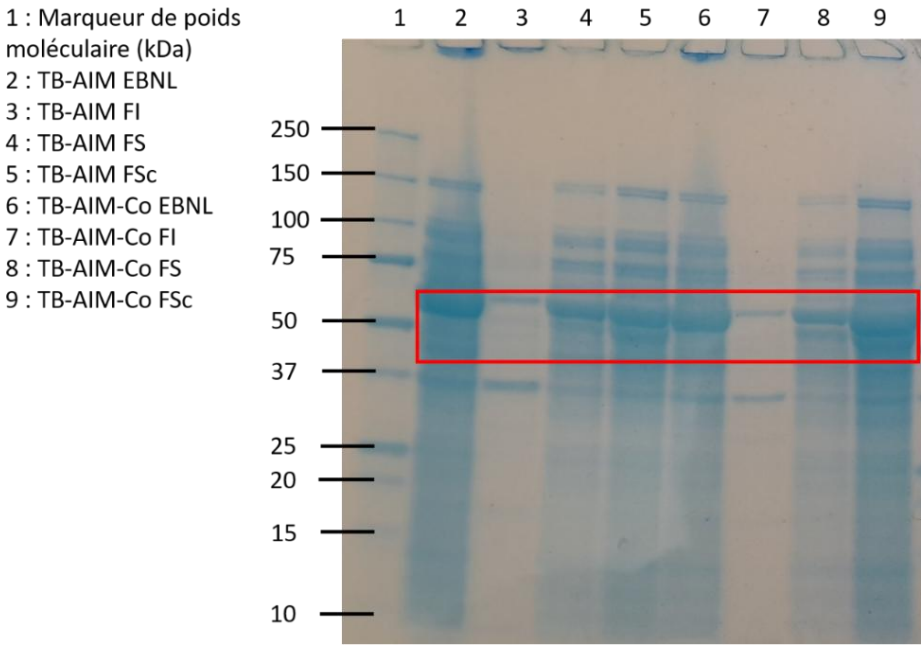


Figure 12 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Arctic Express (DE3) avec le plasmide pET-28a-SamAA StrepTag N-ter en milieu TB-AIM supplémenté ou non en cobalt (0,25 mM). EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée. Le cadre rouge indique la position de SamAA exprimée.

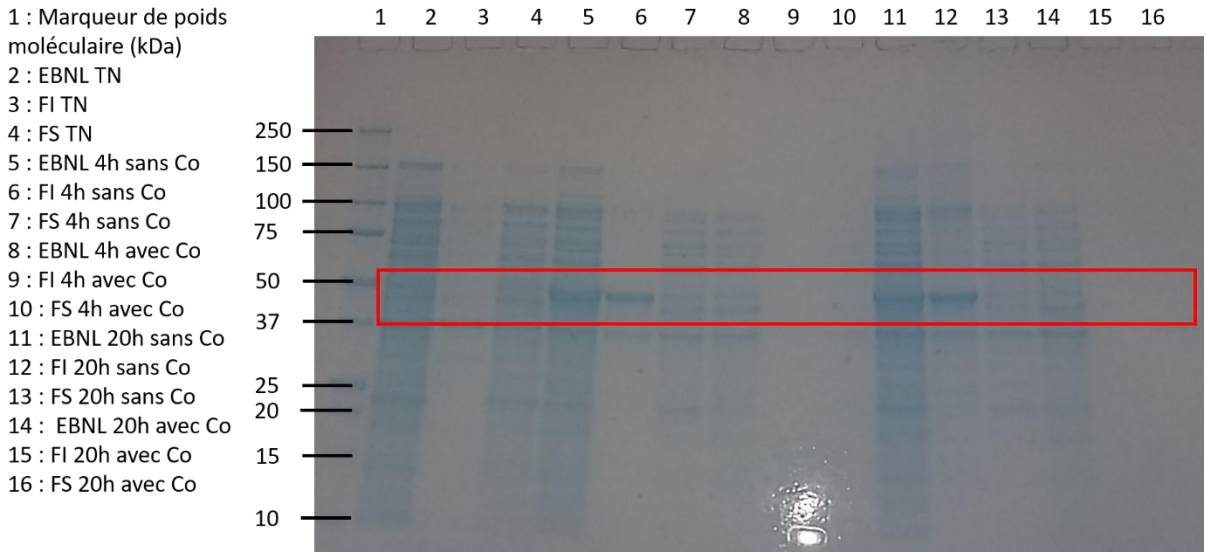


Figure 13 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold pLysS (DE3) avec le plasmide pET-28a-SamAA 6-His C-ter en milieu LB supplémenté ou non en cobalt (0,25 mM). TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

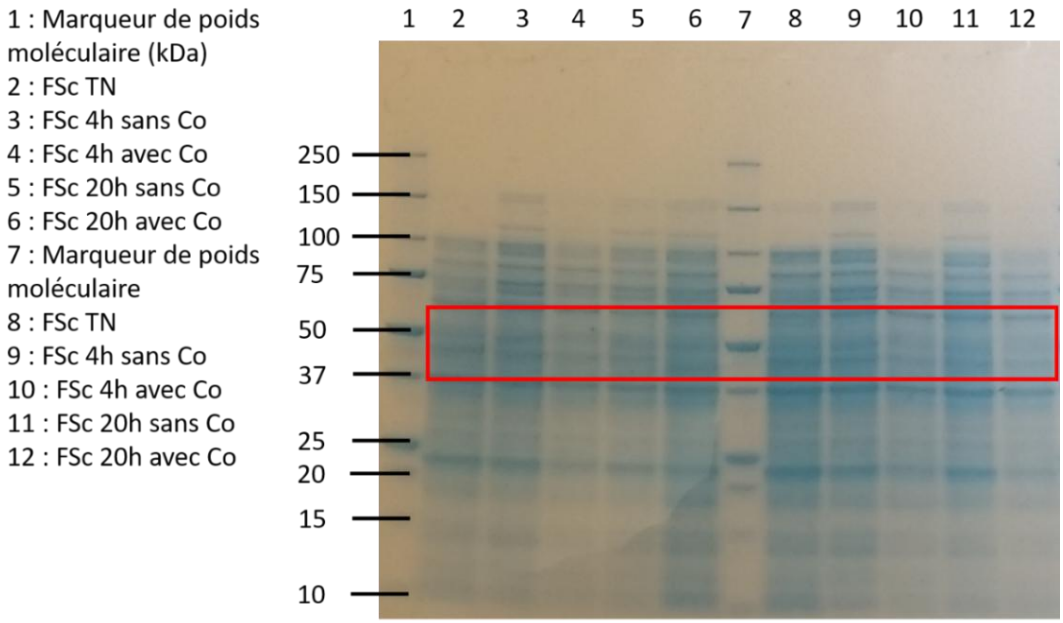


Figure 14 : Analyse par SDS-PAGE des protéines issues des fractions solubles concentrées des cultures d'*E. coli* BL21 Gold pLysS (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter ou C-ter en milieu LB supplémenté ou non en cobalt (0,25 mM). TN : témoin négatif (sans induction) ; FSc : fraction soluble concentrée ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

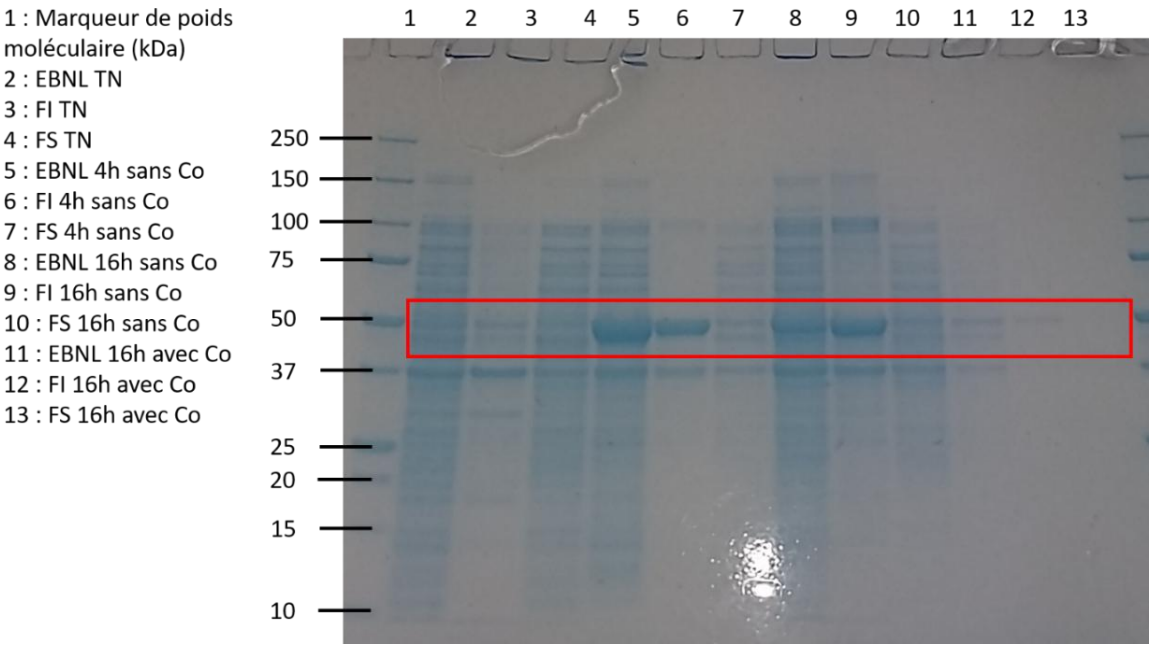


Figure 15 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His C-ter en milieu LB supplémenté ou non en cobalt (0,25 mM). TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

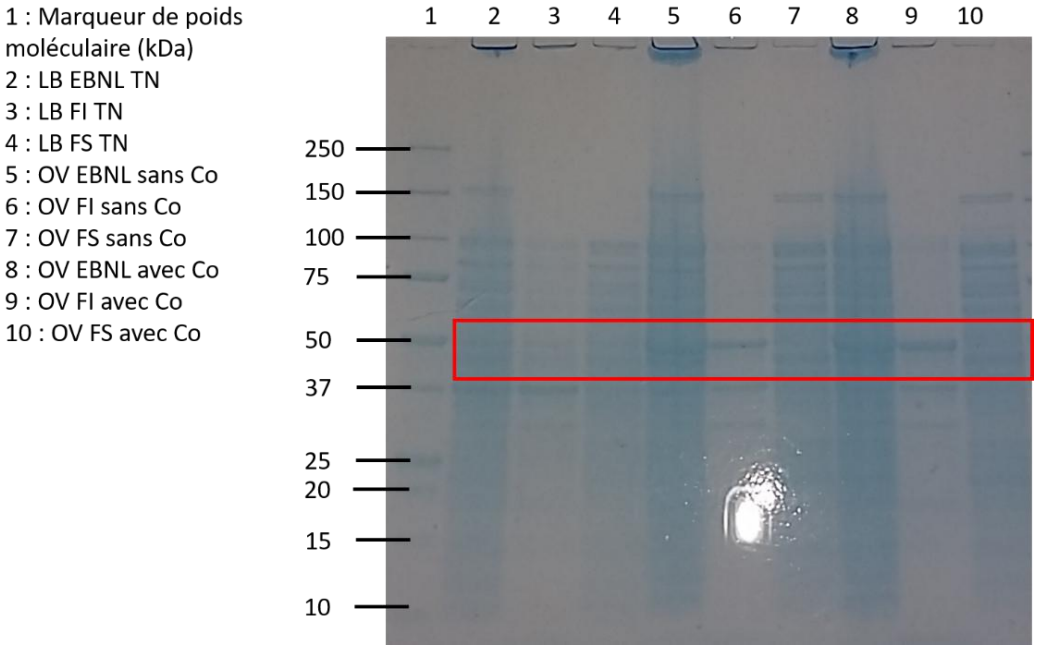


Figure 16 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu OV supplémenté ou non en cobalt (0,25 mM), pendant 16h. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

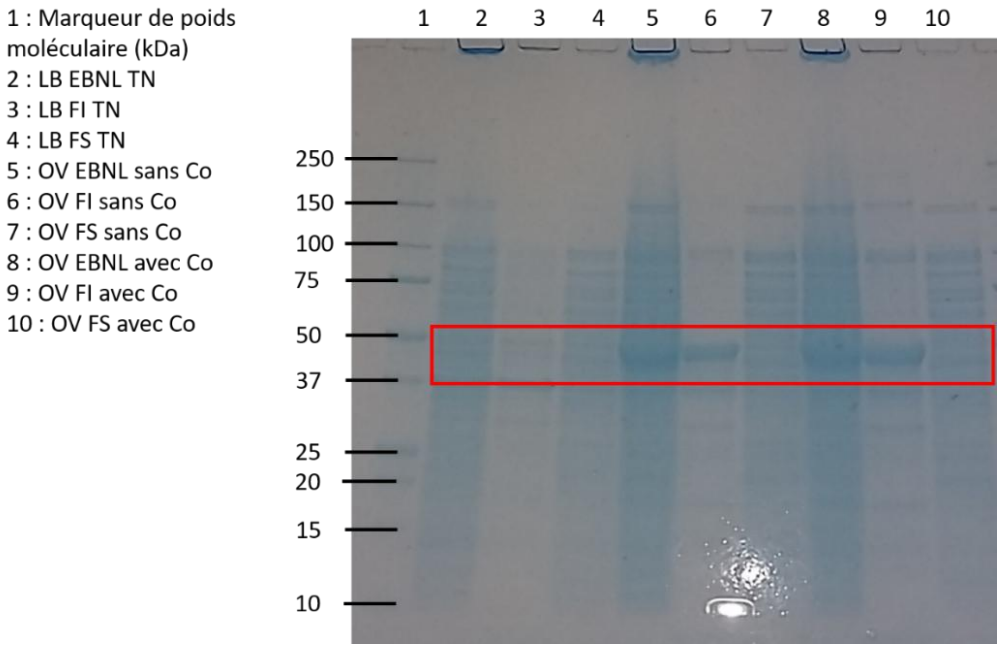


Figure 17 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His C-ter en milieu OV supplémenté ou non en cobalt (0,25 mM), pendant 16h. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

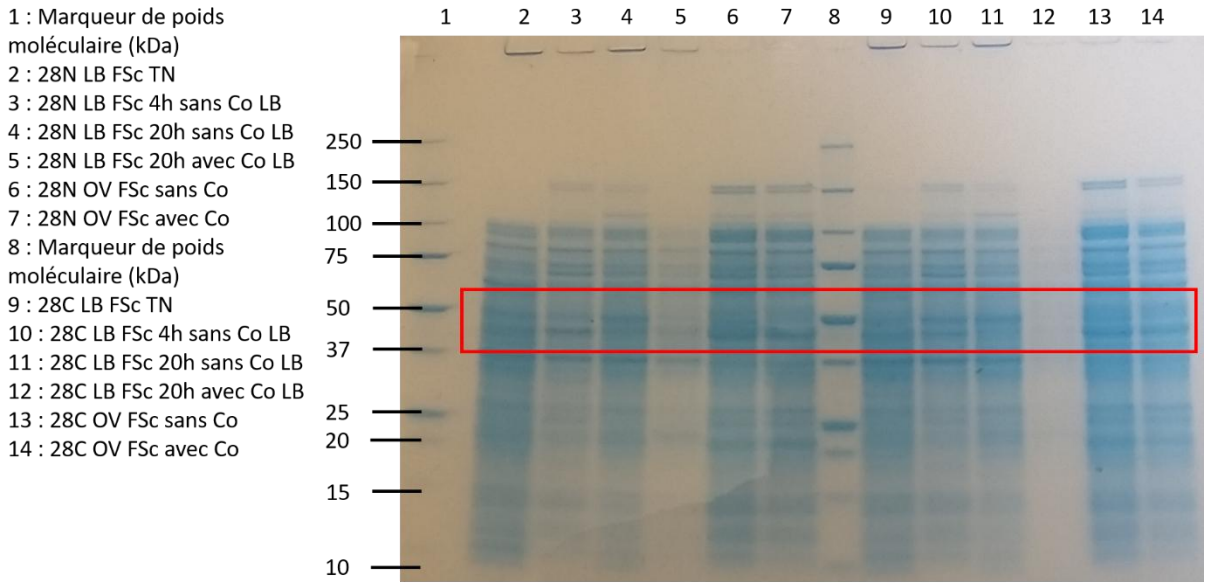


Figure 18 : Analyse par SDS-PAGE des protéines issues des fractions solubles concentrées des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter ou C-ter en milieu LB ou OV supplémenté ou non en cobalt (0,25 mM). FSc : fraction soluble concentrée ; TN : témoin négatif (sans induction) ; Co : cobalt ; OV : Overnight. Le cadre rouge indique la position de SamAA exprimée.

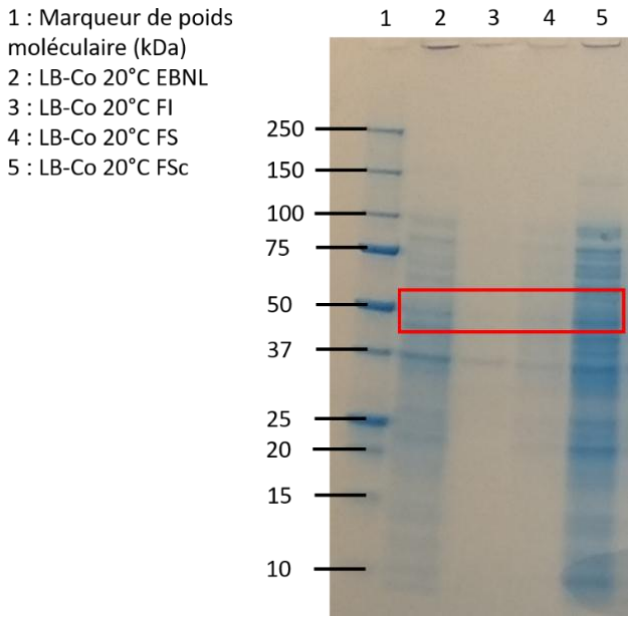


Figure 19 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu LB supplémenté en cobalt (0,25 mM), pendant 24h. EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; Co : cobalt. Le cadre rouge indique la position de SamAA exprimée.

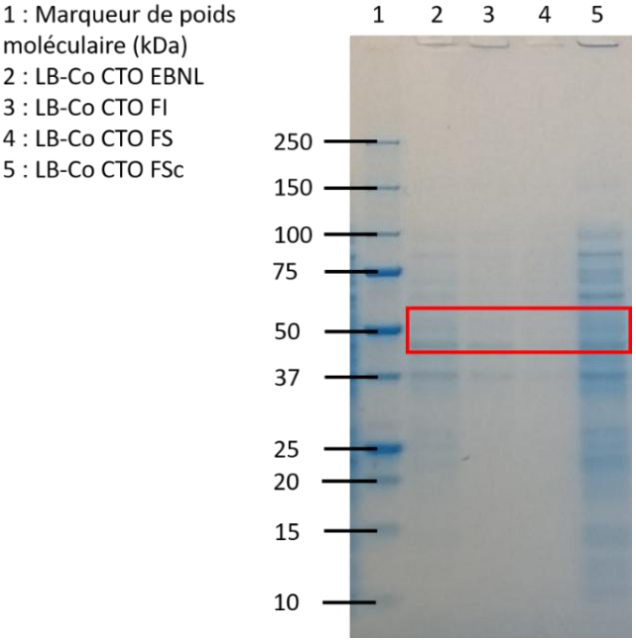


Figure 20 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu LB supplémenté en cobalt (0,25 mM), pendant 24h. EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; Co : cobalt ; CTO : choc thermique et osmotique. Le cadre rouge indique la position de SamAA exprimée.

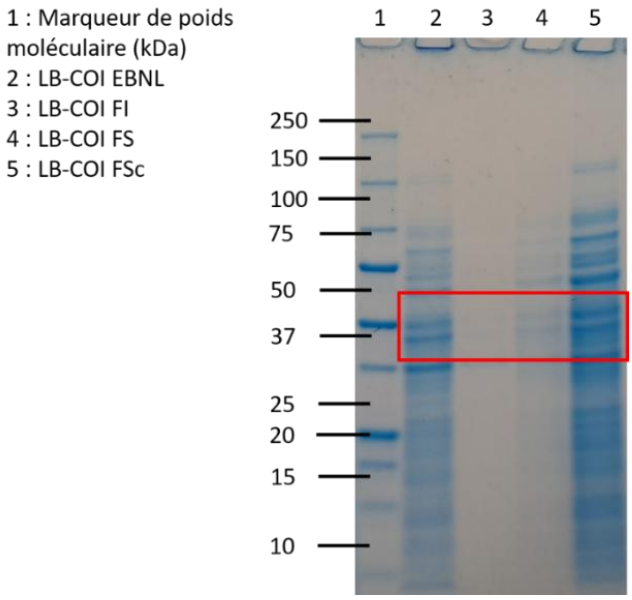


Figure 21 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamAA 6-His N-ter en milieu LB pendant 24h. EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée. Le cadre rouge indique la position de SamAA exprimée.

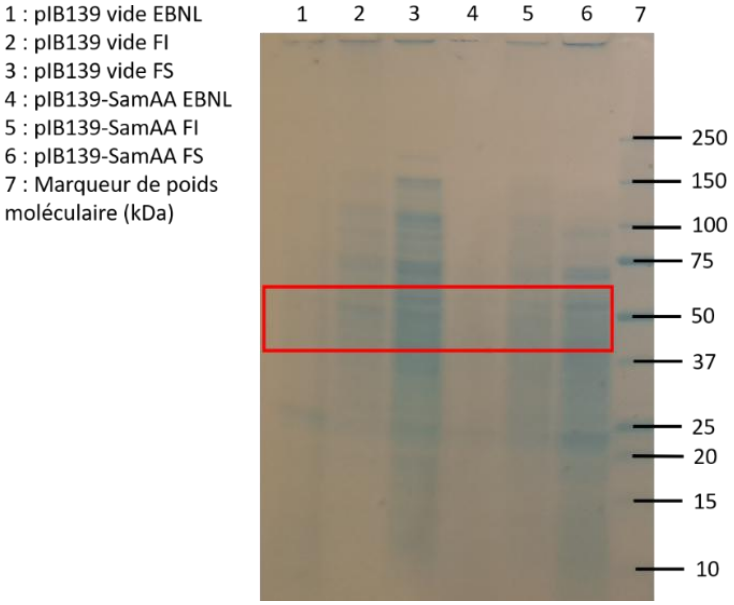


Figure 22 : Analyse par SDS-PAGE des protéines issues des cultures de *S. lividans* avec le plasmide pIB139 vide ou avec SamAA StrepTag N-ter. EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble. Le cadre rouge indique la position de SamAA exprimée.

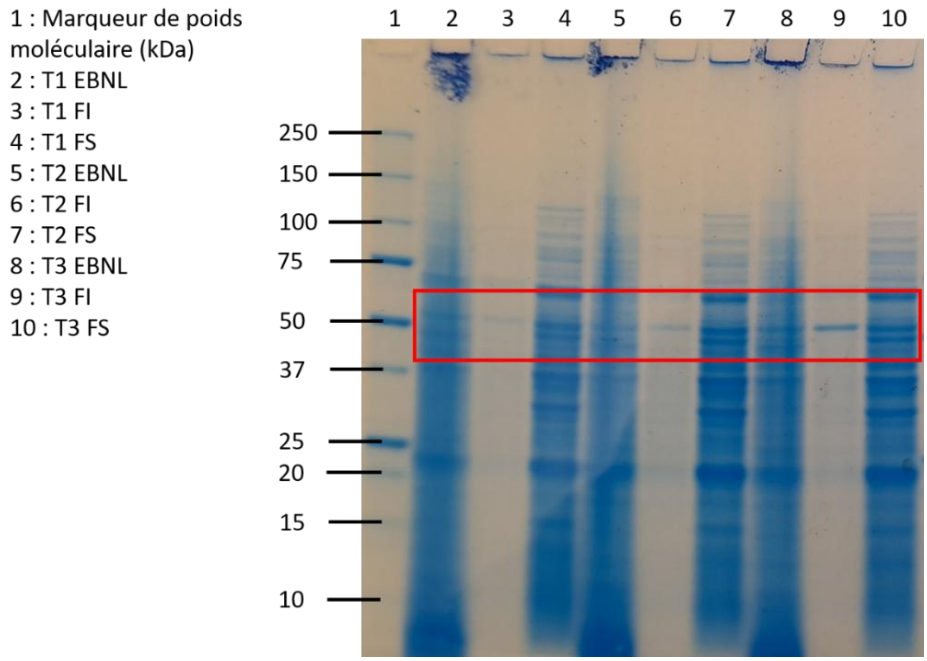


Figure 23 : Analyse par SDS-PAGE des protéines issues des cultures de *S. lividans* avec le plasmide pIB139 vide StrepTag N-ter. T1, T2, T3 : erlenmeyer 1, 2, 3 ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble. Le cadre rouge indique la position de SamAA exprimée.

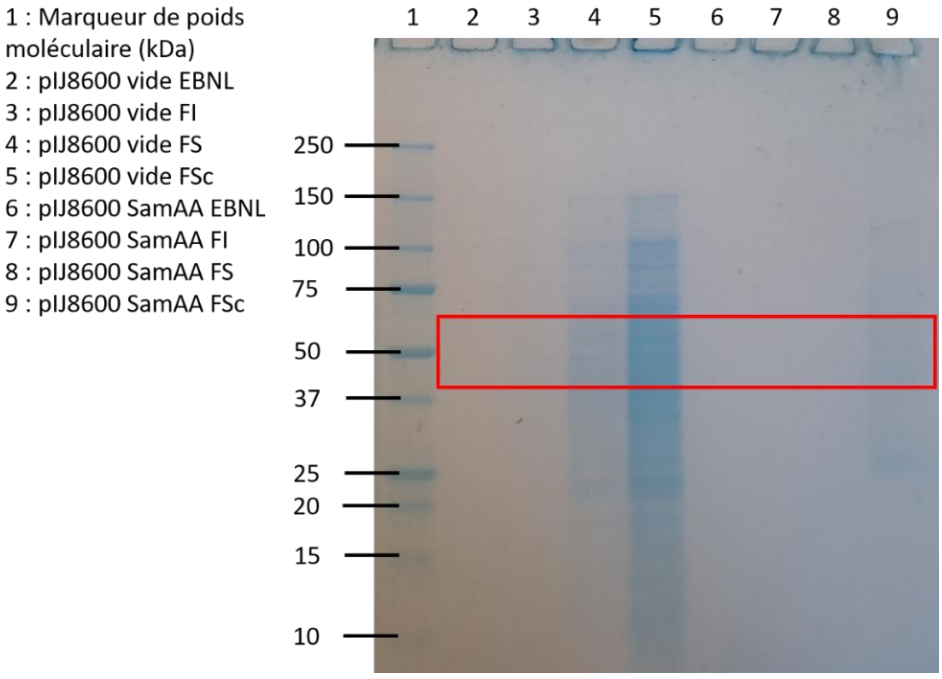


Figure 24 : Analyse par SDS-PAGE des protéines issues des cultures de *S. lividans* avec le plasmide pIJ8600 vide ou avec SamAA StrepTag N-ter. EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble. Le cadre rouge indique la position de SamAA exprimée.

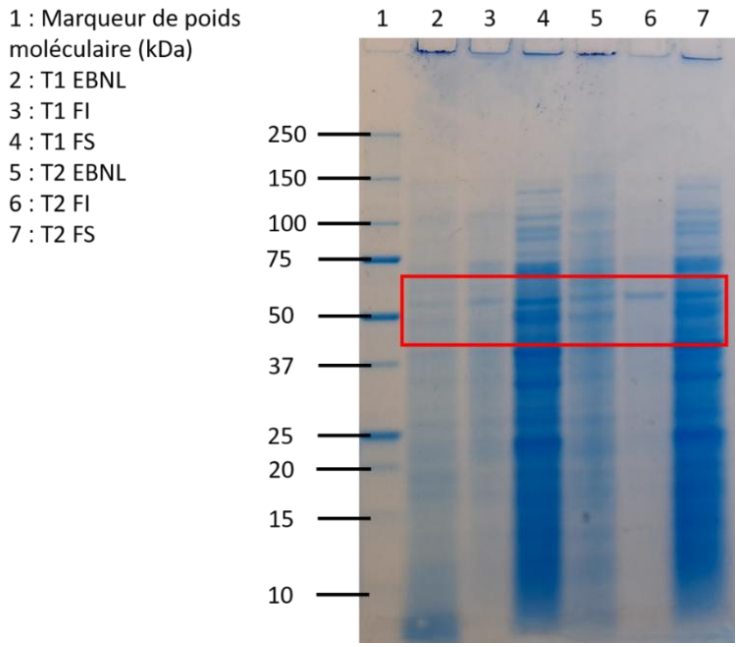


Figure 25 : Analyse par SDS-PAGE des protéines issues des cultures de *S. lividans* avec le plasmide pIJ8600 vide StrepTag N-ter. T1 et T2 : erlenmeyer 1 et 2 ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble. Le cadre rouge indique la position de SamAA exprimée.

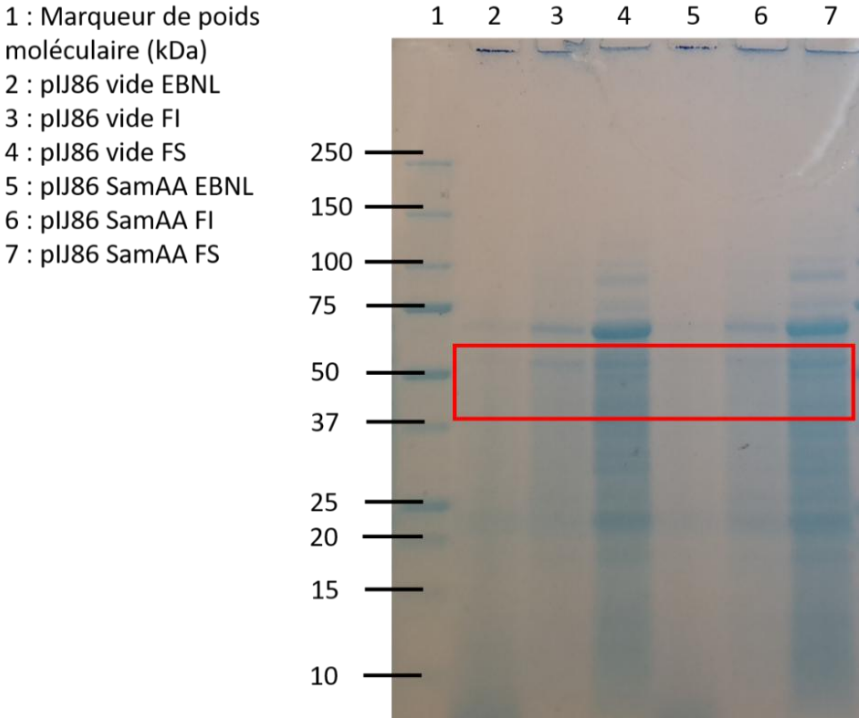


Figure 26 : Analyse par SDS-PAGE des protéines issues des cultures de *S. lividans* avec le plasmide pIJ86 vide ou avec SamAA StrepTag N-ter. EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble. Le cadre rouge indique la position de SamAA exprimée.

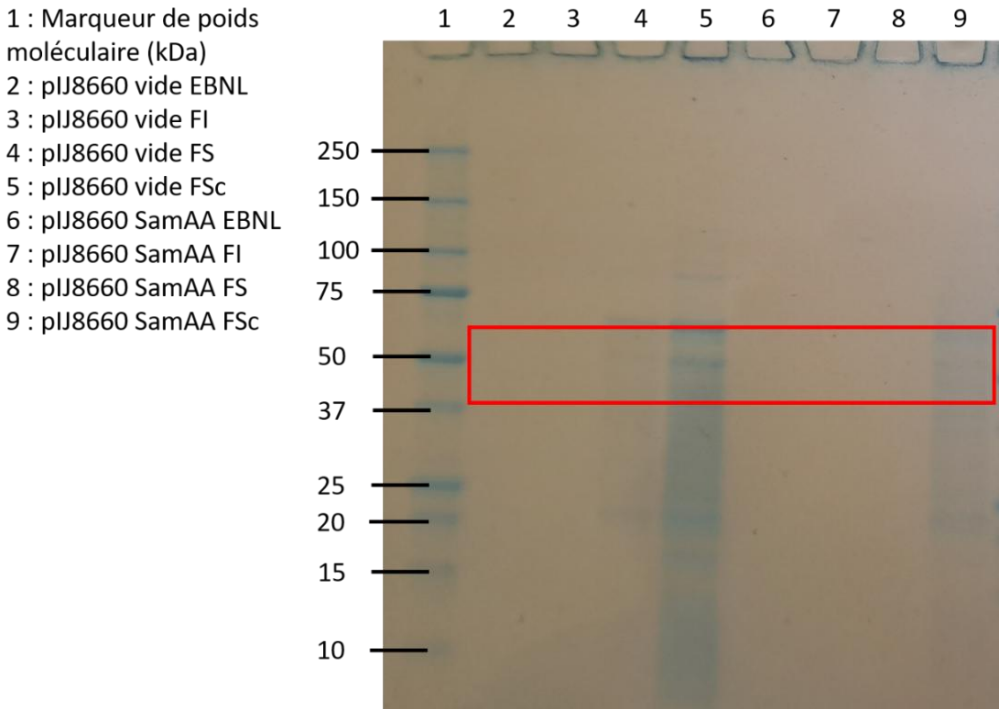


Figure 27 : Analyse par SDS-PAGE des protéines issues des cultures de *S. lividans* avec le plasmide pIJ8660 vide ou avec SamAA StrepTag N-ter. EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble. Le cadre rouge indique la position de SamAA exprimée.

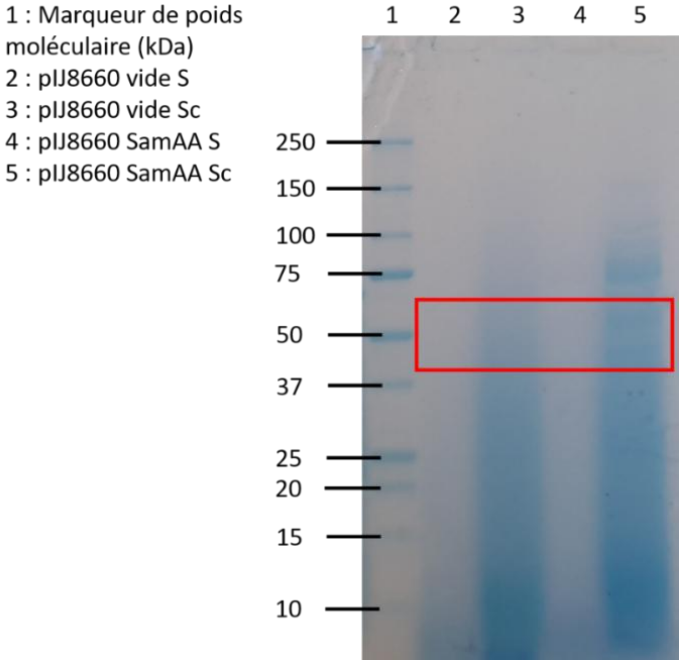


Figure 28 : Analyse par SDS-PAGE des protéines issues des cultures de *S. lividans* avec le plasmide pIJ8660 vide ou avec SamAA StrepTag N-ter. S : surnageant de culture ; Sc : surnageant de culture concentré. Le cadre rouge indique la position de SamAA exprimée.

Annexe 2 : Résultats SDS-PAGE et CCM complémentaires de SamELA

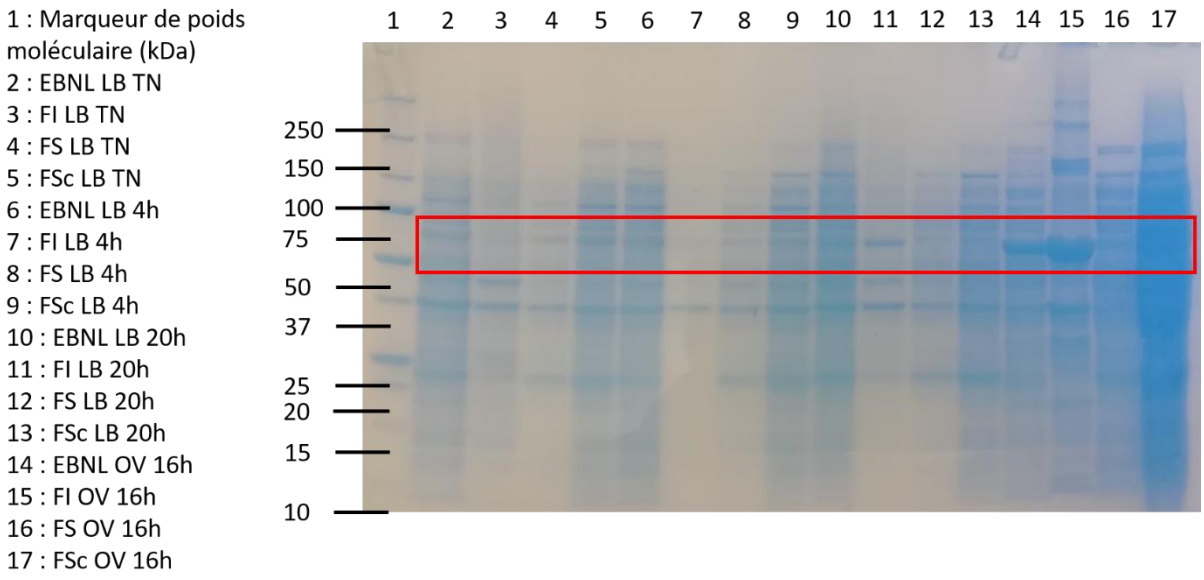


Figure 1 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-15b-SamELA 6-His N-ter en milieu LB ou OV. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight. Le cadre rouge indique la position de SamELA exprimée.

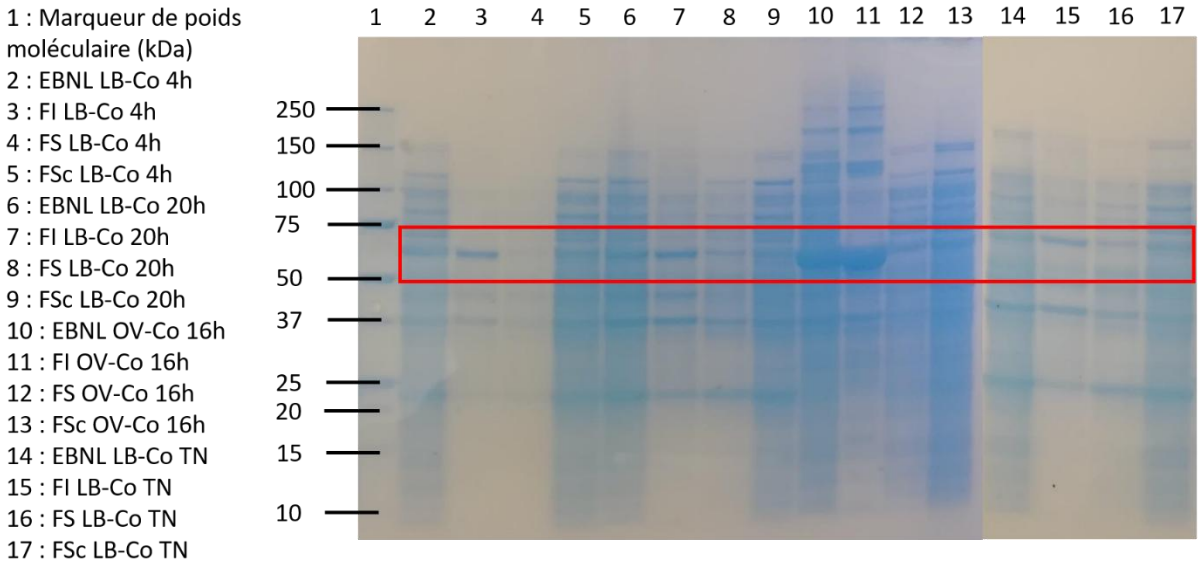


Figure 2 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-15b-SamELA 6-His N-ter en milieu LB-Co ou OV-Co. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt. Le cadre rouge indique la position de SamELA exprimée.

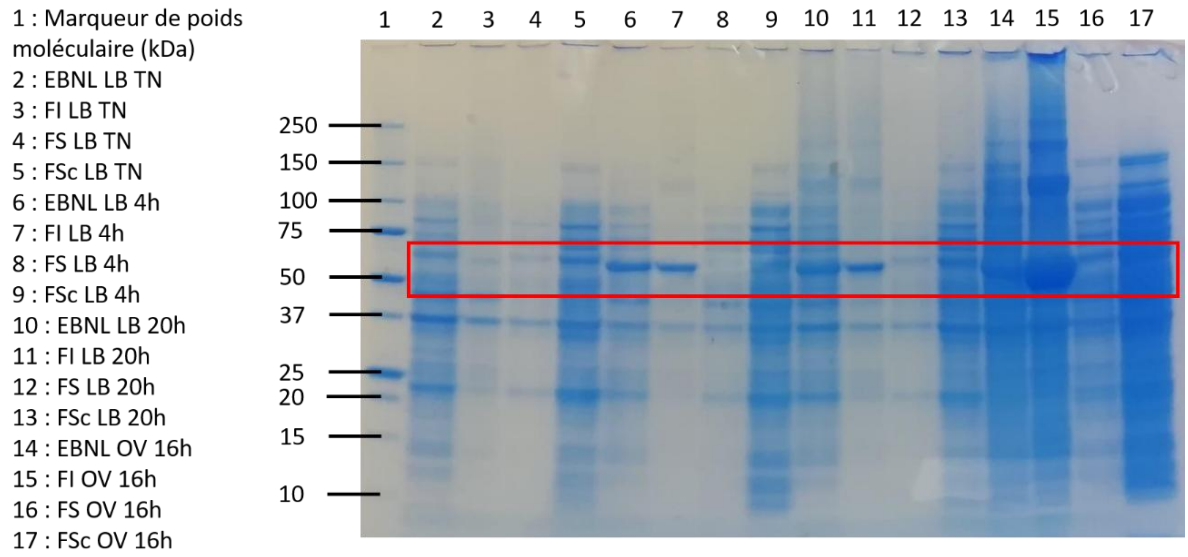


Figure 3 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter en milieu LB ou OV. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight. Le cadre rouge indique la position de SamELA exprimée.

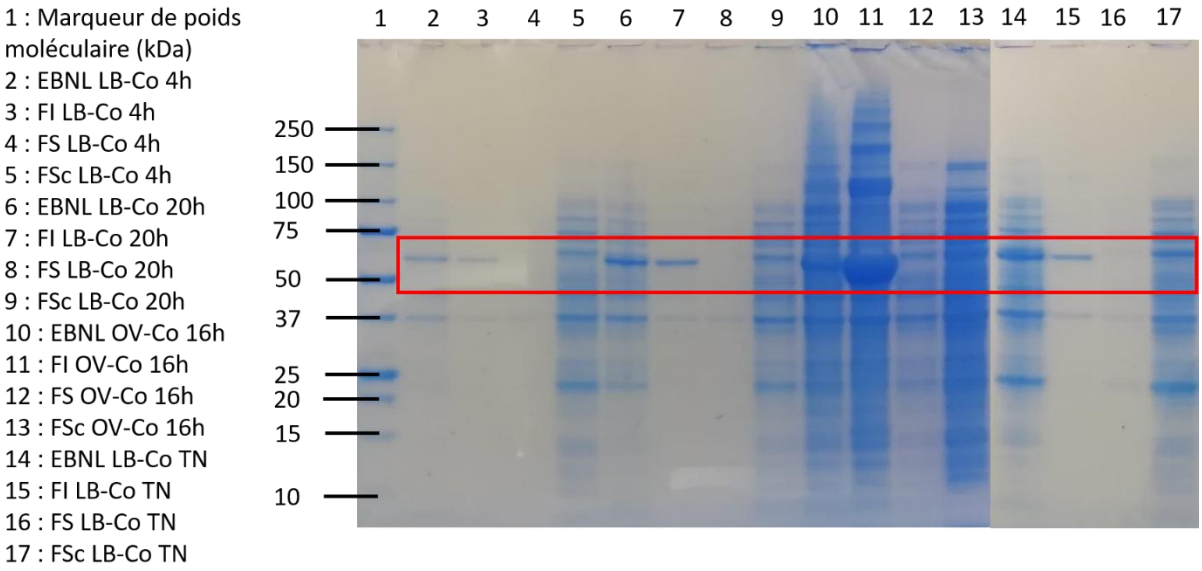


Figure 4 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter en milieu LB-Co ou OV-Co. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt. Le cadre rouge indique la position de SamELA exprimée.

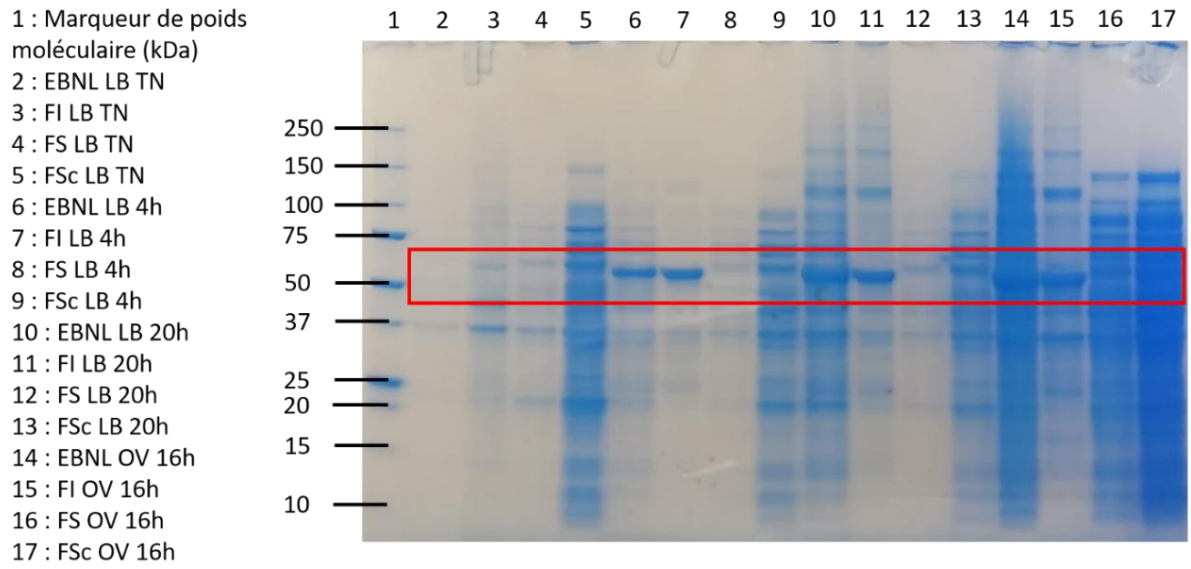


Figure 5 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter en milieu LB ou OV. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight. Le cadre rouge indique la position de SamELA exprimée.

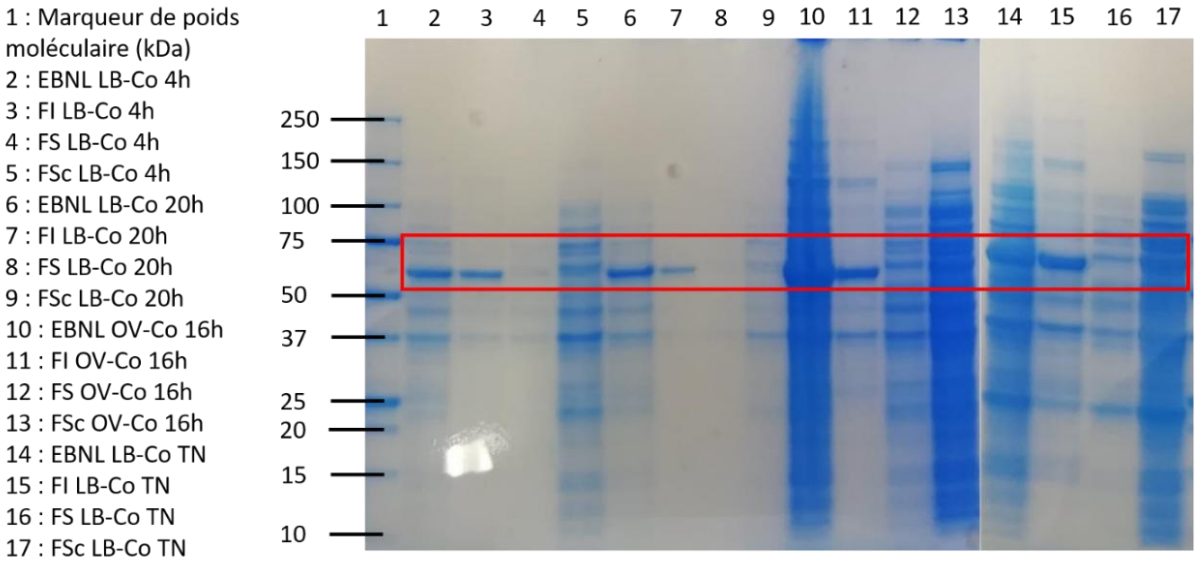


Figure 6 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter en milieu LB-Co ou OV-Co. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt. Le cadre rouge indique la position de SamELA exprimée.

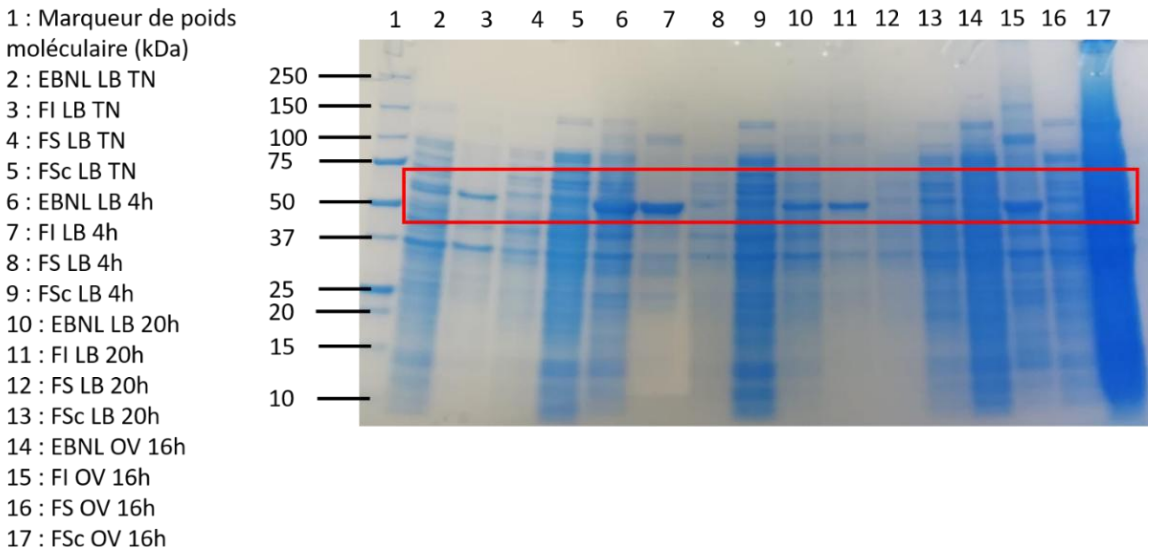


Figure 7 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter en milieu LB ou OV. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight. Le cadre rouge indique la position de SamELA exprimée.

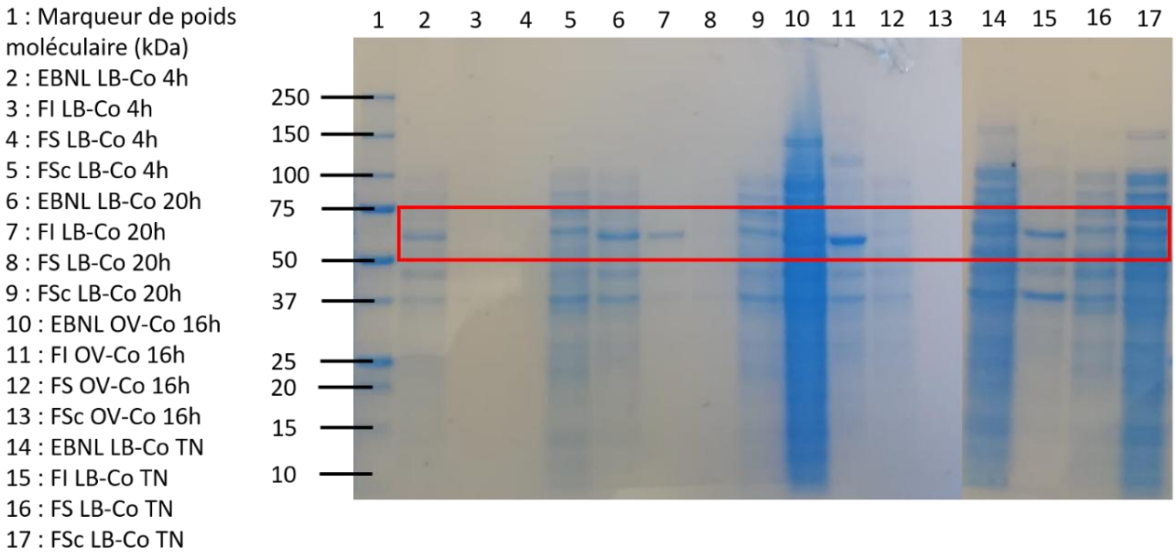


Figure 8 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter en milieu LB-Co ou OV-Co. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt. Le cadre rouge indique la position de SamELA exprimée.

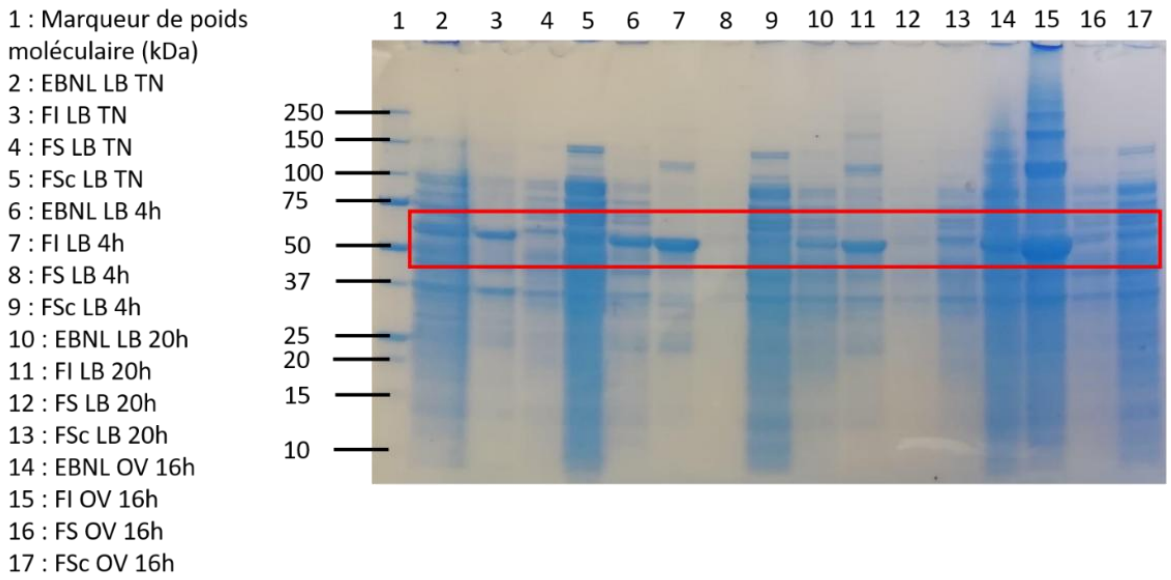


Figure 9 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter en milieu LB ou OV. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight. Le cadre rouge indique la position de SamELA exprimée.

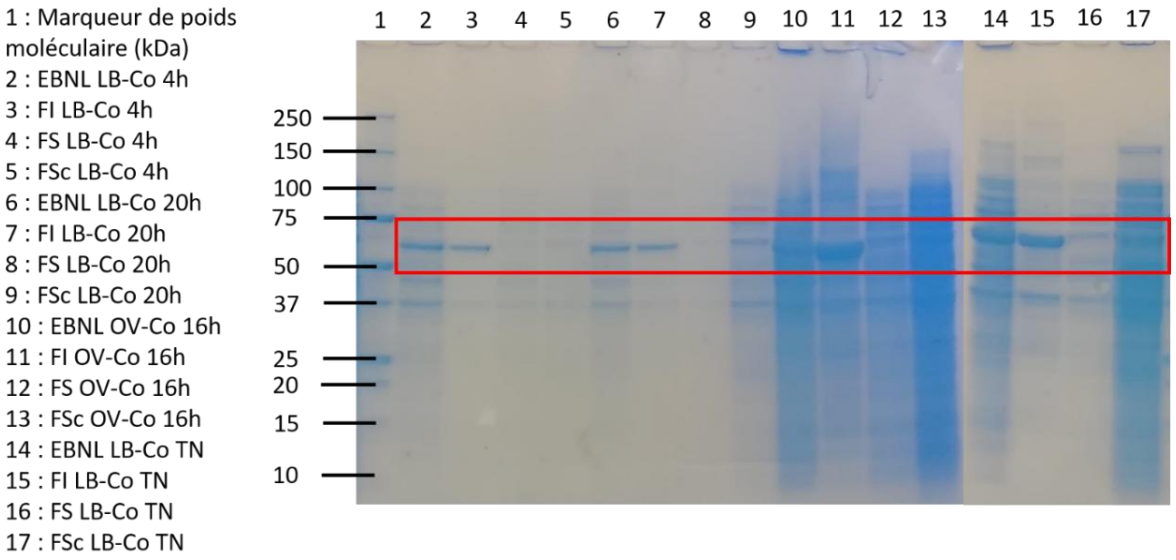


Figure 10 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter en milieu LB-Co ou OV-Co. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt. Le cadre rouge indique la position de SamELA exprimée.

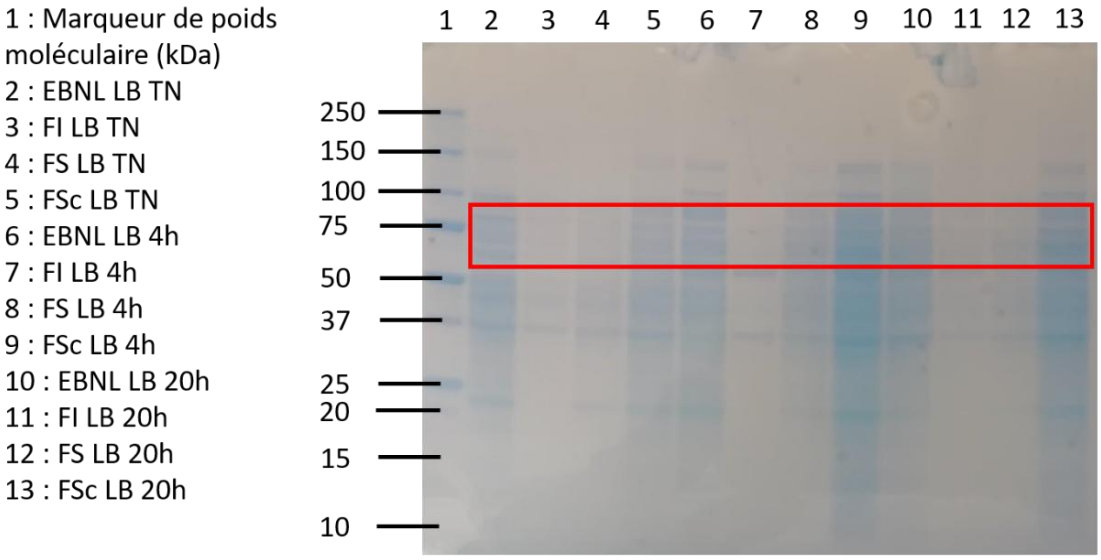


Figure 11 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold pLysS (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter en milieu LB. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée. Le cadre rouge indique la position de SamELA exprimée.

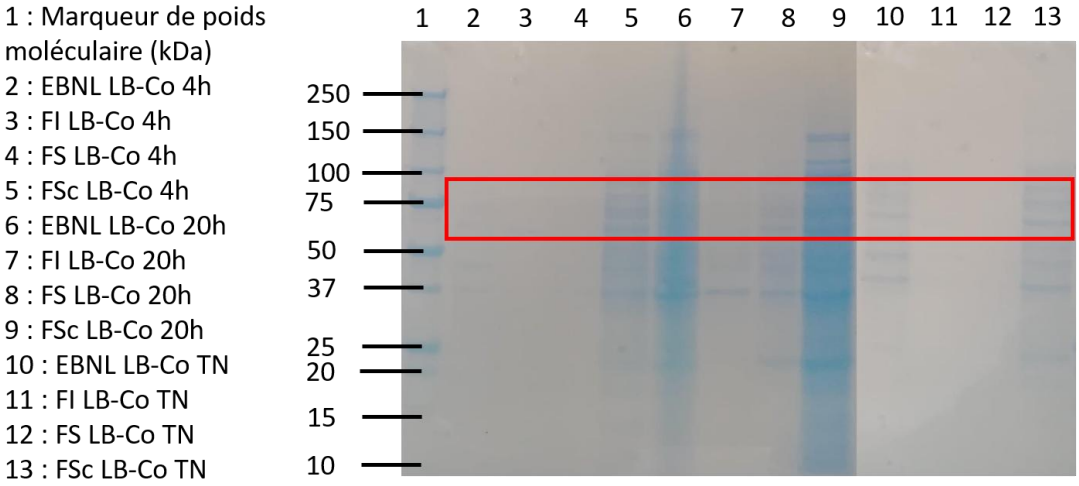


Figure 12 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold pLysS (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter en milieu LB-Co. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; Co : cobalt. Le cadre rouge indique la position de SamELA exprimée.

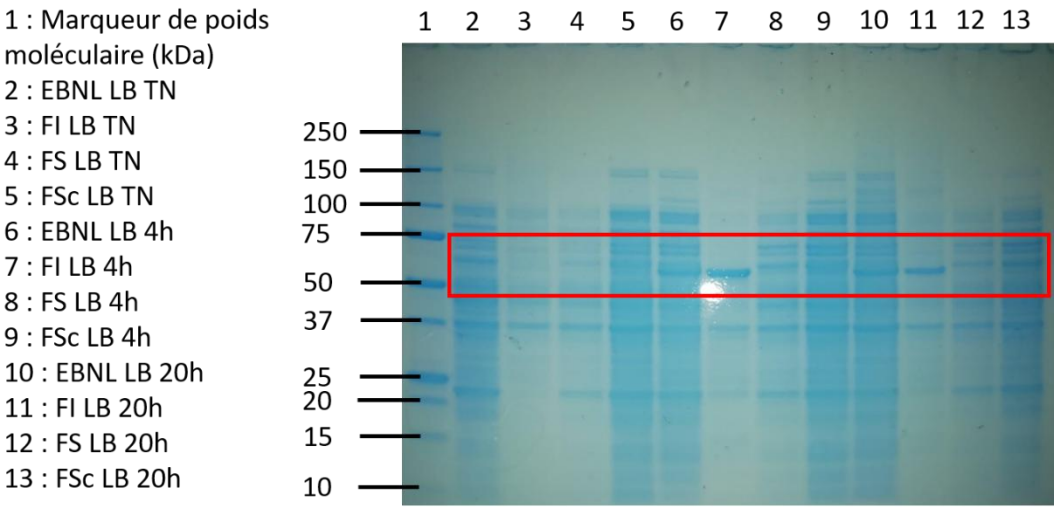


Figure 13 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold pLysS (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter en milieu LB. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée. Le cadre rouge indique la position de SamELA exprimée.

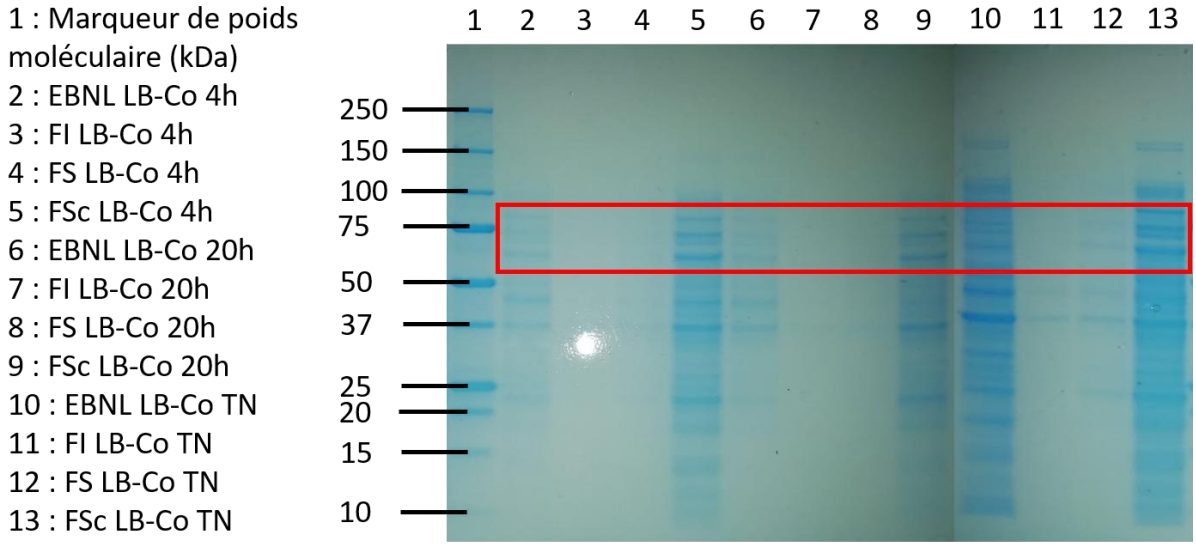


Figure 14 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold pLysS (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter en milieu LB-Co. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; Co : cobalt. Le cadre rouge indique la position de SamELA exprimée.

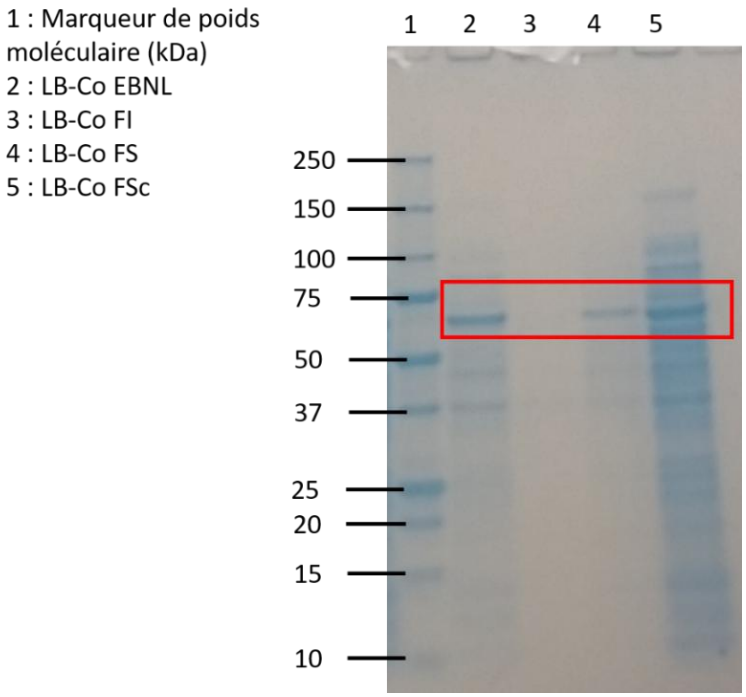


Figure 15 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Arctic Express (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter en milieu LB-Co. EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; Co : cobalt. Le cadre rouge indique la position de SamELA exprimée.

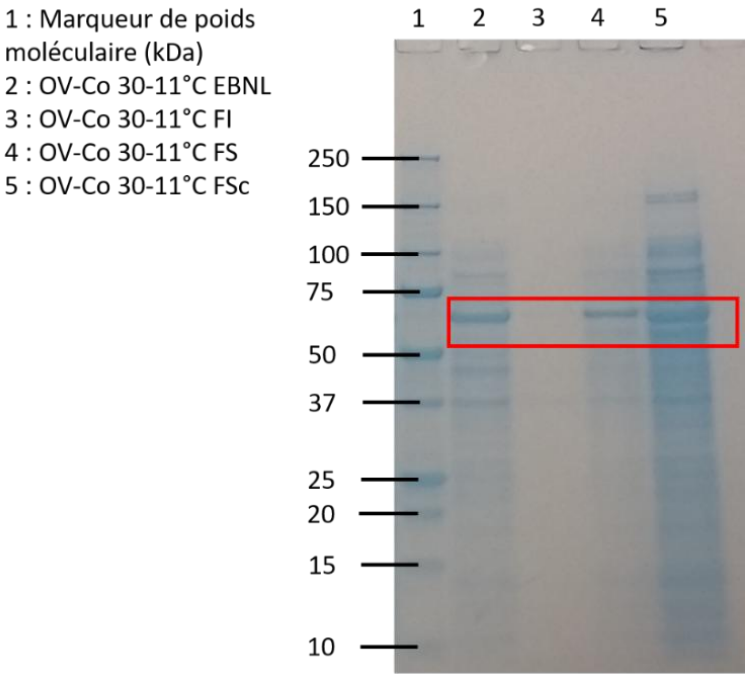


Figure 16 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Arctic Express (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter en milieu OV-Co. EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt. Le cadre rouge indique la position de SamELA exprimée.

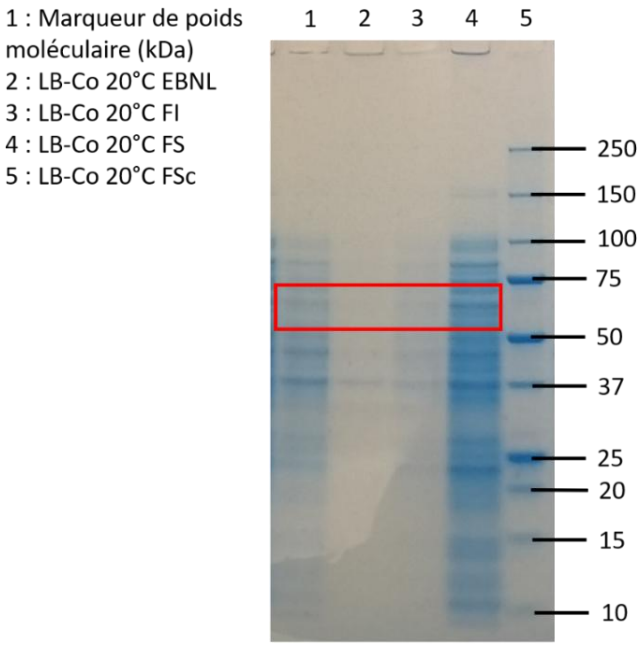


Figure 17 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter en milieu LB-Co. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; Co : cobalt. Le cadre rouge indique la position de SamELA exprimée.

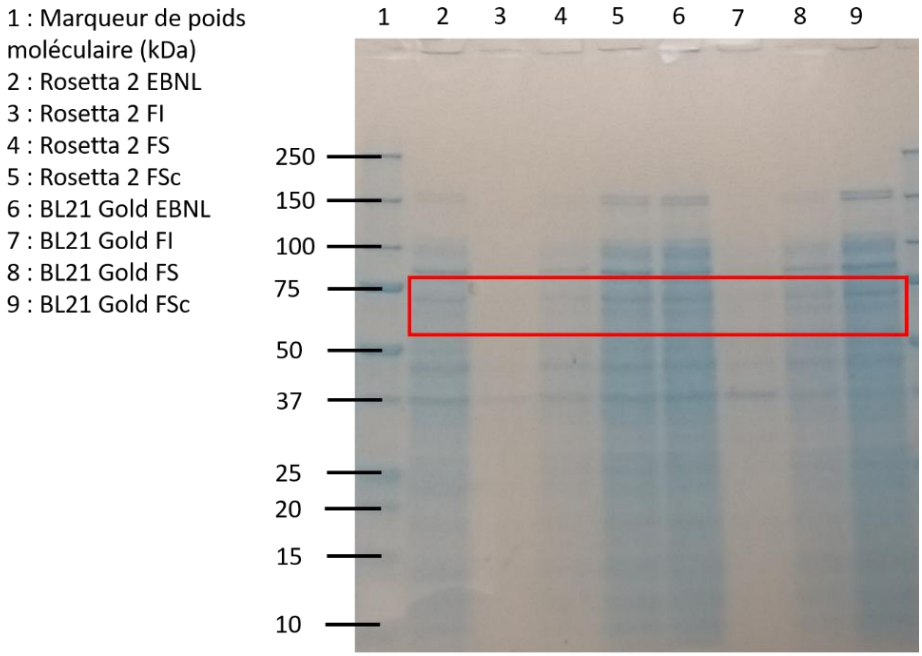


Figure 18 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* Rosetta 2 et BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter en milieu OV-Co. La température des cultures est abaissée à 20°C. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt. Le cadre rouge indique la position de SamELA exprimée.

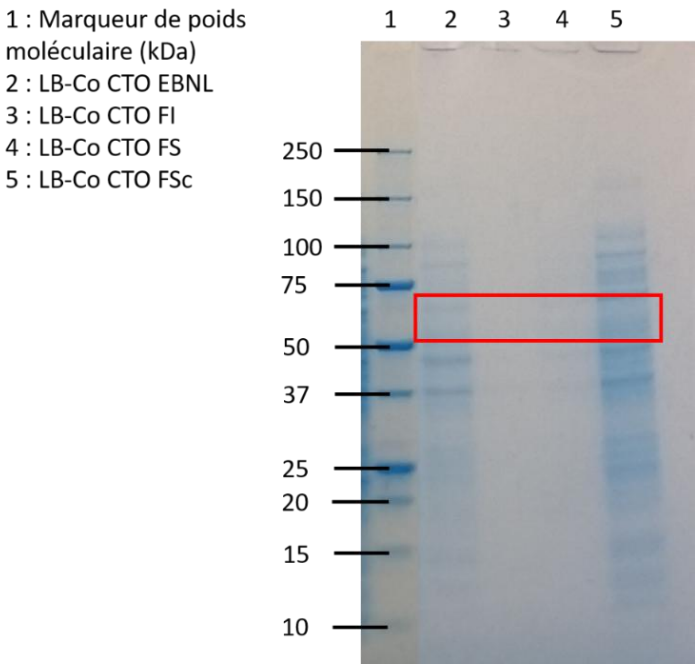


Figure 19 : Analyse par SDS-PAGE des protéines issues des cultures d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter en milieu LB supplémenté en cobalt (0,25 mM), pendant 24h. EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; Co : cobalt ; CTO : choc thermique et osmotique. Le cadre rouge indique la position de SamELA exprimée.

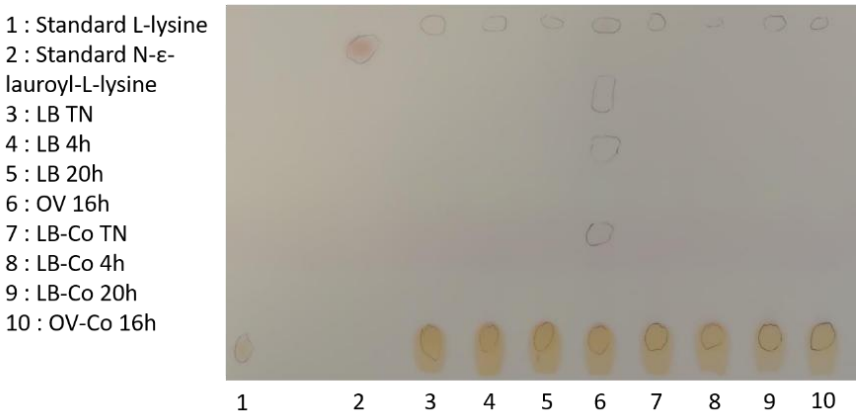


Figure 20 : Analyse par CCM du milieu de synthèse de la N-ε-lauroyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-15b-SamELA 6-His N-ter. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

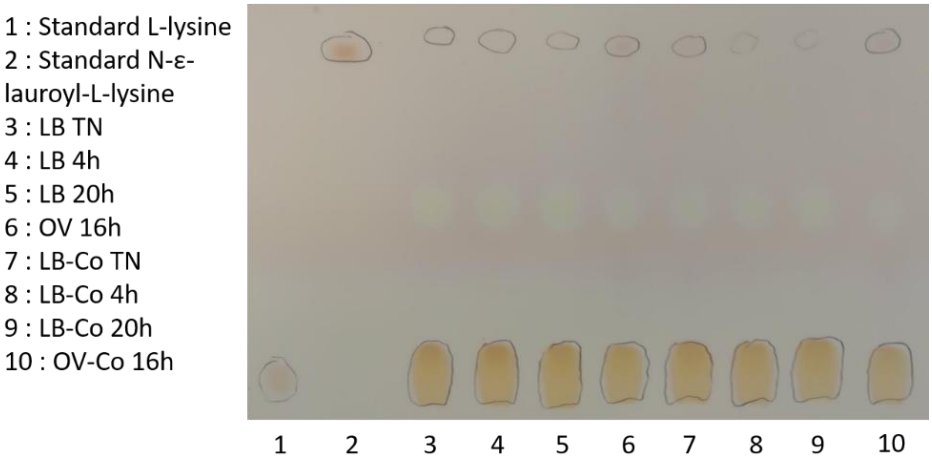


Figure 21 : Analyse par CCM du milieu de synthèse de la N-ε-lauroyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

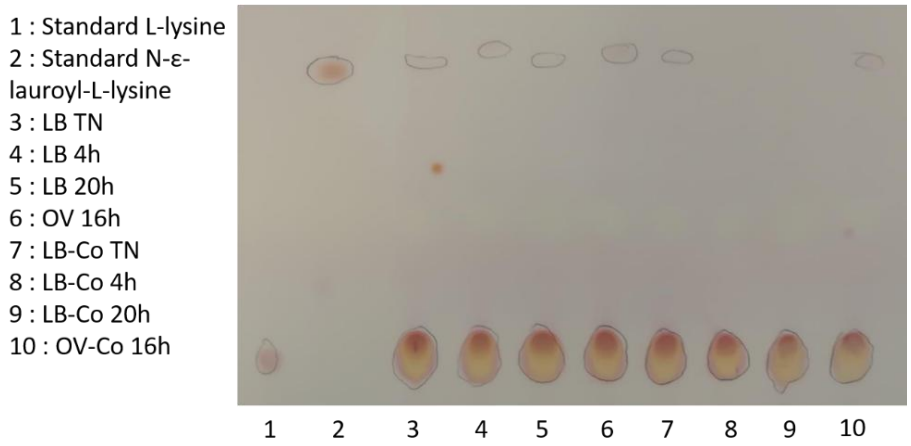


Figure 22 : Analyse par CCM du milieu de synthèse de la N-ε-lauroyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

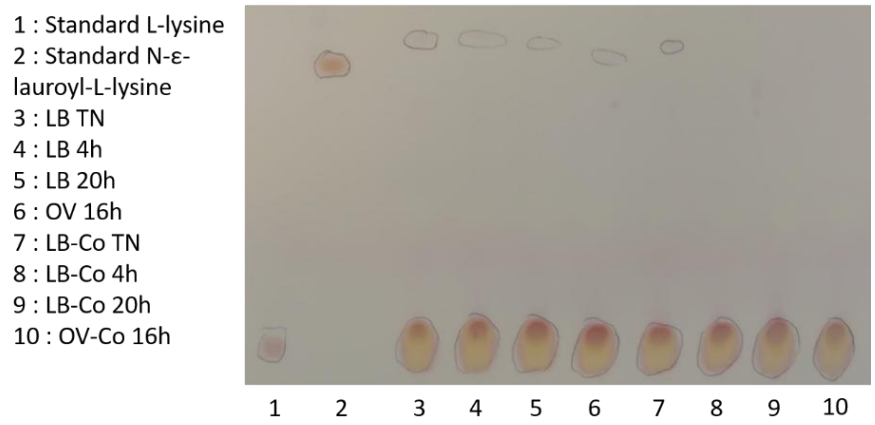


Figure 23 : Analyse par CCM du milieu de synthèse de la N-ε-lauroyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

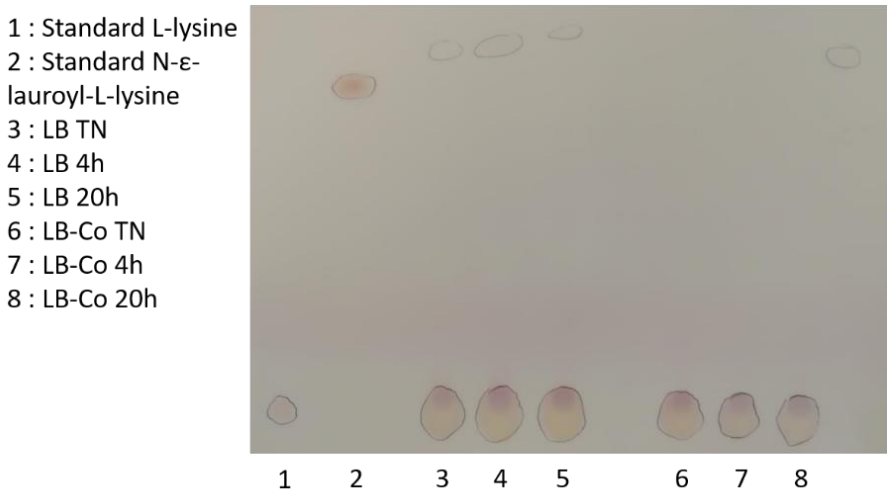


Figure 24 : Analyse par CCM du milieu de synthèse de la N-ε-lauroyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* BL21 Gold pLysS (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

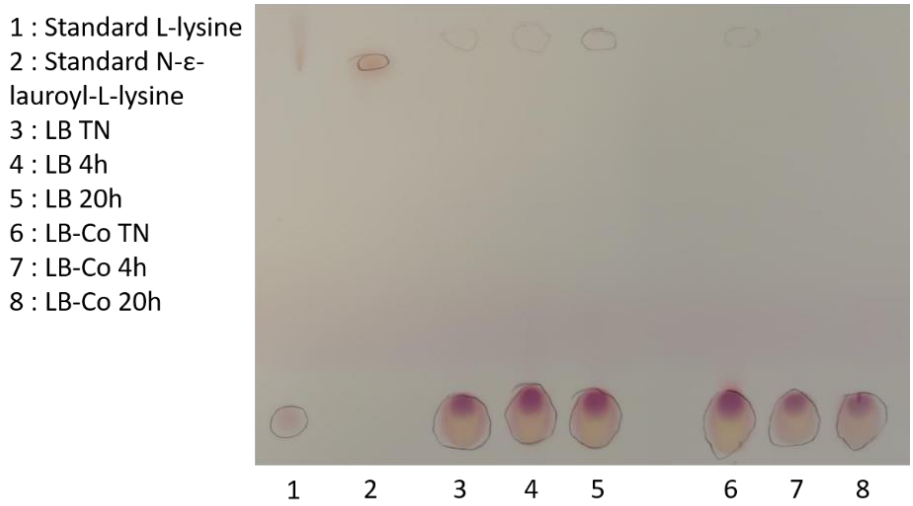


Figure 25 : Analyse par CCM du milieu de synthèse de la N-ε-lauroyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* BL21 Gold pLysS (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

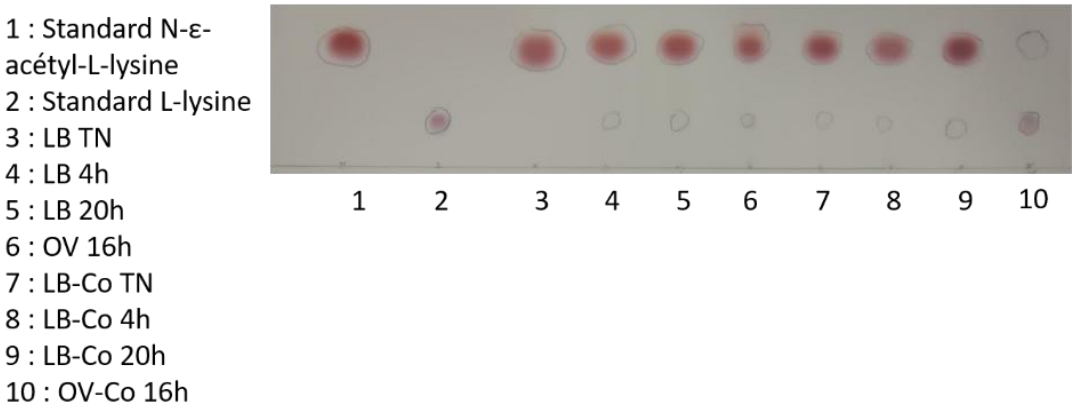


Figure 26 : Analyse par CCM du milieu d'hydrolyse de la N-ε-acétyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-15b-SamELA 6-His N-ter. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

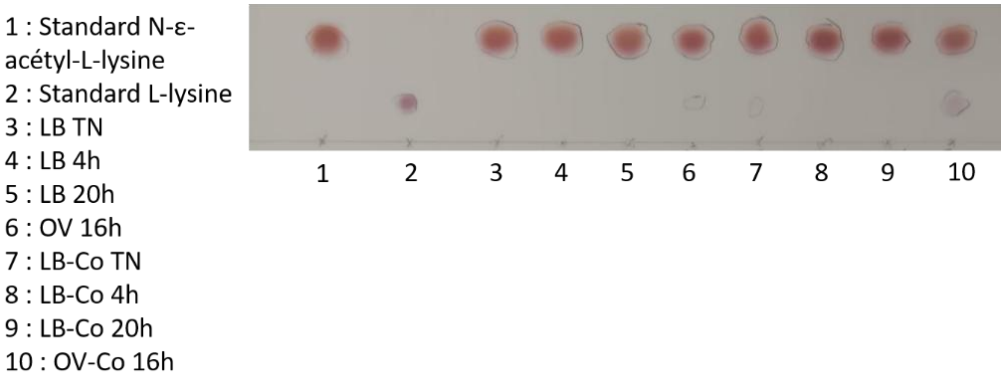


Figure 27 : Analyse par CCM du milieu d'hydrolyse de la N-ε-acétyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* Rosetta 2 (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

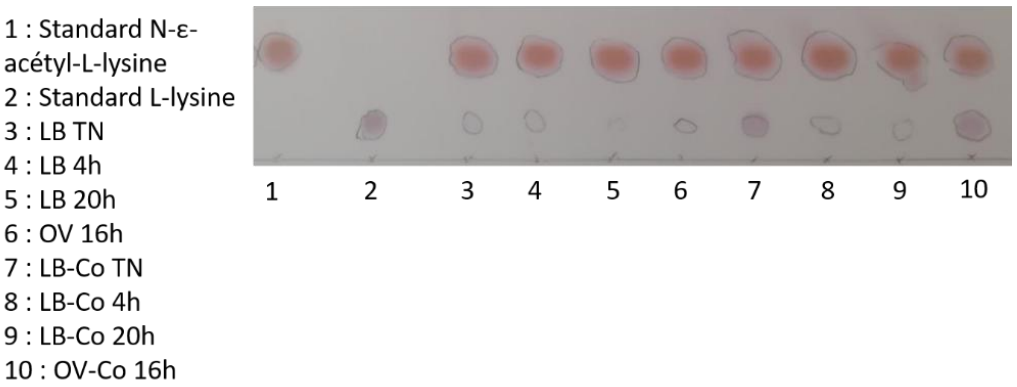


Figure 28 : Analyse par CCM du milieu d'hydrolyse de la N-ε-acétyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

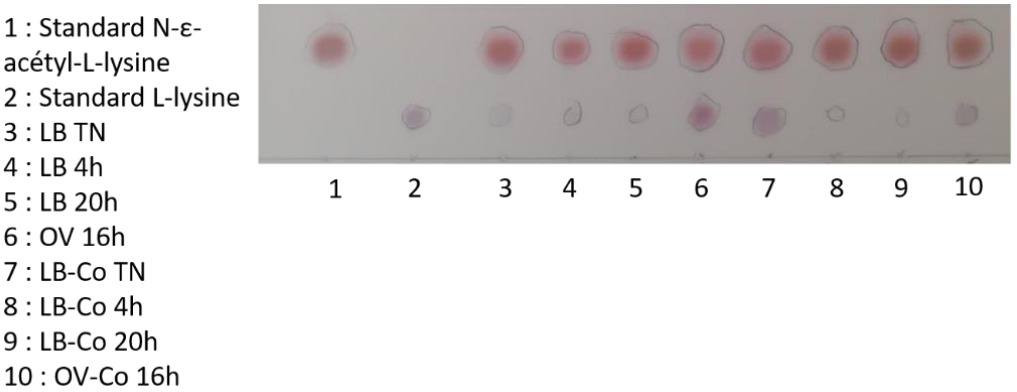


Figure 29 : Analyse par CCM du milieu d'hydrolyse de la N-ε-acétyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* BL21 Gold (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI :

fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

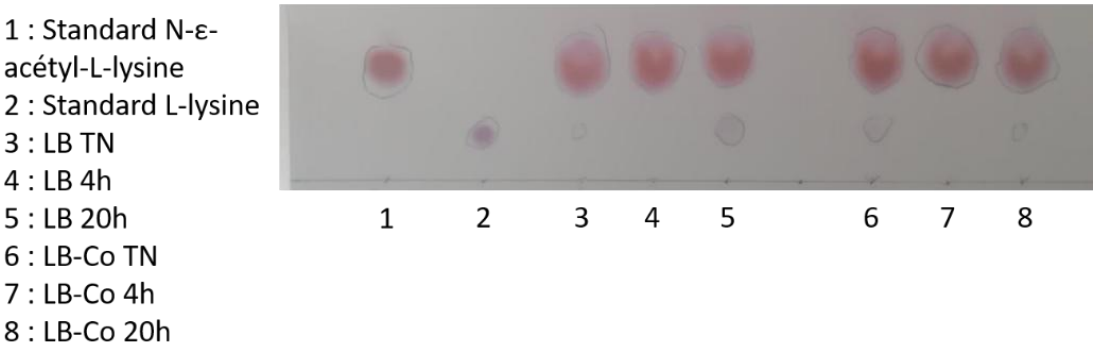


Figure 30 : Analyse par CCM du milieu d'hydrolyse de la N-ε-lauroyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* BL21 Gold pLysS (DE3) avec le plasmide pET-28a-SamELA 6-His N-ter. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

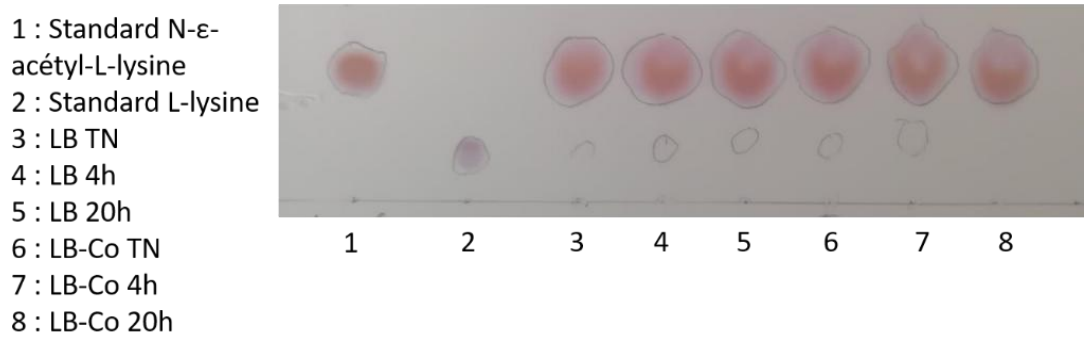


Figure 31 : Analyse par CCM du milieu d'hydrolyse de la N-ε-lauroyl-L-lysine en présence des fractions solubles concentrées issues d'*E. coli* BL21 Gold pLysS (DE3) avec le plasmide pET-28a-SamELA 6-His C-ter. TN : témoin négatif (sans induction) ; EBNL : extrait brut non lysé ; FI : fraction insoluble ; FS : fraction soluble ; FSc : fraction soluble concentrée ; OV : Overnight ; Co : cobalt.

Résumé de la thèse

Les acides aminés acylés constituent une large famille de molécules tensioactives biosourcées. En raison de leurs diverses propriétés techno-fonctionnelles, ces composés présentent un fort potentiel applicatif dans différents secteurs industriels. Ils sont obtenus par une réaction de N-acylation et sont classiquement produits par voie chimique, selon la réaction de Schotten-Baumann. Celle-ci est efficace toutefois ses conditions de mise en œuvre sont impactantes sur le plan environnemental. De plus, elle manque de sélectivité, conduisant à des produits indésirables. Une alternative prometteuse consiste à utiliser des enzymes pour catalyser la réaction. Celles-ci présentent l'avantage d'être sélectives ; par ailleurs leurs conditions de mise en œuvre sont douces et écocompatibles. SamAA et SamELA sont deux des quatre aminoacylases découvertes chez *Streptomyces ambofaciens*. Elles catalysent la N-acylation sélective de la L-lysine, soit au niveau de sa fonction amine α, soit au niveau de sa fonction amine ε. Cette thèse vise d'une part l'étude des éléments responsables de cette sélectivité à l'échelle moléculaire, d'autre part l'identification des conditions de production et de purification des deux enzymes recombinantes.

En l'absence de données cristallographiques, la structure tridimensionnelle des deux enzymes a été prédite par modélisation par homologie. Plusieurs templates ainsi que des résidus conservés ont été identifiés. Des hypothèses sur la structure quaternaire des enzymes, l'organisation de leur site actif et le rôle des résidus conservés dans la catalyse ou la fixation des ions métalliques ont pu être formulées. SamAA serait un homodimère binucléaire, caractérisé par un monomère en forme de sifflet et la présence de deux domaines, tandis que SamELA serait un monomère dont la structure s'organise autour de trois domaines. Des simulations de docking ont été menées pour expliquer la sélectivité des deux aminoacylases observée expérimentalement. Les résultats relatifs à SamAA suggèrent fortement l'acylation/déacylation sélective de la fonction amine α de la L-lysine, tandis que la prise en charge de la fonction amine ε par l'enzyme paraît improbable. Ceci est parfaitement cohérent avec les données expérimentales. Une hypothèse selon laquelle l'enzyme passerait d'une conformation ouverte à une conformation fermée lors de l'arrivée des substrats a été formulée et un mécanisme catalytique a été proposé. Dans le cas de SamELA, les résultats issus des modèles ne corroborent pas les observations expérimentales et ne permettent pas d'expliquer la sélectivité de l'enzyme vis-à-vis de la fonction amine ε de la L-lysine. Ceci pourrait s'expliquer par des imprécisions dans la modélisation d'une boucle à proximité du site actif.

La production des deux aminoacylases recombinantes a été étudiée dans différents systèmes d'expression pour identifier les conditions limitant la formation de corps d'inclusion et favorisant la forme soluble et active des enzymes. Plusieurs souches (*Escherichia coli* et *Streptomyces lividans*) et une combinaison de diverses conditions de culture ont été mises en œuvre. Dans tous les cas, l'absence de surexpression des enzymes sous forme soluble a été constatée. Seule la présence d'une faible quantité de SamAA active a pu être mise en évidence dans certaines fractions solubles capables de catalyser la synthèse de la N-α-lauroyl-L-lysine. L'utilisation de la souche *E. coli* BL21 Gold (DE3) et la supplémentation du milieu de culture en cobalt semblent favoriser la production de SamAA sous forme soluble. L'activité la plus élevée a été obtenue avec *S. lividans* transformée par le vecteur intégratif pIB139-SamAA StrepTag N-ter et cultivée dans le milieu de fermentation à 28 °C pendant trois jours. Pour SamELA, une activité hydrolytique de la N-ε-acétyl-L-lysine a été également détectée dans plusieurs fractions solubles, indiquant la présence de l'enzyme sous forme soluble et active.

Thesis abstract

Acylated amino acids constitute a broad family of bio-based surfactant molecules. Due to their diverse techno-functional properties, these compounds exhibit high application potential in multiple industrial sectors. They are produced *via* N-acylation reactions and are classically synthesised following the Schotten-Baumann reaction. Although effective, this chemical process entails environmentally detrimental conditions and suffers from limited selectivity, leading to undesirable by-products. A promising alternative consists in using enzymes to catalyze the reaction. This way combines high selectivity with mild and environmentally compatible conditions.

SamAA and SamELA are two of the four aminoacylases discovered in *Streptomyces ambofaciens*. These enzymes catalyze the selective N-acylation of lysine, either at its α-amino or ε-amino group. The present work aims, first, to identify the molecular determinants underlying this regioselectivity, and second, to define the conditions enabling the production and purification of the two recombinant enzymes. Due to the lack of crystallographic data, the three-dimensional structures of SamAA and SamELA were predicted by homology modeling. Several templates and conserved residues were identified, enabling hypotheses on the quaternary structure of the enzymes, the active site organization, and the role of conserved residues in catalysis and/or metal ion coordination. SamAA is suggested to be a binuclear homodimer, characterized by a whistle-shaped monomer comprising two domains, whereas SamELA is predicted to be monomeric, with three domains. Docking simulations were performed to explain the experimentally observed selectivity of the two aminoacylases. For SamAA, the models strongly support the selective acylation/deacylation of the L-lysine α-amino group, while acylation of the ε-amino group appears unlikely, in full agreement with experimental data. A conformational change from an open to a closed form upon substrate binding was suggested, and a putative catalytic mechanism was proposed. By contrast, in the case of SamELA, the models were unable to rationalize the experimentally observed selectivity, that may result from inaccuracies or even errors in the modeling of a loop near the active site.

Production of the two recombinant aminoacylases was investigated in various expression systems, with the objective to identify the conditions minimizing inclusion body formation and favoring accumulation of the soluble, active enzymes. Several host strains (*Escherichia coli* and *Streptomyces lividans*) and diverse culture conditions were studied. In all cases, no overexpression of the enzymes in soluble form was observed. Only minor amounts of active

SamAA was detected in some soluble fractions which were capable of catalyzing the synthesis of N-α-lauroyl-L-lysine. Production of soluble SamAA was enhanced by using the *E. coli* BL21 Gold (DE3) strain and supplementing the culture medium with cobalt. The highest activity was obtained with *S. lividans* transformed with the integrative vector pIB139-SamAA StrepTag N-ter, cultivated in fermentation medium at 28 °C for three days. For SamELA, hydrolytic activity toward N-ε-acetyl-L-lysine was detected in several soluble fractions, confirming the presence of the enzyme in an active and soluble state.